TITLE: A Randomized, Double-Blind, Placebo-Controlled Trial of Adjunctive BHV4157 in Obsessive Compulsive Disorder

NCT NUMBER: NCT03299166

DATE: 11-June-2020

Biohaven Pharmaceuticals

Protocol BHV4157-202

A Randomized, Double-Blind, Placebo-Controlled Trial of Adjunctive BHV-4157 in Obsessive Compulsive Disorder

Statistical Analysis Plan

Draft Version 2.0
Date: 11 June 2020

SIGNATURE PAGE

Protocol Title:	A Randomized, Double-Blind, Placebo-Controlled Trial of Adjunctive BHV-4157 in Obsessive Compulsive Disorder
Sponsor:	Biohaven Pharmaceuticals, Inc.
Protocol Number:	BHV4157-202
Document Version/Date:	V2.0/11 June 2020: Revised SAP
Author:	PPD PPD PPD CCI

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

List of Tables

List of Figures

REVISION HISTORY

Version	Description of Change
1.0	Original version based on Protocol Version 5.0
2.0	Revised version based on Protocol Version 7.0. Revision primarily includes changes related to the potential impact of the COVID-19 pandemic crisis. These changes include: • Details on the summarization of data obtained from the CRF collecting information on the impact of COVID-19 pandemic crisis on study visits • Additional sensitivity analyses included to assess the impact of COVID-19 pandemic crisis on the primary and secondary endpoints as well as key safety domains. • Specification on list of subject for safety narratives • Added subgroup of interest based on Concomitant Standard of Care OCD Treatment

ABBREVIATIONS

Abbreviation	Definition
AE	Adverse event
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
ANCOVA	Analysis of covariance
AST	Aspartate aminotransferase
BABS	Brown Assessment of Beliefs Scale
BAI	Beck Anxiety Inventory
BDNF	Brain derived neurotrophic factor
BPD	Borderline Personality Disorder
BQL	Below limit of quantification
BUN	Blood urine nitrogen
CGI-S	Clinical Global Impression of Severity
CI	Confidence interval
CPK	Creatine phosphokinase
CRF	Case Report Form
CSR	Clinical study report
DILI	Drug induced liver injury
ECG	Electrocardiogram
HbA1c	Hemoglobin A1c
HBsAg	Hepatitis B surface antigen
HCV	Hepatitis C virus
HDL	High-density lipoprotein
HIV	Human immunodeficiency virus
GGT	Gamma-glutamyl transferase
ICH	International Conference on Harmonisation
IWRS	Interactive web response system
LDH	Lactate dehydrogenase
LDL	Low-density lipoprotein
LSMeans	Least square means
mITT	Modified Intent to Treat
MAR	Missing at random
MCMC	Markov chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple imputation
MMRM	Mixed model for repeated measures
MMSE	Mini Mental State Exam

Abbreviation	**Definition**
MNAR	Missing not at random
NC=F	Non Completer equals Failure
OCD	Obsessive Compulsive Disorder
PID	Patient Identification Number
PK	Pharmacokinetic
PT	Preferred term
QD	Once daily
QIDS-SR	Quick Inventory of Depressive Symptomatology – Self Report
S-STS	Sheehan-Suicidality Tracking Scale
SAE	Serious adverse event
SAP	Statistical analysis plan
SD	Standard deviation
SDS	Sheehan Disability Scale
SE	Standard error
SI	Standard International
SOC	System organ class
TSH	Thyroid-stimulation hormone
ULN	Upper limit of normal
WHO-DD	World Health Organization-Drug Dictionary
Y-BOCS	Yale-Brown Obsessive Compulsive Scale

1 INTRODUCTION AND OBJECTIVES OF ANALYSIS

1.1 Introduction

This document presents the statistical analysis plan (SAP) for Biohaven Pharmaceuticals, Protocol BHV4157-202: A Randomized, Double-blind, Placebo-controlled Trial of Adjunctive Troriluzole (BHV-4157) in Obsessive Compulsive Disorder (OCD).

This SAP is based on version 7.0 of the protocol dated March 16, 2020. 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.

Randomization completed on 2-Mar-2020 with 248 subjects randomized. At around the time of the COVID-19 National Emergency declaration on 13-Mar-2020 there were approximately 50 subjects still ongoing in the double-blind randomization phase. The COVID-19 pandemic crisis will likely impact assessment of the primary objective by limiting the physical ability to assess ongoing subjects and potentially by influencing the assessments on clinical symptoms and status. Although the principle concern is with regard to the primary endpoint of the change in Y-BOCS total score from baseline to the end of the randomization phase (Week 12) there could be potential impact on other endpoints.

Biohaven believes the potential impact of the COVID-19 pandemic crisis involves both the potential to increase missing data and the unknown effect the crisis may have on subjects' OCD symptoms and level of anxiety. With respect to the potential for missing data, planned study visits and observations are expected to be impacted by (a) site closure or site restricted access, (b) subjects diagnosed with COVID-19 or quarantined due to COVID-19; and (c) subjects who are unwilling or unable to come to the site due to COVID-19.

For the purposes of assessing the impact of COVID-19 on missing data, Biohaven initiated collection of data on each planned study visit to determine if missing observations were due to COVID-19. For analysis purposes, a planned reference date of 13-March-2020 (the date a National Emergency was declared) is used to assess the impact on study results. This date was selected to capture all data that became available from the start of the formally announced COVID-19 crisis. To minimize missing data, the BHV4157-202 study protocol was amended (version 7.0) to allow an additional 6 weeks for the subject to come in for the final Week 12 visit. In addition, the sites will be allowed to administer the Y-BOCS by remote video except for Week 12. These assessments will be clearly marked as being collected via a remote video and will only be used in sensitivity analyses. Other assessments collected remotely will also be clearly marked in the database and be handled as indicated in the SAP.

1.2 Objectives of Statistical Analysis

Primary Objective

- To evaluate the efficacy of Troriluzole as adjunctive therapy in subjects with OCD who have had an inadequate response to SSRI, clomipramine, venlafaxine or desvenlafaxine treatment

Secondary Objectives

- To assess the safety and tolerability of Troriluzole, relative to placebo, in subjects with OCD
- To evaluate the efficacy of Troriluzole compared to placebo on functional disability as measured by the Sheehan Disability Scale (SDS)
- To evaluate the efficacy of Troriluzole compared to placebo on global functioning as measured by the Clinical Global Impression – Severity Scale (CGI-S)
- To evaluate the efficacy of Troriluzole compared to placebo on obsessive symptomatology as measured by the change in the Yale Brown Obsessive Compulsive Scale (Y-BOCS) obsessions subscale

Exploratory Objectives

- To evaluate the efficacy of Troriluzole compared to placebo on depressive symptomatology as measured by the Quick Inventory of Depressive Symptomatology – Self Report (QIDS-SR)
- To evaluate the efficacy of Troriluzole compared to placebo on anxiety symptoms as measured by the Beck Anxiety Inventory (BAI)
- To evaluate the efficacy of Troriluzole compared to placebo on insight regarding obsessional beliefs as measured by the Brown Assessment of Beliefs Scale (BABS)
- To assess pharmacodynamics effects of Troriluzole vs placebo on markers of synaptic plasticity as measured by plasma levels of BDNF and proBDNF
- To characterize the pharmacokinetics (PK) of Troriluzole based on sparse sampling

2 STUDY DESIGN

2.1 Synopsis of Study Design

BHV4157-202 is a Phase 2b/3, multicenter, randomized, double-blind, placebo-controlled, 2-arm study designed to assess safety, tolerability, and efficacy of Troriluzole as adjunctive therapy when added to standard of care treatment in subjects with OCD who failed to respond adequately to prior pharmacotherapy. Current treatment failure is defined by a Y-BOCS score of 19 or greater despite at least 10 weeks of treatment at baseline with an adequate dose of an SSRI, clomipramine, venlafaxine or desvenlafaxine treatment.

Subjects who are stable on Standard of Care medication and having an inadequate response will be randomized to additionally receive placebo (QD) or Troriluzole (200 mg QD, after four weeks at 140 mg QD).

Dosing will continue for approximately 12 weeks with scheduled study visits at weeks 4, 8 and 12. Due to the COVID-19 pandemic crisis subjects might not be able to visit the site for their scheduled visit. The protocol was revised (version 7.0) to extend the time allowable for the week 12 visit by an additional 6 weeks.

Eligible subjects will have the opportunity to continue in a 48 week open label Extension Phase. Those subjects not continuing in the 48 week Extension Phase will return 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 will be 4 weeks after the Week 12 Randomization Phase visit. Then subjects will undergo visits every 12 weeks up to Week 48 of this phase. All subjects will undergo 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 will be randomized in a 1:1 ratio to receive either placebo (QD) or Troriluzole (200 mg QD). Subjects will receive either Troriluzole (140mg) or Placebo for the first four weeks and then will be increased to 200 mg (or matching placebo) for the duration of the study. Treatment assignments will be obtained by the investigator (or designee) via the Interactive Web Reponse 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.

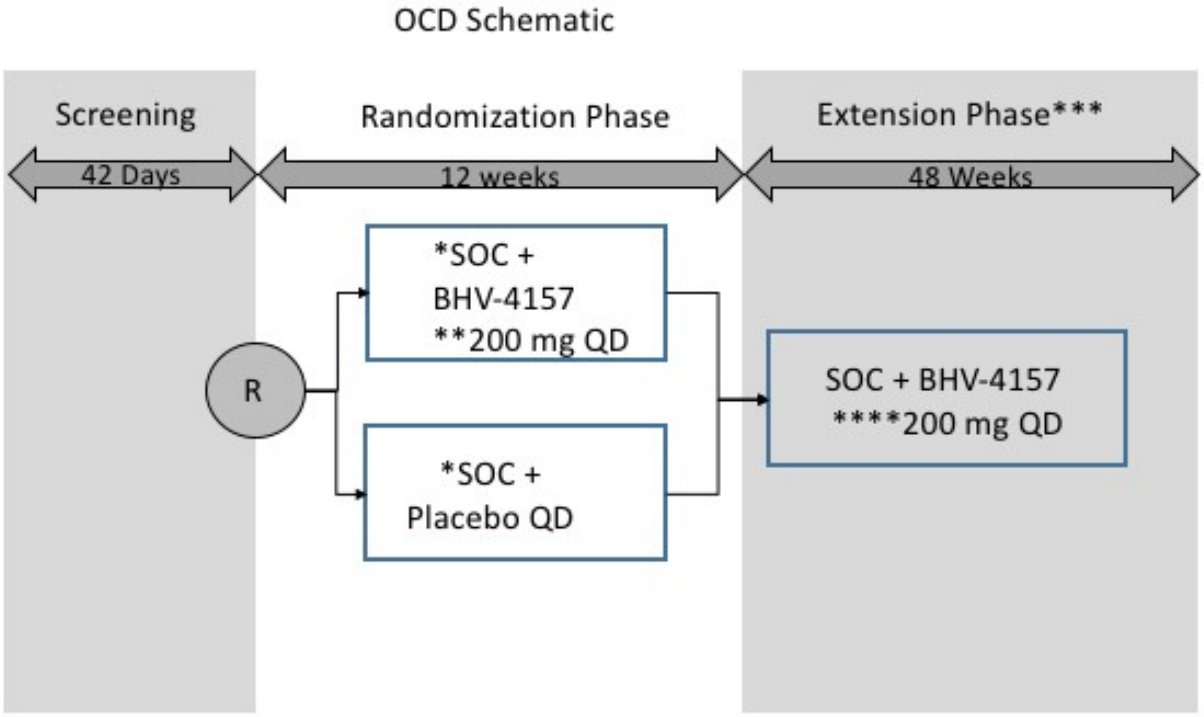
Unblinding will be managed via the IWRS system. A pharmacokineticist, IWRS randomization manager, and pharmacovigilance role may be unblinded before data are more generally unblinded after the Randomized Phase of the study. Except as noted above, other members of the BHV research team will remain blinded.

Although the Extension Phase is open label, the study remains blinded through the first 4 weeks of the Extension Phase. Subjects on placebo during the Randomization Phase are switched in a blinded manner to Troriluzole 140 mg for the first 4 weeks, and then increased to Troriluzole 200 mg for the remainder of the Extension Phase.

Once all subjects have completed the Randomization Phase, the database will be cleaned, locked, and the data will be analyzed in an unblinded manner.

In cases of accidental unblinding, the Medical Monitor will be contacted and it will be ensured that every attempt to preserve the blind is made.

Figure 1: Study Schematic



* Subjects should have been taking an adequate maximum tolerated dose, as defined in Inclusion Criteria 3.b.2, of an SSRI, clomipramine, venlafaxine or desvenlafaxine for at least 8 weeks prior to screening and 10 weeks at Baseline.
** Subjects will receive 140mg for the first four (4) weeks and will then be increased to 200mg for the duration of the study. Down titration will only be allowed to address tolerability issues.
*** Eligible subjects will include those who perceived benefit in earlier phases or for whom the PI believes extended treatment with BHV-4157 would offer an acceptable risk-benefit profile.
****Subjects entering the Extension Phase will continue with the same dose taken at the end of the Randomization Phase. Subjects on placebo in the Randomization Phase will be switched in a blinded manner to 140mg for the first four weeks and then will be increased to 200mg for the duration of the study. Down titration after the first four weeks of the Extension Phase will only be allowed for tolerability purposes. All visits after Week 4 will be open-label.

2.4 Efficacy, Safety, and Other Variables

2.4.1 Primary Endpoint

- Improvement in obsessive-compulsive symptomatology is assessed using the change in the Y-BOCS total score from baseline to the end of the Randomization Phase (Week 12)

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 SDS total score from baseline to the end of the Randomization Phase (Week 12)
- Improvement in global clinical condition is assessed using the CGI-S at the end of the Randomization Phase (Week 12)
- Improvement in obsessive symptomatology is assessed using the change in the Y-BOCS obsessions subscale score from baseline to the end of the Randomization Phase (Week 12)

2.4.3 Exploratory Endpoints

- Improvement in depressive symptomatology is measured by the change in the QIDS-SR from baseline to the end of the Randomization Phase
- Improvement in anxiety is assessed using the change in the BAI from baseline to the end of the Randomization Phase
- Improvement of insight into obsessive-compulsive beliefs is measured by the change in the BABS from baseline to the end of the Randomization Phase
- The impact on markers of synaptic plasticity is assessed using the change in plasma BDNF and proBDNF levels from baseline to the end of the Randomization Phase
- The PK profile of Troriluzole is characterized by blood concentrations observed in treated subjects

3 SUBJECT POPULATIONS

3.1 Population Definitions

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

- Enrolled Subjects: Subjects 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

- Randomized Subjects Completed or Discontinued Prior to COVID-19: Enrolled subjects who received a treatment assignment from the IWRS and completed or discontinued the Randomization Phase prior to March 13, 2020 (the date a national emergency for COVID-19 was declared)

- Randomized Subjects Active During COVID-19: Enrolled subjects who received a treatment assignment from the IWRS and were still active in the Randomization Phase on or after March 13, 2020 (the date a national emergency for COVID-19 was declared)

- Treated Subjects in the Randomization Phase: Enrolled subjects who received at least 1 dose of blinded study therapy (Troriluzole or placebo)

- Treated Subjects with any Troriluzole: Enrolled subjects who received at least 1 dose of blinded Troriluzole or open-label Troriluzole

- Treated Subjects in the Extension Phase: Enrolled subjects who received at least 1 dose of blinded study therapy (Troriluzole or placebo) and at least one dose of Troriluzole in the open-label Extension Phase

- Modified Intent to Treat (mITT) Subjects (Randomized Phase): Randomized subjects that received at least one dose of study therapy and provided a non-missing baseline assessment and at least one non-missing post-baseline efficacy assessment (in person only) in the Randomization Phase

- Modified Intent to Treat (mITT) Subjects (Extension Phase): Randomized subjects that received at least one dose of Troriluzole in the open-label Extension Phase and provided a non-missing extension phase baseline assessment and at least one non-missing extension phase efficacy assessment

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 protocol deviations will be presented in a tabulation and data listing. Protocol deviations related to the COVID-19 pandemic will be indicated and summarized separately where appropriate.

4 STATISTICAL METHODS

4.1 Sample Size Justification

The sample size for this study will be approximately 226 randomized subjects. This accommodates for lost subjects and is based on the following rationale.

Pittenger (2015), based on a 12 week study, concluded that riluzole produced an effect size (Cohen's d) of 0.45 in outpatients. An analysis of the Pittenger data found the correlations between baseline Y-BOCS and the Y-BOCS at 8, 10, and 12 weeks to be 0.50, 0.45, and 0.68, respectively.

Pittenger's results are consistent with those from Emamzadehfard's (2016) study in which riluzole produced an effect size of 0.59 at Week 10. Based on statistics presented in the 2016 paper, the correlation between baseline and Week 10 on the Y-BOCS appears to be roughly 0.18. In both studies, more than 90% of the randomized subjects completed the study.

Assuming that no more than 10% of the subjects are lost by Week 12, a sample size of about 113 per arm will yield roughly 101 subjects per arm. With an effect size of 0.45, a two-sided alpha of 0.05, and a correlation of 0.2 between baseline and Week 12 for Y-BOCS, the sample size provides 90% power to detect a difference between the treatment groups.

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, standard deviation (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, the Randomization Phase and the Extension Phase will be analyzed separately. For subjects receiving Troriluzole during both phases, summary statistics will be provided for data from both phases combined.

By-subject listings will display site-subject ID and "(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 describerace abbreviations as applicable, e.g., "Race: A = Asian, B = Black or African American, I = American Indian or Alaska Native, M = Multiple, N = Native Hawaiian or Other Pacific Islander, W = White". 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, and 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 Medical Dictionary for Regulatory Activities (MedDRA, Version 23.0) 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 AIDS (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 analysis of the primary endpoint and other continuous efficacy endpoints will be adjusted by including the baseline value of the endpoint as a covariate in statistical models. A sensitivity analysis on the primary and secondary endpoints will also be included which adds a covariate indicating whether or not the subject was assessed in the double-blind phase during the COVID-19 pandemic crisis (based on a March 13, 2020 reference date).

The randomization was not stratified.

4.2.4 Multiple Comparisons/Multiplicity

Type 1 error will be controlled for the primary and secondary efficacy endpoints by testing these endpoints with a gate-keeping procedure. The primary endpoint, change from baseline in total Y-BOCS at Week 12, will be tested at a two-sided alpha level of 0.05. If this test is significant, then the secondary efficacy endpoints of change from baseline for SDS, CGI-S and Y-BOCS obsessive subscale at Week 12 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 two-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 baseline severity (Y-BOCS, ≤ median or > median), current Standard of Care OCD treatment (citalopram, escitalopram, fluoxetine, fluvoxamine, paroxetine, sertraline, clomipramine, venlafaxine, desvenlafaxine), sex, race (Asian, Black or African American, White, and all other races combined), and COVID-19 Impact (separating subjects assessed before March 13, 2020 and those with at least one double-blind assessment on or after March 13, 2020).

Only descriptive summaries will be provided for the primary and secondary endpoints for each subgroup.

4.2.6 Withdrawals, Dropouts, and Loss to Follow-up

Subjects who withdraw from the study will not be replaced.

4.2.7 Missing, Unused, and Spurious Data

Unless otherwise noted, efficacy analyses will be based on observed data only. We initially expected the amount of data that will be lost due to early drop-out in the 12-week double blind phase of this study to be small; most likely less than 10% given that two previous studies in 2015 and 2016 by Pittenger and Emamzadehard, respectively, had less than 10% of subjects fail to complete the study. Primarily due to the COVID-19 pandemic crisis as well as an overall higher discontinuation rate seen in the study the final drop-out rate is likely to be approximately 20%. Since the COVID-19 pandemic crisis is external to the study, the resulting missing assessments would be considered missing at random as the impact is independent of the treatment or subject status.

For the main analyses of efficacy for primary, secondary and exploratory endpoints, no imputation will be done to impute data following discontinuation from study. In some cases missing items will be imputed (see individual scale sections below). For the primary and secondary endpoints, a sensitivity analysis using multiple imputation (MI) will be conducted using a jump to reference and a copy reference approach to assess the impact of the missing at random assumption of the analysis model used; refer to Section 4.2.7.1.

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

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.

4.2.7.1 Multiple Imputation Sensitivity Analysis

The primary analysis of the Y-BOCS total score based on a Mixed Model for Repeated Measures (MMRM) model assumes that data is missing at random (MAR) and subjects who discontinue study medication prematurely would have a response profile for the remainder of the Randomization phase similar to subjects completed the 12 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 a “copy reference” approach). These analyses use a “defacto” estimand and are based on the methods described in Carpenter (2) and implemented in the SAS Macros provided by the DIA Working Group section of the missingdata.org.uk web site.

Mean changes from baseline in Y-BOCS total 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 is 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. This is done with the 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 MNAR (missing not at random); part of the model is used to impute values for trailing missing values (e.g. after subject withdrawal 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 this would be using the mean response distribution after withdrawal and for the "copy reference" the whole distribution both before and after withdrawal is assumed to come from the placebo arm. The placebo arm subjects will use the profile under MAR.

For each analysis, the imputed data will consist of 1000 imputed data sets (using a MCMC chain length of 100). Both the parameter-estimation model and the imputation model will include treatment, visit (Weeks 4, 8, and 12), and the treatment-by-visit interaction as fixed effects, and the baseline Y-BOCS total score and visit by baseline interaction terms as covariates.

Each imputed data set will be analyzed using the following method. Change from baseline for the Y-BOCS total score at Week 12 will be based on observed and imputed data. The model will be an analysis of covariance model (ANCOVA) with baseline Y- BOCS total score as a covariate. Treatment group comparison at Week 12 will be based on least square mean (LSMean) difference between Troriluzole and placebo estimated by the analysis model in each of the imputed data sets. Results from the analysis of each imputed data set (LSMean treatment differences and their standard errors), will be combined using Rubin's imputation rules to produce a pooled LSMean estimate of treatment difference, it's 95% confidence interval, and a pooled P-value for the test of null hypothesis of no treatment effect.

The same method detailed above will be employed for the secondary endpoints: SDS, CGI-S, Y-BOCS obsessive symptomatology. For the secondary endpoints, the parameter-estimation model and the imputation model will include the baseline total score of each endpoint, and visit by baseline interaction terms as covariates.

4.2.8 Visit Windows

The protocol-specified visit window is ± 2 days during the Randomization Phase and ± 7 days during the Extension Phase of the study, however, analysis windows will be continuous to include all data. Refer to Table 1 for details on the protocol-specified day of evaluation and associated visit windows used for efficacy and safety analyses and Table 2 and Table 3 (in Appendix 2) for the details on the schedule of assessments during the Randomization Phase and Extension Phase, respectively. The CRF contains a page where all assessments and their acceptable windows will be scheduled at the baseline visit. The baseline visit is expected to be the date of randomization and first dose of randomized treatment.

Table 1: On-Treatment Analysis Windows for Visit

Evaluation	Protocol-Specified Day	Analysis-Specified Interval
Randomization Phase Visit		
Week 4	Day 28	Day 2-42
Week 8	Day 56	Day 43-70
Week 12	Day 84	Day 71-98
Extension Phase Visit**		
Ext. Week 4	Ext. Day 28	Day 2-42
Ext. Week 8	Ext. Day 56	Day 43-70
Ext, Week 12	Ext. Day 84	Day 71-126
Ext. Week 24	Ext. Day 168	Day 127-210
Ext. Week 36	Ext. Day 252	Day 211-294
Ext. Week 48	Ext. Day 336	Day 295-343

Note: Baseline assesment for the Randomization Phase will be defined as the last available assessment on or before first day of Randomization Phase study drug. For the summaries of the Extension Phase, baseline will be the last available assessment on or before first day of Extension Phase study drug. For the safety summaries of the combined troriluzole exposure, baseline will be the last available assessment on or before first day of DB or OL troriluzole . **Relative to First day of Extension Phase

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

Note that in some cases, endpoints are not measured at each of the visits listed in the above table. For example, the BABS and ECG is not measured at Visit Week 4 or Visit Week 8 during the Randomization Phase. These endpoints will still use the above visit definitions but will only be summarized or analyzed at scheduled visits and at a defined “LOCF” endpoint for the Randomization Phase (i.e. last available on-treatment assessment in the Randomization Phase).

4.3 Analysis Periods and Study Phases

Analysis periods and study phases are defined as follows:

- Screening Phase: will include all assessments on or before the first day of study drug.
- On-Treatment in Randomization Phase (Efficacy Assessments): "on-treatment" in the Randomization Phase will include all efficacy assessments after first day of study drug and up to the first day of Extension Phase study drug (for subjects entering extension) or last day of Randomization Phase study drug + 2 days (which ever is earliest)
- On-Treatment in Randomization Phase (Safety Assessments): except where indicated, "on-treatment" in the Randomization Phase will include all assessments after first day of study drug and up to the first day of Extension Phase study drug (for subjects entering extension) or last day of Randomization Phase study drug + 30 days (which ever is earliest). For AEs, “on-treatment” in the Randomization Phase will include first day of dosing of Randomization Phase study drug up to last day of Randomization Phase study drug + 30 days for those not going into the Extension Phase or up to the day before the first day of dosing of Extension Phase medication for those subjects who do.

- Extension Phase (Efficacy Assessments): will include all efficacy assessments after the first day of Extension Phase study drug to the last day of study drug + 2 days. Baseline for the extension will be the last available assessment on or before the first day of Extension Phase medication.

- Extension Phase (Safety Assessments): will include all assessments after the first day of Extension Phase study drug to the last day of study drug + 30 days. Baseline for the extension will be the last available assessment on or before the first day of Extension Phase medication.

- On-Treatment in Study (Efficacy Assessments): except where indicated, "on-treatment" for any Troriluzole treatment (combined Randomization Phase and Extension Phase) will include all efficacy assessments after first day of Troriluzole and up to the last day of Troriluzole treatment + 2 days.

- On-Treatment in Study (Safety Assessments): except where indicated, "on-treatment" for any Troriluzole treatment (combined Randomization Phase and Extension Phase) will include all assessments after first day of Troriluzole and up to the last day of Troriluzole treatment + 30 days. For AE summaries for any Troriluzole treatment (combined Randomization Phase and Extension Phase), On-treatment (Troriluzole) will include AEs with start date on or after first day of dosing with Troriluzole up to last day of dosing with Troriluzole + 30 days.

- Treatment-emergent AEs (TEAEs): AEs with a start date on or after the first dose of study drug and prior to 30 days after the last dose of study drug. For the Randomization Phase, TEAEs will be assessed from the date of first treatment until: 1) the first day of the Extension Phase or 2) if the subject did not continue into the Extension Phase, 30 days after the last dose of study drug. For the Extension Phase, TEAEs will be assessed from the first dose during the Extension Phase until 30 days after the last dose of study drug.

4.4 Planned Analyses

An initial analysis of the data will be conducted after the last subject completes their Week 12 visit, or discontinues from the Randomization Phase. The study will be unblinded and will include all data from the double-blind Randomization Phase of the study, but will not include summaries of efficacy data accumulated from the Extension Phase.

Interim looks at data from the Extension Phase may be performed after the completion and unblinding of the Randomization Phase, particularly for safety purposes.

A final analysis of the study will be completed after the last subject completes their last study visit. This will summarize all efficacy data collected in the open-label Extension Phase as well as all safety, laboratory, and other data collect though the entire study. Change from baseline will be based on the Extension Baseline, unless otherwise specified.

Additional analyses may be conducted during the open label Extension Phase of the study to support regulatory and administrative requirements.

4.5 Subject Disposition

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

- Number of subjects enrolled and signed informed consent form
- Number of enrolled subjects excluded from the study and reason for exclusion
- Number of subjects randomized
- Number of subjects treated in Randomization Phase
- Number of mITT subjects in 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 subjects treated in Extension Phase
- Number of mITT subjects in Extension Phase
- Number of subjects who completed the Extension Phase
- Number of subjects who withdrew from the Extension Phase and reasons for withdrawal

A special form will collect information on whether a visit was impacted by the COVID-19 pandemic crisis and will be used for all visits on or after March 1, 2020. This form will collect the specific impact on the visit (i.e. missed visit, in-person visit not all assessments completed, in-person visit occurring earlier or delayed, remote visit video, and remote visit telephone) as well as how COVID-19 impacted on a given visit (i.e. subject diagnosed/quarantined, site closure/restricted access, subject unwilling/unable to come to site, other). If the subject is terminating the study prematurely the CRF will also collect whether the reason was related to COVID-19 (yes or no).

An additional summary will present the disposition of subjects whose discontinuation was indicated as being related to the COVID-19 pandemic crisis on the COVID-19 CRF.

For the All Randomized Subjects and Randomized Subjects Active During COVID-19, the impact of the COVID-19 pandemic crisis on the study will be summarized by presenting (by treatment and overall):

- number and percentage of randomized subjects active in the double-blind phase during the pandemic crisis as referenced by having at least one visit in the double-blind phase on or after the March 13, 2020 National Emergency declaration.

- number and percentage of randomized subjects reporting (on the CRF) at least one impact on a given visit (missed visit, in-person visit not all assessments completed, in-person visit occurring earlier or delayed, remote visit video, and remote visit telephone) a subject will be included in each category they reported in the period.

    - Summarize at any time in double-blind phase

    - Summarize by week (analysis window)

- number and percentage of randomized subjects reporting (on the CRF) how COVID-19 impacted on a given visit (diagnosed/quarantined, site closure/restricted access, subject unwilling/unable to come to site, other) a subject will be included in each category they reported in the period.

    - Summarize at any time in double-blind phase

    - Summarize by week (analysis window)

- Time of first visit impacted by COVID-19 pandemic crisis (based on the first visit date indicated as impacted from the COVID-19 CRF)

- Duration of time in double-bind phase impacted by COVID-19 pandemic crisis (duration = date of last visit impacted – date of first visit impacted + 1) (based on the first visit date indicated as impacted from the COVID-19 CRF)

A by-subject listing of study completion information for both the Randomization and Extensions Phases, including the reason for withdrawal and impact due to COVID-19, if applicable, will be presented.

A listing of all subjects with at least one visit impacted by COVID-19 pandemic crisis with the impact and how the impact was related to COVID-19.

4.6 Demographic and Baseline Characteristics

Tabulations of demographic and baseline characteristics, including disease and medical history, current standard of care treatment, Mini Mental State Exam (MMSE), Borderline Personality Disorder Module (BPD), and baseline Y-BOCS total score (continuous and categorized as ≤ median or > median), will be made for all treated subjects. For demographics only, a separate set of tabulations will be made for subjects enrolled but not treated. Demographic and baseline characteristics for the treated subjects will be summarized by treatment group and for all treatment groups combined.

In addition to all Treated subjects additional summaries will be provided for:

- Completing or discontinuing the double-blind randomization phase before March 13, 2020

- Still active in the double-blind randomization phase on or after March 13, 2020

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

4.7 Efficacy Evaluation

Unless otherwise noted, all efficacy analyses will be conducted using the mITT population as well as any specified populations for sensitivity analyses described below. 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 for most endpoints.

4.7.1 Primary Endpoint

4.7.1.1 Yale-Brown Obsessive Compulsive Scale (Y-BOCS)

The Y-BOCS is a clinician-administered scale used extensively in research and clinical practice to both rate severity of OCD and to monitor improvement during treatment. It is designed to rate the severity of obsessions and compulsions as well as the type of symptoms in patients with OCD. The scale consists of 10 items, the first 5 items assess obsessions and the last 5 items assess compulsions. Subscale scores can be calculated for obsessions and compulsion, each on a scale of 0 – 20. Each subscale will be calculated providing there is no more the 1 item missing as the mean of the available items times 5. These are summed to create a total score ranging from 0 – 40 to indicate overall severity. The Y-BOCS Symptom Checklist will be used as an aid for identifying current symptoms. Items 1b, 6b, and 11 will be shown in listings, but not included in the total score.

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 (when added to a standard of care therapy) if taken as directed, a de jure or 'hypothetical' efficacy estimand. The primary endpoint will be the change from baseline in the Y-BOCS total score Troriluzole relative to placebo, at week 12 of the Randomization Phase. This treatment effect will be summarized as the difference in change from baseline in the Y-BOCS total score between the Troriluzole and placebo groups. All analyses and summaries will only consider assessments done by the clinician with the subject in person except where indicated in sensitivity analyses.

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 including COVID-19 related issues). 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 Y-BOCS total score through week 12 will be analyzed using the mITT set via a MMRM analysis model. The model will include fixed effects for treatment, visit, and the treatment-by-visit interaction, and the baseline total Y-BOCS score and the interaction of visit and baseline total Y-BOCS score will be included as a covariates. Repeated measurements are made on each subject. The covariance structure for within-subject error ("R" Matrix) will be initially specified as unstructured. In the case that the model fails to converge, a Huynh-Feldt error structure will be attempted, 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.

LSMeans for the change from baseline for each treatment group will be derived for Week 4, Week 8, and Week 12. These will be presented with degrees of freedom, standard errors (SEs), and two-sided 95% confidence intervals (CIs). The difference in change from baseline between the two treatment groups will also be derived for the same time points. These will be presented with degrees of freedom, SEs, two sided 95% CIs, and p-values.

Descriptive statistics for the total Y-BOCS score and subscale total scores, and the change from baseline in the total Y-BOCS score and subscale total scores will be presented by visit for the Randomization and Extension Phase, separately. In addition the Y-BOCS total score and change from baseline (from randomization) will be summarized with Descriptive statistics for the entire study period for those subjects randomized to Troriluzole.

The following sensitivity analyses will be conducted to support the principal analysis for the Randomization Phase only:

- To address the potential impact of COVID-19 to the primary analysis results, the primary MMRM analysis will be repeated excluding assessments done on or after the start of the COVID-19 crisis (March 13, 2020 date of declared national emergency will be used). All other aspects of the model and presentation are similar as described above for the primary analysis.

- The MAR assumption for the MMRM analysis specified for the primary endpoint will be assessed by multiple imputation analyses using referenced-based imputation methods, please Section 4.2.7.1.

- A responder analysis with "response" defined as an improvement of at least 35% of the total Y-BOCS score (Week 12 compared to baseline); this analysis will treat non-completers as failures (NC=F). The data will be evaluated with a Cochran-Mantel Haenszel test, stratified by baseline severity (total Y-BOCS score, categorized as ≤ median or > median), and tested at a two-sided alpha level of 0.05. Success rates for each treatment 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).

- To address the potential impact of COVID-19 to the primary analysis results, a fixed effect term will be added as a covariate to indicate whether the subject had at least one visit in the double-blind phase impacted by COVID-19 (based on March 13, 2020 date of declared national emergency). The fixed effect for patients active during the COVID pandemic will take the same value for visits before the COVID pandemic and visits during/after the COVID pandemic (i.e. this is a patient-based covariate, not a visit-based covariate). All other aspects of the model and presentation are similar as described above for the primary analysis.

- To address the potential impact of COVID-19 limiting the in-person assessment of YBOCS to the primary analysis results the primary MMRM analysis will be repeated to include any remote video assessments of the YBOCS that sites were able to attain. All other aspects of the model and presentation are similar as described above for the primary analysis.

  Although standard psychometric validation of video versus live YBOCS assessment is not feasible with this current study data, blinded change from baseline in total score and mean score at the Week 8 visit will be assessed and compared to mean and change scores among the patients assessed live at this visit. If these mean scores and associated standard deviations are similar then the above sensitivity analysis will be implemented.

Subgroup Analyses

Descriptive statistics for the total Y-BOCS score, and the change from baseline, will be tabulated by visit for the Randomization Phase for the subgroups defined by:

- Baseline severity (total Y-BOCS, categorized as ≤ median and > median)

- Sex

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

- Standard of Care Concomitant OCD Treatment (citalopram, escitalopram, fluoxetine, fluvoxamine, paroxetine, sertraline, clomipramine, venlafaxine, desvenlafaxine)

- Double-blind participation during COVID-19 pandemic crisis (i.e. before and on or after March 13, 2020)

4.7.2 Secondary Efficacy Endpoints

The estimand for the secondary endpoints will be the effect due to the initially randomized treatments (when added to the standard of care therapy) if taken as directed, a “de jure” or “hypothetical” efficacy estimand.

4.7.2.1 Sheehan Disability Scale (SDS)

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 OCD, then the total score will be missing. All items must be non-missing for the SDS total to be calculated.

The secondary endpoint, change from baseline in the SDS total score from baseline to the end of the Randomization Phase, will be estimated using the mITT population comprised of subjects that take the drug, are on a stable standard of care therapy, and have at least one post-baseline efficacy measurement. This treatment effect will be summarized as the difference in change from baseline in the SDS total score between the Troriluzole and placebo groups. All analyses and summaries will only consider assessments done by the clinician with the subject in person.

The change from baseline in the SDS total score will be analyzed using the mITT set via a MMRM analysis model. The model will include fixed effects for treatment, visit, and the treatment-by-visit interaction, and the baseline total SDS and visit by baseline total SDS interaction as covariates. Repeated measurements are made on each subject. The covariance structure (SAS "R" Matrix) will be initially specified as unstructured. In the case that the model fails to converge, a Huynh-Feldt error structure will be attempted, followed by an AR(1) structure. Error degrees of freedom will be calculated in the same manner as the primary endpoint, noted in Section 4.6.1.1.

LSMeans for the change from baseline for each treatment group will be derived for Week 4, Week 8, and Week 12. These will be presented with degrees of freedom, SEs, and two-sided 95% CIs. The difference in change from baseline between the two treatment groups will also be derived for the same time points. These will be presented with degrees of freedom, SEs, two sided 95% CIs, and p-values.

The following sensitivity analyses will be conducted to support the secondary analysis for the Randomization Phase only:

- To address the potential impact of COVID-19 to the primary analysis results, the primary MMRM analysis will be repeated excluding assessments done on or after the start of the COVID-19 crisis (March 13, 2020 date of declared national emergency will be used). All other aspects of the model and presentation are similar as described above for the primary analysis.

- The MAR assumption for the MMRM analysis will be assessed by multiple imputation analyses using a referenced-based imputation methods; refer to Section 4.2.7.1.

- To address the potential impact of COVID-19 to the primary analysis results, a fixed effect term will be added as a covariate to indicate whether the subject had at least one visit in the double-blind phase impacted by COVID-19 (based on March 13, 2020 date of declared national emergency). The fixed effect for patients active during the COVID pandemic will take the same value for visits before the COVID pandemic and visits during/after the COVID pandemic (i.e. this is a patient-based covariate, not a visit-based covariate). All other aspects of the model and presentation are similar as described above for the primary analysis.

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 Phase, separately.

Subgroup Analyses

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

- Baseline severity (total Y-BOCS, categorized as ≤ median and >median)
- Sex
- Race (Asian, Black or African American, White, and all other races combined))
- Standard of Care Concomitant OCD Treatment (citalopram, escitalopram, fluoxetine, fluvoxamine, paroxetine, sertraline, clomipramine, venlafaxine, desvenlafaxine)
- Double-blind participation during COVID-19 pandemic crisis (i.e. before and on or after March 13, 2020)

4.7.2.2 Clinical Global Impression of Severity (CGI-S)

The secondary endpoint, the clinician global impression of severity via the CGI-S will be summarized descriptively by post-baseline visit for the Randomization and Extension Phase, separately, including the number and percentage of subjects in each category:

- Normal, not at all ill
- Borderline ill
- Mildly ill
- Moderately ill
- Markedly ill
- Severely ill
- Among the most extremely ill patients

The change from baseline in the CGI-S score will be analyzed using the mITT set via a MMRM analysis model. In order to utilize this model, the CGI-S scores will first 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”). The model will include fixed effects for treatment, visit, and the treatment-by-visit interaction, and the baseline CGI-S score and visit by baseline CGI-S score interaction as covariates. Repeated measurements are made on each subject. The covariance structure (SAS “R” Matrix) will be initially specified as unstructured. In the case that the model fails to converge, a Huynh-Feldt error structure will be attempted, followed by an AR(1) structure. Error degrees of freedom will be calculated in the same manner as the primary endpoint, noted in Section 4.6.1.1. All analyses and summaries will only consider assessments done by the clinician with the subject in person.

The following sensitivity analyses will be conducted to support the principal analysis for the Randomization Phase only:

- To address the potential impact of COVID-19 to the primary analysis results, the primary MMRM analysis will be repeated excluding assessments done on or after the start of the COVID-19 crisis (March 13, 2020 date of declared national emergency will be used). All other aspects of the model and presentation are similar as described above for the primary analysis.

- The MAR assumption for the MMRM analysis specified for the primary endpoint will be assessed by a multiple imputation analyses using referenced-based imputation methods; refer to Section 4.2.7.1.

- To address the potential impact of COVID-19 to the primary analysis results, a fixed effect term will be added as a covariate to indicate whether the subject had at least one visit in the double-blind phase impacted by COVID-19 (based on March 13, 2020 date of declared national emergency). The fixed effect for patients active during the COVID pandemic will take the same value for visits before the COVID pandemic and visits during/after the COVID pandemic (i.e. this is a patient-based covariate, not a visit-based covariate). All other aspects of the model and presentation are similar as described above for the primary analysis.

Descriptive statistics for the CGI-S scores (both categorical and numeric) and the change from baseline in the CGI-S scores will be presented by visit for the Randomization and Extension Phase, separately.

Subgroup Analyses

Descriptive statistics for the CGI-S score will be tabulated by visit for the Randomization Phase, separately, for the subgroups defined by:

- Baseline severity (total Y-BOCS, categorized as ≤ median and >median)

- Sex

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

- Standard of Care Concomitant OCD Treatment (citalopram, escitalopram, fluoxetine, fluvoxamine, paroxetine, sertraline, clomipramine, venlafaxine, desvenlafaxine)

- Double-blind participation during COVID-19 pandemic crisis (i.e. before and on or after March 13, 2020)

4.7.2.3 Yale-Brown Obsessive Compulsive Scale (Y-BOCS) – Obsessive Symptomatology

The obsessive symptomatology section of the Y-BOCS is a clinician-administered scale used extensively in research and clinical practice to rate severity of OCD obsessive symptoms and to monitor improvement during treatment. It is designed to rate the severity of obsessions and in

patients with OCD. The scale has 5 items to assess obsessions and produces a score ranging from 0 – 20.

The secondary endpoint, change from baseline in the Y-BOCS obsessive symptomatology score from baseline to the end of the randomization phase, will be estimated using the mITT population comprised of subjects that take the drug, are on a stable standard of care therapy, and have at least one post-baseline efficacy measurement. This treatment effect will be summarized as the difference in change from baseline in the Y-BOCS obsessive symptomatology score between the BVH-4157 and placebo groups. All analyses and summaries will only consider assessments done by the clinician with the subject in person except where indicated in sensitivity analyses.

The change from baseline in the Y-BOCS obsessive symptomatology score will be analyzed using the mITT set via a MMRM analysis model. The model will include fixed effects for treatment, visits, and the treatment-by-visit interaction and the baseline Y-BOCS obsessive symptomatology score and visit by baseline Y-BOCS obsessive symptomology score will be included as a covariates. Repeated measurements are made on each subject. The covariance structure ("R" Matrix) will be initially specified as unstructured. In the case that the model fails to converge, a Huynh-Feldt error structure will be attempted, followed by an AR(1) structure. Error degrees of freedom will be calculated in the same manner as the primary endpoint, noted in Section 4.6.1.1.

LSMeans for the change from baseline for each treatment group will be derived for Week 4, Week 8, and Week 12. These will be presented with degrees of freedom, SEs, and two-sided 95% CIs. The difference in change from baseline between the two treatment groups will also be derived for the same time points. These will be presented with degrees of freedom, SEs, two sided 95% CIs, and p-values.

The following sensitivity analyses will be conducted to support the principal analysis for the Randomization Phase only:

- To address the potential impact of COVID-19 to the primary analysis results, the primary MMRM analysis will be repeated excluding assessments done on or after the start of the COVID-19 crisis (March 13, 2020 date of declared national emergency will be used). All other aspects of the model and presentation are similar as described above for the primary analysis.

- The MAR assumption for the MMRM analysis will be assessed by multiple imputation analyses using a referenced-based imputation methods refer to Section 4.2.7.1.

- To address the potential impact of COVID-19 to the primary analysis results, a fixed effect term will be added as a covariate to indicate whether the subject had at least one visit in the double-blind phase impacted by COVID-19. The fixed effect for patients active during the COVID pandemic will take the same value for visits before the COVID pandemic and visits during/after the COVID pandemic (i.e. this is a patient-based covariate, not a visit-based covariate). All other aspects of the model and presentation are similar as described above for the primary analysis.

- To address the potential impact of COVID-19 limiting the in-person assessment of YBOCS to the primary analysis results the primary MMRM analysis will be repeated to include any remote video assessments of the YBOCS that sites were able to attain. All other aspects of the model and presentation are similar as described above for the primary analysis.

Descriptive statistics for the Y-BOCS obsessive symptomatology score and the change from baseline in the Y-BOCS obsessive symptomatology score will be presented by visit for the Randomization and Extension Phase, separately.

Subgroup Analyses

Descriptive statistics for the Y-BOCS obsessive symptomatology score, and the change from baseline, will be tabulated by visit, for the Randomization Phase, for the subgroups defined by:

- Baseline severity (total Y-BOCS, categorized as ≤ median and >median)
- Sex
- Race (Asian, Black or African American, White, and all other races combined)
- Standard of Care Concomitant OCD Treatment (citalopram, escitalopram, fluoxetine, fluvoxamine, paroxetine, sertraline, clomipramine, venlafaxine, desvenlafaxine)
- Double-blind participation during COVID-19 pandemic crisis (i.e. before and on or after March 13, 2020)

4.7.3 Exploratory Efficacy Endpoints

All analyses and summaries will only consider assessments done by the clinician with the subject in person.

4.7.3.1 Quick Inventory of Depressive Symptomatology – Self Report (QIDS-SR)

The QIDS-SR is a self-report, 16 item questionnaire that subjects will use to rate symptoms of depression. Each item is rated on a scale from 0 to 3. For symptom domains that require more than one item, the highest score of the item relevant for each domain is taken. Total scores range from 0 to 27 and are obtained by adding the scores for each of nine symptom domains as follows:

For domain 1, sleep disturbance, take the highest score of items 1 through 4.

For domain 2, sad mood, use the score for item 5.

For domain 3, decrease/increase in appetite/weight, take the highest score of items 6 through 9.

For domain 4, concentration, use the score for item 10.

For domain 5, self-criticism, use the score for item 11.

For domain 6, suicidal ideation, use the score for item 12.

For domain 7, interest, use the score for item 13.

For domain 8, energy/fatigue, use the score for item 14.

For domain 9, psychomotor agitation/retardation, take the highest score of items 15 and 16.

For the total score, sum the scores for each of the 9 domains. All domains must have a non-missing score for the total to be calculated. Higher scores indicate higher levels of depression. Scores ranging from 0 through 5 indicate no depression. Scores ranging from 6 through 10 indicate mild depression. Scores ranging from 11 through 15 indicate moderate depression. Scores ranging from 16 through 20 indicate severe depression. Scores ranging from 21 through 27 indicate very severe depression.

Descriptive statistics (both continuous and using the categories described above) will be summarized by treatment and visit for the Randomization and Extension Phases, separately.

The exploratory endpoint, change from baseline in the QIDS-SR total score from baseline to the end of the randomization phase, will be estimated using the mITT population comprised of subjects that take the drug, are on a stable standard of care therapy, and have at least one post-baseline efficacy measurement. This treatment effect will be summarized as the difference in change from baseline in the QIDS-SR total score between the Troriluzole and placebo groups.

The change from baseline in the QIDS-SR total score will be analyzed using the mITT set via a MMRM analysis model. The model will include fixed effects for treatment, visit, and the treatment-by-visit interaction and the baseline QIDS-SR score and baseline QIDS-SR score by visit interaction will be included as covariates. Repeated measurements are made on each subject. The covariance structure ("R" Matrix) will be initially specified as unstructured. In the case that the model fails to converge, a Huynh-Feldt error structure will be attempted, followed by an AR(1) structure. Error degrees of freedom will be calculated in the same manner as the primary endpoint, noted in Section 4.6.1.1.

LSMeans for the change from baseline for each treatment group will be derived for Week 4, Week 8, and Week 12. These will be presented with degrees of freedom, SEs, and two-sided 95% CIs. The difference in change from baseline between the two treatment groups will also be derived for the same time points. These will be presented with degrees of freedom, SEs, two sided 95% CIs, and p-values.

4.7.3.2 Beck Anxiety Inventory (BAI)

The BAI is a 21 question multiple choice self-report questionnaire that subjects will use to rate symptoms of anxiety using a 4-point Likert Scale. As long as no more than 4 items are missing the total score will be calculated as the mean of available items times 21. Total BAI score ranges from 0 to 63 and higher scores indicate higher levels of anxiety symptoms.

Descriptive statistics will be summarized by treatment and visit for the Randomization and Extension Phases, separately.

The exploratory endpoint, change from baseline in the BAI total score from baseline to the end of the randomization phase, will be estimated using the mITT population comprised of subjects that take the drug, are on a stable standard of care therapy, and have at least one post-baseline efficacy measurement. This treatment effect will be summarized as the difference in change from baseline in the BAI total score between the Troriluzole and placebo groups.

The change from baseline in the BAI total score will be analyzed using the mITT set via a MMRM analysis model. The model will include fixed effects for treatment, visits, and the treatment-by-visit interaction and the baseline BAI score and baseline BAI by visit interaction will be included as covariates. The repeated measurements are made on each subject. The covariance structure ("R" Matrix) will be initially specified as unstructured. In the case that the model fails to converge, a Huynh-Feldt error structure will be attempted, followed by an AR(1) structure. Error degrees of freedom will be calculated in the same manner as the primary endpoint, noted in Section 4.6.1.1.

LSMeans for the change from baseline for each treatment group will be derived for Week 4, Week 8, and Week 12. These will be presented with degrees of freedom, SEs, and two-sided 95% CIs. The difference in change from baseline between the two treatment groups will also be derived for the same time points. These will be presented with degrees of freedom, SEs, two sided 95% CIs, and p-values.

4.7.3.3 Brown Assessment of Beliefs Scale (BABS)

The BABS is a semi-structured, rater-administered scale that assesses insight/delusionality both dimensionally (as a continuum of insight) and categorically (i.e., dichotomously – for example, delusional versus non-delusional) regarding patient beliefs. These beliefs include the delusions as well as the beliefs that may underlie obsessional thinking. The BABS is a 7-item scale that assesses insight during the past week.

BABS items assess the person's conviction that their belief is accurate, perception of others' views of the belief, whether the person could be convinced that the belief is wrong, attempts to disprove the belief, insight (recognition that the belief has a psychiatric/psychological cause), and ideas/delusions of reference related to the belief. The first six items are summed to create a total score that ranges from 0 to 24 where higher scores indicate poorer insight. As long as no more than one item is missing the total score will be calculated as the mean of the available items time 6. Item 7 is not included in the total score, because referential thinking is characteristic of some but not all disorders.

Descriptive statistics will be summarized for the total score and item 7, individually, by treatment and visit for the Randomization and Extension Phases, separately.

The exploratory endpoint, change from baseline in the BABS total score from baseline to the end of the randomization phase, will be estimated using the mITT population comprised of subjects that take the drug, are on a stable standard of care therapy, and have at least one post-

baseline efficacy measurement. This treatment effect will be summarized as the difference in change from baseline in the BABS total score between the Troriluzole and placebo groups.

The change from baseline to Week 12 in the total BABS score (based on observed cases at Week 12 and LOCF endpoint) will be analyzed with a univariate ANCOVA. The model will contain fixed effects for treatment and baseline Y-BOCS score (categorized as ≤median or >median). The baseline BABS score will be entered into the model as a covariate. The covariate adjusted difference in the change from baseline between the drug tested and placebo groups will be tested at a two-sided alpha level of 0.05. Covariate adjusted LSMeans, model based SEs, degrees of freedom, and two-sided 95% CIs will be presented for each treatment group and for the difference between the treatment groups.

4.8 Pharmacokinetic Evaluations

All PK analyses will be conducted using the treated population.

A PK sample will be collected at Weeks 4, 8, and 12 of the Randomization Phase. Additionally, PK samples should be drawn if there are any SAEs that could possibly be drug related or severe AEs that could be drug related. Date and time of doses on the day of visits and day prior will be collected in the case report forms along with the time of last meal. Subjects who are able to schedule a morning visit for Week 4 and Week 8 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 be displayed in listings for the Randomization Phase.

4.9 Safety and Other Analyses

Safety and other exploratory analyses will be conducted on the Treated Subjects Population. All safety and other data will be listed for the Screening, Randomization and Extension Phase, together.

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

Due to the COVID-19 pandemic crisis, approximately the last 50 of the 248 subjects randomized might not be able to be seen for all their visits at the site in person. In order to assess the impact of the COVID-19 pandemic crisis on safety results, select safety summaries will be repeated separating subjects by whether they were active (on-treatment) in the double-blind Randomization phase on or after March 13, 2020 or only active prior to this date.

Other exploratory data, including plasma BDNF and proBDNF, will be listed.

4.9.1 Extent of Exposure and Compliance to Study Treatment

Subjects will receive placebo (QD) or Troriluzole (200 mg QD, after four weeks at 140 mg QD) during the Randomization Phase. Subjects who complete 12 weeks of treatment in the Randomization Phase may be eligible for the open-label Extension Phase.

For the Randomization Phase, the extent of subject exposure (including missed dose days) will be quantified as the number of days on study drug (placebo or Troriluzole) and measured from the time the subject received the first dose until the time the subject received the last dose, either at the end of 12 weeks of treatment or withdrawal from the Randomization Phase (i.e. total days on randomized study medication = last day of double-blind randomized study medication -– Day 1 of double blind randomized study medication + 1).

For the Extension Phase, the extent of subject exposure (including missed dose days) to Troriluzole will be quantified as the number of days on study drug (Troriluzole) in the Extension Phase and measured from the time the subject received the first dose of Troriluzole in the Extension Phase until the time the subject received the last dose, either at the end of the study or withdrawal from the Extension Phase. (i.e. total days on Troriluzole in Extension Phase = last day of Troriluzole study medication – Day 1 of Troriluzole study medication in Extension Phase + 1).

For the combined phase data, the extent of subject exposure (including missed dose days) to Troriluzole during the entire study will be quantified as the number of days on study drug (Troriluzole) and measured from the time the subject received the first dose of Troriluzole until the time the subject received the last dose of Troriluzole. (i.e. total days exposure on Troriluzole during the study = last day of Troriluzole study medication – Day 1 of Troriluzole study medication + 1).

For both the Randomization Phase, Extension Phase and combined phases, extent of subject exposure will also be calculated only including days where number of tablets taken was > 0 (total days on study medication).

For the Randomization Phase, the number of subjects on-treatment and their average daily dose (including minimum and maximum) will be summarized by week (7-day intervals) for the first 12 weeks as well as overall for the randomization phase. For subjects on placebo the number of tablets will be summarized.

For the Randomization Phase, the extent of study medication exposure will also be summarized by a Kaplan-Meier plot presenting proportion of patient still on Randomization Phase study medication (including days of missed dose). Subjects continuing in the Extension Phase should be censored on last day of Randomized Phase study medication.

For the Extension Phase and the entire study, the number of subjects on-treatment and their average daily dose (including minimum and maximum) will be summarized by month (30-day intervals) for the as well as overall for the Extension Phase and the entire study.

Additionally, percent (%) compliance will be calculated and summarized as follows:

- Randomization Phase % compliance = total days on study medication during the Randomization Phase / (last dose date in Randomization Phase – first dose date + 1) × 100.

- Extension Phase % compliance = total days on Troriluzole in Extension Phase / (last dose date – first dose date in Extension Phase + 1) × 100.

- Overall % compliance = total days on Troriluzole during the study / [(last Troriluzole dose date – first Troriluzole dose date + 1) – (number of days with a dosing break in dosing between the Randomization and Extension Phases)].

Summaries on the Randomization Phase will be repeated for Treated subjects completing (or discontinuing) double-blind treatment prior to versus on or after start of COVID-19 pandemic crisis (March 13, 2020).

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

4.9.2 Adverse Events

AEs 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 test 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 just overall for subjects after at least one dose of Troriluzole in any phase.

- TEAEs,

- TEAEs related to treatment,

- Treatment-emergent SAEs,

- TEAEs leading to discontinuation of study treatment,

- TEAEs by highest 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 contribute only once (i.e., the most related occurrence or the most severe occurrence) to each of the incidence rates in the descriptive analysis, regardless of the number of episodes. No formal hypothesis-testing analysis of AEs incidence rates will be performed.

The following summaries on the Randomization Phase will be repeated for Treated subjects completing (or discontinuing) double-blind treatment prior to versus on or after start of COVID-19 pandemic crisis (March 13, 2020).

- TEAEs,
- TEAEs leading to discontinuation of study treatment,

All AEs occurring pre-treatment and during the entire study will be listed. Additional listings will be provided including deaths, SAEs, and AEs leading to discontinuation of study drug. Listings will indicate whether subject was active during COVID-19 pandemic crisis and the date of first visit impacted.

4.9.3 Laboratory Data

Clinical laboratory evaluations include:

- Hematology: hemoglobin, hematocrit, platelets, complete blood count with differential and absolute neutrophil count
- Serum Chemistry: sodium, potassium, chloride, calcium, alanine aminotransferase (ALT), aspartate aminotransferase (AST), lactate dehydrogenase (LDH), alkaline phosphatase (ALP), gamma-glutamyl transferase (GGT), phosphorous, bicarbonate, creatine phosphokinase (CPK), total protein, albumin, total bilirubin (if greater than 2 mg/dl bilirubin will be fractionated), glucose, creatinine, blood urine nitrogen (BUN), and uric acid.
- Urinalysis: macroscopic examination, pH, specific gravity, ketones, nitrites, urobilinogen, leukocyte esterase, protein, creatinine, glucose, occult blood, and microscopic examination (if blood, protein, or leukocytes are positive)
- Serum pregnancy test will be conducted at screening. Urine pregnancy tests will be performed prior to dosing at baseline and at scheduled visits, at study visits where lab assessments are not performed, or at the discretion of the Investigator. Subjects will be provided with urine pregnancy tests to take in between every 3-month office visit during the Extension Phase.

Clinical laboratory values will be expressed using conventional (US) and standard international (SI) units with normal ranges provided. The numeric portion of the results will be used in cases where a value below detectable limit is provided. In the event of repeat values within the same analysis visit, if any measurement has an abnormal result, that measurement will be used for the presentation in by-visit tables. If none of the measurements are 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 and considered for evaluation of potential drug induced liver injury (DILI) or abnormalities.

Due to the COVID-19 pandemic crisis some sites might utilize local laboratories for some visits. These local laboratory values will be combined with the centralized laboratory values by normalizing them to the central laboratory using methods described in Chuang-Stein 1992.

On-treatment laboratory abnormalities are those with an assessment date after the date/time of first dose of study drug and within 30 days after the last dose of study drug. For the Randomization Phase, treatment-emergent laboratory abnormalities will be assessed from the date of first dose of study drug until: 1) the first day of the Extension Phase or 2) if the subject did not continue into the Extension Phase, 30 days after the last dose of study drug. For the Extension Phase, treatment-emergent lab abnormalities will be assessed from the first dose during the Extension Phase until 30 days after the last dose of study drug.

The observed value and change from baseline will be summarized for each continuous laboratory parameter for the Randomization and Extension Phases, separately, in both conventional and SI units.

Clinical laboratory values will be graded according to CTCAE version 5.0 if criteria for test available, otherwise according to DAIDS version 2.1 criteria, if available for test. The number and percent of subjects with at least one on-treatment lab assessment by grade will be summarized for each treatment (regardless of baseline) for the Randomization and Extension Phases, separately, as well as just overall for subjects after at least one dose of Troriluzole in any phase. In addition, the shift from baseline for laboratory abnormalities (based on either normal limits or grading, where available) will be tabulated for the Randomization and Extension Phases, separately, as well as just overall for subjects after at least one dose of Troriluzole in any phase. The shift tables will include only subjects with a baseline assessment and at least one on-treatment assessment.

For the liver function tests, AST, ALT, ALP, and TBILI, the shift from baseline will be presented by visit and to the maximum observed abnormality for the Randomization and Extension Phases, separately, as well as just overall for subjects after at least one dose of troriluzole in any phase. Baseline in the shift tables will be defined as per Section 4.2.8; the last available assessment on or before the first day of study treatment. Shift tables will only include treated subjects with an assessment for the specific test of interest at baseline and on treatment. The following categories will be used to summarize the shift from baseline based on the upper limit of normal (ULN) range for ALT and AST:

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

The following categories will be used to summarize the shift from baseline based on the ULN range for alkaline phosphatase:

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

The following categories will be used to summarize the shift from baseline based on the ULN range for BILI:

- <=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 1) randomization phase and 2) all 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.”

All laboratory data will be presented in data listings. Additional listings will be presented for all abnormal laboratory values considers potentially clinically significant (Grade 3 or 4) for the Randomization and Extension Phases, together. Subjects with a maximum value of ALT or AST >3x ULN and a maximum total bilirubin value >2x ULN observed at any point during the entire study will also be presented in a listing. Note that these abnormalities do not need to occur on concurrent visits. Separate listing will also be provided for subjects meeting just the ALT, AST or total bilirubin criteria.

The following summary on the Randomization Phase will be repeated for Treated subjects completing (or discontinuing) double-blind treatment prior to versus on or after start of COVID-19 pandemic crisis (March 13, 2020).

- The number and percent of subjects with at least one on-treatment lab assessment by grade will be summarized for each treatment (regardless of baseline) for the Randomization Phase

4.9.4 Physical Examinations

Results of 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. The summary will be based on the data after the date/time of first dose of study drug/Extension Phase and within 30 days after the last dose of study drug.

In addition, the number and percentage of subjects with at least one post-treatment vital sign measurement meeting any of these criteria:

- 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 percent of subjects meeting these criteria will be summarized for the Randomization and Extension Phase, separately as well as the overall Troriluzole treated exposure. A subject listing will provide for each vital sign measure listing any subject meeting the criteria with a complete list of the specific test assessments with the abnormal ones flagged.

The criteria summarized for the Randomization Phase will also be repeated for Treated subjects completing (or discontinuing) double-blind treatment prior to versus on or after start of COVID-19 pandemic crisis (March 13, 2020).

4.9.6 Electrocardiogram

ECG readings from the local machines at the sites were entered in the database but the electronic read was sent to a central coding group iCardiac for centralized reading and interpretation. The centralized reading will be the basis for all ECG safety summaries and reporting.

Descriptive statistics for ECG interval data (e.g., QRS, PR, QT, QTcF), and ventricular heart rate will also be reported by visit for the Randomization and Extension Phases, separately. The summary will be based on the data after the date/time of first dose of study drug/Extension Phase and within 30 days after the last dose of study drug.

In addition, the number and percentages with at least one post-treatment QTcF > 450 ms, >480 ms, and >500 ms will be summarized for the Randomization and Extension Phase, separately, as well as the overall Troriluzole treated exposure. Similar number and percentages will be

presented for subjects with at least one post-baseline QTcF Change from baseline >= 30 ms to < 60ms and those with at least one change from baseline >= 60 ms. A subject listing will provide for each QTcF sign measure listing any subject meeting the criteria with a complete list of the QTcF assessments with the abnormal ones flagged.

The criteria summarized for the Randomization Phase will also be repeated for Treated subjects completing (or discontinuing) double-blind treatment prior to versus on or after start of COVID-19 pandemic crisis (March 13, 2020).

4.9.7 Concomitant Medications

Concomitant medications will be coded using the WHO-DD, Sep2016. Results will be tabulated by Anatomic Therapeutic Class (ATC) and PT during the Randomization and Extension Phases, separately.

Concomitant OCD medications will also be shown in a summary table, along with mean, minimum and maximum dosage.

Unless the start date of the medication is after the last study drug dose date, or the end date of the medication is prior to the start date of the study drug, the medication will be considered 'concomitant.'

4.9.8 Sheehan-Suicidality Tracking Scale (S-STS)

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 based on the scoring used for the 2017 version of scale used in this study. In addition, a Behavior subscale score based on the 2019 version of the scoring will also be calculated where the score for a 'Yes' item 20 is 56 and the scores for 18, 19, 21 and 22 are used from the prior visit if applicable.

- Total score: Sum of the ideation and behavior subscale scores. A Total score will be calculated based on both 2017 version and 2019 version

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 at Maximum score in the Randomization and Extension Phases, presented separately. Both versions of the Total score will be summarized. All S-STS assessments will be included in summaries regardless of being completed at study site or remote.

4.9.9 Plasma BDNF and proBDNF

The analysis of the change from baseline to end of phase (either Randomization or Extension) will be analyzed using with a univariate ANCOVA. The model will contain fixed effects for treatment and baseline total Y-BOCS score (categorized as ≤ median or > median). The baseline plasma BDNF (or proBDNF) will be entered into the model as a covariate. The covariate adjusted difference in the change from baseline between the drug treated and placebo groups will be tested at a two-sided alpha of 0.05. Covariate adjusted LSMeans, model based SEs, degrees of freedom, and two-sided 95% CIs will be presented for each treatment group and for the difference between the treatment groups.

Descriptive statistics for the observed value and change from baseline in plasma BDNF and proBDNF will be presented at baseline, Week 12, and Extension Week 48 (if applicable, and separately).

4.9.10 Subjects Identified for Narratives

A safety narrative will be prepared for each subject who received at least one dose of troriluzole 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 drug; 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 (v3) 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 troriluzole treated subjects with the select events as described above.

5 REFERENCES

1. Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. K., … Keller, M. B. (2003). The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biological Psychiatry, 54,* 573-583.
2. Carpenter, James R., Roger, James H., and Kenward, Michael G. (2013). Analysis of longitudinal trials with protocol deviation: A Framework for Relevant, Accessible Assumptions, and Inference via Multiple Imputation. *Journal of Biopharmaceutical Statistics*, 6 (23), 1352 – 1371.
3. Chuang-Stein C. (1992). Summarizing laboratory data with different reference ranges in multi-center clinical trials. *Drug Information Journal,* 26, 77-84.

6 APPENDIX 1

Table 2: Schedule of Assessments – Randomization Phase

Visit	Screening[a]	Baseline	Week 2	Week 4	Week 6	Week 8	Week 12 or early term[b,c]	Week 2 Post Last Dose[c]
Day	-2 to -42	0		28		56	84	98
Eligibility Assessments								
Informed Consent	X							
Pharmacogenetic Informed Consent	X							
Inclusion/Exclusion	X	X						
MINI	X							
MMSE	X							
Borderline Personality Disorder Module (BPD Module)	X							
MGH-TRQ-OCD[d]	X							
Medical History	X							
Demographic Assessment	X							
Disease History	X							
SAFER Interview[e]	X							
Safety Assessments								
Adverse Event Assessment	X	X		X		X	X	X
Telephone Check-in[l] *Includes AE assessment and concomitant medication review*			X		X			
Laboratory Assessments including urinalysis[f]	X	X		X		X	X	
Serology[g]	X							
Pregnancy testing[h]	X	X		X		X	X	
Urine drug test[i]	X	X					X	
Physical Exam	X						X	
Physical Measurements	X						X	
Vital Signs	X	X		X		X	X	X
12-Lead ECG	X	X					X	
Concomitant Medication Review	X	X		X		X	X	X
Sheehan Suicidality Tracking Scale (STS)	X	X		X		X	X	X
Clinical Outcome Assessments								
Yale-Brown Obsessive Compulsive Scale (Y-BOCS)	X	X		X		X	X	
Clinical Global Impressions-Severity Scale (CGI-S)	X	X		X		X	X	
Sheehan Disability Scale (SDS)[m]	X	X		X		X	X	
Quick Inventory of Depressive Symptomatology-Self Report (QIDS-SR)	X	X		X		X	X	
Beck Anxiety Inventory (BAI)	X	X		X		X	X	
Brown Assessment of Beliefs (BABS)	X	X					X	
Biomarker and Other Assessments								
BDNF and proBDNF Blood Sample		X					X	
Pharmacokinetics Blood Sample[j]				X		X	X	
Pharmacogenomics Blood Sample		X					X	
Clinical Drug Supply								
Randomization		X						
Dispense Study Drug[k]		X		X		X	X[b, n]	

Drug Accountability				X		X	X	

*Visit Window is +/- 2 days during the Randomization Phase

[a] Screening window is minimum of 2 days to maximum of 42 days. Screening can be as short as 2 days as long as the subjects has been on at least 8 weeks of their current SOC OCD therapy at an adequate dose at Screening and at least 10 weeks of their current SOC OCD therapy (SSRI, clomipramine, venlafaxine or desvenlafaxine) at an adequate dose by the Baseline Visit.

[b] Study drug will be dispensed at Week 12 if subject is deemed eligible and agreed to participate in the Extension Phase. Subjects will not be allowed to transition to the Extension Phase until they have an in-person Week 12 visit.

[c] Every effort should be made to conduct the Week 12 visit and maintain the +/- 7 day window. However, due to concerns related to the COVID-19 pandemic, the Week12 visit window may be modified beyond the +/- 7 day window, in order to minimize any potential risks to study participant safety and to comply with governmental and local institutional guidance (e.g., study site has a policy that a clinical research visit must be delayed) note the following guidance. Under these circumstances, the last visit window may be extended up to 6 weeks (up to a maximum treatment duration of 18 weeks), but every attempt should be made to conduct the visit as close to the date the visit is due as possible. If the visit window is modified, and the investigator determines that a remote visit offers an acceptable risk-benefit approach and is appropriate for a particular subject, participants should be evaluated remotely for safety only (e.g., via phone) at the time of the scheduled Week 12 visit to perform and document appropriate safety assessments including the Sheehan Suicidality Tracking Scale (STS). If the remote visit requires laboratory testing, local labs must be able to be obtained, reviewed by the site and a redacted copy submitted to the study Medical Monitor. Study medication may be sent to the participant via tracked and certified courier. For any such cases, the investigator should discuss the specific circumstances of each case with the sponsor medical monitor: PPD PPD Cell PPD PPD Only for subjects NOT entering the Extension Phase. Subjects entering the Extension Phase will not require the 2-week post dose visit.

[d] The subject must have an inadequate response to the standard of care treatment, as defined in the protocol. The MGH-TRQ-OCD will be used to capture information on past treatments.

[e] 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.

[f] Laboratory assessments are not required to be fasting.

[g] HBsAg, HCV, HIV antibody, RPR

[h] Serum pregnancy test (b-hcg) conducted at screening. Urine pregnancy test conducted at subsequent visits. To be done prior to dosing at baseline. The site may test a patient at any time if pregnancy is suspected.

[i] Urine drug test to be conducted at screening, baseline and EOS visit and at unscheduled visit at the discretion of the investigator. Reflex confirmatory drug testing will be conducted by the lab vendor for all positive urine drug screen samples.

[j] Plasma samples for PK will be collected at random at Weeks 4, 8 and 12. Date and time of doses on the day of visits and 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 that are possibly drug related. Subjects who are able to schedule a morning visit for Week 4 and Week 8 can be instructed to hold their dose of study drug that morning until after a PK trough sample is obtained, if possible and appropriate.

[k] Study Drug will be dispensed at the baseline visit. Subjects should take the first dose in the morning the day after the baseline visit.

[l] Telephone calls to subjects will be made between visits during the first and second months of the Randomization Phase (Weeks 2 and 6) to monitor subject condition, any new concomitant medications and adverse events.

[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] If the study site needs to send drug overnight via certified and tracked courier and this is acceptable to the institution because a visit is absolutely not possible because of the COVID-19 pandemic, this is permissible per study. The sponsor should be consulted prior to shipping drug.

Table 3: Schedule of Assessments – Extension Phase

Visit	Ext Wk 2	Ext Wk4	Ext Wk 6	Ext Wk 8	Ext Wk 12	Ext Wk 24	Ext Wk 36	Ext Wk 48 or early term	Wk2 Post last dose
Safety Assessments									
Adverse Event Assessment		X		X	X	X	X	X	X
Telephone Check-in[b] *Includes AE assessment and concomitant medication review*	X		X						
Laboratory Assessments		X		X	X	X	X	X	
Pregnancy testing[a]		X		X	X	X	X	X	
Urine drug test		X		X	X	X	X	X	
Physical Exam								X	
Physical Measurements						X		X	
Vital Signs		X		X	X	X	X	X	X
12-Lead ECG		X			X	X	X	X	
Concomitant Medication Review		X		X	X	X	X	X	X
Sheehan Suicidality Tracking Scale (STS)		X		X	X	X	X	X	X
Clinical Outcome Assessments									
Yale-Brown Obsessive Compulsive Scale (Y-BOCS)		X		X	X	X	X	X	
Clinical Global Impressions-Severity Scale (CGI-S)		X		X	X	X	X	X	
Sheehan Disability Scale (SDS)[c]		X		X	X	X	X	X	
Quick Inventory of Depressive Symptomatology-Self Report (QIDS-SR)		X		X	X	X	X	X	
Beck Anxiety Inventory (BAI)		X		X	X	X	X	X	
Brown Assessment of Beliefs (BABS)		X			X				
Biomarker and Other Assessments									
BDNF and proBDNF Blood Sample								X	
Clinical Drug Supply									
Dispense Study Drug		X		X	X	X	X		
Drug Accountability		X		X	X	X	X	X	

Visit window is +/- 7 days during the extension phase.

[a] In addition to urine pregnancy tests conducted at each study visit, subjects will be provided with urine pregnancy tests to take in between every 3-month office visit during the Extension Phase. *Note: test will be sent home at Week 12 and at Week 24 and should be performed once between each 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] Telephone calls to subjects will be made between visits during the first and second months of the Extension Phase (Weeks 2 and 6) to monitor subject condition, any new concomitant medications and adverse events.

[c] 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.