Biohaven Pharmaceuticals

Protocol BHV4157-206

A Phase III, Long-Term, Randomized, Double-blind, Placebo-controlled Trial of Troriluzole in Adult Subjects with Spinocerebellar Ataxia

Statistical Analysis Plan

Version 3.0

Date: 27-April-2022

SIGNATURE PAGE

Protocol Title:	A Phase III, Long-Term, Randomized, Double-blind, Placebo-controlled Trial of Troriluzole in Adult Subjects with Spinocerebellar Ataxia
Sponsor:	Biohaven Pharmaceuticals, Inc.
Document Version:	3.0
Date:	27-April-2022
Author:	[redacted] Cytel, Inc.

Signature:_____

Date:_____

Sponsor Approval

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

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

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

Sponsor Signatories:

[redacted]
Biohaven Pharmaceuticals

Signature:_____

Date:_____

[redacted]
Biohaven Pharmaceuticals

Signature:_____

Date:_____

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

REVISION HISTORY

Version	Description of Change
V1.0	Original Version based on Protocol Version 6 and Version 7 China Amendment
V2.0	Based on Protocol Version 6 and Version 7 China Amendment Section 1: Specified that SAP is based on Version 6 of the protocol Section 3.2.2: Clarified FARS-FUNC only includes a total score Section 6.2.4: Added description for China protocol deviations; reference to new section (9.3) that details programmable deviations; clarified analysis set for reporting of deviations Section 6.2.6.1: Added definition of average daily dose Sections 6.3.1.1 and 6.3.1.2: Removed df determined by Satterthwaite approximation Section 6.3.2.1: Added derivation for pro-rating score if missing scale items Section 6.3.2.2: Added country to the parameter-estimation model; updated tipping point analysis; added sensitivity of primary endpoint to exclude sites in China Sections 6.3.3.1, 6.3.3.2, 6.3.3.3: Added derivation for pro-rating score if missing scale items; added sensitivity of endpoint to exclude sites in China Sections 6.3.4.1, 6.3.4.2, 6.3.4.3: Added derivation for pro-rating score if missing scale items per noted reference Section 6.4.3: Added management of local lab reference ranges Section 7.1: New section - added summary statistics rounding rules Section 7.2: Analysis Periods moved to section 7.2; clarified the periods (start and stop) for efficacy and safety for double-blind randomization phase, OL extension phase and on study Section 7.3: Analysis Visit Windows moved to section 7.3; clarified the baseline assessment window and measurements may have been during the screening phase Section 9.3: New section titled Relevant Protocol Deviations to include details for programmable relevant protocol deviations Section 9.4: New section titled Non-Quest Local Laboratory Normal Ranges – Data Handling Conventions Section 10: Yang Yuan reference added; Neuro-QOL scoring manual added
V3.0	Based on Protocol Version 6 and Version 7 China Amendment Updated list of abbreviations Updated references to sections Fix typos Section 6.1.2: Clarified precision of statistics presented Section 6.4.3: Table 5: eGFR MDRD added and removed incorrect acronym for absolute lymphocyte count; removed ‘at each visit’ in description of summary table of graded laboratory measurements Section 7.3: remove window for baseline assessment; Table 6, corrected start day of Ext Week 96 for Neuro-QOL and CGI-I, PGI-C

	Section 9.2: Table 10, clarified qualifier for absolute neutrophil count for grade 2 and 3; added eGFR MDRD, moved Glucose to Chemistry, added LDL cholesterol, added urine glucose, and updated grade 1 protein (urinalysis). Added notes to table to clarify selection of toxicity grades. Renumbered appendices

ABBREVIATIONS

Abbreviation	Definition
AE	Adverse events
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
BLQ	Below the limit of quantification
BUN	Blood urine nitrogen
CGI-I	Clinical Global Impression - Improvement
CI	Confidence interval
CPK	Creatine phosphokinase
CRO	Contract research organization
CRC	Cognitive Research Corporation
CRF	Case Report Form
CSR	Clinical study report
DB	Double-blind
df	Degrees of freedom
ECG	Electrocardiogram
eDISH	Evaluation of drug-induced serious hepatotoxicity
eGFR MDRD	Estimated glomerular filtration rate Modification of Diet in Renal Disease
FACIT	Functional Assessment of Chronic Illness Therapy
FARS-ADL	Functional Activities of Daily Living Scale from the Friedreich's Ataxia Rating Scale
FARS-FUNC	Functional Staging for Ataxia Scale from the Friedreich's Ataxia Rating Scale
f-SARA	Modified Functional Scale for the Assessment and Rating of Ataxia
GGT	Gamma-glutamyl transferase
HbA1c	Hemoglobin A1c
HDL	High-density lipoprotein
ICH	International Conference on Harmonisation
IEC	Independent Ethics Committee
IWRS	Interactive web response system
IRB	Institutional Review Board
LDH	Lactate dehydrogenase
LDL	Low-density lipoprotein
LFT	Liver function test
MAR	Missing at random

Abbreviation	Definition
MedDRA	Medical Dictionary for Regulatory Activities
MCMC	Markov chain Monte Carlo
MI	Multiple imputation
mITT	Modified Intent to Treat
MMRM	Mixed model for repeated measures
MMSE	Mini Mental State Exam
MNAR	Missing not at random
Neuro-QOL	Neurology Quality of Life
OL	Open-label
PGI-C	Patient Global Impression of Change
PI	Primary investigator
PIFAS	Patient Impression of Function and Activities of Daily Living Scale
PK	Pharmacokinetic
PT	Preferred term
RBC	Red blood cell
SAE	Serious adverse event
SAP	Statistical analysis plan
SCA	Spinocerebellar ataxia
SD	Standard deviation
SE	Standard error
SMQ	Standardized MedDRA query
SOC	System organ class
S-STS	Sheehan Suicidality Tracking Scale
TSH	Thyroid-stimulation hormone
ULN	Upper limit of normal
WBC	White blood cell
WHO-DD	World Health Organization Drug Dictionary

1 INTRODUCTION

This document presents the statistical analysis plan (SAP) for Biohaven Pharmaceuticals, Inc. Protocol BHV4157-206: A Phase 3, Long-Term, Randomized, Double-blind, Placebo-controlled study of Troriluzole in Adult Subjects with Spinocerebellar Ataxia.

This SAP is based on version 6 of the protocol dated March 01, 2021 as well as the version 7 China Amendment dated August 13, 2020. It contains the analysis details and methodology to address the study objectives, including planned summary tables, by-subject listings, and figures, which will provide the basis for the results section of the clinical study report (CSR). Operational aspects related to collection and timing of planned clinical assessments are not repeated in this SAP unless relevant to the planned analyses.

The investigation product is BHV-4157, also known as Troriluzole, and these terms are used interchangeably through the document.

1.1 Research Hypothesis

BHV-4157 monotherapy for 48 weeks is superior to placebo in the treatment of Spinocerebellar Ataxia.

1.2 Schedule of Analyses

The final analysis of the Double-Blind (DB) Randomization Phase will occur after the last subject randomized has completed 48 weeks of blinded treatment or discontinues from the DB Randomization Phase. The study will be unblinded and analyses will include all data from the Randomization Phase.

A final analysis of the study will be completed after the last subject has completed their last study visit at the end of the Open-Label (OL) treatment phase. These analyses will summarize all efficacy data collected in the OL Extension Phase as well as safety, laboratory, and other data collected throughout the entire study.

In addition, data may be locked, analyses conducted, and reports produced as required to support safety monitoring or regulatory requirements.

2 STUDY DESCRIPTION

2.1 Study Design

BHV4157-206 is a Phase 3, multicenter, randomized, DB, 2-arm placebo-controlled, parallel-group study designed to assess safety, tolerability, and efficacy in a population of patients with Spinocerebellar Ataxia (SCA).

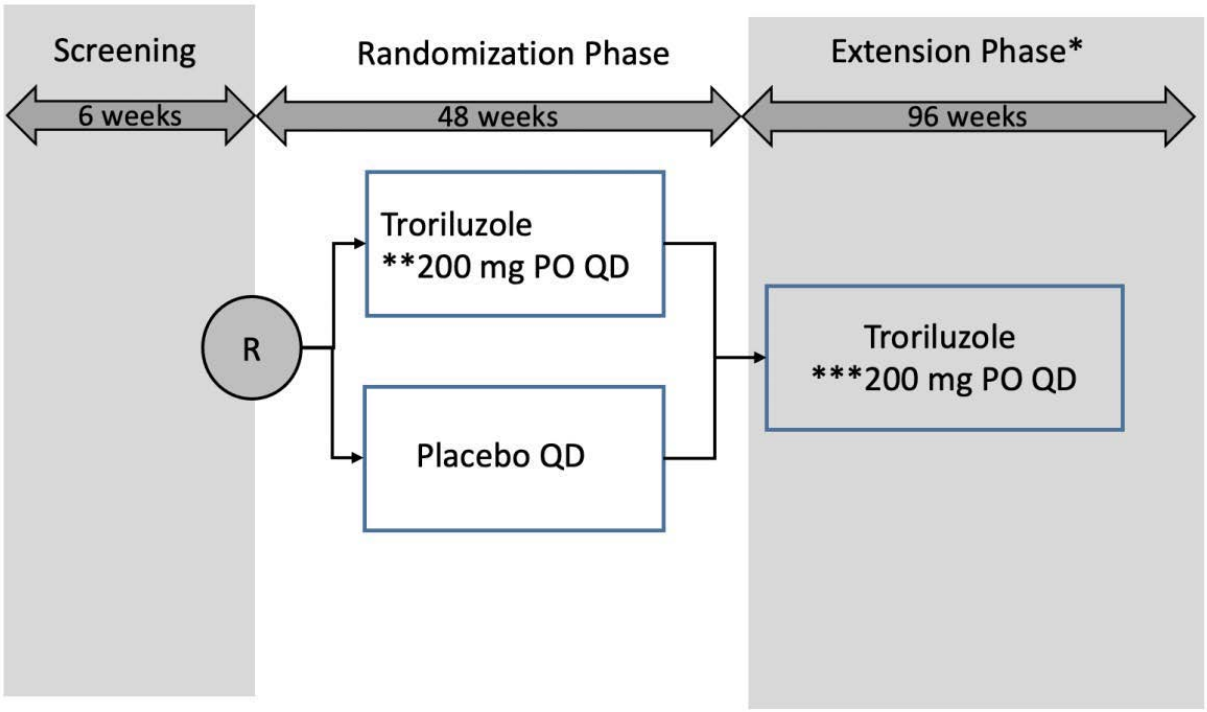
Subjects who are determined to be eligible for the study will enter the Randomization Phase. Dosing will continue for 48 weeks during which subjects will receive either troriluzole (140 mg) or matching placebo for the first 4 weeks and then will be increased to 200 mg (or matching

placebo) for the duration of the study. [redacted] In addition, subjects completing the Randomization Phase will be offered up to 96 weeks of OL treatment provided the Primary Investigator (PI) believes OL treatment offers an acceptable risk-benefit profile. Subjects who agree to enter the Extension Phase will not be required to wash-out the drug or complete the follow-up safety visit, but instead should continue dosing as specified in the Extension Phase.

Subjects entering the Extension Phase will continue with the same dose taken at the end of the Randomization Phase. Subjects on placebo in the Randomization Phase will be switched in a blinded manner to troriluzole 140 mg for the first 4 weeks and then the dose will be increased to 200 mg for the duration of the study. Subjects who enter the Extension Phase on 140 mg due to intolerability issues can be re-challenged to increase to 200 mg after Extension Week 4 at the discretion of the PI. Down titration while on 200 mg will only be allowed for tolerability purposes. All extension visits after Week 4 will be OL. The first extension visit will occur 4 weeks after the Week 48 Randomization Phase visit. Subjects will be assessed at clinic visits per the Schedule of Assessments (Appendix, Section 9.2, Table 8 and Table 9).

Figure 1 displays the study design.

Figure 1: Study Schematic



R signifies randomization; randomization will be stratified by 3 SCA subgroups: SCA1 & 2; SCA3; and SCA 6, 7, 8, 10.

* Eligible subjects for the Extension Phase will include those for whom the PI believes extended treatment with troriluzole would offer an acceptable risk-benefit profile.
**Subjects will receive 140 mg for the first four (4) weeks and will then be increased to 200 mg for the duration of the study. Down titration to 140 mg will only be allowed to address tolerability issues.
***Subjects entering the Extension Phase will continue with the same dose taken at the end of the Randomization Phase. Subjects on placebo in the Randomization Phase will be switched in a blinded manner to 140 mg QD for the first four weeks and then will be increased to 200 mg QD for the duration of the study. Down titration after the first four weeks of the Extension Phase will only be allowed for tolerability purposes. All visits after Week 4 will be open label.

2.2 Treatment Assignment

Approximately 270 subjects will be screened in order to randomize approximately 210 subjects.

Subjects will be randomized to receive placebo (QD) or troriluzole (200 mg QD) and will be stratified by diagnosis (SCA genotype). Subjects will receive either troriluzole (140 mg) or placebo for the first 4 weeks and then the dose will be increased to 200 mg (or matching placebo) for the duration of the study.

Subjects with SCA1 and SCA2 genotypes will comprise approximately 80%-90% of the randomized subjects. Subjects with SCA3 genotype will be limited to approximately 10% of the total number of randomized subjects; and subjects with SCA6, SCA7, SCA8, and SCA10 genotypes will comprise approximately 10% of the randomized subjects in total. Randomization was stratified by: (1) SCA1 and SCA2; (2) SCA3; and (3) SCA6, SCA7, SCA8, and SCA10.

2.3 Blinding and Unblinding

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

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

A pharmacokineticist, bioanalytical scientists, IWRS randomization manager, a pharmacovigilance role, and executive committee members of the DMC may be unblinded during the course of the study in order to perform specific study related activities (i.e. PK assessment, randomization coordination, unblinded safety assessments) before data are unblinded for the primary end point and all subjects complete the study. In order to minimize unnecessary analysis of placebo blood PK samples, the bioanalytical scientist will be unblinded to treatment prior to unblinding for the primary endpoint. Results of the blood concentration assay will be kept secure until database lock unblinding for the primary endpoint.

Additionally, an independent statistician will review the unblinded randomization assignments when 5% and 10% of the subjects have been randomized to ensure the IWRS system is assigning as specified. If enrollment is evenly distributed across the SCA strata, 10% will show approximately 2 blocks filled per strata. If at least 1 block per strata is not filled, then a review will occur when 15% of the subjects are randomized. Except as noted above, other members of the sponsor and CRO research team will remain blinded until the database is locked for the primary analysis.

Although the Extension Phase is OL, subjects on placebo during the Randomization Phase will be switched in a blinded manner to BHV-4157 140 mg for the first 4 weeks, and then increased to BHV-4157 200 mg for the remainder of the Extension Phase. In cases of accidental unblinding, contact the Medical Monitor and ensure every attempt to preserve the blind is made.

3 STUDY OBJECTIVES AND ESTIMANDS

3.1 Objectives

3.1.1 Primary Objective

To compare the efficacy of troriluzole (200 mg once daily) versus placebo on ataxia symptoms in subjects with spinocerebellar ataxia (SCA) after 48 weeks of treatment as measured by the total score on the Modified Functional Scale for the Assessment and Rating of Ataxia [f-SARA]

3.1.2 Secondary Objectives

1. To compare the efficacy of troriluzole versus placebo on patient impression of benefit via use of the Patient Impression of Function and Activities of Daily Living Scale (PIFAS), an internally developed twelve-item instrument modeled after the FACIT-Fatigues Scale.

2. To compare the efficacy of troriluzole versus placebo on activities of daily living as measured by the Activities of Daily Living Scale from the Friedreich's Ataxia Rating Scale (FARS-ADL).

3. To compare the efficacy of troriluzole versus placebo on daily functioning as measured by the Functional Staging for Ataxia Scale from the Friedrich's Ataxia Rating Scale (FARS-FUNC).

4. To assess the safety and tolerability of troriluzole in subjects with SCA during the Randomization and Extension Phases.

3.1.3 Exploratory Objectives

1. To compare the efficacy of troriluzole versus placebo on lower extremity mobility and activities as measured by the Neuro-QOL Lower extremity mobility scale.

2. To compare the efficacy of troriluzole versus placebo on upper extremity mobility and activities as measured by the Neuro-QOL Upper extremity mobility scale.

3. To compare the efficacy of troriluzole versus placebo on daily fatigue and activities as measured by the Neuro-QOL Fatigue scale.

4. To compare the efficacy of troriluzole versus placebo on clinician impression of global functioning via use of the Clinical Global Impression-Global Improvement Scale (CGI-I).

5. To compare the efficacy of troriluzole versus placebo on patient impression of global functioning as measure by the Patient Global Impression Scale (PGI)

[redacted]

3.2 Estimands

An estimand is the target of estimation to address the scientific question of interest posed by a study objective. The 4 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.

Population of Interest

The population of interest for this study is adult, male and female patients, with SCA. Protocol Section 5 provides a detailed description of inclusion and exclusion criteria for this study. Refer to Section 4.1 for analysis sets that are used to assess endpoints.

Intercurrent Events

Intercurrent events are those that occur after treatment initiation and either preclude observation of the endpoint or affect its interpretation.

Note that efficacy and safety data should be collected for subjects who discontinue the study early at an early discontinuation visit as well as the 2-Week Post Dose Visit. The 2-Week Post Dose Visit would not need to occur if the subject discontinued dosing more than 2 weeks prior to the early discontinuation visit.

Subjects who withdrew from the study are not to be replaced.

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

For the main analyses of efficacy for primary, secondary, and exploratory endpoints, no imputation will be performed to impute data following discontinuation from study. In some cases, missing items will be imputed (see individual scales in efficacy section below).

Sensitivity analyses using multiple imputation (MI) will be conducted on the primary endpoint using jump to reference and copy increment reference approaches to assess the impact of the missing at random (MAR) assumption (see Section 6.3.2.2). Additionally, a tipping point analysis will be performed to test how robust the primary analysis method is to departure from the MAR assumption. A complete-case analysis will also be performed as a sensitivity analysis for the primary endpoint to assess any potential difference between subjects who attended all of their scheduled visits and those who did not.

Data Sources for Endpoints

f-SARA scores are from the f-SARA electronic case report form (eCRF).

PIFAS scores are from the PIFAS eCRF.

FARS-ADL and FUNC scores are from the FARS eCRF.

Neuro-QOL scores are from the Neuro-QOL eCRFs.

CGI-I scores are from the C-CGI eCRF.

PGI scores are from the PGI eCRF.

AEs (i.e., non-serious AEs or SAEs) are from AE/SAE eCRFs.

Laboratory test results are from a central laboratory as well as laboratory test eCRFs.

PK concentrations and PK parameters are from an external source.

3.2.1 Primary Objective Estimand

Table 1 presents the estimand for the primary objective.

Table 1: Primary Objective Estimand

Objective	**To compare the efficacy of troriluzole versus placebo on ataxia symptoms in subjects with SCA after 48 weeks of treatment as measured by the total score on the f-SARA**
Efficacy Endpoint	The total score on the Modified Functional Scale for the Assessment and Rating of Ataxia (f-SARA) at Week 48
Summary	Using the efficacy analysis set (mITT, see section 4.1): • Descriptive statistics for f-SARA (total score and individual items) and change from baseline (total score) as a continuous variable by treatment group at each visit • Difference in mean change from baseline in f-SARA scores between treatment groups at randomization Week 48 estimated and tested using a mixed-effect model for repeated measures (MMRM)
Intercurrent Events	1) Main Analysis – Observed data only 2) Sensitivity Analysis – Multiple imputation with jump to reference and copy increment reference 3) Sensitivity Analysis – Tipping Point Analysis 4) Sensitivity Analysis – Complete Case Analysis

3.2.2 Secondary Objective Estimands

Table 2 presents the estimands for secondary objectives.

Table 2: Secondary Objective Estimands

Objective 1	**To compare the efficacy of troriluzole versus placebo on patient impression of benefit via use of the Patient Impression of Function and Activities of Daily Living Scales (PIFAS)**
Efficacy Endpoint	Patient Impression of Function and Activities of Daily Living Scale (PIFAS)
Summary	Using the mITT analysis set: • Descriptive statistics for PIFAS (total score and individual items) and change from baseline (total score and individual items) as a continuous variable by treatment group at each visit • Difference in mean change from baseline in PIFAS scores between treatment groups at randomization week 48 estimated and tested using a mixed-effect model for repeated measures (MMRM)
Intercurrent Events	1) Observed data only 2) Sensitivity Analysis – Multiple imputation with jump to reference and copy increment reference
Objective 2	**To compare the efficacy of troriluzole versus placebo on activities of daily living as measured by the Activities of Daily Living Scale from Friedreich's Ataxia Rating Scale (FARS-ADL)**
Efficacy Endpoint	Activities of Daily Living Scale from the Friedrich's Ataxia Rating Scale (FARS-ADL)
Summary	Using the mITT analysis set: • Descriptive statistics for FARS-ADL (total score and individual items) and change from baseline (total score and individual items) as a continuous variable by treatment group at each visit • Difference in mean change from baseline in FARS-ADL scores between treatment groups at randomization week 48 estimated and tested using a mixed-effect model for repeated measures (MMRM)
Intercurrent Events	1) Observed data only 2) Sensitivity Analysis – Multiple imputation with jump to reference and copy increment reference

Objective 3	**To compare the efficacy of troriluzole versus placebo on daily functioning as measured by the Activities of Daily Living Scale from Friedreich's Ataxia Rating Scale (FARS-FUNC)**
Efficacy Endpoint	Functional Staging for Ataxia (FARS-FUNC)
Summary	Using the mITT analysis set: • Descriptive statistics for FARS-FUNC (total score) and change from baseline (total score) as a continuous variable by treatment group at each visit • Difference in mean change from baseline in FARS-FUNC scores between treatment groups at randomization week 48 estimated and tested using a mixed-effect model for repeated measures (MMRM)
Intercurrent Events	Observed data only
Objective 4	**To assess the safety and tolerability of troriluzole in subjects with SCA**
Safety Endpoint	Frequency and severity of adverse events and discontinuations due to adverse events
Summary	Using the safety analysis set: • Number and percentage of subjects with these events or findings by treatment group and overall
Intercurrent Events	Observed data only

3.2.3 Exploratory Objective Estimands

Table 3 presents the estimands for exploratory objectives.

Table 3: Exploratory Objective Estimands

Objective 1	**To compare the efficacy of troriluzole versus placebo on lower extremity mobility and activities as measured by the Neuro-QOL Lower extremity mobility scale**
Efficacy Endpoint	Neuro-QOL Lower Extremity Function score
Summary	Using the mITT analysis set: • Descriptive statistics for Neuro-QOL lower extremity function score (total score) and change from baseline (total score) as a continuous variable by treatment group at each visit • Difference in mean change from baseline in Neuro-QOL lower extremity function score between treatment groups at randomization week 48 estimated and tested using a mixed-effect model for repeated measures (MMRM)
Intercurrent Events	Observed data only

Objective 2	**To compare the efficacy of troriluzole versus placebo on upper extremity mobility and activities as measured by the Neuro-QOL Upper extremity mobility scale**
Efficacy Endpoint	Neuro-QOL Upper Extremity Function score
Summary	Using the mITT analysis set: • Descriptive statistics for Neuro-QOL upper extremity function score (total score) and change from baseline (total score) as a continuous variable by treatment group at each visit • Difference in mean change from baseline in Neuro-QOL upper extremity function score between treatment groups at randomization week 48 estimated and tested using a mixed-effect model for repeated measures (MMRM)
Intercurrent Events	Observed data only
Objective 3	**To compare the efficacy of troriluzole versus placebo on daily fatigue and activities as measured by the Neuro-QOL Fatigue scale**
Efficacy Endpoint	Neuro-QOL Fatigue score
Summary	Using the mITT analysis set: • Descriptive statistics for Neuro-QOL fatigue score (total score) and change from baseline (total score) as a continuous variable by treatment group at each visit • Difference in mean change from baseline in Neuro-QOL fatigue score between treatment groups at randomization week 48 estimated and tested using a mixed-effect model for repeated measures (MMRM)
Intercurrent Events	Observed data only
Objective 4	**To compare the efficacy of troriluzole versus placebo on clinician impression of global functioning via use of the Clinical Global Impression-Global Improvement Scale (CGI-I)**
Efficacy Endpoint	CGI-I score
Summary	Using the mITT analysis set: • Descriptive statistics for CGI-I score as a categorical variable by treatment group at each visit • Difference in CGI-score between treatment groups at randomization week 48 tested using mixed model for repeated measures (MMRM)
Intercurrent Events	Observed data only
Objective 5	**To compare the efficacy of troriluzole versus placebo on patient impression of global functioning as measure by the Patient Global Impression Scale (PGI)**
Efficacy Endpoint	PGI-C score
Summary	Using the mITT analysis set: • Descriptive statistics for PGI-C score as a categorical variable by treatment group at each visit • Difference in PGI-C score between treatment groups at randomization week 48 tested using a mixed model for repeated measures (MMRM)
Intercurrent Events	Observed data only

4 ANALYSIS SETS, TREATMENT GROUPS, AND SUBGROUPS

4.1 Analysis Sets

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

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

- Modified intent-to-treat (mITT) subjects in the Open Label Extension Phase: Randomized subjects who receive at least one dose of OLE study medication (troriluzole) during the open label extension phase and provided a non-missing baseline measurement and at least one non-missing efficacy assessment during the open-label extension phase.

4.2 Treatment Groups

The 2 treatment groups are troriluzole and placebo.

Populations of treated subjects will be assessed by the as-treated treatment group; i.e., by the treatment actually received. Otherwise, all other populations will be assessed by the as-randomized treatment group. The enrolled population will be assessed overall.

4.3 Subgroups

The following subgroups are of interest:

- Gender (Female, Male)
- Race (Asian, Black or African American, White, and all other races combined)
- age group (based on age at informed consent): < 40, ≥40 and < 65, ≥65 years.
- SCA genotype randomization strata: SCA 1 & 2; SCA 3; and SCA 6, 7, 8, 10
- SCA genotype
- Country (USA, China)

[redacted]

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

5 SAMPLE SIZE, POWER, AND TYPE I ERROR

5.1 Sample Size and Power

At full recruitment, the sample size for this study will be approximately 210 randomized subjects to accommodate for dropouts and was determined based on the rationale that follows.

As the f-SARA is a new scale with no data to estimate changes and variance in this patient population, a modified SARA scale from study BHV4157-201 is used as a proxy. Both scales have the same structure, consisting of 4 items with 5 points/item. The mean changes from randomized baseline using a modified SARA were -0.3 (SD=1.35) and 0.00 (SD=1.44) at OL Weeks 24 and 48, respectively, for all genotypes. Because this study expects an increase on the f-SARA for the non-treated patients, the Week 48 difference in the change from baseline in

f-SARA between treatment groups (troriluzole – placebo) is estimated to be -0.75. For a single fixed test, on the primary endpoint, 210 mITT subjects will provide 90% power, based on a two-sample t-test, assuming a two-sided alpha of 0.05, SD of 1.44, a delta of 0.75 and 25% dropout.

5.2 Type I Error

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

[redacted]

6 STATISTICAL ANALYSES

6.1 General

6.1.1 Programmed Output

A separate document contains the list of all TLFs, along with corresponding templates, attributes, and programming notes.

All output will be sorted and labeled according to the International Conference on Harmonisation (ICH) recommendations and formatted to the appropriate page size(s).

Medical history, AEs, and specify text for other procedures are coded using Medical Dictionary for Regulatory Activities (MedDRA) Version 24, updated in March 2021.

Concomitant medications are coded using World Health Organization Drug Dictionary (WHO-DD), Version B3 March 2021.

All statistical analyses will be performed using SAS statistical software (Version 9.4 or higher), unless otherwise noted.

6.1.1.1 Tables

Unless otherwise specified, the Randomization Phase and Extension Phase will be analyzed separately. For tables summarizing subjects who received at least one dose of DB or OL study medication (troriluzole), summary statistics will be calculated from both phases combined for subjects randomized to troriluzole and from the OL Extension Phase for subjects randomized to placebo.

6.1.1.2 By-Subject Listings

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

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

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

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

6.1.2 Statistical Methods

Categorical variables will be tabulated with the count and percentage within each category (with a 'Missing' category if applicable). Continuous variables will be summarized with univariate statistics (e.g., n, mean, median, standard deviation (SD), minimum, and maximum). The median, minimum, and maximum values will be presented with the same precision as the data. The mean and percentiles will be presented with the precision of the data + 2 decimal place for efficacy and + 1 decimal place for safety. Measures of variance (eg, SD, SE, CI) will be presented with the precision of the data + 2 decimal places.

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

6.1.3 Handling of Missing Data

Unless otherwise noted, efficacy analyses will be based on observed data only. For the main analyses of efficacy for primary, secondary, and exploratory endpoints, no imputation will be performed to impute data following discontinuation from study. In some cases, missing items will be imputed (see individual scales in efficacy section below).

Sensitivity analyses using multiple imputation (MI) will be conducted on the primary and secondary endpoints using a jump to reference and copy increment reference approach to assess the impact of the missing at random (MAR) assumption (see Section 6.3.2.2).

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

[redacted]

6.1.4 Adjustments for Covariates and Stratification

The stratification variable, diagnosis (SCA1 & 2; SCA3; SCA6, 7, 8, 10), will be included as a blocking factor in the analysis of primary and other continuous efficacy endpoints. Unless otherwise specified, these statistical models will also be adjusted by the baseline value of the dependent variable, e.g. the model for f-SARA total score will include f-SARA total score at baseline as a covariate.

6.2 Study Population

Results are presented by treatment group and overall, unless specified otherwise.

6.2.1 Analysis Sets

The number of subjects in each analysis set described in Section 4.1 is tabulated by as-randomized or as-treated treatment group (see Section 4.2), not randomized, and overall.

Inclusion and exclusion from the efficacy analysis set is tabulated by treatment group and overall, as the number and percentage of subjects in the randomized analysis set in both categories:

- Included in efficacy analysis set
- Excluded from efficacy analysis set (e.g., not treated with study drug, randomized more than once).

A by-subject listing of analysis sets is provided for the enrolled analysis set. The listing identifies subjects in analysis sets and includes as-treated treatment group.

A by-subject listing of subjects excluded from the efficacy analysis is provided for subjects in the randomized analysis set who are not in the efficacy analysis set, including the reason for exclusion (i.e., not treated with study drug, randomized more than once). Treated subjects may have > 1 exclusion reason.

A by-subject administrative listing of randomization scheme and codes is 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, and displays the randomization number, block number, site-subject ID, treatment group, randomized age group, and randomization date.

6.2.2 Enrollment

Enrollment by (1) country and site and (2) age group are tabulated overall for the enrolled analysis set, where age group is based on age at informed consent.

A frequency table of accrual over time is also provided for the randomized analysis set. Time is randomization month and year based on the IWRS randomization date.

6.2.3 Subject Disposition

A summary of subject disposition will be tabulated for all subjects by treatment group and overall for the Randomization and Extension Phases, separately, including:

- Number of subjects enrolled and signed the informed consent form
- Number of subjects enrolled by country and site
- Number of subjects enrolled by age group (based on age at informed consent)
- Number of enrolled subjects excluded from the study and reason for exclusion
- Number of subjects randomized
- Number of subjects in each analysis population in the Randomization Phase

- Number of subjects who completed the Randomization Phase
- Number of subjects who prematurely withdrew from the Randomization Phase and reasons for withdrawal
- Number of subjects who entered the Extension Phase
- Number of subjects in each analysis population in the Extension Phase
- Number of subjects who completed the Extension Phase
- Number of subjects who withdrew from the Extension Phase and reasons for withdrawal

A by-subject listing of study completion information for both the Randomization and Extension Phases, including the reason for withdrawal, if applicable, will be presented.

6.2.4 Protocol Deviations

Any event that does not comply with the inclusion/exclusion criteria, study conduct, or study procedures will be documented as a deviation.

[redacted]
and its designation (per Medical Lead) as a significant or non-significant protocol deviation. Deviations were determined by CRO clinical monitor, CRC, and reviewed by Medical Lead in the US. Similarly, deviations at sites in China were determined by the LabCorp clinical monitor and reviewed and agreed upon by the US Medical Lead.

Relevant protocol deviations as defined in section 9.3 are programmable deviations and include any event that does not comply with the inclusion/exclusion criteria, study conduct, or study procedures, and that significantly impacts the completeness, reliability, and interpretability of the study data.

Both relevant and significant protocol deviations will be tabulated for the enrolled analysis set for the randomization and extension phases, separately, with all deviations presented in a by-subject listing.

6.2.5 Baseline Characteristics

Baseline characteristics include (1) demographics and other relevant baseline characteristics, (2) smoking history, (3) Mini Mental State Exam (MMSE) results, (4) medical history, and (5) prior non-study medications.

Baseline characteristics are tabulated for treated subjects and for subjects enrolled but not randomized.

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

Baseline is defined according to analysis set as follows:

- Enrolled but not randomized: Last non-missing value
- Treated subjects: Last non-missing value at or before the study drug start date/time.

6.2.6 Extent of Exposure and Compliance to Study Treatment

Exposure and compliance are presented by as-treated treatment group and overall.

6.2.6.1 Study Therapy

Randomization Phase

For the extent of subject exposure, treatment duration, total days of study medication and average daily dose will be summarized by treatment group (troriluzole 200 mg or placebo) for the Randomization Phase. Treatment duration will be quantified as the number of days on study drug and measured from the time the subject received first dose until the time the subject received the last dose, either at the end of 48 weeks of treatment or withdrawal from the Randomization Phase. Double-blind treatment duration includes missed dose days; i.e., last dose date of DB randomized study medication (with non-zero number of doses) – first dose date of DB randomized study medication + 1. Total days on study medication will be calculated to exclude days when subjects missed a dose (i.e., treatment duration – number of missed doses of the DB randomized study medication). Average daily dose is the total mg taken by a subject in DB divided by DB treatment duration.

The average daily dose and the number of subjects on-treatment will also be summarized by 4-week (28-day intervals) of the Randomization Phase. If a subject took at least one dose of study medication during the 4-week interval, then the subject will be considered as exposed to study medication during that 4-week period.

Additionally, the percent (%) compliance (i.e., total days on study medication in the Randomization Phase / treatment duration in the Randomization Phase × 100) and total number of capsules will be summarized.

The number and percentage of subjects who had dose titrations during the randomization phase will also be summarized by time point and dose.

Extension Phase

For the Extension Phase, treatment duration, total days on study medication, number of subjects on-treatment during each 4-week interval, average daily dose, % compliance, and total number of capsules will be summarized by treatment group (troriluzole 140 mg or 200 mg) using the following definitions:

- Treatment duration: last dose date of troriluzole study medication in the OL Extension Phase (with non-zero number of doses)– first dose date of troriluzole study medication in the OL Extension Phase + 1
- Total days on study medication: treatment duration – number of missed doses of the OL Extension Phase study medication
- Average daily dose: total mg in OL/treatment duration in OL
- % compliance: total days on study medication in the OL Extension Phase / treatment duration in the OL Extension Phase × 100

The number and percentage of subjects who had dose titrations during the Extension Phase will also be summarized by time point and dose.

Combined Phases

For the combined phase data, treatment duration, total days on troriluzole, average daily dose, % compliance, and total number of capsules will be summarized for all subjects exposed to troriluzole during the entire study using the following definitions:

- Treatment duration: last dose day of troriluzole study medication – first dose date of troriluzole study medication + 1
- Total days on troriluzole: treatment duration – number of missed doses of study medication
- Average daily dose: total mg of troriluzole/treatment duration on troriluzole
- % compliance: total days on troriluzole / treatment duration × 100

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

Additionally, a listing will be created which includes subjects with a dose titration recorded as a result of an AE and/or as a decreased dose from the dosing records. The listing will present the subject ID, site, start and stop dates of decreased dose, associated AE term, and associated AE start and stop dates.

6.2.6.2 Prior and Concomitant Non-study Medications

Concomitant medications will be coded using the WHO-DD (see section 6.1.1 for version). Results will be tabulated by Anatomic Therapeutic Class (ATC) and PT for treated subjects for the Randomization and Extension Phases.

Imputed medication start and stop dates are used to assign non-study medication type (previous, concomitant, and/or follow-up) to all non-study medications. 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.

All prior, concomitant, and follow-up medications will be listed.

6.3 Efficacy

The efficacy analyses will be conducted on mITT subjects in the Randomization Phase. Efficacy endpoints will not be tested in the extension phase and data will be presented only descriptively. All efficacy data will also be included in listings by subject, study medication, and visit (as applicable). Randomization and Extension Phase data will be shown within the same listing unless otherwise specified. A summary of estimands and planned analyses for all efficacy endpoints are provided in Table 4.

Table 4: Primary, Secondary, and Exploratory Efficacy Estimands and Analyses

Index	Description	Variable Type	Treatment Group Comparison	On-Study Data
Primary Estimand				
P01	Difference between treatment groups in change from baseline in the f-SARA total score at Week 48 of the Randomization Phase	Continuous	MMRM	Weeks 4, 8, 12, 24, 36, and 48
Secondary Estimand				
S01	Difference between treatment groups in change from baseline in PIFAS score at Week 48 of the Randomization Phase	Continuous	MMRM	Weeks 4, 8, 12, 24, 36, and 48
S02	Difference between treatment groups in change from baseline in FARS-ADL score at Week 48 of the Randomization Phase	Continuous	MMRM	Weeks 4, 8, 12, 24, 36, and 48
S03	Difference between treatment groups in change from baseline in FARS-FUNC score at Week 48 of the Randomization Phase	Continuous	MMRM	Weeks 4, 8, 12, 24, 36, and 48

Index	Description	Variable Type	Treatment Group Comparison	On-Study Data
Exploratory Estimand				
E01	Difference between treatment groups in change from baseline in Neuro-QOL Lower Extremity Function score at Week 48 of the Randomization Phase	Continuous	MMRM	Weeks 8, 12, 24, 36, and 48
E02	Difference between treatment groups in change from baseline in Neuro-QOL Upper Extremity Function score at Week 48 of the Randomization Phase	Continuous	MMRM	Weeks 8, 12, 24, 36, and 48
E03	Difference between treatment groups in change from baseline in Neuro-QOL Fatigue score at Week 48 of the Randomization Phase	Continuous	MMRM	Weeks 8, 12, 24, 36, and 48
E04	Difference between treatment groups in CGI-I score at Week 48 of the Randomization Phase	Ordinal, Tested as Continuous	MMRM	Weeks 8, 24, and 48
E05	Difference between treatment groups in PGI score at Week 48 of the Randomization Phase	Ordinal, Tested as Continuous	MMRM	Weeks 8, 24, and 48

MMRM = Mixed effects model for repeated measures

6.3.1 Efficacy Endpoints

6.3.1.1 Continuous Efficacy Endpoints

Measurements are slotted into analysis visits as described in Section 7.2.

Descriptive Analyses

Summary statistics for parameter values are tabulated at select time points by treatment group during the randomization and extension phases, separately.

Treatment Group Comparison Using MMRM

Treatment groups are compared and assessed using MMRM with the following variables:

- Dependent variable, e.g., f-SARA score
- Fixed effects (categorical variables): treatment group, visit, treatment group-by-visit interaction, country, and diagnosis (SCA 1 & 2; SCA 3; SCA 6, 7, 8, 10). Visit represents on-study analysis visit.
- Covariate (continuous variable): baseline value of dependent variable

The covariance structure for within subject error ("R" matrix in SAS proc mixed) will be initially specified as unstructured. If the model fails to converge, then a Huynh-Feldt structure may be used, followed by an AR(1) structure, and lastly, a CS structure. Error degrees of freedom (df) 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.

The following statistics are tabulated from the MMRM at select analysis visits:

- Least-squares mean with corresponding SE with 95% CI of values and changes from baseline by treatment group and visit.
- Difference in least-squares means between treatment groups (troriluzole – placebo) with corresponding SE, 95% CI, and p-value.

6.3.1.2 Ordinal Endpoints, Treated as Continuous

Measurements are slotted into analysis visits as described in Section 3.

Descriptive Analyses

Descriptive statistics will be presented at select time points by treatment group during the randomization and extension phases, separately.

Treatment Group Comparison Using MMRM

Treatment groups are compared and assessed using MMRM with the following variables:

- Dependent variable, e.g., CGI-I score
- Fixed effects (categorical variables): treatment group, visit, treatment group-by-visit interaction, country, and diagnosis (SCA 1 & 2; SCA 3; SCA 6, 7, 8, 10). Visit represents on-study analysis visit.

The covariance structure for within subject error ("R" matrix in SAS proc mixed) will be initially specified as unstructured. If the model fails to converge, then a Huynh-Feldt structure may be used, followed by an AR(1) structure, and lastly, a CS structure. Error degrees of freedom (df) 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.

The following statistics are tabulated from the MMRM at select analysis visits:

- Least-squares mean with corresponding SE with 95% CI of values by treatment group and visit.

- Difference in least-squares means between treatment groups (troriluzole – placebo) with corresponding SE, 95% CI, and p-value.

6.3.2 Primary Endpoint: f-SARA

6.3.2.1 Main Analyses

The f-SARA is a modified version of the SARA scale, designed to create new response categories that reflect clinically meaningful changes in patient function. Response categories for each item range from 0 to 4, where 0 is normal (no impairment), 1 is mildly impaired function but no assistance required, 2 is moderately impaired function but needs assistance for certain parts of the task, 3 is severely impaired function to the degree that assistance is needed for all parts of the task, and 4 is unable to perform function. The total score is derived as the sum of 4 individual items, which include: gait, stance, sitting, and speech disturbance. If 1 item is missing for this scale, pro-rated score will be imputed. Pro-rated score is defined as the total score based on non-missing items divided by the number of non- missing items and multiplied by the total number of items of the scale. If 2 or more items are missing, the total score will be considered as missing.

The primary endpoint is the change from baseline in the f-SARA total score at Week 48 of the Randomization Phase. The treatment comparison of troriluzole versus placebo will use a mixed model for repeated measures (MMRM). The model will include fixed effects for treatment group, randomization stratum (SCA 1 & 2; SCA 3; SCA 6, 7, 8, 10), visit, treatment group-by-visit interaction, country (US; China), and baseline f-SARA score as a covariate. Repeated measurements are made on each subject. The covariance structure for within subject error ("R" matrix in SAS proc mixed) will be initially specified as unstructured. If the model fails to converge, then a Huynh-Feldt structure may be used, followed by an AR(1) structure, and lastly, a CS structure. Error degrees of freedom (df) will be calculated using Kenward-Roger approximation if an unstructured covariance structure fits appropriately; otherwise, a sandwich estimator will be utilized to estimate the covariance structure and degrees of freedom will be calculated using the between-within method.

Least square means (LSMeans) of the change from baseline for each treatment group will be derived for Week 4, Week 8, Week 12, Week 24, Week 36, and Week 48. These estimates will be presented along with df, standard error (SE), and 95% confidence intervals. The difference in the change from baseline between treatment groups (troriluzole – placebo) will also be derived for the same visits, and presented with df, SE, 95% confidence interval, and p-value.

In addition, descriptive statistics for the total f-SARA score (and individual items) and the change from baseline in the total f-SARA score (and individual items) will be presented by visit for the Randomization and Extension Phases, separately.

[redacted]

[redacted]

[redacted]:

6.3.3 Secondary Efficacy Endpoints

If the main analysis of the primary endpoint is significant, then the main analyses of secondary endpoints will be tested using the Hochberg procedure, as described in section 5.2.

6.3.3.1 S01: Patient Impression of Function and Activities of Daily Living Scale

PIFAS is a 17-item instrument designed to assess the level of functional disability. The scale is rater administered and the items were selected to encompass domains of relevance to SCA (i.e., mobility, speech/swallowing, and fatigue), and to capture function and activities of daily living within these domains. Statements are rated on a 5-point Likert scale ranging from "0" reflecting "not at all" to "4" reflecting "very much." The total score is derived as the sum of the individual items. If 4 or less items are missing for this scale), pro-rated score will be imputed. If 5 or more of the seventeen items are missing, the total score will be considered as missing.

The estimand is the change from baseline in the total PIFAS score Week 48 of the Randomization Phase.

Main Analysis

The change from baseline in the total PIFAS score will be analyzed with the MMRM methodology described in Section 6.3.1.1.

Descriptive statistics for the total PIFAS score (and individual items) and the change from baseline in the total PIFAS score (and individual items) will also be presented as described in Section 6.3.1.1.

[redacted]

6.3.3.2 S02: Activities of Daily Living from Friedreich's Ataxia Rating Scale

The FARS is a 9-item scale designed to assess neurological domains affected in Friedreich's ataxia, another hereditary cerebellar ataxia disorder. The FARS-ADL subscale, validated in

Friedreich's ataxia, is a measure of functional disability that correlates with total SARA scores. It assesses 9 areas of activities of daily living with response categories rated on a 5-point scale with "0" reflecting "normal" and "4" reflecting an inability to perform the specific function. The total score is derived as the sum of the individual items. If 2 or less items are missing for this scale, pro-rated score will be imputed. If 3 or more of the nine items are missing, the total score will be considered as missing.

The estimand is the change from baseline in the total FARS-ADL score at Week 48 of the Randomization Phase.

Main Analysis

The change from baseline in the total FARS-ADL score will be analyzed with the MMRM methodology described in Section 6.3.1.1.

Descriptive statistics for the total FARS-ADL score (and individual items) and the change from baseline in the total FARS-ADL score (and individual items) will also be presented as described in Section 6.3.1.1.

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

6.3.3.3 S03: Functional Staging for Ataxia from Friedreich's Ataxia Rating Scale

The FARS-FUNC is a subscale of the FARS designed to provide functional staging for ataxia in which clinicians are asked to assess function on a 6-stage staging system with "0" reflecting "normal" and "6" reflecting a stage in which the patient is "total disability." Severity of neurologic disabilities as expressed by both the FARS-ADL and FARS-FUNC have been shown to correlate with disease progression and duration in Friedreich's ataxia.

The estimand is the change from baseline in the FARS-FUNC total score at Week 48 of the Randomization Phase.

Main Analysis

The change from baseline in the total FARS-FUNC score during the Randomization Phase will be analyzed with the MMRM methodology described in Section 6.3.1.1.

Descriptive statistics for the total FARS-FUNC score and the change from baseline in the total FARS-FUNC score will also be presented as described in Section 6.3.1.1.

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

6.3.4 Exploratory Efficacy Endpoints

6.3.4.1 Neuro-QOL Lower Extremity Function

The Neuro-QOL lower extremity function scale is designed to assess fine motor skills of the lower extremities by asking specific questions about activities of daily living. This is a 19-item scale with questions rated on a 5-point Likert scale with "5" reflecting "no difficulty" and "1" reflecting "unable to do" said activity. The total score is derived as the sum of the individual items. If 50% or less of the items is missing (i.e. 9 or less items missing for this scale), pro-rated score will be imputed. If more than 50% of the nineteen items are missing, the total score will be considered as missing (Neuro Qol, Scoring Manual 2021).

The estimand is the change from baseline in the Neuro-QOL lower extremity function score at Week 48 of the Randomization Phase. The change from baseline in the Neuro-QOL lower extremity function score will be analyzed with the MMRM methodology described in Section 6.3.1.1.

Descriptive statistics for the Neuro-QOL lower extremity function score (and individual items) and the change from baseline in the Neuro-QOL lower extremity function score (and individual items) will also be presented as described in Section 6.3.1.1.

6.3.4.2 Neuro-QOL Upper Extremity Function

The Neuro-QOL upper extremity function scale is designed to assess fine motor skills of the upper extremities by asking specific questions about activities of daily living. This is a 20-item scale with questions rated on a 5-point Likert scale with "5" reflecting "no difficulty" and "1"

reflecting "unable to do" said activity. The total score is derived as the sum of the individual items. If 50% or less of the items is missing (i.e. 10 or less items missing for this scale), pro-rated score will be imputed. If more than 50% of the twenty items are missing, the total score will be considered as missing (Neuro Qol, Scoring Manual 2021).

The estimand is the change from baseline in the Neuro-QOL upper extremity function score at Week 48 of the Randomization Phase. The change from baseline in the Neuro-QOL upper extremity function score will be analyzed with the MMRM methodology described in Section 6.3.1.1.

Descriptive statistics for the Neuro-QOL upper extremity function score (and individual items) and the change from baseline in the Neuro-QOL upper extremity function score (and individual items) will also be presented as described in Section 6.3.1.1.

6.3.4.3 Neuro-QOL Fatigue

The Neuro-QOL fatigue scale is designed to assess the level of fatigue in patients with neurologic disorders. This is a 19-item, patient-rated scale designed to rate a patient's fatigue over the past 7 days. Patients are asked to rate answers to these items on a 5-point Likert scale with "1" reflecting "never" and "5" reflecting "always". The total score is derived as the sum of the individual items. If 50% or less of the items is missing (i.e. 9 or less items missing for this scale), pro-rated score will be imputed. If more than 50% of the nineteen items are missing, the total score will be considered as missing (Neuro Qol, Scoring Manual 2021).

The estimand is the change from baseline in the Neuro-QOL fatigue score at Week 48 of the Randomization Phase. The change from baseline in the Neuro-QOL fatigue score will be analyzed with the MMRM methodology described in Section 6.3.1.1.

Descriptive statistics for the Neuro-QOL fatigue score (and individual items) and the change from baseline in the Neuro-QOL fatigue score (and individual items) will also be presented as described in Section 6.3.1.1.

6.3.4.4 Clinical Global Impression – Global Improvement

The CGI-I is a 7-point scale that requires the clinician to assess how much the subject's illness has improved or worsened relative to the baseline visit and it is rated on a 7-point Likert-based scale with "1" reflecting "very much improved" and "7" reflecting "very much worse".

The estimand is the CGI-I score at Week 48 of the Randomization Visit. Descriptive statistics for the clinician impression of benefit via the CGI-I will be presented as described in Section 6.3.1.2.

For the Randomization Phase only, the difference between treatment groups will be assessed as described in Section 6.3.1.2.

6.3.4.5 Patient Global Impression of Change

The PGI-C is a 7-point patient self-reported Likert-based scale used to assess the response of a condition to a therapy. The response categories of the scale range from "no change (or condition has gotten worse)" to "a great deal better and a considerable improvement that has made all the difference."

The estimand is the PGI-C score at Week 48 of the Randomization Phase. Descriptive statistics for the patient impression of benefit via the PGI-C will be presented as described in Section 6.3.1.2.

For the Randomization Phase only, the difference between treatment groups will be assessed described in Section 6.3.1.2.

6.4 Safety

Safety and other analyses will be conducted on treated subjects using their actual treatment received. For the Randomization Phase outputs, treated subjects in the Randomization Phase will be used (summarized by treatment, but not overall); for the Extension Phase outputs, treated subjects in the Extension Phase will be used (summarized by previous treatment and overall); and for the combined phase outputs, DB and OL troriluzole treated subjects will be used (presented as a single column of all subjects who have received at least one dose of active treatment). All safety and other data will be listed for the entire study with screening, Randomization Phase, and Extension Phase data presented together.

Safety outcome measures include extent of exposure, compliance to study treatment, AEs, laboratory assessments, vital signs, ECGs, concomitant medications, and the S-STS questionnaire.

6.4.1 Adverse Events

Adverse events will be coded using MedDRA (see section 6.1.1 for version) and displayed in tables and listings using system organ class (SOC) and preferred term (PT).

Analyses of AEs will be performed for those events that are considered treatment-emergent. Any AE that first occurs pre-dose but worsens in severity after the first study drug administration will also be considered a TEAE. Non-TEAEs are those that occur prior to the first administration of the study medication and resolved prior to dosing or that first occur prior to the first study drug administration, but do not worsen in severity after dosing.

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

- All TEAEs
- TEAEs related to treatment (i.e., possibly related or related)
- Treatment-emergent SAEs
- TEAEs leading to treatment discontinuation
- TEAEs by highest severity (mild, moderate, severe)
- TEAEs related to treatment (i.e., possibly related, or related) by highest severity (mild, moderate, severe)

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

All AEs occurring pre-treatment and during the entire study will be listed with a flag designating which phase the AE had onset in (i.e., screening, Randomization Phase, or Extension Phase). Additional by-subject listings will be provided for deaths, SAEs, AEs leading to treatment discontinuation and, and AEs of Special Interest:

- TEAEs indicating potential interstitial lung disease (using Standardized MedDRA Queries [SMQ]Interstitial Lung Disease including Eosinphilic Pneumonia and Hypersensitivity Pneumonitis)

- Suicidality AEs: PTs based on Suicide/self-injury SMQ

6.4.2 Exposure-Adjusted Multiple Occurrences of Unique AEs

Exposure-adjusted multiple occurrences of unique AEs in the Randomization Phase period will be summarized by SOC and PT with patient-years, the number and percentage of subjects with AEs, number of multiple occurrences of unique AEs, exposure-adjusted event rates, and exposure-adjusted multiple occurrence rates which are defined as follows:

- Number and percentage of subject with AEs by SOC and PT.

- Number of multiple occurrences of unique AEs: Total number of multiple occurrences of unique AEs in the on-treatment period summed across all subjects. For a given subject and PT, an occurrence of a unique AE is determined only by distinct AE start date. Example for SAEs related to study drug:

    - Subject has 3 records of gastroenteritis that are serious and related to study drug in the on-treatment period with the following attributes:

        - AE start date = 01JAN2021 and end date = continuing

        - AE start date = 01JAN2021 and end date = 10JAN2021

        - AE start date = 09JAN2021 and end date = 12JAN2021.

          Then this subject has 2 occurrences of unique gastroenteritis.

          For a given PT, the multiple occurrence count must be ≥ the number of subjects with an AE. Note that this is only for the purpose of counting occurrences, no collapsing of records for overlapping and contiguous AEs, because these are expected to be cleaned prior to database lock.

- Patient-years: total exposure in the on-treatment period summed across all subjects, where a subject's total exposure in the on-treatment period is defined as either of the following:

  If the subject did not have any between study phase gap that is > 30 days:

    - (reference last date – first dosing date + 1)/365.25, if the reference last date is not missing

    - (reference end date – first dosing date + 1)/365.25 otherwise.

  If the subject had between study phase gap that is > 30 days:

    - (reference last date – first dosing date + 1- sum(gap length-30))/365.25, if the reference last date is not missing, where the sum(gap length-30) only includes the gaps with length > 30 days for that subject.

- (reference end date – first dosing date + 1- sum(gap length-30))/365.25 if the reference last date is missing, where the sum(gap length-30) only includes the gaps with length > 30 days for that subject.

- Exposure-adjusted rate per 100 patient-years: 100 x (number of subjects with AE)/patient-years.

- Exposure-adjusted multiple occurrence rate per 100 patient-years: 100 x (number of multiple occurrences of unique AEs)/patient-years.

Reference last date is the date defining the end of the on-treatment period for subjects who have discontinued the treatment corresponding to the safety analysis period, e.g., study drug last date + 30.

Reference end date is the date defining the end of the on-treatment period for subjects who have not discontinued the treatment corresponding to the safety analysis period, e.g., study drug end date.

The following exposure-adjusted multiple occurrences of unique AEs will be summarized:

- Exposure-adjusted multiple occurrences of unique AEs by SOC and PT

- Exposure-adjusted multiple occurrences of unique SAEs by SOC and PT

- Exposure-adjusted multiple occurrences of unique SAEs related to study drug by SOC and PT

- Exposure-adjusted multiple occurrences of unique non-SAEs by SOC and PT occurring with ≥ 5% frequency.

6.4.3 Laboratory Tests

The following clinical laboratory parameters will be collected and graded (if applicable) by the Common Technical Criteria for Adverse Events (CTCAE) version 5.0 (2017) or the Division of Acquired Immunodeficiency Syndrome (DAIDS) Table for Grading the Severity of Adult and Pediatric Adverse Events Corrected version 2.1 (2017).

Table 5: Clinical Laboratory Parameters and Criteria for Grading

Panel	Not Graded	CTCAE	DAIDS
Hematology	Absolute basophil count	Absolute lymphocyte count	
	Absolute eosinophil count	Absolute neutrophil count	
	Absolute monocyte count	Hemoglobin	
	Hematocrit	Platelet count	

Panel	Not Graded	CTCAE	DAIDS
	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)	
	Gamma-glutamyl transferase (GGT)	Aspartate aminotransferase (AST)	
	Hemoglobin A1c (HbA1c)*	Bicarbonate (CO2)	
	High-density lipoprotein (HDL)*	Calcium	
	Lipase*	Creatine kinase	
	P-Amylase*	Creatinine	
	Phosphorous	eGFR MDRD	
	Pregnancy testing	Lactate dehydrogenase (LDH)	
	T4*	Potassium	
	Thyroid-stimulation hormone (TSH)*	Sodium	
	Total protein	Total bilirubin	
		Total cholesterol*	
		Triglycerides*	
		Uric acid	
Urinalysis			
	Ketones		Protein
	Leukocyte esterase		Glucose
	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.

Additional tests include:

- Urine pregnancy tests to be performed 72 hours prior to dosing at baseline, at study visits where lab assessments are not performed, or at the discretion of the Investigator

- HIV, HBsAg, and HCV antibody at screening

If a subject was unable to come to the study site safety labs may have been conducted locally. For the local labs drawn by Quest, a Quest file of laboratory reference ranges was applied, for all other local labs the entered reference ranges were used and the data handling specifications in Section 9.5 were applied.

Clinical laboratory values will be expressed using conventional (US) and standard international (SI) units, unless otherwise specified. In the event of repeat values within the same analysis visit, if any measurement has an abnormal result, then the "worst" measurement, maximum for abnormalities above the upper reference and 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 as well as shift from baseline. If all values are in normal range, then the latest measurement in the analysis visit interval will be used for presentation in the by-visit tables. All observed values will be presented in listings and considered for evaluation of potential drug induced liver injury or abnormalities.

The observed value and change from baseline by visit will be summarized for each continuous laboratory parameter in both US and SI units for treated subjects in the Randomization and Extension Phases.

Grading will be derived based on criteria noted in Table 5 and detailed in Appendix, Section 9.2. 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 at any time post-baseline. 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 in the Randomization and Extension Phases. Shift tables will only include treated subjects with an assessment for the specific test of interest at baseline and on treatment.

In addition, the shift from baseline for liver function test abnormalities (based on either normal limits or grading where available) will be tabulated by grade (if applicable) and visit for treated subjects in the Randomization and Extension phases. The shift tables will only include subjects with a baseline assessment and at least one on-treatment assessment at the summarized visit. For laboratory tests that are not graded, abnormality criteria are specified below (same as used for the listings):

- Low (<lower limit of normal; LLN)

- Normal
- High (>upper limit of normal; ULN)

For ALT and AST, the following categories will be used to summarize the shift from baseline based on the ULN:

- ≤ULN
- >ULN to ≤ 3x ULN
- >3x ULN to ≤ 5x ULN
- >5x ULN to ≤ 10x ULN
- >10x ULN to ≤ 20x ULN
- >20x ULN

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

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

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

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

For total bilirubin, the following categories will be used to summarize the shift from baseline based on the ULN:

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

- >2x ULN

An evaluation of drug-induced serious hepatotoxicity (eDISH) scatter plot will display the maximum total bilirubin 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. 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." The eDISH plot will be produced for treated subjects in the Randomization Phase and DB and OL troriluzole treated subjects.

Additional listings will be presented for the following:

- All abnormal laboratory values considered potentially clinically significant (Grade 3 or 4) for the Randomization and Extension phases combined

- Subjects with a maximum value of ALT or AST >3x ULN or a maximum total bilirubin value >2x ULN observed at any point during the entire study, but not necessarily on concurrent visits

6.4.4 Vital Signs, Physical Measurements, and Neurological Assessment

All physical examination and neurological assessment findings will be presented in data listings.

The observed value and change from baseline for each vital sign taken from the sitting position (i.e., heart rate, systolic/diastolic blood pressure, temperature, weight, and BMI) will be summarized by visit. In the event of repeat values within the same analysis visit, then results the measurement with the greatest change from baseline will be used.

In addition, the number and percentage of subjects who meet the following criteria will be summarized:

- Systolic Blood Pressure: <90 mmHg, >140 mmHg, >160 mmHg

- Diastolic Blood Pressure: <50 mmHg, >90 mmHg, >100 mmHg

- Pulse Rate: <60 bpm, >100 bpm

- Body Weight: decrease of ≥7% from baseline and increase of ≥7% from baseline

- Temperature: >38.0 °C, <36.0 °C

Additionally, all vital sign and physical measurements will be presented in data listings.

6.4.5 Electrocardiogram

Descriptive statistics for the ECG interval data (e.g., QRS, PR, QT, QTcF), and ventricular heart rate will be reported by visit. In the event of repeat values within the same analysis visit, then results the measurement with the greatest change from baseline will be used.

In addition, the number and percentage of subjects who meet the following criteria will be summarized:

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

All ECG data for each subject will be provided in data listings. A subject listing of ECG results will identify abnormal QTcF assessments using the above criteria.

6.4.6 Sheehan-Suicidality Tracking Scale

The S-STS is a prospective, self-reported rating scale that contains 16 questions to track both treatment-emergent suicidal ideation and behaviors. In the event the subject is unavailable, the S-STS clinician-administered rating scale will be completed that contains 6 yes/no questions (Questions 17 to 22). Note that only 1 of the 6 clinician-reported questions should be answered positively ("yes") if patient-reported Questions 1 to 16 cannot be completed.

S-STS scores will be calculated as follows:

- If ≥ 1 of the patient-reported Questions 1 to 16 has a non-missing response:
    - Ideation subscale score: Sum of scores (0 to 4) for Questions 2 to 11
    - Behavior subscale score: Sum of scores (0 to 4) for the following questions:
        - Question 1a, only if Question 1b has a "yes" response
        - Highest score of Question 12 or any row of Question 16 ("How serious was each preparation?")
        - Highest score of Question 14 or any row of Question 15 ("How serious was each attempt?")
    - Total score: Sum of the ideation and behavior subscale scores.
- Otherwise, if clinician-reported Question 17 or 20 has a "yes" response, then the behavior subscale and total scores are set to the higher score of Question 17 or 20.

- Otherwise, if clinician-reported Question 18, 19, 21, or 22 has a "yes" response, then the ideation subscale, behavior subscale, and total scores are set respectively to the last ideation subscale, last behavior subscale, and last total scores derived from patient-reported questions before the assessment.

- Otherwise, scores are missing.

The self-reported S-STS ideation subscale, behavior subscale, and total score will be summarized as the change from baseline (i.e., <-1, -1, no change, 1, >1) at each visit for treated subjects and troriluzole treated subjects.

A by-subject listing will be provided for subjects with change from baseline ≥ 1.

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 mm3

- LFT abnormalities:

  - ALT or AST > 3x ULN

  - ALT or AST > 3x ULN, and serum total bilirubin ≥ 2x ULN

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

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.

[redacted]

[redacted]

7 CONVENTIONS

7.1 Rounding Rules

Continuous variables are tabulated using summary statistics (e.g., n, mean, median, standard deviation (SD) or standard error (SE), minimum, and maximum), and are presented as follows:

- Values and changes from baseline
    - Minimum and maximum: Same precision as the data collected
    - Mean and percentiles: 1 decimal place more than the data collected (2 decimals places for efficacy questionnaires)
    - Measures of variance (e.g., SD, SE, CI): 2 decimal places more than the data collected
- Percent changes from baseline
    - Minimum and maximum: 0 decimal places
    - Mean and percentiles: 1 decimal place
    - Measures of variance: 2 decimal places.

P-values < 0.0001 are presented as "<0.0001". Otherwise, p-values are presented to 4 decimal places.

Templates specified in a TLF document may supersede these conventions.

7.2 Analysis Periods

Analysis periods are defined as follows:

- Screening: assessment date/time on or before the study drug start date/time
- On-treatment efficacy:
    - Randomization Phase: measurement date after the DB study drug start date through study drug last date + 14 days or (≤ start of OL) for those who enter OL, whichever is earlier. Subjects who take medication in the p.m. may have started OL treatment the same day as their Week 48 visit, therefore the start date of OLE is included as part of the

Randomization Phase to not exclude the efficacy measurements taken at the Week 48 visit.

- Extension Phase: Measurement date after the OL study drug start date +1 through study drug last date + 14 days.

- On study: measurement date after the troriluzole start date through study drug last date + 14 days.

- On-treatment safety:

  - Randomization Phase: Measurement date after the DB study drug start date through study drug last date + 30 days or, for those who enter OL, ≤ start of OL study drug for laboratory and ECG measurements and < OL for adverse event reports, whichever is earlier. Subjects who take medication in the p.m. may have started OL treatment the same day as their Week 48 visit, therefore the start date of OLE is included as part the Randomization Phase so as not to exclude the laboratory or ECG measurements that would have been conducted at the Week 48 visit.

    - On-treatment laboratory abnormalities in the Randomization Phase: to include any abnormality with an assessment date after the date of the first dose of DB study drug (DB start date < assessment date)

    - Treatment-emergent AEs (TEAEs) in the Randomization Phase: Any AE that first occurs pre-dose but worsens in severity after the first study drug administration will also be considered a TEAE (DB start date ≤ assessment date)

    - Non-TEAEs are those that occur prior to the first administration of the study medication and resolved prior to dosing or that first occur prior to the first study drug administration, but do not worsen in severity after dosing.

  - OL Extension Phase: Measurement date after the OL study drug start date through study drug last date + 30 days for subjects who enter the extension phase.

    - On-treatment laboratory and ECG measurements in the Extension Phase: to include any measurement with an assessment date after the date of the first dose of OL study drug (OL start date < assessment date)

    - Treatment-emergent AEs (TEAEs) in the Extension Phase: Any AE that first occurs pre-OL-dose but worsens in severity after the first OL study drug administration will also be considered a TEAE (OL start date≤ assessment date)

    - Non-TEAEs are those that occur prior to the first administration of the OL study medication and resolved prior to dosing or that first occur prior to the first OL study drug administration, but do not worsen in severity after dosing.

- On study: measurement date or imputed after the study drug start date through study drug last date + 30 days.

7.3 Analysis Visit Windows

For the Randomization Phase, the protocol specified window is +/- 5 days for Weeks 4, 8, 12, and 16, +/- 7 days for Weeks 24 and 36, and +/- 2 weeks for Week 48. For the Extension Phase, the protocol-specified visit window is +/- 7 days. However, to include as much data as possible, the visit windows for the Randomization and Extension Phases will be adjusted, depending on the endpoint, to exclude any gaps in study days. [redacted] Unless otherwise noted, the interval in the second column carries to subsequent columns. Shaded cells indicate that the assessment was not done at the visit.

For the Randomization Phase, the baseline assessment is defined as the last available assessment on or before the first day of the Randomization Phase study drug (this assessment could have been done during the Screening or Randomization phase). For the Extension Phase (unless otherwise noted), the baseline assessment is defined as the last available assessment on or before the first day of Extension Phase study drug.

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

Measurements from subjects who withdraw early from the study will be included in the 'End of Study' visit in the summary tables and may or may not fall into an evaluation interval. The last assessment for a subject, excluding the 2-week post-last dose visit, will be considered End of Study. Final assessments, whether they occur at a discontinuation visit or not, will be marked in by-subject listings to indicate they are End of Study.

[redacted]

[redacted]

Table 7: [redacted]

[3]2-Week Post-Last Dose liver function tests and other labs only occurs for subjects not entering Extension Phase or who discontinued early. The 2-Week Post Dose Visit would not need to occur if the subject discontinued dosing more than 2 weeks prior to the early discontinuation visit
[4]Protocol-specified visit window: +/- 7 days. Only summarize Extension Phase visits if the assessment date is after the start date of Extension Phase dosing.
[5]All subjects who discontinue study treatment early should complete an early discontinuation visit as well as the 2-Week Post Dose Visit. The 2-Week Post Dose Visit would not need to occur if the subject discontinued dosing more than 2 weeks prior to the early discontinuation visit
Shaded cells indicate no assessment.

8 CONTENT OF REPORTS

The efficacy data from the double-blind phase of the study and the available safety data will be summarized in a CSR after the last subject has completed or withdrawn from the double-blind randomization phase.

The final CSR will be written after open-label extension and 2-week follow-up visit have been completed.

All analyses described in this SAP are included.

9 APPENDICES

9.1 Schedule of Assessments and Events

Table 8: Schedule of Assessments and Events – Randomization Phase

	Screen	Baseline	Week 4 (+/-5)	Week 8 (+/-5)	Week 12 (+/-5)	Week 16[2] (+/-5)	Week 24 (+/-7)	Week 36 (+/-7)	Week 48[3, 5] Or Early DC	Follow-up Week 2 (ONLY for subjects NOT entering extension phase or who discontinued early)[4]	Additional Instruction
Visit	**Visit 1**	**Visit 2**	**Visit 3**	**Visit 4**	**Visit 5**	**Visit 6**	**Visit 7**	**Visit 8**	**Visit 9**	**Visit 10**	
Eligibility Assessments											
Informed Consent	X										
Inclusion/Exclusion	X										Should be re-reviewed at Baseline
Medical History	X										
Demographic Assessment	X										
Disease History	X										
Neurological Exam	X			X							
Mini Mental State Exam (MMSE)	X										
Pregnancy Test for WOCBP (serum)	X								X		

	Screen	**Baseline**	**Week 4 (+/-5)**	**Week 8 (+/-5)**	**Week 12 (+/-5)**	**Week 16[2] (+/-5)**	**Week 24 (+/-7)**	**Week 36 (+/-7)**	**Week 48[3, 5] Or Early DC**	**Follow-up Week 2 (ONLY for subjects NOT entering extension phase or who discontinued early)[4]**	**Additional Instruction**
Visit	**Visit 1**	**Visit 2**	**Visit 3**	**Visit 4**	**Visit 5**	**Visit 6**	**Visit 7**	**Visit 8**	**Visit 9**	**Visit 10**	
Pregnancy Test for WOCBP (urine)1		X	X		X		X	X			The site may test a patient at any time if they suspect the patient may be pregnant.
Safety Assessments											
Laboratory Assessments including chemistry, hematology & urinalysis	X	X	X	X	X	X	X	X	X	X	No fasting required
Pharmacokinetics Sample			X	X	X		X	X	X		PK samples should also be drawn when there are any SAEs or severe AEs that are possibly drug related
Pharmacogenetics Sample	X										
Physical Exam	X	X			X		X		X		PE will be done at W4 if warranted by emergence of new AE's
Physical Measurements	X	X	X	X	X		X	X	X		

	Screen	**Baseline**	**Week 4 (+/-5)**	**Week 8 (+/-5)**	**Week 12 (+/-5)**	**Week 16[2] (+/-5)**	**Week 24 (+/-7)**	**Week 36 (+/-7)**	**Week 48[3, 5] Or Early DC**	**Follow-up Week 2 (ONLY for subjects NOT entering extension phase or who discontinued early)[4]**	**Additional Instruction**
Visit	**Visit 1**	**Visit 2**	**Visit 3**	**Visit 4**	**Visit 5**	**Visit 6**	**Visit 7**	**Visit 8**	**Visit 9**	**Visit 10**	
Vital Signs	X	X	X	X	X		X	X	X		
12-Lead ECG	X			X	X		X		X		
Concomitant Medication Review	X	X	X	X	X		X	X	X	X	
Adverse Event Assessments		X	X	X	X		X	X	X	X	
Sheehan Suicidality Tracking Scale (STS)	X	X	X	X	X		X	X	X	X	
Clinical Outcome Assessments											
f-SARA	X	X	X	X	X		X	X	X	X	This should be the first clinical outcome assessment done at each visit. Subjects with a 2-point change or greater on the Modified Functional SARA between Screening and Baseline will be not be eligible for randomization.

	Screen	**Baseline**	**Week 4 (+/-5)**	**Week 8 (+/-5)**	**Week 12 (+/-5)**	**Week 16[2] (+/-5)**	**Week 24 (+/-7)**	**Week 36 (+/-7)**	**Week 48[3, 5] Or Early DC**	**Follow-up Week 2 (ONLY for subjects NOT entering extension phase or who discontinued early)[4]**	**Additional Instruction**
Visit	**Visit 1**	**Visit 2**	**Visit 3**	**Visit 4**	**Visit 5**	**Visit 6**	**Visit 7**	**Visit 8**	**Visit 9**	**Visit 10**	
PIFAS		X	X	X	X		X	X	X		
FARS-ADL		X	X	X	X		X	X	X		
FARS-FUNC		X	X	X	X		X	X	X		
Neuro-QOL Lower Extremity		X		X	X		X	X	X		
Neuro-QOL Upper Extremity		X		X	X		X	X	X		
Neuro-QOL Fatigue		X		X	X		X	X	X		
CGI-I				X			X		X		
PGI				X			X		X		
Clinical Drug Supply											
Randomization		X									
Dispense Study Drug[6]		X	X	X	X		X	X	X		Dispense at Wk 48 only if entering Extension Phase
Drug Accountability			X	X	X		X	X	X		

1 Subjects will be provided with pregnancy tests to take in between every 3-month office visit. Subjects should be instructed to contact the site if they become pregnant at any time during the study. Pregnancies should be reported to PPD.
2 The week 16 Visit is only for routine labs, which may be done locally
3 The week 48 Visit is +/- 2 Weeks to ensure the primary f-SARA rater is present at the 48 Week Visit. Visit Window during the Randomization Phase is +/- 5 Days for Weeks 4, 8, 12, and 16; and +/- 7 days for Weeks 24 and 36
4 All subjects who discontinue study treatment early should complete an early discontinuation visit as well as the 2-Week Post Dose Visit. The 2-Week Post Dose Visit would not need to occur if the subject discontinued dosing more than 2 weeks prior to the early discontinuation visit
5 If absolutely necessary, treatment duration may be extended due to the COVID-19 pandemic. Under these circumstances, the sponsor medical monitor should be consulted and must approve the request to change the treatment duration (may be extended if necessitated by the COVID-19 pandemic from 48 weeks up to maximum of 60 weeks).
6 If the study site needs to send drug overnight via certified and tracked courier and this is acceptable to the institution because a visit is absolutely not possible because of the COVID-19 pandemic, this is permissible per study. The sponsor should be consulted prior to shipping drug.

Table 9: Schedule of Assessments and Events – Extension Phase

	Extension Week 4	Extension Week 8	Extension Week 12	Extension Week 16[2]	Extension Week 24	Extension Week 36	Extension Week 48	Extension Week 60	Extension Week 72	Extension Week 84	Extension Week 96 or Early DC	2-Wk Post Last Dose[4]
Visit	**Ext. Visit 1**	**Ext. Visit 2**	**Ext. Visit 3**	**Ext. Visit 4**	**Ext. Visit 5**	**Ext. Visit 6**	**Ext. Visit 7**	**Ext. Visit 8**	**Ext. Visit 9**	**Ext. Visit 10**	**Ext. Visit 11**	**Ext. Visit 12**
Eligibility Assessments												
Neurological Exam					X		X		X		X	
Pregnancy Test for WOCBP (serum)					X		X	X	X	X	X	X
Pregnancy Test for WOCBP (urine)[1]	X	X	X			X						
Safety Assessments												
Laboratory Assessments including urinalysis	X				X		X		X		X	
Pharmacokinetic Sample[3]												
Lab: LFT tests only (ALT, AST, BILI, GGT)		X	X	X		X		X		X		X
Physical Exam					X		X		X		X	
Physical Measurements	X	X	X		X	X	X	X	X	X	X	
Vital Signs	X	X	X		X	X	X	X	X	X	X	
12-Lead ECG					X		X		X		X	
Concomitant Medication Review	X	X	X		X	X	X	X	X	X	X	X
Adverse Event Assessments	X	X	X		X	X	X	X	X	X	X	X
Sheehan Suicidality Tracking Scale (STS)	X	X	X		X	X	X	X	X	X	X	X

	Extension Week 4	Extension Week 8	Extension Week 12	Extension Week 16[2]	Extension Week 24	Extension Week 36	Extension Week 48	Extension Week 60	Extension Week 72	Extension Week 84	Extension Week 96 or Early DC	2-Wk Post Last Dose4
Visit	**Ext. Visit 1**	**Ext. Visit 2**	**Ext. Visit 3**	**Ext. Visit 4**	**Ext. Visit 5**	**Ext. Visit 6**	**Ext. Visit 7**	**Ext. Visit 8**	**Ext. Visit 9**	**Ext. Visit 10**	**Ext. Visit 11**	**Ext. Visit 12**
Clinical Outcome Assessments												
f-SARA	X	X	X		X	X	X	X	X	X	X	X
PIFAS	X	X	X		X		X		X		X	
FARS-ADL	X	X	X		X		X				X	
FARS-FUNC	X	X	X		X		X				X	
Neuro-QOL Lower Extremity		X			X		X				X	
Neuro-QOL Upper Extremity		X			X		X				X	
Neuro-QOL Fatigue					X		X				X	
CGI-I					X		X				X	
PGI					X		X				X	
Clinical Drug Supply												
Dispense Study Drug	X	X	X		X	X	X	X	X	X		
Drug Accountability	X	X	X		X	X	X	X	X	X	X	

[1] Subjects will be provided with pregnancy tests to take in between every 3-month office visit. Subjects should be instructed to contact CRC if they become pregnant at any time during the study.
[2] The Extension Week 16 Visit is only for routine labs, which may be done locally. Visit window is +/- 7 days during the extension phase.
[3] PK samples to be drawn only if there is an AE determined to be related to study drug.
[4]All subjects who discontinue study treatment early should complete an early discontinuation visit as well as the 2-Week Post Dose Visit. The 2-Week Post Dose Visit would not need to occur if the subject discontinued dosing more than 2 weeks prior to the early discontinuation visit.

9.3 Laboratory Test Toxicity Grades

Table 10: 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 (CO2)	<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	

Panel[1]	**Grade 1 (Mild)**	**Grade 2 (Moderate)**	**Grade 3 (Severe)**	**Grade 4 (Potentially Life-Threatening)**
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
eGFR MDRD	≥60 to < LLN mLmin/1.73 m2	≥30 to < 60 mL/min/1.73 m2	≥15 to < 30 mL/min/1.73 m2	< 15 mL/min/1.73 m2
Glucose (mmol/L) - Low	3.05 - <3.55	2.22 - <3.05	1.67 - <2.22	<1.67
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			
LDL cholesterol, age ≥ 18 years	≥ 130 to < 160 mg/dL; ≥ 3.37 to < 4.12 mmol/L	≥ 160 to < 190 mg/dL; ≥ 4.12 to < 4.90 mmol/L	≥ 190 mg/dL; ≥ 4.90 mmol/L	
LDL cholesterol, age > 2 to < 18 years	≥ 110 to < 130 mg/dL; ≥ 2.85 to < 3.37 mmol/L	≥ 130 to < 190 mg/dL; ≥ 3.37 to < 4.90 mmol/L	≥ 190 mg/dL; ≥ 4.90 mmol/L	
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

Panel[1]	**Grade 1 (Mild)**	**Grade 2 (Moderate)**	**Grade 3 (Severe)**	**Grade 4 (Potentially Life-Threatening)**
Urinalysis				
Urine glucose	Trace or 1+; > 180 to ≤ 250 mg/dL	2+; > 250 to ≤ 500 mg/dL	3+ or higher; > 500 mg/dL	
Protein	Trace or 1+	2+	3+ or higher	

[1]Graded by CTCAE Version 5.0 (2017) or DAIDS Version 2.1 (2017)
If a value falls into >1 toxicity grade category, then the highest toxicity grade is chosen.
If fasting status is yes, then fasting toxicity grades are used, else non-fasting toxicity grades are used.

9.4 Relevant Protocol Deviations

Relevant eligibility protocol deviations include the following categories:

- Subject <18 or >75 years old
- SCA genotype
    - Subjects who do not have SCA1, SCA2, SCA3, SCA6, SCA7, SCA8 or SCA10
    - SCA type doesn't match randomization strata
- Screening/Baseline f-SARA
    - Screening f-SARA score <3
    - Screening or baseline f-SARA gait score<1
    - A ≥2-point difference between screening and baseline f-SARA score
    - A score of 4 on any of the individual f-SARA scores (screening or baseline)
- MMSE<24
- Screening laboratory test abnormalities
    - Hepatic
        - TSH> ULN and T4< LLN
        - AST, ALT or GGT >1.5xULN
        - TBIL>2xULN
    - P-amylase or lipase >1.5xULN
    - HbA1c≥8%
    - 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). [Scr: Standardized serum creatinine]

- Hematologic
  - Hemoglobin <10 g/dL
  - WBC <3.0 x 103/mm3
  - Platelet count <100,000/mm3

- Positive HIV test at screening (positive Western Blot)
- HBsAg or HCV positive at screening
- QTc (Bazett's) and QTc (Fridericia) interval > 480 msec at screening or baseline and confirmed by repeat measurement or uncontrolled arrhythmia or frequent premature ventricular contraction (PVCs) (> 5/minute) or Mobitz Type II second or third degree atrioventricular (AV) block or evidence of acute or sub-acute myocardial infarction or ischemia
- History of Riluzole
- Prohibited medications
  - Chlorzoxazone within 30 days prior to randomization (baseline visit) or during the Randomization Phase of the study
  - Aminopyridine is within 30 days prior to randomization (baseline visit) or during the Randomization Phase of the study
  - Acetylcholinesterase inhibitors within 30 days prior to randomization (baseline visit) or during the Randomization Phase of the study
  - Memantine; topiramate, lamotrigine, N-acetylcysteine, ketamine, sodium valproate within 30 days prior to randomization (baseline visit) or during the Randomization Phase of the study
  - Varenicline (Chantix®) within 30 days prior to randomization (baseline visit) or during the Randomization Phase of the study
  - Tricyclic antidepressants and mono-amine-oxidase (MAO) inhibitors within 30 days prior to randomization (baseline visit) or during the study
- Study drug compliance <80% or >110%

[redacted]

[redacted]

[redacted]

[redacted]

10 REFERENCES

Carpenter, J.R., Roger, J.H., and Kenward, M.G. (2013). Analysis of longitudinal trials with protocol deviation: A Framework for Relevant, Accessible Assumptions, and Inference via

Multiple Imputation. *J Biopharmaceutical Statistics, 6 (23)*, 1352 – 1371.

Yang Yuan. (2014). Sensitivity Analysis in Multiple Imputation for Missing Data. Paper SAS270.

Neuro Qol, Quality of Life Neurological Disorders, Scoring Manual, version 3.0, April 2021.