Biohaven Pharmaceuticals

BHV-4157-303

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

Statistical Analysis Plan

Version 1.0

Date: 16-Aug-2023

SIGNATURE PAGE

Protocol Title:	A Randomized, Double-Blind, Placebo-Controlled Trial of Adjuntive Troriluzole in Obsessive Compulsive Disorder
Sponsor:	Biohaven Pharmaceuticals, Inc.
Document Version:	Version 1.0
Date:	16-Aug-2023
Author:	[redacted]
	Signature:_____
	Date:_____

Sponsor Approval

By signing this document, I acknowledge that I have read the document and approve of the planned statistical analyses described herein. I agree that the planned statistical analyses are appropriate for this study/project, are in accordance with the study/project objectives, and are consistent with the statistical methodology described in the protocol, clinical development plan, and all applicable regulatory 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 BHV4157-303 Clinical Study Report.

Sponsor Signatories:

[redacted]	Signature:_____
Biohaven Pharmaceuticals	Date:_____
[redacted]	Signature:_____
Biohaven Pharmaceuticals	Date:_____

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

REVISION HISTORY

Version	**Description of Change**
V1.0	Original Version (based on BHV4157-303 Protocol Version 3.0 Amendment 2 (US and Canada: 06-Apr-2022, China: 26-Apr-2022, UK and EU: 04-May-2022).

ABBREVIATIONS

Abbreviation	Definition
AE	Adverse event
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical
BABS	Brown Assessment of Beliefs Scale
BAI	Beck Anxiety Inventory
C-SSRS	Columbia-Suicide Severity Rating Scale
CGI-I	Clinical Global Impression- Improvement Scale
CGI-S	Clinical Global Impression-Severity Scale
CI	Confidence interval
CRF	Case report form
CSR	Clinical study report
CTCAE	Common Technical Criteria for Adverse Events
DOCS	Dimensional Obsessive Compulsive Scale
eDISH	Evaluation of drug-induced serious hepatotoxicity
eGFR	Estimated glomerular filtration rate
ICH	International Conference on Harmonization
IRT	Interactive response technology
LFT	Liver function test
MedDRA	Medical Dictionary for Regulatory Activities
OCD	Obsessive compulsive disorder
PK	Pharmacokinetics
PT	Preferred term
QIDS-SR	Quick Inventory of Depressive Symptomatology-Self Report
SAE	Serious adverse event
SAP	Statistical analysis plan
SD	Standard deviation
SDS	Sheehan Disability Scale
SE	Standard error
SI	Systeme Internationale
SMQ	Standardized MedDRA Queries

SOC	standard of care
TEAE	Treatment-emergent adverse event
TBL	Total bilirubin
TLF	Table listing figure
ULN	Upper limit of normal
US	United States
WHO-DD	World Health Organization-Drug Dictionary
Y-BOCS	Yale-Brown Obsessive-Compulsive Scale

1 BACKGROUND AND RATIONALE

Biohaven Pharmaceuticals, Inc. (Biohaven) is developing a new drug, troriluzole, for the treatment of Obsessive Compulsive Disorder (OCD) as well as for other neurologic and psychiatric disorders.

Troriluzole is a novel [redacted] prodrug of the glutamatergic agent riluzole. The FDA originally approved riluzole (RILUTEK®) 50 mg twice-a-day (NDA #20-599) for the treatment of patients with Amyotrophic Lateral Sclerosis (ALS). Riluzole is only indicated for ALS and has a number of non-desirable attributes that have limited its clinical use.

Troriluzole is a tripeptide prodrug of the glutamate modulating agent riluzole [redacted] for improved bioavailability, pharmacokinetics (PK) and dosing. The proposed study in OCD is based on recent preclinical, clinical and neuroimaging studies that implicate glutamatergic hyperactivity in the pathogenesis of OCD (Carlsson 2000, Coric, Milanovic et al. 2003, Coric, Taskiran et al. 2005, Pittenger, Bloch et al. 2011, Pittenger, Bloch et al. 2015). Additionally, preliminary efficacy findings from BHV4157-202, a proof of concept study, indicated troriluzole 200 mg, administered once daily as adjunctive therapy in subjects with OCD who had an inadequate response to standard of care (SOC) treatment resulted in numerically greater improvement versus placebo in the total Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) score in the randomization phase.

Biohaven hypothesizes that the pleiotropic effects of riluzole (e.g., glutamate modulation) may target mechanisms underlying pathologic brain function that is associated with OCD, and thus provide symptomatic benefit in patients with OCD.

1.1 Research Hypothesis

Troriluzole is superior to placebo as adjunctive therapy when added to SOC treatment over a 10 week period in subjects with OCD who have had an inadequate response to their current SOC treatment.

[redacted]

[redacted]

2 STUDY DESCRIPTION

2.1 Study Design

BHV4157-303 is a Phase 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 SOC treatment in subjects with OCD who failed to respond adequately to their current pharmacotherapy. Current inadequate response is defined by a Y-BOCS score of 22 or greater despite at least 8 weeks of treatment at screening and at least 12 weeks of treatment at baseline with an adequate dose of an SSRI (with the exception of fluvoxamine), clomipramine, venlafaxine or desvenlafaxine.

Subjects who are stable on SOC medication and having an inadequate response (as defined above) will be randomized to additionally receive placebo (QD) or troriluzole (280 mg QD, after two weeks at 200 mg QD). [redacted]

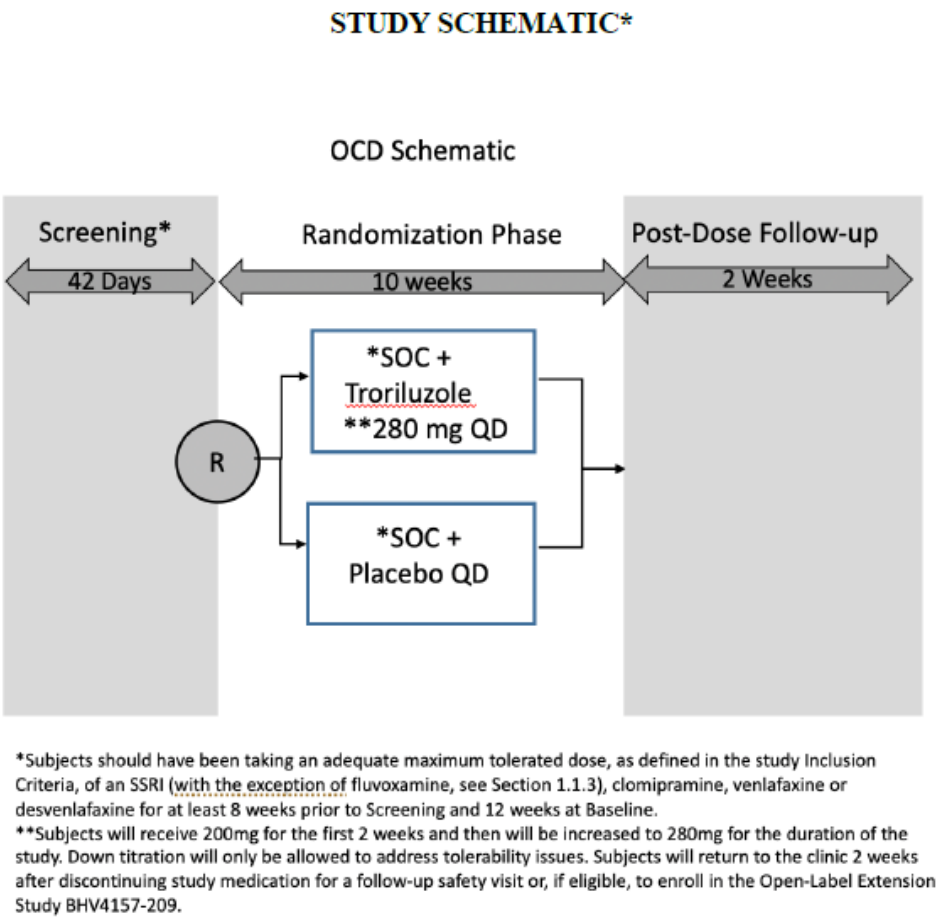
[redacted]

The total treatment period will be 10 weeks, after which subjects will return to the clinic 2 weeks after discontinuing study medication for a follow-up safety visit or, if eligible, will participate in the Open Label Extension Study, BHV4157-209.

[redacted]

Figure 1 illustrates the study schematic. The schedule of assessments is provided in Table 8 of Appendix 9.1.

Figure 1: Study Schematic



2.2 Treatment Assignment

Subjects who are stable on SOC medication and having an inadequate response (as defined above) will be randomized, in a 1:1 ratio, to additionally receive placebo (QD) or troriluzole (280 mg QD, after two weeks at 200 mg QD). Treatment assignments will be obtained by the investigator (or designee) via the Interactive Response Technology (IRT) system. [redacted]

2.3 Blinding and Unblinding

This is a double-blind study. [redacted]

[redacted]

2.4 Protocol and Protocol Amendments

BHV4157-303 SAP Version 1.0 is based on BHV4157-303 Protocol Version 3.0 Amendment 2 (US and Canada: 06-Apr-2022, China: 26-Apr-2022, UK and EU: 04-May-2022).

3 STUDY OBJECTIVES AND ESTIMANDS

3.1 Objectives

3.1.1 Primary Objectives

- The primary objective of the study is to evaluate the efficacy of troriluzole as adjunctive therapy compared to placebo in subjects with OCD who have had an inadequate response to their current SOC treatment, based on the change in their post-baseline Y-BOCS score.

3.1.2 Secondary Objectives

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

3.1.3 Exploratory Objectives

- Evaluate the efficacy of troriluzole compared to placebo on global functioning as measured by the Clinical Global Impression-Improvement Scale (CGI-I)
- Evaluate the efficacy of troriluzole compared to placebo on obsessive and compulsive symptomatology as measured by the change in the Y-BOCS obsession and compulsion subscales
- Evaluate the efficacy of troriluzole compared to placebo on depressive symptomatology as measured by the Quick Inventory of Depressive Symptomatology-Self Report (QIDS-SR)
- Evaluate the efficacy of troriluzole compared to placebo on anxiety symptoms as measured by the Beck Anxiety Inventory (BAI)
- Evaluate the efficacy of troriluzole compared to placebo on insight regarding obsessional beliefs as measured by the Brown Assessment of Beliefs Scale (BABS)

- Evaluate the efficacy of troriluzole compared to placebo on obsessive and compulsive symptomatology as measured by the change in the Dimensional Obsessive Compulsive Scale (DOCS)
- To characterize the pharmacokinetics of troriluzole from sparse sampling, using an existing population PK modeling, and explore population PK/PD relationships

3.2 Estimands

An estimand is the target of estimation to address the scientific question of interest posed by a study objective. The four attributes of an estimand include the population of interest, endpoint of interest, summary of the endpoint, and specification of how intercurrent events are reflected in the scientific question of interest.

For all objectives, the population of interest is defined through appropriate inclusion/exclusion criteria to reflect the targeted patient population for approval. Refer to the protocol for inclusion/exclusion criteria.

The estimand for the primary and secondary efficacy endpoints will be the effect due to the initially randomized treatments (when added to the SOC therapy), a "treatment policy" efficacy estimand. Please see specific sections below for additional detail on estimands for each endpoint.

3.2.1 Primary Objective Estimand

As the primary objective of this study is based on the evaluation of severity of subjects' symptomology, the estimand for the primary endpoint will be the effect due to the initially randomized treatments (when added to a SOC therapy) as taken, a treatment policy efficacy estimand. [redacted]

The primary endpoint will be the change from baseline to Week 8 in the Y-BOCS total score, troriluzole relative to placebo. This treatment effect will be summarized as the difference in change from baseline in the Y-BOCS between the groups receiving troriluzole and placebo.

Since the primary intent of this trial is to evaluate the effect of the drug as taken, a treatment policy strategy will be employed. All available assessments on the subject will be used regardless of treatment discontinuation, treatment non-compliance, protocol allowed dose adjustments, or initiation or adjustment of concomitant medications related to other symptoms.

Table 1: Primary Objective Estimand

Objective	Evaluate the efficacy of troriluzole as adjunctive therapy compared to placebo in subjects with OCD who have had an inadequate response to their current SOC treatment based on the change in their Y-BOCS score.
Efficacy Endpoint	Change from baseline to Week 8 in the Y-BOCS total score
Summary	Difference in change from baseline in the Y-BOCS total score between the groups receiving troriluzole and placebo using a Mixed Model for Repeated Measures (MMRM) analysis model based on the mITT population[redacted]
Intercurrent Events	Treatment policy strategy: All available assessments on the subject will be used regardless of treatment discontinuation, treatment non-compliance, protocol allowed dose adjustments, or initiation or adjustment of concomitant medications related to other symptoms.

3.2.2 Secondary Objective Estimands

Table 2: Secondary Objective Estimands

Objective	To assess the safety and tolerability of troriluzole, relative to placebo, in subjects with OCD
Safety Endpoint	Safety and tolerability are assessed using the frequency and percent of unique treated subjects with: AEs, SAEs; AEs leading to discontinuation; AEs judged to be related to study medication; and ECG QTc abnormalities (Section 6.4.3), vital sign abnormalities (Section 6.4.2) and clinically significant laboratory abnormalities (≥ Grade 3 per Section 6.4.5 as well as ALT, AST > 3x ULN or Total BILI > 2x ULN) that are observed during the double-blind randomization phase
Summary	Frequencies and percentages of all treated subjects with a specified event
Intercurrent Events	Study treatment discontinuation (due to any reason): "While on-treatment" strategy will be used such that events on or within 30 days of last day of medication will be included.

Objective	Evaluate the efficacy of troriluzole compared to placebo on functional disability as measured by the Sheehan Disability Scale (SDS)
Efficacy Endpoint	Change from baseline to Week 8 in the Sheehan Disability Scale (SDS) total score
Summary	Same as for primary estimand.
Intercurrent Events	Same as for primary estimand.
Objective	Evaluate the efficacy of troriluzole compared to placebo on global clinical condition as measured by the Clinical Global Impression-Severity Scale (CGI-S)
Efficacy Endpoint	Change from baseline to Week 8 in CGI-S total score
Summary	Same as for primary estimand.
Intercurrent Events	Same as for primary estimand.

3.2.3 Exploratory Objective Estimands

Table 3: Exploratory Objective Estimands

Objective	Evaluate the efficacy of troriluzole compared to placebo on global functioning as measured by the Clinical Global Impression- Improvement Scale (CGI-I).
Efficacy Endpoint	Response at Week 8 as defined by 'Much improved' or 'Very much improved' on the CGI-I scale;
Summary	The proportion of randomized subjects with a response will be analyzed using a Cochran-Mantel-Haenszel test [redacted]
Intercurrent Events	Same as for primary estimand.
Objective	Evaluate the efficacy of troriluzole compared to placebo on obsessive and compulsive symptomatology as measured by the change in the Y-BOCS obsession and compulsion subscales;
Efficacy Endpoint	Change from baseline to Week 8 in Y-BOCS obsession and compulsion subscales
Summary	Same as for primary estimand.
Intercurrent Events	Same as for primary estimand.
Objective	Evaluate the efficacy of troriluzole compared to placebo on depressive symptomatology as measured by the Quick Inventory of Depressive Symptomatology-Self Report (QIDS-SR)
Efficacy Endpoint	Change from baseline to Week 8 in Quick Inventory of Depressive Symptomatology-Self Report (QIDS-SR) total score
Summary	Same as for primary estimand.
Intercurrent Events	Same as for primary estimand.
Objective	Evaluate the efficacy of troriluzole compared to placebo based on anxiety symptoms as measured by the Beck Anxiety Inventory (BAI);
Efficacy Endpoint	Change from baseline to Week 8 in the Beck Anxiety Inventory (BAI) total score
Summary	Same as for primary estimand.
Intercurrent Events	Same as for primary estimand.
Objective	Evaluate the efficacy of troriluzole compared to placebo on insight regarding obsessional beliefs as measured by the Brown Assessment of Beliefs Scale (BABS)
Efficacy Endpoint	Change from baseline to Week 10 in Brown Assessment of Beliefs Scale (BABS) total score
Summary	Difference in change from baseline in the BABS total score between the groups receiving troriluzole and placebo using an ANCOVA analysis model based on the mITT population [redacted]
Intercurrent Events	Same as for primary estimand.
Objective	Evaluate the efficacy of troriluzole compared to placebo based on obsessive and compulsive symptology as measured by the Dimensional Obessessive Compulsive Scale (DOCS);
Efficacy Endpoint	Change from baseline to Week 8 in the Dimensional Obsessive Compulsive Scale (DOCS) total score

Summary	Same as for primary estimand.
Intercurrent Events	Same as for primary estimand.
Objective	To characterize the pharmacokinetics of troriluzole from sparse sampling, using an existing population PK modeling, and explore population PK/PD relationships.
Efficacy Endpoint	The pharmacokinetic profile of troriluzole is characterized from sparse plasma concentrations observed in treated subjects using an existing population PK model. Further exploration of PK/PD relationships for both efficacy and toxicity are explored as warranted.
Summary	Individual concentrations will be summarized by visit.
Intercurrent Events	Same as for primary estimand.

4 ANALYSIS SETS, TREATMENT GROUPS, AND SUBGROUPS

4.1 Analysis Sets

The following analysis sets are used for statistical analyses:

- Enrolled: Subjects who signed an informed consent form and were assigned a Subject Identification Number (SID), i.e., non-missing informed consent date

- Randomized: Enrolled subjects who received a treatment assignment from the Interactive Response Technology (IRT) system.

[redacted]

- Modified Intent to Treat (mITT) Population: Randomized subjects who received at least 1 dose of study therapy, and provided a non-missing baseline assessment and at least one non missing post-baseline on-treatment efficacy assessment (based on YBOCS in person only) during the randomization phase.

[redacted]

- Safety Population: Enrolled subjects who received at least one dose of blinded study therapy (troriluzole or placebo), i.e., non-missing study drug start date.

[redacted]

- Coronavirus Disease 2019 (COVID-19) impacted: Subjects in the enrolled analysis set who have at least one visit impacted by COVID-19. This analysis set is used to identify COVID-19 impacted subjects in listings.

4.2 Treatment Groups

The two treatment groups are:

- placebo (QD) or
- troriluzole (280 mg QD, after two weeks at 200 mg QD)

The safety population is assessed by as-treated treatment group (i.e., actual treatment received), the randomized and mITT populations are assessed by as-randomized treatment group, and the enrolled analysis set is assessed overall. If a subject receives at least 1 dose of troriluzole, then that subject has an as-treated treatment group of troriluzole.

4.3 Subgroups

For the efficacy analysis set, the following efficacy subgroups are of interest:

- Age: (<= median, > median)
- Sex (female, male)
- Race (Asian, Black or African American, White, and all other races combined)
- Ethnicity: Hispanic or Latino, Not Hispanic or Latino
- Baseline body mass index (BMI; kg/m^2): < 25, ≥ 25 to < 30, ≥ 30
- Geographic region (North America, China and Rest of World)
- SOC Concomitant Treatment (citalopram, escitalopram, fluoxetine, paroxetine, sertraline, clomipramine, venlafaxine, desvenlafaxine) based on actual treatment as collected per CRF
- Number of Prior Treatments for OCD (out of citalopram, escitalopram, fluoxetine, paroxetine, sertraline, clomipramine, venlafaxine, desvenlafaxine): (0, 1, 2, > 2)
- Number of Prior Treatments for OCD (based on any treatment for OCD): (0, 1, 2, > 2)

Subgroup analyses will be performed for the primary efficacy endpoint and secondary efficacy endpoints. [redacted] For all other subgroup analyses, summary statistics only will be presented. Efficacy subgroup tables present results by subgroup level for subjects with non-missing subgroup level data, unless specified otherwise.

Subgroup levels may be redefined or combined based on the availability of data.

5 SAMPLE SIZE, POWER, AND TYPE I ERROR

The sample size for this study will be up to approximately 700 randomized subjects. [redacted] Assuming a standard deviation of 5.8, a 2-sided alpha of 0.05, this sample size provides 90% power to detect a difference on the Y-BOCS total score change from baseline of 1.9 between the treatment groups. [redacted]

Type 1 error will be controlled for the primary and secondary efficacy endpoints by testing them with a gate-keeping procedure. The primary endpoint, change from baseline to Week 8 in the total Y-BOCS, will be tested at a two-sided alpha level [redacted] 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 this result.

No attempt will be made to adjust for multiplicity when testing the exploratory endpoints. Any exploratory endpoints subjected to significance testing will have a p-value presented for descriptive purposes only.

[redacted]

6 STATISTICAL ANALYSES

6.1 General

6.1.1 Programmed Output

A list of TLFs and corresponding templates, attributes, and programming notes are presented separately in a mock TLF document corresponding to this SAP.

All programmed TLFs are formatted and numbered according to the latest version of Biohaven Standard Outputs for CSRs.

6.1.1.1 Tables

Tables present results by treatment group (i.e., troriluzole and placebo) with the following exceptions: (1) results for the enrolled analysis set are presented only by overall, without treatment group; and (2) results for some tables such as study population parameters displays also include overall in addition to treatment group.

6.1.1.2 Listings

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 describe race 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.

Unless specified otherwise, by-subject listings are sorted by site-subject ID, and additional variables such as time points, as applicable.

All listings except listings of batch numbers identify subjects who are impacted by COVID-19.

6.1.2 Statistical Methods

Summaries will be presented overall [redacted] unless otherwise specified.

[redacted] The primary presentation for the safety analysis will include all treated subjects [redacted]

[redacted]

6.1.2.1 Descriptive Statistics in Summary Tables

Categorical variables are tabulated as the number and percentage of subjects within each category, including a "not reported" category for missing data if applicable. Percentages are displayed as follows:

- 0 < percentage < 0.1 as "<0.1"
- 99.9 < percentage < 100 as ">99.9"
- 0 count without a percentage
- Rounded to 1 decimal place otherwise.

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 5 decimal places.

6.1.2.2 Counting Rules in Frequency Tables

Tabulations of the following endpoints present the number and percentage of unique subjects with an event: protocol deviations; medical history; non-study medications; AEs; laboratory test worst abnormalities, laboratory elevations and shifts from baseline; vital sign and physical measurement abnormalities; ECG shifts from baseline and abnormalities; and suicidality categories. Thus, for these endpoints, multiple occurrences of the same event are counted only once per subject.

6.1.2.3 Rounding Rules

For frequency tables with cuts on incidences (e.g., AEs related to study drug occurring in ≥ 2% of subjects), the determination for inclusion in the table based on incidence is determined prior to rounding. For example, an AE occurring at an incidence < the cut-point percentage prior to rounding is not included in a table of "AEs occurring in ≥ {cut-point percentage}subjects" even if the incidence rounds up to the cut-point percentage; however, an AE occurring at an incidence ≥ the cut-point percentage prior to rounding is included in such a table.

Incidence rates are compared to the cut-point percentage prior to rounding but displayed after rounding using the rules above.

Similarly, subgroup levels that are derived from continuous variables (e.g., body mass index [BMI]) are based on the actual, non-rounded calculations.

6.1.3 Handling of Missing Data

The primary analysis of the primary endpoint will be based on observed data using MMRM based on the missing at random (MAR) assumption. As a sensitivity analysis to assess missing data assumptions of the MMRM model, missing data will be multiply imputed for the primary endpoint using missing not at random (MNAR) methods (pattern mixture model and "jump to reference") and "tipping point" methods (see Section 6.3.2.4).

All other analyses are based on observed data unless specified otherwise.

6.2 Study Population

6.2.1 Analysis Sets

The number of subjects in each analysis set described in Section 4.1 is tabulated. A by-subject listing of analysis sets is provided for the enrolled analysis set.

A by-subject administrative listing of randomization scheme and codes are provided for all randomization numbers and block numbers, even those not assigned to a subject. This listing is sorted by randomization number and block number.

6.2.2 Enrollment

Enrollment by (1) country and site and (2) age group (i.e Age groups are defined in the Trial information workbook of EudraCT result related data Dictionary_V2_for publication.xlsx 12-17,18-64, >= 65) is tabulated for the enrolled analysis set. The enrollment by country and site table also displays results for the Safety Population.

Accrual by month and year of treatment start is tabulated as the number and percentage of subjects in each time category (i.e., month and year of study drug start date) for the Safety Population. See Section 7.1 for derived dates.

6.2.3 Subject Disposition

Subject disposition is based on the Double-Blind Study Discontinuation/Completion case report form (CRF) , Screen Failure CRF or Inclusion/Exclusion Eligibility Criteria CRF as specified below.

Subject disposition from enrollment to the randomization is tabulated for the enrolled analysis set as the number and percentage of subjects in the following categories:

- Number of subjects enrolled and signed informed consent form
- Number of enrolled subjects excluded from the study and reason for screen failure/exclusion based on the Screen Failure and the Inclusion/Exclusion Eligibility Criteria CRF.
- Number of subjects randomized

Subject disposition during Randomization Phase is tabulated for randomized subjects by treatment group and overall as the number and percentage of subjects in the following categories:

- Number of subjects treated
- Number of mITT subjects
- Number of subjects who completed the Randomization Phase
- Number who completed Randomization Phase and rolled over into open label extension study
- Number who completed Randomization Phase and did not enter the open label extension study

- Number of subjects who prematurely withdrew from the Randomization Phase and reasons for withdrawal

A by-subject listing of subject disposition is presented for randomized subjects based on the Double-Blind Study Discontinuation/Completion CRF. Where applicable, the listing will include whether any reason for withdrawal was due to COVID-1.

A by-subject listing of eligibility with inclusion and exclusion criteria is provided for all subjects in the enrolled analysis set, not just those who have non-missing criteria. This is based on the Screen Failure and Inclusion/Exclusion Eligibility Criteria CRF.

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

For all randomized subjects, the impact of the COVID-19 pandemic crisis on the study will be summarized by presenting (by treatment and overall)

- Number of subjects who prematurely withdrew from the Randomization Phase due to COVID-19

- Number and percentage of subjects impacted by COVID-19 pandemic crisis by visit with subjects categorized by the first visit impacted.

- Summary of the impact including type of visit (e.g. missed, remote telephone, etc.) and relation to COVID-19 by visit based on the COVID-19 Visit Impact CRF.

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

6.2.4 Protocol Deviations

A significant protocol deviation is study conduct that differs significantly from the current protocol, including Good Clinical Practice (GCP) non-compliance, which is identified during monitoring of the study (on-site or in-house) and at minimum includes the following categories:

- Inclusion or exclusion deviations

- Incorrect dosing or treatment assignment

- Use of prohibited concomitant medications

- Failure to obtain written informed consent prior to each subject's participation in the study

- Failure to report all Serious Adverse Events (SAEs) in accordance with the time period required by GCP, the protocol, and applicable regulations

- Implementation of protocol changes prior to review by Institutional Review Board/Independent Ethics Committee IRB/IEC (except when necessary to eliminate an immediate hazard to trial subjects) or failure to implement an IRB/IEC approved amendment

- Additional study-specific deviations defined as Significant by the protocol team and documented in the Site Monitoring Plan.

A listing of significant protocol deviations will be provided for the enrolled analysis set including whether the deviation was due to Covid-19. Protocol deviations must be determined prior to database lock and recorded on the protocol deviation CRF.

A relevant protocol deviation is a deviation from the protocol which is programmed from the database and which could potentially affect the interpretability of the study results. This type of deviation includes but is not limited to:

- Eligibility (i.e., some inclusion or exclusion criteria)

- Subject management, such as:

- Incorrect dosing or study treatment assignment

- Use of prohibited concomitant medications

Subjects not withdrawn from treatment and/or study despite having met specified criteria for withdrawal.

Relevant protocol deviations are tabulated by treatment group as the number and percentage of subjects with deviations by deviation type (eligibility, subject management), category, and subcategory in the order specified in Appendix 9.4 for the Randomized Population.

Results for all relevant protocol deviation categories and subcategories are displayed, even those with 0 counts, unless specified otherwise.

A by-subject listing of relevant protocol deviations is provided for the enrolled analysis set. This includes deviation type, category, and subcategory.

6.2.5 Baseline Characteristics

Tabulations of demographic and baseline characteristics will be made for the Safety Populations including tabulations for OCD history, medical history, current standard of care treatment, Borderline Personality Disorder Module (BPD), and baseline Y-BOCS total score (continuous and categorized as ≤ median or > median). A summary of prior OCD medications based on the MGH-TRQ-OCD and the Other Prior OCD Treatment CRF pages will also be provided.

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.

The baseline value for a parameter (e.g., weight) is defined as the last non-missing value in the pre-treatment analysis period; see Section 7.2. "Last" is determined by the measurement date/time; other criteria such as last entry date/time or laboratory test source (i.e., central versus local laboratory) may be applied to break ties as needed.

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

6.2.6 Exposure

6.2.6.1 Study Therapy

Subjects will receive placebo (QD) or troriluzole (280 mg QD, after first two weeks at 200 mg QD) during the Randomization Phase. Subjects who complete 10 weeks of treatment in the Randomization Phase may be eligible for the Open Label Extension Study, BHV4157-209.

The extent of subject exposure (including missed dose days) will be quantified as treatment duration (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 10 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).

The extent of subject exposure will also be calculated only including days where number of tablets taken was > 0 (total days on study medication).

The number of subjects on-treatment and their average daily dose (including minimum and maximum) will be summarized by week (7-day intervals) as well as overall for the randomization phase. For subjects on placebo, "exposure" will be summarized using pseudocapsule strength based on reported capsule strength taken..

The extent of study medication exposure will also be summarized by a Kaplan-Meier plot presenting proportion of subjects still on Randomization Phase study medication (including days of missed dose). Subjects who elect to continue in the open label extension phase should be censored on the day of last dose in the double-blind study.

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

% compliance = total days on study medication/(last dose date – first dose date + 1) × 100.

The number and percent of treated subjects down-titrating to 200 mg toriluzole (or matching placebo) after one or more days at 280mg troriluzole (or matching placebo) will be presented, as

will the number and percent of treated subject never reaching 280 mg troriluzole (or matching placebo).

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

6.2.6.2 Concomitant Medications

Non-study concomitant medications are tabulated by therapeutic class (ATC4) and preferred name for the Safety Population.

A summary of concomitant SOC Treatments (citalopram, escitalopram, fluoxetine, paroxetine, sertraline, clomipramine, venlafaxine, desvenlafaxine) will be presented.

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.' If the end date of the medication is prior to the start date of the study drug, the medication will be considered 'prior'. Imputed medication start and stop dates are used to assign non-study medication type as 'prior' or 'concomitant'.

Non-study medications are identified from those reported on the Concomitant Medication CRF, which links medical history and AE terms respectively to the Medical History and AE CRFs using numeric identification variables. Medications are displayed in descending order of overall frequency within therapeutic class and preferred name.

A by-subject listing of non-study medications is provided by therapeutic class and preferred name for the enrolled analysis set.

6.3 Efficacy

6.3.1 Overall Efficacy Summary

Table 4 provides a summary of the statistical methods used for the primary, secondary, and exploratory endpoints.

Unless otherwise noted, all efficacy analyses will be conducted using the mITT population. All efficacy data will be included in listings by subject, treatment group, and visit (as applicable).

Table 4: Efficacy Endpoints and Analysis Methods

Efficacy Endpoints	MMRM	ANCOVA	Cochran Mantel Haenszel Test
Primary Change from baseline to Week 8 in Yale-Brown Obsessive Compulsive Scale (Y-BOCS) total score	X		
Secondary			
Change from baseline to Week 8 in Sheehan Disability Scale (SDS) Total Score	X		
Change from baseline to Week 8 in Clinical Global Impression of Severity (CGI-S) Total Score	X		
Exploratory			
Response as defined by 'Much improved' or 'Very much improved' on the CGI-I scale using all randomized subjects			X
Change from baseline in Yale-Brown Obsessive Compulsive Scale (Y-BOCS) obsession and compulsion subscales	X		
Change from baseline in Quick Inventory of Depressive Symptomatology-Self Report (QIDS-SR)	X		
Change from baseline in Beck Anxiety Inventory (BAI) Total Score	X		
Change from baseline in Brown Assessment of Beliefs (BABS)		X	
Change from baseline in Dimensional Obsessive Compulsive Scale (DOCS)	X		

6.3.2 Primary Endpoint

6.3.2.1 Change from baseline to Week 8 in Yale-Brown Obsessive Compulsive Scale (Y-BOCS) total score

The primary endpoint is change from baseline to Week 8 in Yale-Brown Obsessive Compulsive Scale (Y-BOCS) total score and will be analyzed as described below. [redacted]

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 than 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. For this study, the Y-BOCS assessment is scheduled to be completed at Screening, Baseline, and at Weeks 4, 8, and 10 or early termination. All analyses and summaries will consider only assessments done by the clinician with the subject in person.

The change from baseline in the Y-BOCS total score through Week 10 will be analyzed using a MMRM analysis model. The model will include mITT subjects [redacted] 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 and finally by a compound symmetry 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 10. These will be presented with degrees of freedom, standard errors (SEs), and two-sided 1-α 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 1-α 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, separately.

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

6.3.2.2.2 Final Analysis

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

6.3.2.3 *Analysis after Sample Size Re-estimation*

[redacted]

6.3.2.4 Sensitivity Analyses

The following sensitivity analyses will only be conducted [redacted]

Sensitivity Analyses Addressing Missing Data Due to Drop Outs

The primary analysis of the Y-BOCS total score based on a Mixed Model for Repeated Measures (MMRM) model assumes that data are 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 that completed the 10 weeks of the Randomization phase. Sensitivity analyses of the primary efficacy endpoint will be conducted to evaluate the robustness of study results under different assumptions and imputation algorithms. For multiple

imputation, the number of imputations used will be 100. For each of these sensitivity analyses, the final p-value will be calculated based on the CWH procedure, using the formula specified above.

The following sensitivity analyses will be conducted to support the primary analysis:

- **Sensitivity 1:** Multiple imputation; missing not at random (MNAR), Jump to Reference

This sensitivity analysis will be conducted under the assumption that subjects who discontinue troriluzole prematurely have a response profile similar to those subjects on placebo. 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 using the mean response distribution after withdrawal. The placebo arm subjects will use the profile under MAR.

  - Intermittent missing Y-BOCS scores will first be imputed using the Markov Chain Monte Carlo (MCMC) method(Schafer 1997) with the SAS MI procedure, which is appropriate for non-monotonic missing data.

  - For subjects who discontinued prematurely on the troriluzole arm, Y-BOCS scores missing after discontinuation will be imputed using the mean response distribution after withdrawal on placebo. (MNAR)

  - The change from baseline to each visit through Week 10 in total Y-BOCS score will be calculated for each imputed dataset.

  - The analysis of the primary endpoint will be performed for each complete, imputed dataset.

  - Results of the MMRM on the multiply imputed data sets will be combined to generate an overall estimate and associated variance using Rubin's rules(Rubin 1987).

- **Sensitivity 2:** Multiple imputation; missing not at random (MNAR), Pattern Mixture Model

This sensitivity analysis is an implementation of a pattern mixture model where subjects are grouped based on the reason for withdrawal and subjects in the same group share the same pattern of missing data. The assumption is that the AE and lack of efficacy(LOE) dropouts would revert to their baseline values for the unobserved time points.

  - Intermittent missing Y-BOCS scores will first be imputed using the MCMC method (Schafer 1997) with the SAS MI procedure, which is appropriate for non-monotonic missing data.

- For subjects who discontinued study drug due to adverse events, Y-BOCS scores missing after discontinuation will be imputed using the distribution of the baseline value of all subjects Y-BOCS score in the mITT population.. (MNAR)

- For subjects who discontinued study drug due to lack of efficacy, Y-BOCS scores missing after discontinuation will be imputed using the distribution of the baseline value of all subjects Y-BOCS score in the mITT population.. (MNAR)

- For subjects who discontinue due to reasons other than adverse event or lack of efficacy, missing Y-BOCS scores after subjects discontinue study drug early will be multiply imputed using multiple calls of the SAS MI procedure using data from subjects within the same treatment group that have complete data at that time, including subjects that discontinued due to adverse event and subjects that discontinued due to lack of efficacy. [redacted]

- The change from baseline to each visit through Week 10 in total Y-BOCS score will be calculated for each imputed dataset.

- The analysis of the primary endpoint will be performed for each complete, imputed dataset.

- Results of the MMRM on the multiply imputed data sets will be combined to generate an overall estimate and associated variance using Rubin's rules.(Rubin 1987)

- **Sensitivity 3:** Tipping point analysis

Multiple imputation with mixed missing data mechanisms (MNAR for a BHV4157 and MAR for placebo) will be used to assess the robustness of the MAR assumption. This sensitivity analysis is to investigate the departure from MAR assumption by progressively decreasing the treatment differences with respect to the Y-BOCS total scores over the missing visits in active treatment group until conclusion from the primary analysis is overturned. This shift will be applied to all values up to Week 10. This will not be performed should the initial primary results fail to achieve significance. The MI procedure includes the following steps:

- Intermittent missing Y-BOCS scores will first be imputed using the MCMC method (Schafer 1997) with the SAS MI procedure, which is appropriate for non-monotonic missing data.
- The monotone missing Y-BOCS values will then be multiply imputed with the SAS MI procedure using the monotone regression method.

[redacted]

- For subjects in the active treatment group, a shift parameter will be progressively applied

to impute the missing data at Week 8, until the p-value is >0.05.

Other Sensitivity Analyses

In order to account for prior treatment failures a fixed effect will be added to the model as specified in Section 6.3.2.1 to adjust for the number of prior treatment failures with the levels being no prior failures, one prior failure, two prior failures and more than 2 prior failures.

In order 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 study impacted by COVID-19. All other aspects of the model and presentation are similar as described above for the primary analysis. This analyses will be performed if ≥ 10% of mITT subjects had at least one visit in the study impacted by COVID-19 at the time of analysis.

A responder analysis with "response" defined as an improvement of at least 25% of the total Y-BOCS score (Week 8 compared to baseline) will be performed using the mITT population [redacted] Subjects who discontinue the study prior to Week 8 will be treated as failures. The data will be evaluated with a Cochran-Mantel Haenszel test, [redacted] Response rates for each treatment will be presented with exact (Clopper-Pearson) 1-α% CIs. [redacted]

6.3.3 Secondary Efficacy Endpoints

[redacted]

The estimand for the secondary endpoints will be the effect due to the initially randomized treatments (when added to the standard of care therapy), a "treatment policy" efficacy estimand. The change from baseline on the SDS total score and the CGI Severity score will be analyzed using the same methodology as the primary endpoint.

The secondary endpoints are:

- Change from baseline to Week 8 in Sheehan Disability Scale (SDS) total score;
- Change from baseline to Week 8 in Clinical Global Impression of Severity (CGI-S) total score

6.3.3.1 Change from baseline to Week 8 in Sheehan Disability Scale (SDS) Total Score

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, will be estimated using the mITT population. 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 the same MMRM analysis model as used for the primary analysis except the baseline assessment used in the model would be the baseline SDS total score. [redacted] 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 and finally by a compound symmetry structure. Error degrees of freedom will be calculated in the same manner as the primary endpoint, noted in Section 6.3.2.1.

LSMeans for the change from baseline for each treatment group will be derived for Week 4, Week 8, and Week 10. These will be presented with degrees of freedom, SEs, and two-sided 1-α% CIs (where α is the appropriate alpha level based on Sections 6.3.2.2 and 6.3.2.3). 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 1-α% CIs, and p-values.

Given that secondary efficacy analyses are Type 1 error controlled, the CHW procedure will also be performed for SDS total score if the study continues to the final analysis and the sample size is re-estimated.

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.

These analyses will also be performed for the subgroups presented in Section 4.3.

[redacted]

6.3.3.2 Change from baseline to Week 8 in Clinical Global Impression of Severity (CGI-S) Total Score

The secondary endpoint, the clinician global impression of severity via the CGI-S will be summarized descriptively by post-baseline visit 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 the same MMRM analysis model as used for the primary analysis except the baseline assessment used in the model would be the baseline CGI-S total score. 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"). [redacted] 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 and finally by a compound symmetry structure. Error degrees of freedom will be calculated in the same manner as the primary endpoint, noted in Section 6.3.2.1. All analyses and summaries will only consider assessments done by the clinician with the subject in person.

Given that secondary efficacy analyses are Type 1 error controlled, the CHW procedure will also be performed for CGI-S score if the study continues to the final analysis and the sample size is re-estimated.

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.

These analyses will also be performed for the subgroups presented in Section 4.3.

[redacted]

6.3.4 Exploratory Efficacy Endpoints

[redacted]

6.3.4.1 Response as defined by 'Much improved' or 'Very much improved' on the CGI-I scale

The CGI-I is a clinician rated assessment of the subject's improvement in global functioning on a 7 point scale. The CGI-I will be conducted at study visits as indicated in the Schedule of Assessments and Events and the subject's improvement will be compared to baseline.

For analysis, CGI-I categories of interest for the numerator are categorized as "response" (which included the categories of "very much improved" and "much improved").

The number and proportion of randomized subjects with a response at each specific week will be analyzed using a Cochran-Mantel-Haenszel test [redacted]

6.3.4.2 Change from baseline in Yale-Brown Obsessive Compulsive Scale (Y-BOCS) obsession and compulsion subscales

The change from baseline in Yale-Brown Obsessive Compulsive Scale (Y-BOCS) obsession and compulsion subscales will be analyzed in the same manner as described for the primary endpoint in Section 6.3.2.1.

6.3.4.3 Change from baseline in QIDS-SR score

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.

The exploratory endpoint, change from baseline in the QIDS-SR total score, will be estimated using the mITT population. 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 the same MMRM analysis model as used for the primary analysis except the baseline assessment used in the model would be the baseline QIDS-SR total score. 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 and finally by a compound symmetry structure. Error degrees of freedom will be calculated in the same manner as the primary endpoint, noted in Section 6.3.2.1.

LSMeans for the change from baseline for each treatment group will be derived for Week 4, Week 8, and Week 10. 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.

6.3.4.4 Change from baseline in Beck Anxiety Inventory (BAI) total score

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.

The exploratory endpoint, change from baseline in the BAI total score, will be estimated using the mITT population. 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 the same MMRM analysis model as used for the primary analysis except the baseline assessment used in the model would be the baseline BAI total score. 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 and finally by a compound symmetry structure. Error degrees of freedom will be calculated in the same manner as the primary endpoint, noted in Section 6.3.2.1.

LSMeans for the change from baseline for each treatment group will be derived for Week 4, Week 8, and Week 10. 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.

6.3.4.5 Change from baseline in Brown Assessment of Beliefs Scale (BABS) total score

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 times 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.

The exploratory endpoint, change from baseline in the BABS total score from baseline to week 10, will be estimated using the mITT population. 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 10 in the total BABS score (based on observed cases at Week 10 and LOCF endpoint where endpoint assessment must be a post-baseline assessment) will be analyzed with a univariate ANCOVA [redacted] The covariate adjusted difference in the change from baseline between the troriluzole 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.

6.3.4.6 Dimensional Obsessive Compulsive Scale (DOCS)

The DOCS is a 20-item self report assessment based on 4 different dimensions. These dimensions are based on contamination, responsibility for harm and mistakes, incompleteness/symmetry and unacceptable thoughts.

Within each of the 4 dimensions, the five item scores are summed to produce a subscale score (range = 0-20). The four subscale scores can be summed to produce an overall DOCS total score (range = 0-80).

Descriptive statistics for the DOCS total score and 4 subscale scores will be summarized by treatment and visit.

The exploratory endpoint, change from baseline in the DOCS total score, will be estimated using the mITT population. This treatment effect will be summarized as the difference in change from baseline in the DOCS total score between the troriluzole and placebo groups.

The change from baseline in the DOCS total score will be analyzed using the mITT set via the same MMRM analysis model as used for the primary analysis except the baseline assessment used in the model would be the baseline DOCS total score. 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 and finally by a compound symmetry structure. Error degrees of freedom will be calculated in the same manner as the primary endpoint, noted in Section 6.3.2.1.

LSMeans for the change from baseline for each treatment group will be derived for Week 4, Week 8, and Week 10. 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.

6.3.4.7 Pharmacokinetic Evaluations

All PK analyses will be conducted using the Safety Population.

A pharmacokinetic sample will be collected at Week 4, Week 8, and Week 10. 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 the last dose and whether a meal was eaten within 2 hours of last dose will be collected. Subjects should be instructed to take their dose at their routine time on the days for the next visit, unless the visit is scheduled in the morning, in which case subject will be instructed to hold their dose and only take the dose after the blood sample is collected at the clinical site.

PK concentrations will be summarized descriptively by visit as continuous variables using the following summary statistics: n, arithmetic mean, median, Q1, Q3, coefficient of variation (CV%), SD, geometric mean, geometric CV%, minimum, and maximum. 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.

6.4 Safety

Safety analyses are based on the Safety Population, unless otherwise noted. Safety measures include: deaths, AEs, laboratory tests, vital signs, physical measurements, electrocardiograms (ECGs), and Columbia-Suicide Severity Rating Scale.

Safety summaries will be done for the overall Safety Population [redacted] The primary presentation of safety will be based on the overall Safety Population.

Select safety parameters are tabulated descriptively as continuous variables at baseline and each scheduled visit over time during the on-treatment period. Measurements are slotted into analysis periods and analysis visits using the following steps:

1) Measurements are slotted into the pre-treatment and on-treatment periods.
2) Measurements are slotted into analysis visits in the analysis periods listed in the previous step.

For tables of parameters summarized over time, if a subject has multiple values in an analysis visit window in an analysis period and any of them is abnormal, then the "worst" measurement (maximum for abnormalities above the upper reference, minimum for abnormalities below the lower reference, and maximum absolute change from baseline for measurements with abnormality criteria both upper and lower) will be used for presentation in the by-visit tables for summary statistics. If all values are in the normal range, then the latest non-missing value in the visit window is used; in the case of a tie, the last value collected (from the central laboratory, if

applicable) is used. See Sections 6.2.5, 7.2, and 7.3 for definitions of baseline, analysis periods, and analysis visit windows, respectively.

By-subject listings of safety parameters are described in subsections.

The term "on-treatment" applies to the on-treatment period.

6.4.1 Adverse Events

AEs will be coded using the most current version of MedDRA at the time of database lock (i.e. Version 24.0 or later) 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 (TEAE), where treatment-emergent is defined as any AE that developed, worsened, or became serious after first dose of test treatment. Programmatically TEAE is based on onset date relative to first day of medication (see sections 7.1 and 7.2)

The number and percentage of subjects with the following AEs will be summarized by treatment group.

- 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)
- Non-serious TEAEs
- Non-serious TEAEs with >5% incidence in either treatment group
- TEAEs indicating potential Interstitial Lung Disease (based on the interstitial lung disease Narrow Standardized MedDRA Queries (SMQ)

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 TEAEs summary by SOC and PT will also include the number of unique AE occurrences; unique occurrences mean that if a subject had an AE twice, both will be reported (i.e., the subject will contribute twice to the count of events). However, if the end of the first occurrence and the start of the second occurrence are overlapping, or the end of the first occurrence indicates "ongoing", then this would be counted as only one occurrence.

[redacted]

All AEs occurring pre-treatment and during the entire study will be listed with those indicated as TEAE flagged. Additional listings will be provided including deaths, SAEs, AEs leading to discontinuation of study drug, and TEAEs indicating potential Interstitial Lung Disease.

6.4.2 Vital Signs and Physical Measurements

The observed value and change from baseline in vital signs and physical measurements will be summarized at each visit. The summary will be based on the data after the date/time of first dose of study drug and within 30 days of the last dose of study drug. (see Section 7.2) In addition, the number and percentage of subjects with at least one on-treatment vital sign measurement meeting the specified criteria (each criteria is considered separately) will be presented:

- 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. A subject listing will be provided for each vital sign measure listing any subject meeting the criteria with a complete list of all of the specific test assessments with the abnormal ones flagged.

6.4.3 Electrocardiograms

ECG readings from the local machines at the sites were sent to a central coding group 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. The summary will be based on the data after the date/time of first dose of study drug and within 30 days after the last dose of study drug. (see Section 7.2)

In addition, the number and percentages with at least one on-treatment QTcF > 450 ms, >480 ms, and >500 ms will be summarized. 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 measure listing any subject meeting the criteria with a complete list of the QTcF assessments with the abnormal ones flagged (where "abnormal" is meeting one of the above criteria).

6.4.4 Physical Exam

Results of Physical Examination will be included as a listing.

6.4.5 Laboratory Tests

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, estimated glomerular filtration rate (eGFR) according to the re- expressed abbreviated (four-variable) Modification of Diet in Renal Disease (MDRD) Study equation < 30 ml/min/ 1.73m2; The MDRD estimation is calculated as follows: eGFR (mL/min/1.73m2) = 175 x (standardized Scr)$^{-1.154}$ x (Age)$^{-0.203}$ x (0.742 if female) x (1.212 if Black), blood urine nitrogen (BUN), and uric acid.

- Urinalysis: 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.

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. All measurements will be presented in listings and considered for evaluation of potential drug induced liver injury (DILI) or abnormalities.

Sites from China utilize a separate central lab vendor. The values from the central lab for sites in China will be combined with the centralized laboratory values from the double-blind phase lab by normalizing them to the central laboratory during the double-blind phase 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. (see Section 7.2)

The observed value and change from baseline will be summarized for each continuous laboratory parameter, 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.

Table 5: Clinical Laboratory Parameters and Criteria for Grading

Panel	Not Graded	CTCAE	DAIDS
Hematology	Absolute basophil count	Absolute lymphocyte count (ALT)	
	Absolute eosinophil count	Absolute neutrophil count	
	Absolute monocyte count	Hemoglobin	
	Hematocrit	Platelet count	
	Red blood cell count	White blood cell count	
Serum Chemistry	Blood urine nitrogen (BUN)	Alanine aminotransferase (ALT)	Glucose
	Chloride	Albumin	Low-density lipoprotein (LDL)*
	Folate*	Alkaline phosphatase (ALP)	Uric acid
	Gamma-glutamyl transferase (GGT)	Aspartate aminotransferase (AST)	
	Hemoglobin A1c (HbA1c)*	Bicarbonate	
	High-density lipoprotein (HDL)*	Calcium	
	Lipase*	Creatine kinase	
	P-Amylase*	Creatinine	
	Phosphorous	Estimated glomerular filtration rate (eGFR)	
	Pregnancy testing	Lactate dehydrogenase (LDH)	
	T4*	Potassium	
	Thyroid-stimulation hormone	Sodium	
	Total protein	Total bilirubin	
		Total cholesterol*	
		Triglycerides*	
Urinalysis	Creatinine		Protein
	Ketones		Glucose
	Leukocyte esterase		
	Macroscopic examination		
	Microscopic examination (completed only if any part of the urinalysis is not negative)		
	Nitrites		
	Occult blood		
	pH		
	Specific gravity		

	Urobilinogen		

*Performed at screening only – not included in summary tables.

Grading will be derived based on criteria noted in Table 5 and detailed in Appendix 9.5. The number and percentage of treated subjects with at least one on-treatment laboratory assessment will be summarized by treatment group (regardless of baseline) for Grade 0, Grade 1 to 2, Grade 3 to 4, Grade 3, and Grade 4. Note that Grade 3 to 4 abnormalities are considered clinically significant.

Shift from baseline to maximum observed laboratory abnormality (based on either normal limits or grading where available) will be tabulated by test for treated subjects. Shift tables will only include treated subjects with an assessment for the specific test of interest at baseline and on treatment.

For the liver function tests, AST, ALT, ALP, and TBILI, the shift from baseline will be presented based the maximum observed on-treatment assessment. Baseline in the shift tables will be defined as per Section 6.2.5; 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 overall incidence at maximum as well as 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 overall incidence at maximum as well as 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 overall incidence at maximum as well as 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

The following categories will be used to summarize the overall incidence at maximum as well as the shift from baseline based on the ULN range for GGT:

- ≤ULN
- >ULN to ≤ 2.5x ULN
- >2.5x 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 the Safety Population. 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, by lab test, will be presented for subjects with an abnormal laboratory value considered potentially clinically significant (i.e. Grade 3 or 4) as well as for 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 (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. For these abnormality listing all of the subject lab values for that test will be listed with the ones meeting criteria flagged.

6.4.6 Columbia-Suicide Severity Rating Scale

The Columbia-Suicide Severity Rating Scale (C-SSRS) is an assessment tool that evaluates suicidal ideation and behavior(Posner, Brown et al. 2011). This tool, through a series of simple, plain-language questions that anyone can ask supports in the assessment of an individual's suicide risk. The answers to this series of questions help users identify whether someone is at risk for suicide, assess the severity and immediacy of that risk, and gauge the level of support that the person needs.

There are two versions that will be used for the trial: "Baseline (to include lifetime assessment and within the last six (6) months) and Since last visit. The C-SSRS is administered by a certified rater. The C-SSRS will be completed on a paper form at the site. Subjects who have any positive

("yes") C-SSRS response to suicidal ideations in the last 6 months at screening and/or (before dosing) baseline will be excluded from participation in the trial.

For each of the 11 individual questions of the C-SSRS, the number and percentage of subjects reporting a 'yes' at any time during treatment (i.e after first day of dosing and up to 30 days after last dose of study medication) (see Section 7.2) will be reported.

The C-SSRS assessment date is used to slot measurements into safety analysis periods. C-SSRS questions are as follows:

- Suicidal ideation section: 5 questions with yes/no responses: (1) wish to be dead; (2) nonspecific active suicidal thoughts; (3) active suicidal ideation with any methods (not plan) without intent to act; (4) active suicidal ideation with some intent to act, without specific plan; and (5) active suicidal ideation with specific plan and intent

- If the responses to Questions 1 and 2 are both "no", then Questions 3 through 5 and the intensity of ideation section are not completed. If the response to Question 2 is "yes", then Questions 3 through 5 are completed. If the response to Question 1 or 2 is "yes", then the intensity of ideation section is completed.

- Intensity of ideation section

    - Most severe ideation, rated as 1 (wish to be dead) to 5 (active suicidal ideation with specific plan and intent)

    - 5 questions rated as 0 to 5: (1) frequency; (2) duration; (3) controllability; (4) deterrents; and (5) reasons for ideation

- Suicidal behavior section: 6 questions with yes/no responses about suicidal behaviors: (1) actual attempt; (2) non-suicidal self-injurious behavior; (3) interrupted attempt; (4) aborted attempt; (5) preparatory acts or behavior; and (6) suicidal behavior.

- Actual attempts section: for the most recent, most lethal, and first/initial attempts, the following are specified:

- Actual lethality/medical damage: rated on a scale of 0 (no physical damage) to 5 (death)

    - Potential lethality (only answer if actual lethality = 0): rated on a scale of 0 to 2.

C-SSRS Parameters

C-SSRS parameters are derived as follows:

- Number of subjects with C-SSRS data: non-missing response to any of the 5 suicidal ideation questions or 6 suicidal behavior questions

- Suicidal ideation score: derived from the 5 suicidal ideation questions as follows:
  - If the response to Question 5 is "yes", then the score is 5;
  - Otherwise, if the response to Question 4 is "yes", then the score is 4;
  - Otherwise, if the response to Question 3 is "yes", then the score is 3;
  - Otherwise, if the response to Question 2 is "yes", then the score is 2;
  - Otherwise, if the response to Question 1 is "yes", then the score is 1;
  - Otherwise, if the response to Question 1 is "no", then the score is 0.
- Suicidal ideation: being in any of the 5 suicidal ideation subcategories below
  - Wish to be dead: "yes" response to suicidal ideation Question 1
  - Non-specific active suicidal thoughts: "yes" response to suicidal ideation Question 2
  - Active suicidal ideation with any methods (not plan) without intent to act: "yes" response to suicidal ideation Question 3
  - Active suicidal ideation with some intent to act, without specific plan: "yes" response to suicidal ideation Question 4
  - Active suicidal ideation with specific plan and intent: "yes" response to suicidal ideation Question 5
- Suicidal behavior: being in any of the 5 suicidal behavior subcategories below
  - Preparatory acts or behavior: "yes" response to suicidal behavior Question 5
  - Aborted attempt: "yes" response to suicidal behavior Question 4
  - Interrupted attempt: "yes" response to suicidal behavior Question 3
  - Non-fatal actual attempt: (a) "yes" response to suicidal behavior Question 1, and (b) actual lethality/medical damage rating not equal to 5 (death) for all attempts (most recent, most lethal, first/initial)
  - Completed suicide: actual lethality/medical damage rating of 5 for any attempt
- Non-suicidal self-injurious behavior: "yes" response to suicidal behavior Question 2.

C-SSRS Suicidality Frequency Table

C-SSRS suicidality is tabulated as the number and percentage of subjects with C-SSRS data in safety analysis periods in the following categories:

- Suicidal ideation
    - Subcategories: wish to be dead; non-specific active suicidal thoughts; active suicidal ideation with any methods (not plan) without intent to act; active suicidal ideation with some intent to act, without specific plan; and active suicidal ideation with specific plan and intent.
- Suicidal behavior
    - Subcategories: preparatory acts or behavior; aborted attempt; interrupted attempt; nonfatal actual attempt; and completed suicide.
- Suicidal ideation or behavior
- Non-suicidal self-injurious behavior.

Results for all categories and subcategories are presented, even those with 0 counts.

C-SSRS Listing

A by-subject C-SSRS listing displays the suicidal ideation score and responses to all questions for subjects with suicidal ideation, suicidal behavior, or non-suicidal self-injurious behavior at any time point. Otherwise, the listing only displays the suicidal ideation score.

6.4.7 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”) in subjects who have received at least one dose of troriluzole

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 Narrow SMQ

These select events are described in the current version (v4) 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 subjects who received at least one dose of troriluzole with the select events as described above.

7 CONVENTIONS

7.1 Derived Dates

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.

7.2 Analysis Periods

Analysis periods and study phases are defined as follows:

- Pre-Treatment: will include all assessments on or before the first day of study drug.
- On-Treatment (Efficacy Assessments): "on-treatment" will include all efficacy assessments after first day of study drug
- On-Treatment (Safety Assessments excluding AEs): except where indicated, "on-treatment" will include all assessments after first day of study drug and up to the last day of study drug + 30 days.
- Treatment-emergent AEs (TEAEs): AEs with a start date on or after the first dose of study drug and up to 30 days after the last dose of study drug. TEAEs will be assessed from the date of first treatment until 30 days after the last dose of study drug.

7.3 Analysis visit windows

The protocol-specified visit window is ± 3 days during the Randomization Phase of the study, however, analysis windows will be continuous to include all data. Refer to Table 6 and Table 7 for details on the protocol-specified day of evaluation and associated visit windows used for efficacy and safety analyses. The baseline visit is expected to be the date of randomization and first dose of randomized treatment.

Table 6: On-Treatment Analysis Windows For Visit (Vital Signs and C-SSRS)

Evaluation	Protocol Specified Day	Analysis Specified Interval
Week 2	Day 14	Day 2-21
Week 4	Day 28	Day 22-42
Week 8	Day 56	Day 43-63
Week 10	Day 70	Day 64-87

Note: Baseline will be defined as the last available assessment on or before first day of study drug.

Table 7: On-Treatment Analysis Windows For Visit (All Other Assessments)

Evaluation	Protocol Specified Day	Analysis Specified Interval
Week 4	Day 28	Day 2-42
Week 8	Day 56	Day 43-63
Week 10	Day 70	Day 64-87

Note: Baseline will be defined as the last available assessment on or before first day of study drug.

If a subject has more than one record within an analysis window, then the latest non-missing value in the window will be used, except as noted in Section 6.4 for safety assessments.

Note that in some cases, endpoints are not measured at each of the visits listed in the above table. For example, the BABS is not measured at Visit Week 4 or Visit Week 8 during the Randomization Phase and ECG is not measured at 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, for BABS) for any change from baseline summaries. Abnomality summaries would include all assessments onsidered "on-treatment".

8 CONTENT OF REPORTS

Planned analyses described in Section 1.1 focus on the following endpoints:

[redacted]

[redacted]

9 APPENDICES

9.1 Schedule of Assessments

Table 8: Schedule of Assessments and Events

Visit	Screening	Baseline	Week 2	Week 4	Week 8	Week 10 or early term	Week 2 Post Last Dose	Comments
Day	**up to -42**	**0**	**14[a]**	**28[a]**	**56[a]**	**70[a]**	**84[a]**	
Eligibility Assessments								
Informed Consent	X							
Pharmacogenetic Informed Consent	X							
Inclusion/Exclusion	X	X						
MINI	X							
Borderline Personality Disorder Module (BPD Module)	X							
MGH-TRQ-OCD	X							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
Medical History	X							To include: smoking history, cardiovascular disease, family, OCD hx and hx of tic disorder.
Demographic Assessment	X							
Disease History	X							

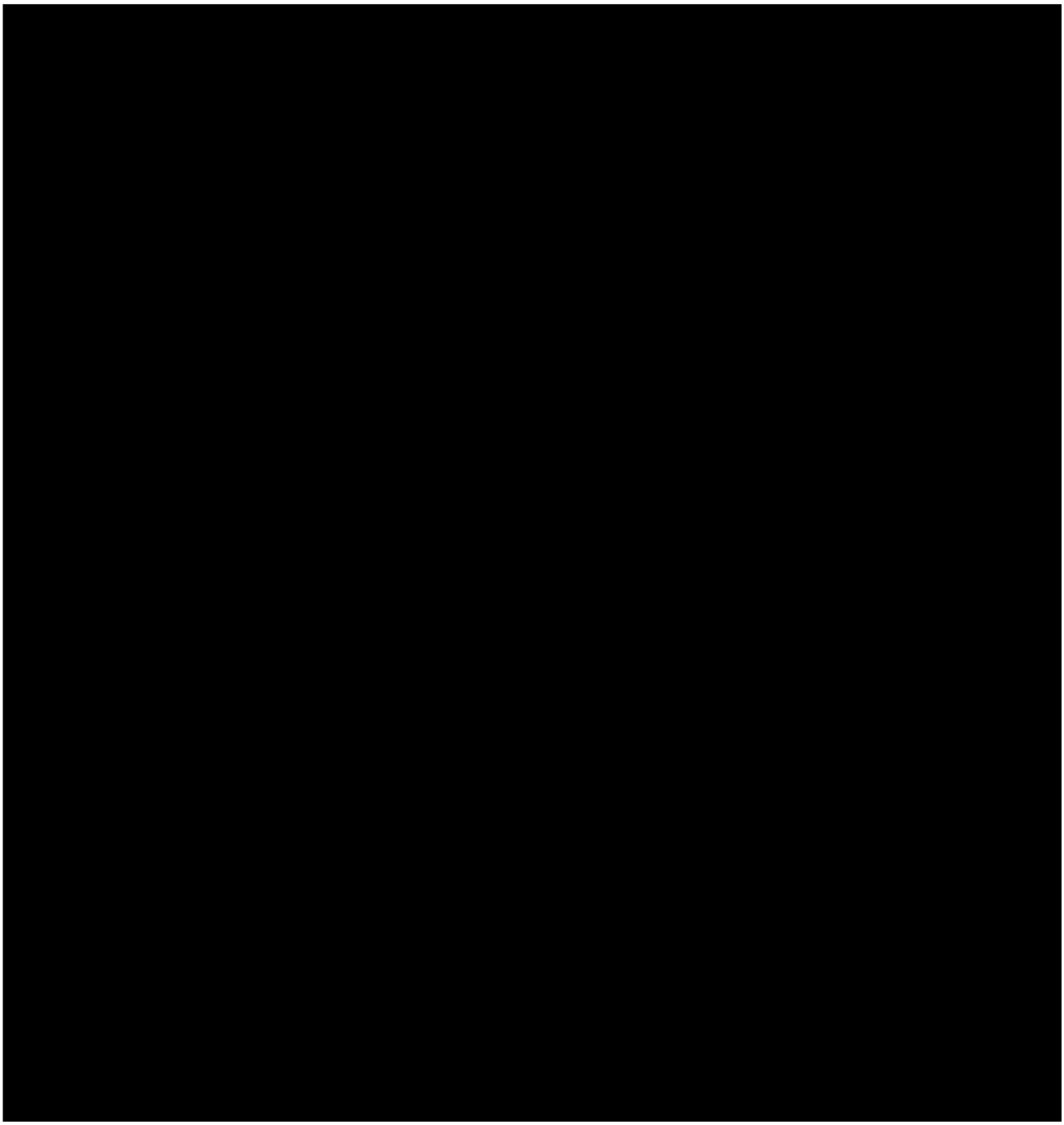
Visit	Screening	Baseline	Week 2	Week 4	Week 8	Week 10 or early term	Week 2 Post Last Dose	Comments
Day	**up to -42**	**0**	**14[a]**	**28[a]**	**56[a]**	**70[a]**	**84[a]**	
SAFER Interview	X							The SAFER Interview will be conducted remotely with the subject by a Rater at CTNI shortly after the screening visit and prior to Baseline. A SAFER pass is necessary for randomization.
Safety Assessments								
Adverse Event Assessment	X	X	X	X	X	X	X	Non-serious AEs will be collected from the initiation of study drug. SAEs will be collected from subject's written consent to within 30 days of discontinuation of dosing.
Laboratory Assessments including urinalysis	X	X		X	X	X		Laboratory assessments are not required to be fasting.
Serology	X							Serology includes; HBsAg, HCV, HIV antibody and RPR testing
Pregnancy testing (serum)	X	X		X	X	X		The site may test a subject at any time if pregnancy is suspected.
Pregnancy testing (urine)		X	X	X	X	X		WOCBP subjects must have a urine pregnancy tests performed prior to the baseline dose. The site may test a subject at any time if pregnancy is suspected.
Urine drug test	X	X				X		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

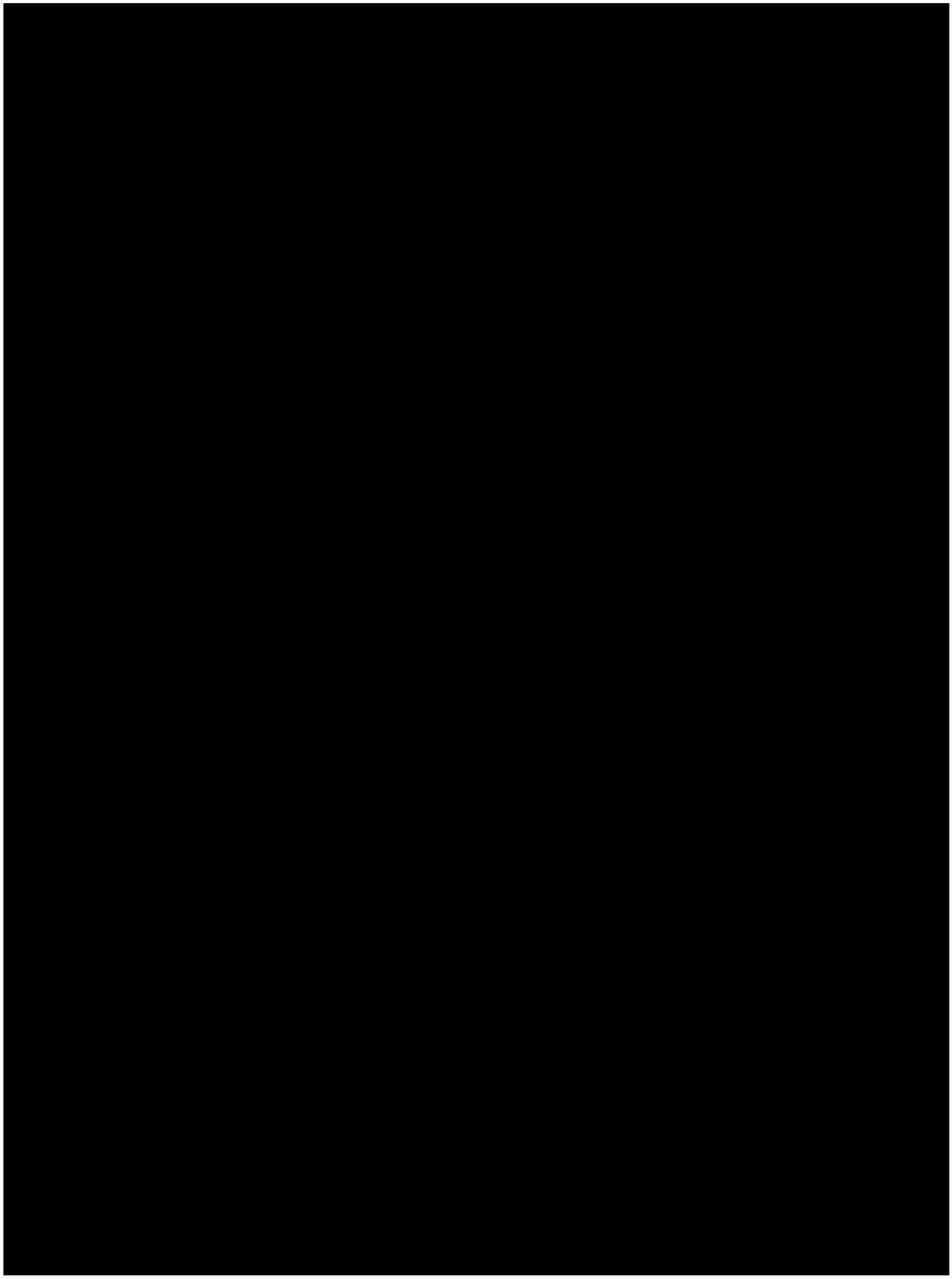
Visit	Screening	Baseline	Week 2	Week 4	Week 8	Week 10 or early term	Week 2 Post Last Dose	Comments
Day	**up to -42**	**0**	**14[a]**	**28[a]**	**56[a]**	**70[a]**	**84[a]**	
								vendor for all positive urine drug screen samples
Complete Physical Exam	X					X		To include: HEENT, neck, lymph nodes, lungs, cardi., abdomen, skin and musculoskeletal.
Physical Measurements	X					X		Height Weight: The same scale should be used for a given subject throughout their study participation. Subject should void just before being weighed. Should be recorded before a meal and at approx. the same time each day Subject should be minimally clothed (no shoes or heavy garments).
Vital Signs	X	X	X	X	X	X	X	Vital signs (Temp., BP, HR)
12-Lead ECG	X	X		X		X		
Concomitant Medication Review	X	X	X	X	X	X	X	
Columbia-Suicide Severity Rating Scale (C-SSRS)	X	X	X	X	X	X	X	A certified clinician rated assessment.
Clinical Outcome Assessments								
Placebo-Control Reminder Script (PCRS)		X		X	X	X		The PCRS should be read by a clinician to a subject prior to every administration of the Y-BOCS at every visit (except screening) for every subject.

Visit	Screening	Baseline	Week 2	Week 4	Week 8	Week 10 or early term	Week 2 Post Last Dose	Comments
Day	**up to -42**	**0**	**14[a]**	**28[a]**	**56[a]**	**70[a]**	**84[a]**	
Yale-Brown Obsessive-Compulsive Scale (Y-BOCS)	X	X		X	X	X		A certified clinician administered scale
Clinical Global Impressions-Improvement Scale (CGI-I)				X	X	X		A certified clinician rated assessment.
Clinical Global Impressions-Severity Scale (CGI-S)	X	X		X	X	X		A certified clinician rated assessment.
Sheehan Disability Scale (SDS)	X	X		X	X	X		A subject-rated measure. Note on scale completion: 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. Please refer to protocol section 6.4.3.
Quick Inventory of Depressive Symptomatology-Self Report (QIDS-SR)	X	X		X	X	X		A subject "self" report. The QIDS-SR should be reviewed prior to the administration of the C-SSRS.
Beck Anxiety Inventory (BAI)	X	X		X	X	X		A subject "self" report.
Brown Assessment of Beliefs (BABS)	X	X				X		A certified clinician rated assessment.
Dimensional Obsessive Complusive Scale (DOCS)		X		X	X	X		A subject "self" report
Biomarker and Other Assessments								
Pharmacokinetics Blood Sample				X	X	X		1. PK samples should also be drawn when there are any SAEs or severe AEs that are possibly drug related. 2. Subjects who are able to schedule a morning visit for Week 4, Week 8 and Week 10 can be instructed to

Visit	Screening	Baseline	Week 2	Week 4	Week 8	Week 10 or early term	Week 2 Post Last Dose	Comments
Day	**up to -42**	**0**	**14[a]**	**28[a]**	**56[a]**	**70[a]**	**84[a]**	
								hold their dose of study drug that morning until after a PK trough sample is obtained, if possible and appropriate
Pharmacogenomics Blood Sample		X				X		
Clinical Drug Supply								
Randomization		X						Study Drug will be dispensed at the baseline visit. Subjects should take the first dose in the morning the day after the baseline visit
Dispense Study Drug		X	X	X	X			The first dose should be taken the day after the baseline visit. Study medication should be administered in the morning.
Drug Accountability			X	X	X	X		

[a] Visit window +/- 3 days





[redacted]

9.4 Relevant Protocol Deviations

Relevant **eligibility protocol deviations** include the following categories:

- Subjects did not provide a written signed informed consent form/forms (IRB/EC specific) prior to the initiation of any protocol required procedures.
- Subject not between the ages of 18 - 65, inclusive
- Primary diagnosis of OCD not confirmed as per Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition by the MINI at Screening; The duration of the subject's illness must be ≥ 1 year;
- Subjects was not currently experiencing non-response or inadequate response to their current SOC medication defined as: a. Subjects Y-BOCS total score must be ≥ 22 at Screening and Baseline, reflecting moderate or severe OCD symptoms.
- Subject not currently on an SSRI (with the exception of fluvoxamine, see Section 1.1.3 of the protocol), or clomipramine, venlafaxine or desvenlafaxine monotherapy treatment for an adequate duration and at an adequate dose defined as:
    - Adequate Duration: At least 8 weeks at Screening and 12 weeks at Baseline of SSRI (with the exception of fluvoxamine, see Section 1.1.3 of the protocol), clomipramine, venlafaxine or desvenlafaxine;
    - Adequate Dose: Defined by the table in protocol, section 5.2.
- CGI-S score is not ≥ 4 at screening and baseline;
- Subject did not "pass" the SAFER interview performed by CTNI (Rater Training vendor) prior to randomization;
- Woman of child bearing potential does not have a negative serum pregnancy test at screening and a negative urine pregnancy test prior to dosing at Baseline;
- For subjects consented to version 1 of protocol only: Subjects with a history of more than two (2) previous failed or inadequate treatment trials of SSRIs, clomipramine, venlafaxine, or desvenlafaxine, (not including the current SSRI trial) given for an adequate duration at an adequate dose as defined by the following criteria taken from the MGH-TRQ-OCD as:
    - Treatment non—response / inadequate treatment response: As per the MGH-TRQ-OCD, there has been minimal or no meaningful clinical benefit as perceived by the subject despite an adequate dose and duration of treatment as defined by;
        - Adequate duration: At least 12 weeks of treatment
        - Adequate dose: Defined by the table in protocol section 5.3.

For protocol version 2 and greater, this exclusion criteria based on Exclusion criteria #1 was no longer considered programable.

- Current or prior history, per DSM-5 criteria, of bipolar I or II disorder, schizophrenia or other psychotic disorders, schizoaffective disorder, autism or autistic spectrum disorders, borderline personality disorder, antisocial personality disorder, body dysmorphic disorder, hoarding disorder (symptoms of hoarding disorder as part of the OCD diagnosis are allowed, but a primary diagnosis of hoarding disorder is excluded); a current diagnosis of Tourette's disorder is also excluded;

- Subject has an eating disorder within the last 12 months;

- Primary active major depressive episode or primary active anxiety disorder within the past 6 months. Note: Subjects on a stable maintenance dose of a non-tricyclic, non-monoamine oxidase inhibitor (MAOI) antidepressant medication may be eligible if the subject has been treated with a stable dose for at least 3 months prior to randomization and no dose changes are expected throughout the randomization phase of the study;

- Any positive ("yes") C-SSRS response to questions 1-5 in last 6 months at screening and/or (before dosing) baseline

- Total BABS score >17 at screening and baseline;

- History of psychosurgery, Deep Brain Stimulation (DBS) or Electroconvulsive Therapy (ECT).

- Positive urine drug screening for cannabis (both medical and recreational use of cannabis are prohibited; subjects will be expected to refrain from use during the period of the study), amphetamines (including MDMA/ecstasy), cocaine, barbiturate, PCP, and/or opiates during screening (subjects with a positive result for cannabis, that agree to refrain from use during the period of the study and test negative upon retest during the screening phase may participate);

- Clinical history of stroke or seizure disorder.

- Body mass index >40 kg/m2;

- Vitamin B12 or folate deficiency Note: Subjects with a B12 deficiency can participate in the study if they are on stable Vitamin B12 replacement for at least 3 months prior to randomization and their B12 levels are within normal limits prior to randomization;

- Positive syphilis serology including rapid plasma reagin [RPR] test or the toluidine red unheated serum test (TRUST) for subjects in China) and positive confirmatory testing;

- Aspartate Aminotransferase (AST), Alanine Aminotransferase (ALT) or GGT > 1.5 times the upper limit of normal at baseline

- Total bilirubin > 2 times the upper limit of normal at baseline and no documented history of Gilbert's Syndrome. For subjects consented to version 2 of protocol or later, if documented history of Gilbert's subject must have total bilirubin < 5 mg/dL

- Neutrophils, Absolute < 1000/mm3 at baseline (for subjects consented to version 1 of protocol), <1500/mm^3 (for subject consented to version 2 and later versions of protocol)

- QTcF(Fridericia) interval ≥470 msec during the screening or baseline period

- Positive pregnancy test post baseline

Relevant **subject management deviations** include the following categories:

- Randomization error
    - Randomized to incorrect strata
- Study drug dosing error, defined as any of the following subcategories:
    - Study drug compliance < 80% as defined in Section 6.2.6.1
    - Dosing > 280 mg per day for at least one day
    - Incorrect medication dispensed
    - >= four consecutive days without exposure to study drug
- Prohibited non-study medications, defined as any of the following subcategories:
    - Previous treatment with riluzole
    - Use of tricyclic antidepressants and mono-amine-oxidase (MAO) inhibitors are prohibited 30 days prior to randomization (baseline visit) and during the study (with the exception of clomipramine);
    - Use of a stimulant, neuroleptic (antipsychotic), mood stabilizer and glutamate agent (e.g. gabapentin, pregabalin, topiramate, lamotrigine, N- acetylcysteine, ketamine, memantine, sodium valproate, lithium) is prohibited within the 4 weeks prior to screening and during the study;
    - Use of varenicline is prohibited 30 days prior to randomization (baseline visit) and during the randomization phase of the study;
    - Herbal medication and herbal supplement use within 30 days of randomization and during the course of the study is prohibited;
    - Transcranial Magnetic Stimulation (TMS) is prohibited within three months prior to screening and during the study.
- Management of subject discontinuations

- Subject who per protocol potentially should have been discontinued given occurrence of an AE, ECG, and laboratory abnormalities (confirmed by repeat testing if appropriate) or intercurrent illness included in the list in Section 6.5.1 of the protocol.

9.5 Clinical Laboratory Grading

Table 11: Clinical Laboratory Evaluations by Grade Level

Panel[1]	Grade 1 (Mild)	Grade 2 (Moderate)	Grade 3 (Severe)	Grade 4 (Potentially Life-Threatening)
Hematology				
Absolute lymphocyte count (10^9/L) - *Low*	>=0.8 to <LLN	>=0.5 to <0.8	>=0.2 to <0.5	<0.2
Absolute lymphocyte count (10^9/L) - *High*		>4 - 20	>20	
Absolute neutrophil count (10^9/L)	>=1.5 to <LLN	>=1.0 to <1.5	>=0.5 to <1.0	<0.5
Hemoglobin (g/L)	>=100 to <LLN	>=80 to <100	<80	
Platelet Count (10^9/L)	>=75 to <LLN	>=50 to <75	>=25 to <50	<25
White blood cell count (10^9/L)	>=3 to <LLN	>=2 to <3	>=1 to <2	<1
Serum Chemistry				
ALT	>ULN - 3.0*ULN if baseline was normal; >=1.5-3.0*baseline if baseline was abnormal	>3.0 - 5.0*ULN if baseline was normal; >3.0 - 5.0*baseline if baseline was abnormal	>5.0 - 20.0*ULN if baseline was normal; >5.0 - 20.0*baseline if baseline was abnormal	>20.0*ULN if baseline was normal; >20.0*baseline if baseline was abnormal
Albumin (g/L)	>=30 to <LLN	>=20 to <30	<20	
ALP	>ULN - 2.5*ULN if baseline was normal; >=2.0-2.5*baseline if baseline was abnormal	>2.5 - 5.0*ULN if baseline was normal; >2.5 - 5.0*baseline if baseline was abnormal	>5.0 - 20.0*ULN if baseline was normal; >5.0 - 20.0*baseline if baseline was abnormal	>20.0*ULN if baseline was normal; >20.0*baseline if baseline was abnormal
AST	>ULN – 3.0*ULN if baseline was normal; >=1.5-3.0*baseline if baseline was abnormal	>3.0 - 5.0*ULN if baseline was normal; >3.0 - 5.0*baseline if baseline was abnormal	>5.0 - 20.0*ULN if baseline was normal; >5.0 - 20.0*baseline if baseline was abnormal	>20.0*ULN if baseline was normal; >20.0*baseline if baseline was abnormal
Bicarbonate	<LLN			
Calcium (mmol/L) - *Low*	≥ 2.0 to < LLN mmol/L	≥ 1.75 to < 2.0 mmol/L	≥ 1.5 to < 1.75 mmol/L	
Calcium (mmol/L) - *High*	> ULN to ≤ 2.9	> 2.9 to ≤ 3.1	> 3.1 to ≤ 3.4	> 3.4
Creatine kinase	> ULN to ≤ 2.5 x ULN	> 2.5 to ≤ 5 x ULN	> 5 to 10 x ULN	> 10 x ULN
Creatinine	> ULN to ≤ 1.5 x ULN	{> 1.5 to ≤ 3.0 x ULN} or {> 1.5 to ≤ 3.0 x baseline}	{> 3.0 to ≤ 6.0 x ULN} or {> 3.0 x baseline}	> 6.0 x ULN
Glucose (mmol/L) - *Low*	3.05 - <3.55	2.22 - <3.05	1.67 - <2.22	<1.67

Panel[1]	**Grade 1 (Mild)**	**Grade 2 (Moderate)**	**Grade 3 (Severe)**	**Grade 4 (Potentially Life-Threatening)**
Glucose (mmol/L) - *High Fasting*	6.11 - <6.95	6.95 - <13.89	13.89 - <27.75	>=27.75
Glucose (mmol/L) - *High Not Fasting*	6.44 - <8.89	8.89 - <13.89	13.89 - <27.75	>=27.75
Lactate dehydrogenase	>ULN			
Potassium (mmol/L) - *Low*	≥ 3.0 to < LLN		≥ 2.5 to < 3.0	< 2.5
Potassium (mmol/L) - *High*	> ULN to ≤ 5.5	> 5.5 to ≤ 6.0	> 6.0 to ≤ 7.0	> 7.0
Sodium (mmol/L) - *Low*	≥ 130 to < LLN	≥ 125 to < 130	≥ 120 to < 125	< 120
Sodium (mmol/L) - *High*	> ULN to ≤ 150	> 150 to ≤ 155	> 155 to ≤ 160	> 160
Total bilirubin	>ULN - 1.5*ULN if baseline was normal; >1.0-1.5*baseline if baseline was abnormal	>1.5 - 3.0*ULN if baseline was normal; >1.5 - 3.0*baseline if baseline was abnormal	>3.0 - 10.0*ULN if baseline was normal; >3.0 - 10.0*baseline if baseline was abnormal	>10.0*ULN if baseline was normal; >10.0*baseline if baseline was abnormal
Total cholesterol (mmol/L)	> ULN to ≤ 7.75	> 7.75 to ≤ 10.34	> 10.34 to ≤ 12.92	> 12.92
Triglycerides (mmol/L)	≥ 1.71 to ≤ 3.42	> 3.42 to ≤ 5.7	> 5.7 to ≤ 11.4	> 11.4
Uric acid (umol/L)	450 - <590	590 - <710	710 - <890	>=890
Urinalysis				
Protein	Trace or 1+ ≥ 10 to < 100 mg/dL	2+ ≥ 100 to ≤ 300 mg/dL	3+ or higher ≥ 300 mg/dL	
Urine Glycosuria	Trace or 1+; > 180 to ≤ 250 mg/dL	2+; > 250 to ≤ 500 mg/dL	3+ or higher; > 500 mg/dL	

[1]Graded by CTCAE Version 5.0 (2017) or DAIDS Version 2.1 (2017)

10 REFERENCES

Carlsson, M. L. (2000). "On the role of cortical glutamate in obsessive-compulsive disorder and attention-deficit hyperactivity disorder, two phenomenologically antithetical conditions." Acta Psychiatr Scand **102**(6): 401-413.

Coric, V., S. Milanovic, S. Wasylink, P. Patel, R. Malison and J. H. Krystal (2003). "Beneficial effects of the antiglutamatergic agent riluzole in a patient diagnosed with obsessive-compulsive disorder and major depressive disorder." Psychopharmacology **167**: 219-220.

Coric, V., S. Taskiran, C. Pittenger, S. Wasylink, D. H. Mathalon, G. Valentine, J. Saska, Y. T. Wu, R. Gueorguleva, G. Sanacora, R. Malison and J. H. Krystal (2005). "Riluzole augmentation in treatment-resistant obsessive-compulsive disorder: an open-label trial." Biological Psychiatry **58**: 424-428.

[redacted]

Pittenger, C., M. H. Bloch, S. Wasylink, E. Billingslea, R. Simpson, E. Jakubovski, B. Kelmendi, G. Sanacora and V. Coric (2015). "Riluzole augmentation in treatment-refractory obsessive-compulsive disorder: a pilot randomized placebo-controlled trial." J Clin Psychiatry **76**(8): 1075-1084.

Pittenger, C., M. H. Bloch and K. Williams (2011). "Glutamate abnormalities in obsessive compulsive disorder: neurobiology, pathophysiology, and treatment." Pharmacol Ther **132**(3): 314-332.

Posner, K., G. K. Brown, B. Stanley, D. A. Brent, K. V. Yershova, M. A. Oquendo, G. W. Currier, G. A. Melvin, L. Greenhill, S. Shen and J. J. Mann (2011). "The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults." Am J Psychiatry **168**(12): 1266-1277.

Rubin, D. B. (1987). Multiple Imputation for Nonresponse in Surveys. New York, NY USA, John Wiley & Sons Inc.

Schafer, J. L. (1997). Analysis of Incomplete Multivariate Data. New York, NY USA, Chapman and Hall.