TITLE: Randomized, Double-Blind, Placebo-Controlled, Parallel-Group Study to Evaluate the Efficacy and Safety of BHV-3241 in Subjects with Multiple System Atrophy (M-STAR Study)

NCT NUMBER: NCT03952806

DATE: 01-Jun-2021

Biohaven Pharmaceuticals

Protocol BHV3241-301

Randomized, Double-Blind, Placebo-Controlled, Parallel-Group Study to Evaluate the Efficacy and Safety of BHV-3241 in Subjects with Multiple System Atrophy (M-STAR Study)

Statistical Analysis Plan

Version 1.0

Date: 01-June-2021

SIGNATURE PAGE

Protocol Title:	Randomized, Double-Blind, Placebo-Controlled, Parallel-Group Study to Evaluate the Efficacy and Safety of BHV-3241 in Subjects with Multiple System Atrophy (M-STAR Study)
Sponsor:	Biohaven Pharmaceuticals, Inc.
Document Version:	1.0
Date:	01-June-2021
Author:	PPD PPD CCI
	Signature:_____
	Date:_____

Sponsor Approval

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

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

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

Sponsor Signatories:

PPD PPD PPD Biohaven Pharmaceuticals	Signature:_____ Date:_____
PPD PPD PPD PPD Biohaven Pharmaceuticals	Signature:_____ Date:_____

TABLE OF CONTENTS

CCI

[redacted]

LIST OF TABLES

LIST OF FIGURES

REVISION HISTORY

Version	Description of Change
V1.0	Original Version

ABBREVIATIONS

Abbreviation	Definition
AE	Adverse Event
ALP	Alkaline Phosphatase
ALT	Alanine Aminotransferase
ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
AST	Aspartate Aminotransferase
CCI	[redacted]
BID	Twice Daily
BUN	Blood Urea Nitrogen
BP	Blood pressure
CCI	[redacted]
CGI-I	Clinical Global Impression of Improvement
CGI-S	Clinical Global Impression of Severity
CI	Confidence Interval
CCI	[redacted]
CPK	Creatine Phosphokinase
CSR	Clinical Study Report
DILI	Drug Induced Liver Injury
DMC	Data Monitoring Committee
ECG	Electrocardiogram
EGCG	Epigallocatechingallate
CCI	[redacted]
GGT	Gamma-Glutamyl Transferase
HbA1c	Hemoglobin A1c
HDL	High-density Lipoprotein
ICH	International Conference on Harmonisation
IWRS	Interactive Web-Based Response System
KM	Kaplan-Meier
LDH	Lactate Dehydrogenase
LDL	Low-Density Lipoprotein
LFT	Liver Function Test
LSM	Least-square Mean
MAR	Missing at Random
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple Imputation

Abbreviation	Definition
mITT	Modified Intent to Treat
MMRM	Mixed Model for Repeated Measures
MNAR	Missing Not at Random
[redacted]	
modUMSARS	Modified UMSARS
MPO	Myeloperoxidase
[redacted]	
MSA	Multiple System Atrophy
MSA-C	Multiple System Atrophy with Cerebellar Ataxia
MSA-P	Multiple System Atrophy with Parkinsonian Features
MSA-QoL	Multiple System Atrophy Quality of Life Scale
OLE	Open-Label Extension
PD	Pharmacodynamic
PGI-S	Patient Global Impression of Severity
PID	Patient Identification Number
PK	Pharmacokinetic
[redacted]	
PT	Preferred Term
QD	Once Daily
RBC	Red Blood Cell
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Standard Deviation
SE	Standard Error
[redacted]	
SMQ	Standardized MedDRA Query
SOC	System Organ Class
S-STS	Sheehan Suicidality Tracking Scale
TEAE	Treatment Emergent Adverse Event
TSH	Thyroid-Stimulation Hormone
ULN	Upper Limit of Normal
UMSARS	Unified Multiple System Atrophy Rating Scale
VAS	Visual Analog Scale
WBC	White Blood Cell
WHO	World Health Organization
[redacted]	

1 INTRODUCTION

This document presents the statistical analysis plan (SAP) for Biohaven Pharmaceuticals, Protocol BHV3241-301: Randomized, Double-Blind, Placebo-Controlled, Parallel-Group Study to Evaluate the Efficacy and Safety of BHV-324 in Subjects with Multiple System Atrophy (MSA) (M-STAR Study).

This SAP is based on protocol V7.0 dated January 20, 2021. It contains the analysis details and methodology to answer the study objectives, including planned summary tables, by-subject listings, and figures, which will provide the basis for the results section of the clinical study report (CSR). Operational aspects related to collection and timing of planned clinical assessments are not repeated in this SAP unless relevant to the planned analyses.

1.1 Research Hypothesis

BHV-3241 (Verdiperstat) monotherapy for 48 weeks is superior to placebo in the treatment of MSA.

1.2 Schedule of Analyses

The first scheduled unblinded analysis will be the primary analysis, which will be conducted after the last subject completes their Week 48 visit and the database has been locked. This will summarize all efficacy, safety, laboratory and other data collected through the entire study. In addition, data may be locked, blinded analyses conducted, and reports produced as required to support safety monitoring or regulatory requirements.

In addition to the above, a Data Monitoring Committee will review efficacy and safety data on a periodic basis, with activities fully described in a DMC Charter.

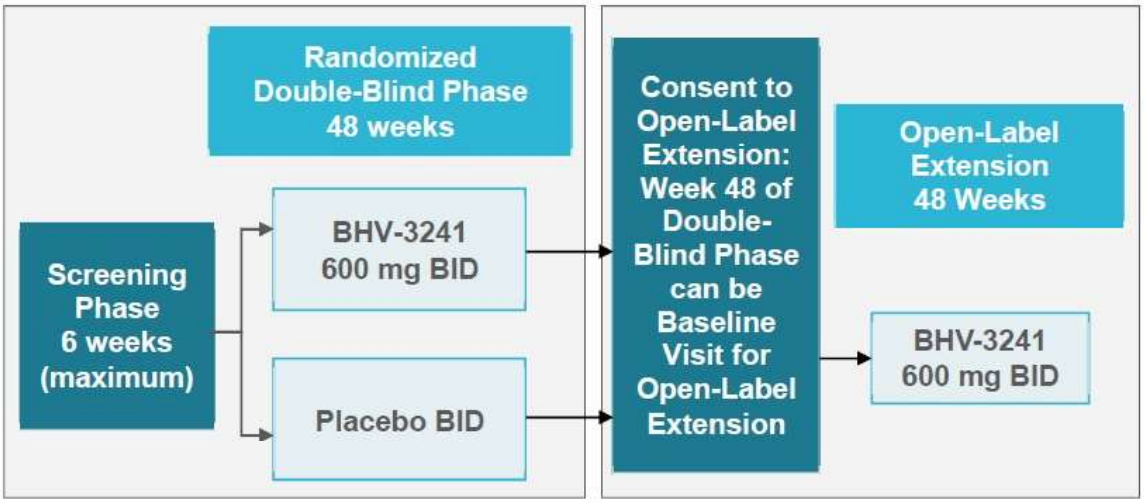
A final analysis will be conducted at the end of the open-label extension (OLE) phase.

2 STUDY DESCRIPTION

2.1 Study Design

Verdiperstat is a Phase 3, multicenter, randomized, double-blind, 2-arm placebo-controlled parallel-group study designed to evaluate the efficacy and safety of Verdiperstat in a population of patients with MSA (MSA-C and MSA-P). The study is planned to consist of a Screening phase lasting a maximum of approximately 6 weeks and a randomized double-blind treatment phase of approximately 48 weeks, per Figure 1. It is anticipated that the Randomization phase will include a dose titration period of approximately 2 weeks followed by a full dose period of approximately 46 weeks. Approximately 325 subjects in total are planned to be randomized in a 1:1 ratio to receive either Verdiperstat 600 mg BID, or matching placebo BID. It is anticipated that subjects will be assessed at clinic visits, per the Schedule of Assessments & Events (Appendix I: Schedule of Assessments and Events – Randomized Double-blind Phase). In addition, subjects who are completing or have completed Week 48 of the double-blind treatment phase may be offered the opportunity to enroll in an OLE phase, in which subjects will receive open-label treatment with Verdiperstat for approximately 48 weeks (Appendix II: Schedule of Assessments and Events – Open-Label Extension Phase).

Figure 1: Study Schematic



Screening Phase

The purpose of the Screening phase is to ensure that the appropriate subjects are entered into the trial and remain stable during the pre-treatment period. It is estimated that approximately 360 subjects will enter the Screening phase of the trial. The Investigator will determine that the subject meets eligibility criteria and will collect demographic and medical data presenting a full characterization of the subject. All attempts should be made to obtain medical and pharmacy records to confirm the subject's medical and medication treatment history. During Screening,

Biohaven (or designee) may review relevant clinical information to confirm the diagnosis and subtype of MSA, for study eligibility and stratification prior to randomization. Screening will range from a minimum of 1 day to a maximum of 42 days.

If initial Screening lab sample results do not meet protocol-specified eligibility criteria and the Investigator determines that a repeat sample may meet the range required for eligibility, a repeat lab sample may be collected (one repeat collection/analysis allowed).

Rescreening

After obtaining Sponsor approval, a subject who does not meet protocol required eligibility criteria during Screening but who may potentially become eligible (e.g., logistical reasons such as exceeding screening window to accommodate time required for adjustment/stability of concomitant medication, washout of prohibited medication or repeat procedures to allow for recheck/confirmation of an abnormality that may meet the required range in a reasonable amount of time), may be entered as a Screen Failure in the Interactive Web Response System (IWRS) and then rescreened one time with Sponsor approval (entered as a Screen Failure and Rescreen in IWRS and the electronic case report form [eCRF]). If the Sponsor approves Rescreening for a subject, instructions will be provided to the site regarding the specific Screening procedures/evaluations that will need to be repeated for Rescreening, as this is dependent on the amount of time elapsed.

Randomization Phase

Subjects who are determined to be eligible for the study will be randomized into the trial. It is estimated that approximately 325 subjects will be randomized and assigned to receive Verdiperstat (600 mg BID) or matching placebo (BID) in a 1:1 ratio, with randomization stratified by disease subtype of either MSA-P vs. MSA-C, diagnostic category of either possible MSA vs. probable MSA, and country.

Subjects should be administered the Day 1/first dose of study medication (300 mg QD or matching placebo QD) while in the office/clinic on the day of the Baseline visit.

Dose titration period: For the rest of Week 1, subjects will ingest either 300 mg QD of Verdiperstat or matching placebo QD. From start to end of Week 2, subjects will ingest either 300 mg BID of Verdiperstat or matching placebo BID. Subjects should take the study drug twice a day; dosing in the morning and evening approximately 12 hours apart. If there is a delay in dosing, the interval between two doses should be no less than 6 hours.

An end of Week 2 assessment of compliance and tolerance to this dose titration schedule will be conducted. Per Investigator judgment, additional evaluations/procedures may be conducted at this time point as appropriate. If tolerability issues are experienced with 300 mg QD or 300 mg BID dosing, the titration schedule may be modified.

Full dose period: Starting with Week 3 and for the remainder of the study, subjects will ingest either 600 mg BID of Verdiperstat or matching placebo BID. Subjects should take the study drug twice a day; dosing in the mornings and evenings approximately 12 hours apart.

At Weeks 4, 12, 24, 36 and 48, the morning dose of study medication should not be taken on the day of the study visit until it is administered by site personnel during the study visit, so that one pre-dose (trough) and one post-dose pharmacokinetic (PK) sample can be collected approximately 2-4 hours post-dose. Study medication can be taken without regard to meals, however, the last meal time prior to collection of PK samples should be obtained. If subjects have difficulty tolerating Verdiperstat 600 mg BID, then the dose may be adjusted.

End of Treatment

Patients will have a Week 48/Early Discontinuation visit when they either complete the study at Week 48 or discontinue from study medication early (prior to Week 48). The Week 48/Early Discontinuation assessments/procedures should be conducted as an in person visit at the study site. Every effort should be made to conduct the Week 48 visit and maintain a +/- 7 day window. However, due to the COVID-19 public health emergency, the Week 48 visit window may be expanded beyond the +/- 7 day window up to an additional 12 weeks (maximum treatment duration of 60 weeks). However, every attempt should be made to conduct the visit as close as possible to the date that the visit is due. Under these circumstances, the Investigator should discuss the specific circumstances of these cases with the sponsor medical monitor (or designee), who must approve the request to extend the treatment duration, prior to any modification of the visit window.

If the Week 48 visit window is modified, participants should have safety assessments evaluated remotely and safety labs conducted at the time of the scheduled Week 48 visit.

For subjects who discontinue early, depending on the reason for early discontinuation, some procedures may not be conducted at this visit. Conduct of all procedures, if possible, is encouraged. If possible, the subjects who discontinue from study medication prior to Week 48 should be contacted by a follow up phone call at the approximate time of their Week 48 visit, to collect information on clinical status (including Time to Event measures of disease progression, and if applicable, AEs, SAES).

Open-Label Extension Phase

The OLE Phase is an optional 48 week open-label treatment phase following the double-blind phase of the study. All subjects who enter the OLE phase will sign a new informed consent form.

All participants must have their Week 48 visit in-person prior to enrolling in the OLE phase. Due to COVID-19, the timing of this visit may be delayed.

Participants who early-terminated from the double-blind phase of the trial or who discontinued the study medication for any reason will not be eligible for enrollment in the OLE phase.

When a site is open for recruitment to the OLE Phase, participants completing the double-blind treatment phase who choose to participate in the OLE phase and are eligible to continue, must consent to the OLE phase and directly enter the OLE phase starting at Week 48 visit of the double-blind treatment phase. The Week 48 visit will serve as the Baseline-Extension visit in the OLE phase for participants who move directly from the double-blind treatment phase to the

OLE. Thereafter, participants will undergo visits at Week 2-Extension, Week 4-Extension, Week 8-Extension, Week 12-Extension, and then every 12 weeks of the OLE phase as outlined in Appendix II: Schedule of Assessments and Events – Open-Label Extension Phase.

A participant who has completed the double-blind treatment phase of the study prior to the OLE phase of the study being approved and open for recruitment at the study center where they are participating, can reconsent and start the OLE phase as long as the investigator determines that open-label treatment offers an acceptable risk-benefit profile for the participant. Once the study site is open for recruitment for the OLE, study participants who previously completed the double-blind phase of the study will have up to approximately 12 weeks (from the date of OLE opening at the site) to enroll and complete the OLE Baseline-Extension Visit, regardless of their date of double-blind phase completion.

For participants who have been off of the study drug for less than or equal to 4 weeks, the subject should attend an Abbreviated Drug Dispensation Visit at the investigative site, when medical and/or medication changes will be reviewed prior to dispensing open-label Verdiperstat. In this case, for these participants, the Week 48 visit will also serve as the Baseline-Extension visit in the OLE Phase. Thereafter, participants will undergo visits at Week 2-Extension, Week 4-Extension, Week 8-Extension, Week 12-Extension, and then every 12 weeks of the Extension Phase as outlined in Appendix II: Schedule of Assessments and Events – Open-Label Extension Phase.

For participants who have been off of the study drug for greater than 4 weeks, these participants should be re-assessed and re-consented by the Investigator during an OLE Baseline-Extension Visit. From the date that the OLE is open for recruitment at their investigative site, these participants will have up to approximately 12 weeks to complete the OLE Baseline-Extension Visit. Thereafter, participants will undergo visits at Week 2-Extension, Week 4-Extension, Week 8-Extension, Week 12-Extension, and then every 12 weeks of the Extension Phase as outlined in Appendix II: Schedule of Assessments and Events – Open-Label Extension Phase.

Dose Titration: Subjects entering the OLE phase of the study will receive open-label active Verdiperstat and will follow the same dose titration schedule as the titration in the double-blind phase of the study, to ensure that all participants safely continue from either active Verdiperstat or placebo to open-label Verdiperstat. Subjects should be administered the Day 1/first dose of open-label Verdiperstat (300 mg QD) while in the office/clinic on the day of the Baseline-Extension visit.

For the rest of Week 1-Extension of the OLE phase, subjects will ingest one tablet of Verdiperstat (300 mg) per day in the morning for one week. From start to end of Week 2-Extension, subjects will ingest 300 mg BID of Verdiperstat. Subjects should take one 300 mg tablet of Verdiperstat twice a day (600 mg per day); in the mornings and evenings, approximately 12 hours apart. If there is a delay in dosing, the interval between two doses should be no less than 6 hours.

An end of Week 2-Extension assessment of compliance and tolerance to this dose titration schedule will be conducted. Per Investigator judgement, additional evaluations/procedures may be conducted at this timepoint as appropriate. If tolerability issues are experienced with 300 mg QD or 300 mg BID dosing, the titration schedule may be modified.

Starting with Week 3-Extension and for the remainder of the OLE phase of the study (additional 46 weeks), subjects will ingest 600 mg BID of Verdiperstat. Subjects should take two 300 mg tablets of Verdiperstat twice a day (1200 mg per day); in the mornings and evenings, approximately 12 hours apart.

Subjects who were taking a reduced dose of 300 mg BID (one tablet twice a day) of the blinded study medication due to tolerability issues when they completed the double-blind phase of the protocol, may continue taking 300 mg BID during the OLE.

2.2 Treatment Assignment

Subjects will be randomized, in a 1:1 ratio to receive either placebo (BID) or Verdiperstat (600 mg BID). Randomization will be stratified by disease subtype of either MSA-P vs. MSA-C, diagnostic category of possible MSA vs. probable MSA, and country. Container assignments will be obtained by the Investigator (or designee) via the IWRS system.

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 without the need for unblinding, by assuming that the subject is receiving active product.

Unblinding will be managed via the IWRS system. The clinical supply manager, PK and pharmacodynamic (PD) vendors, IWRS vendor, pharmacovigilance and DMC roles may be unblinded before data are unblinded for the primary endpoint and all subjects complete the study. Except as noted above, other members of the Biohaven research team will remain blinded. In cases of accidental unblinding, the Medical Monitor will be contacted and every attempt will be made to preserve the blind.

An independent DMC will monitor efficacy and safety throughout the treatment phase and will be unblinded to treatment groups. The tables, listings, and figures prepared for the closed reports will be completely unblinded revealing actual treatment groups. The unblinded statistical team will have access to treatment codes prior to database lock in order to prepare the closed reports and the DMC will have access to the unblinded reports on a secure file sharing platform. Further information regarding DMC blinding and unblinding is documented in a separate DMC Charter and DMC SAP.

During the OLE phase, all participants will not be blinded and will be receiving active Verdiperstat.

2.4 Protocol and Protocol Amendments

BHV3241-301 SAP Version 1.0 is based on BHV241-301 Protocol Version 7.0.

3 STUDY OBJECTIVES AND ESTIMANDS

3.1 Objectives

3.1.1 Primary Objectives

- To evaluate the efficacy of Verdiperstat, compared to placebo, as measured by a change from baseline in a modified Unified MSA Rating Scale (UMSARS), consisting of a subset of items from Part I and Part II, at Week 48 (see Section 6.3.2.1 for derivation of modified UMSARS).

- To assess the safety and tolerability of Verdiperstat, relative to placebo, in subjects with MSA.

3.1.2 Secondary Objectives

3.1.2.1 Key Secondary Objectives

- To evaluate the efficacy of Verdiperstat, compared to placebo, as measured by the Clinical Global Impression of Improvement (CGI-I) score at Week 48.

- To evaluate the impact of Verdiperstat on quality of life, compared to placebo, as measured by a change from baseline in the motor subscale of the MSA-Quality of Life (MSA-QoL) scale at Week 48.

- To evaluate the impact of Verdiperstat on quality of life, compared to placebo, as measured by a change from baseline in the non-motor subscale of the MSA-QoL scale at Week 48.

- To evaluate the efficacy of Verdiperstat, compared to placebo, as measured by a change from baseline in the UMSARS Part I and Part II total score at Week 48.

3.1.2.2 Other Secondary Objectives

- To assess the impact of Verdiperstat, relative to placebo, as measured by a change from baseline at Week 48 in the following instruments:
    - Patient Global Impression of Severity (PGI-S),
    - Clinical Global Impression of Severity (CGI-S),
    - UMSARS Part III,
    - UMSARS Part IV.

CCI

CCI

[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 male and female patients, ≥40 to ≤80 years of age, with a diagnosis of possible or probable MSA according to consensus clinical criteria[1], including subjects with either MSA subtype, MSA-Parkinsonism (MSA-P) or MSA-Cerebellar (MSA-C). Protocol Section 5 provides a detailed description of inclusion and exclusion criteria for this study.

Endpoint

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.

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

Unless otherwise noted, efficacy analyses will be based on observed data only. For the main analyses of efficacy for primary, secondary, and CCI [redacted] 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). Hence all types of intercurrent events are incorporated using a treatment policy strategy, which evaluates the treatment as assigned and taken including missed or modified doses, drug discontinuation, and concurrent treatments. Study dropout including dropouts and missed visits due to COVID-19 pandemic (and corresponding discontinuation of assessments) will be handled by a hypothetical strategy to estimate what the outcome would have been at the designated time point if all subjects were continued to be assessed.

Sensitivity analyses using multiple imputation (MI) will be conducted on the primary and key secondary endpoints using a jump to reference and copy increment from reference approach to assess the impact of the missing at random (MAR) assumption (see Section 6.3.2.3).

Population-level Summary

Refer to Table 4 in Section 6.3.

Data Sources for Endpoints

UMSARS scores are from the electronic case report form (eCRF) Unified MSA Rating Scale Parts I-IV.

- CGI-I scores are from the CGI-I eCRF.
- MSA-QoL scores are from the MSA relevant Time to Event Clinical Milestones eCRF.
- PGI-S scores are from the PGI-S eCR.
- CGI-S scores are from the CGI-S eCRF.
- CCI
- CCI
- CCI
- CCI
- CCI
- CCI
- CCI
- AEs (i.e., non-serious AEs or SAEs) are from AE eCRF.
- Central laboratory test results are from an external source.
- Local laboratory test results are from the local lab results eCRF.
- CCI
- PK concentrations and PK parameters are from an external source.

3.2.1 Primary Objective Estimand

The estimands corresponding to the primary objectives for this study are shown in Table 1.

Table 1: Primary Objective Estimands

Objective	**To evaluate the efficacy of Verdiperstat, compared to placebo, as measured by a change from baseline in a modified UMSARS score at Week 48**
Efficacy Endpoint	Change from baseline in a modified UMSARS score at Week 48
Summary	Using the mITT population: • Descriptive statistics for modified UMSARS score and change from baseline as a continuous variable by treatment group at each visit • Difference in mean change from baseline in modified UMSARS scores between treatment groups at week 48 estimated and tested using a mixed-effect model for repeated measures (MMRM)
Intercurrent Events	1) Main Analysis - Observed data only 2) Supplementary Analysis 1 – While on-treatment strategy 3) Supplementary Analysis 2 – Compliant subjects hypothetical strategy 4) Sensitivity Analysis – Multiple imputation jump to reference and copy increment from reference 5) Sensitivity Analysis – Observed data only using narrower visit window for Week 48
Objective	**To assess the safety and tolerability of Verdiperstat, relative to placebo, in subjects with MSA**
Safety Endpoint	Frequency of unique subjects with serious adverse events (SAEs), AEs leading to study drug discontinuation, AEs related to study drug, clinically significant ECG abnormalities, and clinically significant laboratory test abnormalities
Summary	Number and percentage of subjects with safety events and findings for the safety population
Intercurrent Events	Treatment policy strategy, i.e., no accounting for intercurrent event

3.2.2 Secondary Objective Estimands

The estimands corresponding to the secondary objectives are shown in Table 2.

Table 2: Secondary Objective Estimands

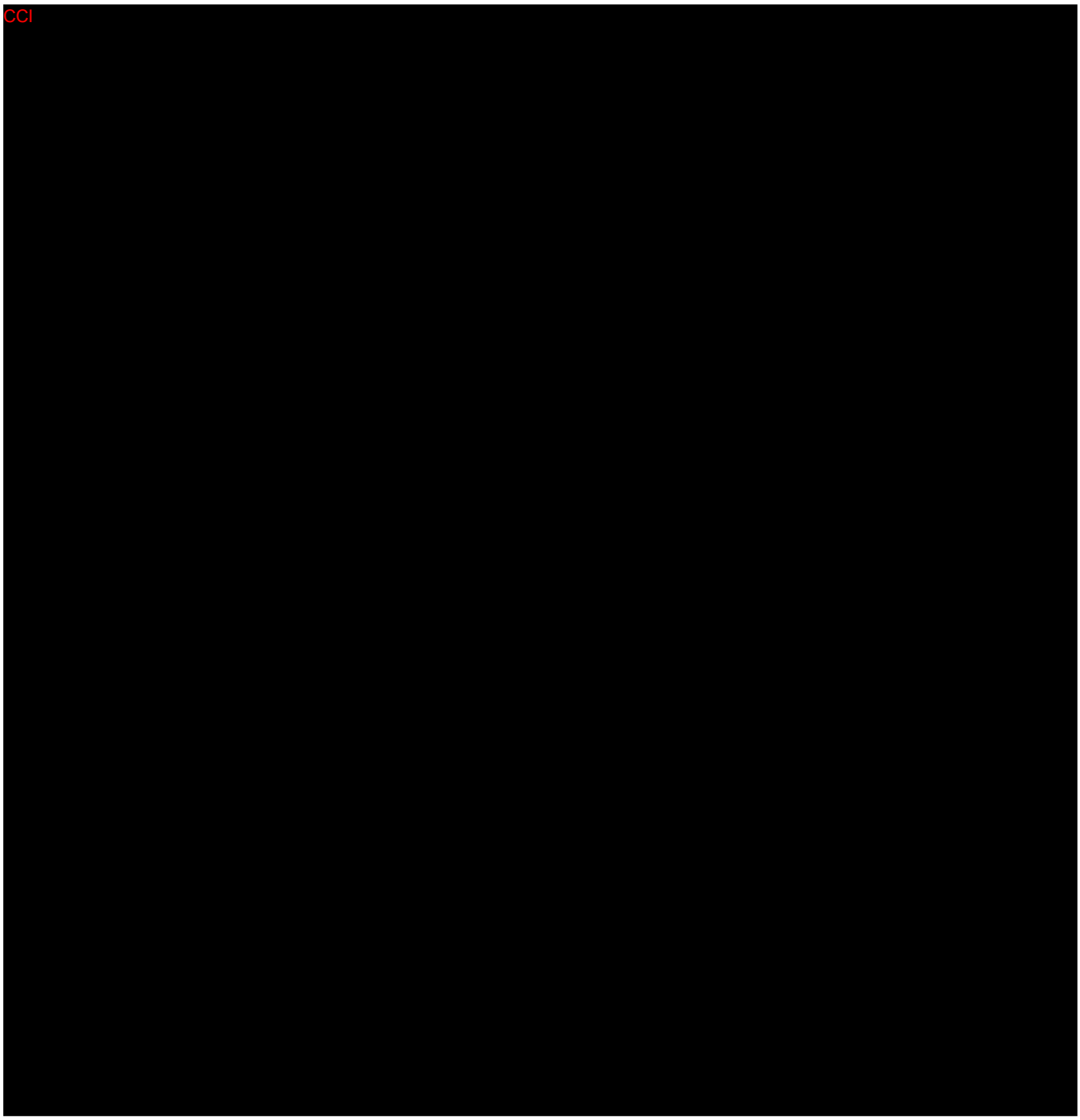
Objective 1	**To evaluate the efficacy of Verdiperstat, compared to placebo, as measured by the CGI-I score at Week 48**
Efficacy Endpoint	CGI-I score at Week 48
Summary	Using the mITT population: • Descriptive statistics and number and percentage of subjects in each response category by treatment group at each visit • Difference in mean CGI-I score between treatment groups at Week 48 estimated and tested using MMRM
Intercurrent Events	1) Main Analysis – Observed data only 2) Sensitivity Analysis – Multiple imputation jump to reference and copy increment from reference
Objective 2	**To evaluate the impact of Verdiperstat on quality of life, compared to placebo, as measured by a change from baseline in the motor subscale of the MSA-QoL scale at Week 48**
Efficacy Endpoint	MSA-QoL motor subscale change from baseline at Week 48
Summary	Using the mITT population: • Descriptive statistics for the MSA-QoL motor subscale scores and the change from baseline in the subscale scores by treatment group at each visit • Difference in mean change from baseline in the MSA-QoL motor subscales scores at Week 48 estimated and tested using MMRM
Intercurrent Events	1) Main Analysis – Observed data only 2) Sensitivity Analysis – Multiple imputation jump to reference and copy increment from reference
Objective 3	**To evaluate the impact of Verdiperstat on quality of life, compared to placebo, as measured by a change from baseline in the non-motor subscale of the MSA-QoL scale at Week 48**
Efficacy Endpoint	MSA-QoL non-motor subscale change from baseline at Week 48
Summary	Using the mITT population: • Descriptive statistics for the MSA-QoL non-motor subscale scores and the change from baseline in the subscale scores by treatment group at each visit • Difference in mean change from baseline in the MSA-QoL non-motor subscales scores at Week 48 estimated and tested using MMRM
Intercurrent Events	1) Main Analysis – Observed data only 2) Sensitivity Analysis – Multiple imputation jump to reference and copy increment from reference

Objective 4	**To evaluate the efficacy of Verdiperstat, compared to placebo, as measured by a change from baseline in the UMSARS Part I and Part II total score at Week 48**
Efficacy Endpoint	UMSARS Part I and II total score change from baseline at Week 48
Summary	Using the mITT population: • Descriptive statistics for the UMSARS Part I and Part II total scores and the change from baseline in the total scores by treatment group at each visit • Difference in mean change from baseline in the UMSARS Part I and Part II total scores at Week 48 estimated and tested using MMRM
Intercurrent Events	1) Main Analysis – Observed data only 2) Sensitivity Analysis – Multiple imputation jump to reference and copy increment from reference
Objective 5	**To assess the impact of Verdiperstat, relative to placebo, as measured by a change from baseline at Week 48 in PGI-S**
Efficacy Endpoint	PGI-S change from baseline at Week 48
Summary	Using the mITT population: • Descriptive statistics for the PGI-S score and the change from baseline score by treatment group at each visit • Difference in mean change from baseline in the PGI-S score at Week 48 estimated and tested using MMRM
Intercurrent Events	Observed data only
Objective 6	**To assess the impact of Verdiperstat, relative to placebo, as measured by a change from baseline at Week 48 in CGI-S**
Efficacy Endpoint	CGI-S change from baseline at Week 48
Summary	Using the mITT population: • Descriptive statistics for the CGI-S scores and the change from baseline in the scores by treatment group at each visit • Difference in mean change from baseline in the CGI-S scores at Week 48 estimated and tested using MMRM
Intercurrent Events	Observed data only
Objective 7	**To assess the impact of Verdiperstat, relative to placebo, as measured by a change from baseline at Week 48 in UMSARS Part III**
Efficacy Endpoint	UMSARS Part III (orthostatic change in systolic blood pressure (BP), diastolic BP, and heart rate) change from baseline at Week 48
Summary	Using the mITT population: • Descriptive statistics for the UMSARS Part III results and the change from baseline in results by parameter and treatment group at each visit • Difference in mean change from baseline in the UMSARS Part III results at Week 48 estimated and tested using MMRM
Intercurrent Events	Observed data only

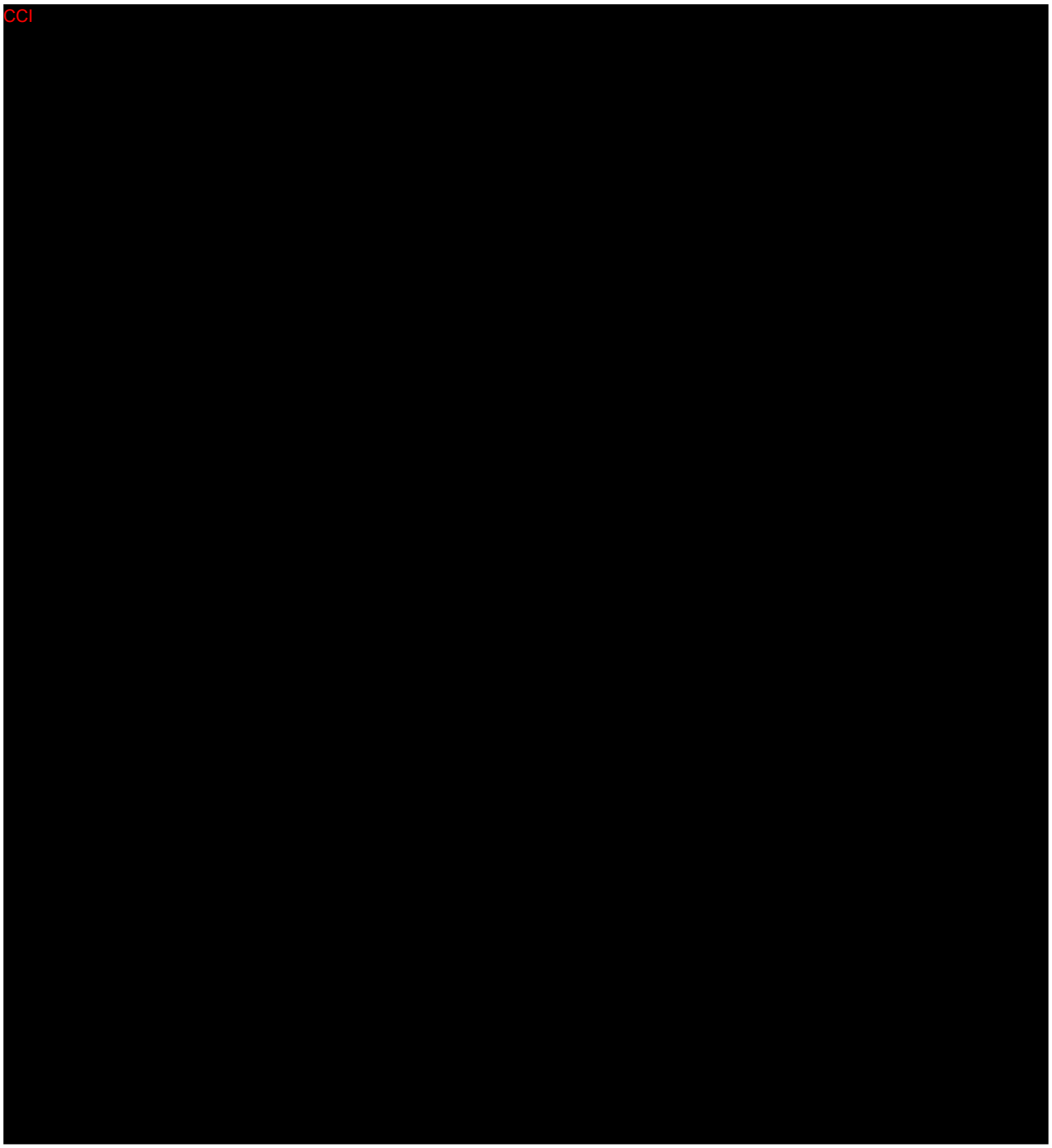
Objective 8	**To assess the impact of Verdiperstat, relative to placebo, as measured by a change from baseline at Week 48 in UMSARS Part IV**
Efficacy Endpoint	UMSARS Part IV change from baseline at Week 48
Summary	Using the mITT population: • Descriptive statistics for the UMSARS Part IV scores and the change from baseline in scores by treatment group at each visit • Difference in mean change from baseline in the UMSARS Part IV scores at Week 48 estimated and tested using MMRM
Intercurrent Events	Observed data only

CCI

[redacted]


CCI


CCI


CCI

CCI

[redacted]

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: Subjects who signed an informed consent form and were assigned a Patient Identification number (PID).
- Randomized subjects: Enrolled subjects who received a treatment assignment from the IWRS.
- Treated/Safety: Enrolled subjects who received at least 1 dose of blinded study therapy (Verdiperstat or placebo).
- Modified Intent-to-Treat (mITT) subjects: Randomized subjects who received at least one dose of blinded study therapy and provided a baseline and at least one post-baseline efficacy assessment.
- Open-Label Extension Treated subjects: Enrolled subjects in the extension phase who received at least 1 dose of open-label study therapy (Verdiperstat).
- Verdiperstat Treated Participants: Enrolled subjects who received at least 1 dose of Verdiperstat (blinded or open-label)

4.2 Treatment Groups

The 2 treatment groups are Verdiperstat 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:

- Sex (male, female)
- Race (Asian, Black or African American, White, and Other)
- Age (<65, ≥65 years)
- Smoking status (current, previous, never)
- Disease subtype (MSA-P, MSA-C)
- MSA diagnostic category (Possible, Probable)
- Country
- Concomitant background therapy use for Parkinson's Disease (yes, no)

5 SAMPLE SIZE, POWER, AND TYPE I ERROR

5.1 Sample Size and Power

The sample size for this study will be approximately 325 randomized subjects.

Based on unpublished data from the Promesa (2016) study.[2] studying EGCG (epigallocatechingallate) vs placebo with a treatment duration of 48 weeks, the placebo group had a 4.88 point increase with a standard deviation of 4.85 in the modified UMSARS Part I and Part II total score. With an expected improvement of 40% at Week 48 for Verdiperstat compared to placebo, a total of 260 subjects (130 per group) provides 90% power based on a 2-sample, two-sided t-test. With a drop-out rate of ~20%, the study will randomize approximately 325 subjects.

5.2 Type I Error

Type-1 error will be controlled for the primary and key secondary efficacy endpoints by testing them with a gate-keeping procedure. The primary endpoint, the change from baseline in the modified UMSARS score measured at the Week 48 visit, will be tested at a 2-sided alpha level of 0.05. If this test is significant, then the key secondary efficacy endpoints of change from baseline in CGI-I, MSA-QoL motor subscale and non-motor subscale scores, and UMSARS Part I and Part II total score at week 48 will be tested using Hochberg's procedure in which the p-values from the test of each secondary endpoint will be ranked from lowest to highest, and all tests for the null hypothesis of each secondary endpoint with a p-value less than the test with the highest p-value below its critical p-value (defined as 0.05/(k-j+1) where j is the rank of the test and k is the total number of tests) can be rejected.

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

No attempt will be made to adjust for multiplicity when testing the non-key secondary or CCI endpoints. Any endpoints subjected to significance testing are evaluated at an unadjusted two-sided alpha level of 0.05.

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 and AEs are coded using Medical Dictionary for Regulatory Activities (MedDRA) Version 24.

Concomitant medications are coded using World Health Organization Drug Dictionary (WHO-DD), released 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 DB Phase and OLE Phase will be analyzed separately. For tables summarizing subjects who received at least one dose of DB or OLE study medication (Verdiperstat), summary statistics will be calculated from both phases combined for subjects randomized to Verdiperstat and from the OL Extension Phase for subjects randomized to placebo.

Treatment Group Presentation

Tables present results by treatment group (i.e., Verdiperstat and placebo) with the following exceptions:

- Results for the enrolled analysis set are presented only by overall, without treatment group.
- Results for study population parameters (see Section 6.2) and pre-treatment safety also include overall.

- OLE tabulations will include the sequence of the treatment, i.e., Verdiperstat/ Verdiperstat or Placebo/ Verdiperstat.

Time-to-event Tables

Time-to-event endpoints are summarized with KM tables displaying the following at each time interval:

- Number of subjects at risk, defined as those who did not have the event just prior to the start of the interval and who were not censored prior to the interval
- Number of subjects with events, defined as those who had the event for the first time during the interval
- Number of subjects censored, defined as those who did not have the event and whose censoring date/time (e.g., last contact date, last available measurement date/time) is in the interval
- Cumulative probability of having the event with lower and upper 95% CI limits, based on the KM product limit method.

Time-to-event distributions of endpoints are tabulated with the following descriptive statistics: number and percentage of subjects with events; number and percentage of subjects censored in or before the last time interval; number and percentage of subjects censored in the last time interval; time-to-event median, first quartile, and third quartile, with 95% CI. The 95% CI for the median is estimated using the method of Brookmeyer and Crowley.

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 as 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.

6.1.1.3 Figures

Time-to-event endpoints are summarized graphically with KM plots that display cumulative percentage of having the event (100 × cumulative probability; see Section 6.1.1.1) on the y-axis versus time on the x-axis. Censored observations are denoted with symbols, and the number of subjects at risk at the start of each time interval are annotated below the x-axis.

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 + 1 decimal place. The SD 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 principal analysis of efficacy for primary, secondary and CCI endpoints, no imputation will be performed on missing data following discontinuation from the study.

If the primary and secondary endpoints results are found to be statistically significant, then sensitivity analyses using multiple imputation (MI) will be conducted on the primary and key secondary endpoints using the copy increment from reference and jump to reference methods to assess the impact of the missing at random (MAR) assumption (see Section 6.3.2.3).

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

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

The following general rules will be used:

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

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

- If the start date is completely missing, an AE will be considered treatment-emergent unless the stop date is before study drug administration.

6.1.4 Adjustments for Covariates and Stratification

The randomization was stratified by disease subtype (MSA-P vs. MSA-C), diagnostic category (possible MSA vs. probable MSA), and country. Hence, treatment group comparisons of continuous efficacy endpoints will be adjusted by disease subtype, diagnostic category, and country. If there are sparse data within a stratum, then stratum may be combined.

The analysis of the primary endpoint and other continuous efficacy endpoints will also be adjusted by the baseline value of the endpoint as a covariate in statistical models.

6.2 Study Population

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 mITT population sample

- Excluded from mITT population sample (e.g., not treated, no baseline or no post-baseline efficacy data).

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 mITT population who are not in the efficacy analysis set, including the reason for exclusion (i.e., not treated, no baseline or post-baseline efficacy data). 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, randomization strata (i.e., disease subtype, diagnostic category, country), 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

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

- Number of screened subjects
- Number of subjects who failed screening and reasons for screen failure (e.g., withdrawal of consent, adverse event). For subjects whose reason for screen failure is due to inclusion/exclusion criteria not met, the inclusion/exclusion criteria will be included
- Number of rescreened subjects
- Number of subjects who failed rescreening and reasons for rescreen failures (e.g., withdrawal of consent, adverse event). For subjects whose reason for screen failure is due to inclusion/exclusion criteria not met, the inclusion/exclusion criteria will be included
- Number of randomized subjects.

Subject disposition for the double-blind phase is tabulated for the randomized subjects as the number and percentage of subjects in the following categories:

- Number of subjects who completed the double-blind phase
- Number of subjects who prematurely withdrew from the double-blind phase and reasons for withdrawal
- Number of subjects who continued into the OLE phase.

Subject disposition for the OLE phase is tabulated for the randomized subjects as the number and percentage of subjects in the following categories

- Number of subjects who continued into the OLE phase
- Number of subjects who completed the OLE phase
- Number of subjects who prematurely withdrew from the OLE phase and reasons for withdrawal.

A by-subject listing of study completion information for both the double-blind and OLE phase, including the reason for premature study 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.

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

Relevant protocol deviations 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. Relevant protocol deviations will be tabulated for the randomization and OLE phases, separately, and all relevant deviations will be presented in a by-subject listing.

A Microsoft Excel file of protocol deviations is extracted from the clinical trial management system (CTMS) by Biohaven Clinical Operations. This file serves as the raw data source of protocol deviations and classifies deviation severity as key and non-key. Significant protocol deviations are defined as those categorized as key deviations.

6.2.5 Baseline Characteristics

Baseline characteristics include (1) demographics and other relevant baseline characteristics, (2) medical and psychiatric history, (3) prior non-study medications, and (4) prior Parkinson's medications.

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

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 treated: 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

Double-Blind Phase

For the double-blind phase, the extent of subject exposure (i.e., treatment duration) will be quantified as the number of days on study medication and measured from the time the subject received the first dose of study medication until the time the subject received the last dose of study medication. Treatment duration includes missed dose days; i.e., last dose date of double-blind randomized study medication – first dose date of double-blind randomized study medication + 1.

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

Additionally, percent (%) compliance will be calculated as (total days on double-blind randomized study medication) / (treatment duration of double-blind randomized study medication) × 100.

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

Open-Label Extension Phase

For the OLE Phase, the extent of subject exposure (i.e., treatment duration) will be quantified as the number of days on study medication and measured from the time the subject received the first dose of study medication in the OLE Phase until the time the subject received the last dose, either at the end of the study or withdrawal from the OLE Phase. Treatment duration includes missed dose days, i.e., last dose date of OLE study medication – first dose date of OLE study medication + 1.

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

Additionally, percent (%) compliance will be calculated as (total days on open-label Verdiperstat study medication) / (treatment duration of open-label Verdiperstat study medication) × 100.

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

Double-Blind and Open-Label Extension Phases Combined

For the combined phase data, the extent of subject exposure (i.e., treatment duration) to Verdiperstat during the entire study will be quantified as the number of days on study medication (Verdiperstat) and measured from the time the subject received the first dose of Verdiperstat until the time the subject received the last dose of Verdiperstat, either at the end of the study or withdrawal from the double-blind or OLE phases. Treatment Duration includes missed dose days within a study phase and excludes the dosing break between study phases (i.e., (last dose day of Verdiperstat study medication – first dose date of Verdiperstat study medication + 1) – (number of days with a dosing break between the double-blind and OLE phases)).

The extent of subject exposure (i.e., total days on Verdiperstat) during the entire study will also be calculated to exclude days when subjects missed a dose (i.e., Treatment Duration – number of days a subject missed all doses of study medication).

Additionally, percent (%) compliance will be calculated as (total days on Verdiperstat study medication) / (Treatment Duration of Verdiperstat study medication) × 100.

6.2.6.2 Prior and Concomitant Non-study Medications

Prior and concomitant medications will be coded using the WHO Drug Global B3 March 2021. Results will be tabulated by Anatomic Therapeutic Class (ATC) and PT by treatment group and overall.

Prior medications are defined as those start or stop date < IWRS enrollment date – 14 days. Unless the start date of the medication is after the last DB 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.' For medication with partial start/stop date, if it cannot be definitively shown that the medication was not taken during the DB treatment period, the medication is included as concomitant.

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

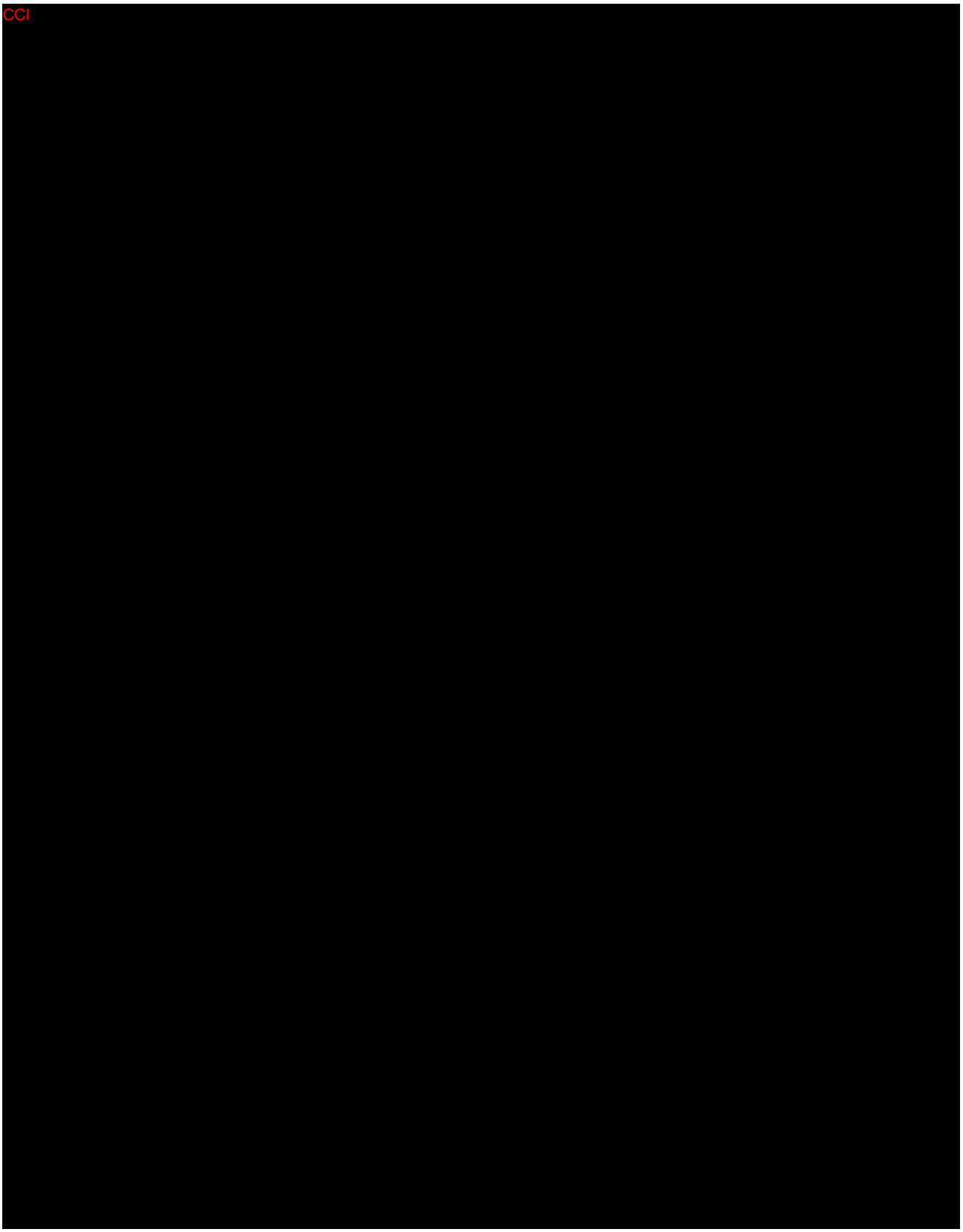
6.3 Efficacy

Unless otherwise noted, the primary and secondary efficacy analyses will be conducted on the mITT population. All efficacy data will be included in listings by subject, treatment group, and visit (as applicable). When applicable, the efficacy analyses will be presented for the double-blind and OLE phases, separately and combined. For the OLE phase, no inferential statistics will be derived.

For the double-blind phase, baseline is considered as the last available assessment on or before the first day of study treatment. For the OLE Phase, baseline is considered as (1) the double-blind Week 48 visit assessment for subjects who have been off of the study drug for less than or equal to 4 weeks, or (2) the OLE Baseline-Extension visit assessment for subjects who have been off of the study drug for greater than 4 weeks.

Table 4: Primary, Secondary, and [redacted] 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 modified UMSARS scores at Week 48 of the DB Phase	Continuous	MMRM	Screening, Baseline, Weeks 4, 12, 24, 36, 48,
Secondary Estimand				
S01	Difference between treatment groups in change from baseline in CGI-I score at Week 48 of the DB Phase	Continuous	MMRM	Weeks 4, 12, 24, 36, 48
S02	Difference between treatment groups in change from baseline in MSA-QoL motor subscales score at Week 48 of the DB Phase	Continuous	MMRM	Baseline, Weeks 24, 48, Baseline Ext, Weeks 12 Ext, 24 Ext, 36 Ext, 48 Ext
S03	Difference between treatment groups in change from baseline in MSA-QoL non-motor subscales score at Week 48 of the DB Phase	Continuous	MMRM	Baseline, Weeks 24, 48, Baseline Ext, Weeks 12 Ext, 24 Ext, 36 Ext, 48 Ext
S04	Difference between treatment groups in change from baseline in UMSARS Part I and II total score at Week 48 of the DB Phase	Continuous	MMRM	Screening, Baseline, Weeks 4, 12, 24, 36, 48, Baseline Ext, Weeks 12 Ext, 24 Ext, 36 Ext, 48 Ext
S05	Difference between treatment groups in change from baseline in PGI-S score at Week 48 of the DB Phase	Continuous	MMRM	Baseline, Weeks 4, 12, 24, 36, 48
S06	Difference between treatment groups in change from baseline in CGI-S score at Week 48 of the DB Phase	Continuous	MMRM	Baseline, Weeks 4, 12, 24, 36, 48
S07	Difference between treatment groups in change from baseline in UMSARS Part III (orthostatic change in systolic BP, diastolic BP, and heart rate) at Week 48 of the DB Phase	Continuous	MMRM	Screening, Baseline, Weeks 4, 12, 24, 36, 48, Baseline Ext, Weeks 12 Ext, 24 Ext, 36 Ext, 48 Ext
S08	Difference between treatment groups in change from baseline in UMSARS Part IV score at Week 48 of the DB Phase	Continuous	MMRM	Screening, Baseline, Weeks 4, 12, 24, 36, 48, Baseline Ext, Weeks 12 Ext, 24 Ext, 36 Ext, 48 Ext



MMRM = Mixed effects model for repeated measures; ANOVA = Analysis of variance, KM = Kaplan-Meier

6.3.1 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 DB and OLE phases, separately.

Treatment Group Comparison Using MMRM

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

- Dependent variable, e.g., modified UMSARS score
- Fixed effects (categorical variables): treatment group, visit, treatment group-by-visit interaction, randomization strata (i.e., disease subtype, diagnostic category, and country). Visit represents on-study analysis visit.
- Covariate (continuous variable): baseline value of dependent variable, baseline neurofilament level
- Random effect: subject

Error degrees of freedom are determined by the Kenward-Roger approximation.

The covariance structure for within subject error will be initially modeled using the unstructured covariance matrix. If the model fails to converge, then a Huynh-Feldt structure will be used, followed by an AR(1) structure.

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

- Least-squares mean with standard error (SE) and 95% CI by treatment group
- Difference in least-squares means between treatment groups (Verdiperstat – placebo) with df, SE, 95% CI, and p-value.

Treatment Group Comparison Using ANOVA

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

- Dependent variable, e.g., brain volume
- Fixed effects (categorical variables): treatment group, baseline value of the dependent variable, randomization strata (i.e., disease subtype, diagnostic category, and country)

The following statistics are tabulated from the ANOVA results:

- Least-squares mean with standard error (SE) and 95% CI by treatment group
- Difference in least-squares means between treatment groups (Verdiperstat – placebo) with df, SE, 95% CI, and p-value.

6.3.2 Primary Endpoint: Unified Multiple System Atrophy Rating Scale (UMSARS)

6.3.2.1 Main Analyses

The UMSARS is a clinician-rated scale comprised of four parts. Part I, Historical Review, is an assessment of functioning across various areas. Part II is a Motor Examination. Part III is an Autonomic Examination, and includes supine and standing vital signs, orthostatic change (derived by subtracting standing measurements from corresponding supine measurements), and orthostatic symptoms. Part IV is a Global Disability Scale. Although the UMSARS will be used to collect data in this study, the primary outcome variable for the study will be a modified UMSARS, composed of a subset of items from UMSARS Part I and Part II, with the items selected based on feedback from Health Authorities, expert opinion, and analysis of data from previous clinical trials in MSA.

The UMSARS is assessed during the Screening phase to determine eligibility for study entry and will subsequently be administered at Baseline, and Weeks 4, 12, 24, 36 and 48, Baseline-Extension, Week 12-Extension, Week 24-Extension, Week 36-Extension, Week 48-Extension, and/or early discontinuation.

The primary endpoint is the change from baseline in the modified UMSARS score measured at the Week 48 visit. The items in the modified UMSARS will contain the following Part I items: Speech, Cutting Food and Handling Utensils, Dressing, Hygiene, Walking and Urinary Function. The following Part II items will also be included: Arising from Chair, Posture and Gait. In the analysis of all items, the response categories of 0 and 1 will be combined into a single category of 0 and then subsequent scores recoded to from 2 to 1, 3 to 2 and 4 to 3, reducing the range of scoring options to 0 to 3, before summing the individual items to derive the modified UMSARS.

Summary of primary estimand:

a. The population of interest for the primary estimand are subjects diagnosed with possible or probable MSA as well as meeting the other inclusion/exclusion criteria of the protocol.

b. Efficacy is measured by the change from baseline to 48 weeks for the modified UMSARS endpoint.

c. The treatment evaluated is the treatment as assigned to study subjects. Hence all types of intercurrent events are incorporated using a treatment policy strategy, which evaluates the treatment as assigned and taken including missed or modified doses, drug discontinuation,

and concurrent treatments. Study dropout including dropouts and missed visits due to COVID-19 pandemic (and corresponding discontinuation of assessments) will be handled by a hypothetical strategy to estimate what the outcome would have been at the designated time point if all subjects were continued to be assessed.

d. The mean change from baseline between treatment and placebo on the primary endpoint will be compared at 48 weeks. A significant difference with an error rate of 5% will be required for success.

The summary will be computed using a Mixed Model for Repeated Measures (MMRM) that will include fixed effect factors for treatment group, randomization strata (i.e., disease subtype, diagnostic category, and country), visit, and the treatment group by visit interaction. Baseline modified UMSARS score and the baseline neurofilament level will enter the model as covariates, and subject will be a random effect. 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. Error degrees of freedom will be calculated using Kenward-Roger approximation if an unstructured covariance structure fits appropriately; otherwise, a sandwich estimator will be utilized to estimate the covariance structure and degrees of freedom will be calculated using the between-within method.

MMRM-based estimates (i.e., least-squares mean [LSM] with corresponding SD and 95% CI) of values and changes from baseline will be presented by treatment group and week. In addition, the LSM difference in change from baseline between treatment groups (Verdiperstat – placebo) at Week 48 with corresponding SD, 95% CI, and p-value from MMRM will be presented.

Observed values and changes from baseline in the modified UMSARS score will be summarized using descriptive statistics over time by treatment group.

Effect sizes (Cohen's D) of the individual items within the modified UMSARS will also be summarized to assess the strength of the effect of Verdiperstat on each individual item.

6.3.2.2 Supplementary Analyses

The estimand for the first supplementary analysis will pursue a while on-treatment strategy. This will use the mITT population. The endpoint will be the change from baseline in the modified UMSARS score at the last measured time point. For subjects that stay on study, this will be the Week 48 visit. For subjects that discontinue early, the last on-treatment visit will be used as the endpoint. The population summary will be the difference between placebo and Verdiperstat treated subjects at their last visit. The summary will be computed using an Analysis of Covariance (ANCOVA), estimated by Restricted Maximum Likelihood (REML) that will include fixed effect factors for treatment group and randomization strata. Baseline modified UMSARS score will be used as a covariate.

The estimand for the second supplementary analysis will pursue a hypothetical strategy that estimates the effect of treatment in compliant subjects. The endpoint and population summary will be the same as used for the primary estimand. However, the population and handling of intercurrent events will be handled differently. The population will consist of subjects that are at least 80% compliant with the drug regimen and complete the Week 48 visit. In terms of intercurrent events, subjects that discontinue treatment, or are lost to follow-up, before Week 48 are not included in the analysis.

6.3.2.3 Sensitivity Analyses

For the primary endpoint, sensitivity analyses using multiple imputation (MI) will be conducted using a copy increment from reference and jump to reference approach to assess the impact of the MAR assumption of the analysis model used.

The primary analysis of the modified UMSARS score based on a MMRM model assumes that data is MAR and subjects who discontinue study medication prematurely have a response profile for the remainder of the treatment phase similar to subjects who completed the 48 weeks of the treatment phase (an analysis of a "de jure" or "hypothetical" based estimand). In order to explore the impact of these assumptions, a sensitivity analysis will be conducted where subjects who discontinue Verdiperstat prematurely have a response profile similar to those subjects on placebo using both a copy increment from reference and jump to reference approach. These analyses use a "defacto" estimand and are based on the methods described in.[3]

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

For each analysis, the imputed data will consist of 1000 imputed data sets (using a MCMC length of 100). Both the parameter-estimation model and the imputation model will include fixed effect factors for treatment group, visit, and the treatment group by visit interaction. The baseline modified UMSARS score and visit by baseline interaction terms will be entered as covariates.

For each imputed dataset, the change from baseline for the modified UMSARS score at Week 48 will be based on observed and imputed data. An ANCOVA model will be used with baseline

modified UMSARS score and randomization strata (i.e., disease subtype, diagnostic category, and country) as covariates, and treatment groups will be compared at Week 48 based on LSM difference between Verdiperstat and placebo in each of the imputed data sets. Results from the analysis of each imputed dataset (LSM treatment differences and standard errors [SE]), will be combined using Rubin's imputation rules to produce a pooled LSM estimate of treatment difference, 95% CI, and a pooled p-value for the test of null hypothesis of no treatment effect.

An additional sensitivity analysis will be conducted on the primary endpoint using the same main analysis described in Section 6.3.2.1 but with a narrower visit window for Week 48 (See Section 7.2).

6.3.2.4 Subgroup Analyses

Only descriptive summaries will be provided for the primary endpoint for each subgroup. See Section 4.3 for list of subgroups of interest.

6.3.3 Secondary Efficacy Endpoints

If the main analysis of the primary endpoint is significant, then the main analyses of the key secondary endpoints are tested using the Hochberg procedure, as described in Section 5.2. The key secondary endpoints include: CGI-I; MSA-QoL motor subscale; MSA-QoL non-motor subscale; UMSARS Part I and Part II Total Score.

6.3.3.1 S01: Clinical Global Impression of Improvement (CGI-I)

The CGI-I scale is a clinician-rated scale measuring the change in the subject's total improvement from the time of study start. It is scored on a 7- point scale, ranging from 1 (very much improved) to 7 (very much worse), with a score of 4 indicating no change. The CGI-I should be assessed relative to the CGI-S at the Week 0 (Day 1) visit as the reference. The CGI-I will be administered at Weeks 4, 12, 24, 36 and 48 and/or early discontinuation.

The estimand for this endpoint will pursue a hypothetical strategy of estimating the effect of Verdiperstat relative to placebo had all subjects stayed on treatment. Intercurrent events will be handled by only using the on-treatment data.

The number and percentage of subjects in each category will be summarized: very much improved, much improved, minimally improved, no change, minimally worse, much worse, and very much worse. A similar summary table will be generated for CGI-I with combined categories: 1-3 (improved), 4 (no change), 5-7 (worse).

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

The transformed CGI-I score at Week 48 will be analyzed similar to the MMRM methodology described in Section 6.3.1, but with baseline CGI-S score as the covariate. If found to be statistically significant, similar methods detailed in Section 6.3.2.3 will be employed.

6.3.3.2 S02: MSA-Quality of Life (MSA-QoL) Motor Subscale

The MSA-QoL is a subject-rated scale that was designed to measure health-related quality of life specifically in subjects with MSA. It assesses activities of daily living and has subscales for motor, non-motor, and emotional/social domains. The MSA-QoL will be administered at Baseline, Weeks 24, 48, Baseline-Extension, Week 12-Extension, Week 24-Extension, Week 36-Extension, Week 48-Extension, and/or early discontinuation.

Change from baseline in the motor subscale score of the MSA-QoL at Week 48 will be analyzed with the MMRM methodology described in Section 6.3.1. If found to be statistically significant, the same methods detailed in Section 6.3.2.3 will be employed.

Descriptive statistics for the MSA-QoL motor subscale score and the change from baseline in the subscale score will be presented by treatment group and visit as described in Section 6.3.1.

6.3.3.3 S03: MSA-Quality of Life (MSA-QoL) Non-Motor Subscale

The MSA-QoL is described in Section 6.3.3.2.

Change from baseline in the non-motor subscale score of the MSA-QoL at Week 48 will be analyzed with the MMRM methodology described in Section 6.3.1. If found to be statistically significant, the same methods detailed in Section 6.3.2.3 will be employed.

Descriptive statistics for the MSA-QoL non-motor subscale score and the change from baseline in the subscale score will be presented by treatment group and visit as described in Section 6.3.1.

6.3.3.4 S04: UMSARS Part I and Part II Total Score

As described in Section 6.3.2.1, the UMSARS is a clinician-rated scale comprised of four parts. The Part I (Historical Review) and Part II (Motor Examination) total score will be assessed as a secondary endpoint.

Change from baseline in the UMSARS Part I and Part II total score at Week 48 will be analyzed with the MMRM methodology described in Section 6.3.1. If found to be statistically significant, the same methods detailed in Section 6.3.2.3 will be employed. The UMSARS Part I and Part II total score will exclude item 11 from Part I (sexual function).

Descriptive statistics for the UMSARS Part I and Part II total score and the change from baseline total score will also be presented by treatment group and visit as described in Section 6.3.1.

6.3.3.5 S05: Patient Global Impression of Severity (PGI-S)

The PGI-S is a patient-rated scale and will be administered at Baseline and Weeks 4, 12, 24, 36 and 48 and/or early discontinuation. The patient will rate how they perceive the severity of their illness. The PGI-S is a 1-item questionnaire on a 4-point scale scored as: “normal” (1), “mild” (2), “moderate” (3), or “severe” (4).

Change from baseline in the PGI-S score at Week 48 will be analyzed with the MMRM methodology described in Section 6.3.1.

Descriptive statistics for the PGI-S score and the change from baseline score will be presented by treatment group and visit as described in Section 6.3.1.

Frequency table for PGI-S score shift from baseline to any post-baseline score will be presented as well.

6.3.3.6 S06: Clinical Global Impression of Severity (CGI-S)

The CGI-S scale is a clinician-rated scale measuring the severity of the subject's illness. It is scored on a 7- point scale ranging from 1 (normal, not ill at all) to 7 (among the most extremely ill patients). The CGI-S will be administered at Baseline and Weeks 4, 12, 24, 36 and 48 and/or early discontinuation.

The change from baseline in the CGI-S score will also be analyzed using a MMRM analysis model, similar to the analysis described in Section 6.3.1.

The number and percentage of subjects in each category will be summarized: normal, not at all ill; borderline ill; mildly ill; moderately ill; markedly ill; severely ill; and among the most extremely ill patients.

In addition, the CGI-S scores will be transformed into numeric values ranging from 1 to 7 (1 indicating "normal, not at all ill" and 7 indicating "among the most extremely ill patients"). Descriptive statistics for the CGI-S score and the change from baseline will be presented by treatment group and visit as described in Section 6.3.1.

Frequency table for CGI-S score shift from baseline to any post-baseline score will be presented as well.

6.3.3.7 S07: UMSARS Part III

As described in Section 6.3.2.1, the UMSARS is a clinician-rated scale comprised of four parts. The Part III (Autonomic Examination) score includes supine and standing vital signs, orthostatic change, and orthostatic symptoms and will be assessed as a secondary endpoint.

Change from baseline in the UMSARS Part III results (orthostatic change in systolic BP, diastolic BP, and heart rate) at Week 48 will be analyzed with the MMRM methodology described in Section 6.3.1.

Descriptive statistics for orthostatic change in systolic BP, diastolic BP, and heart rate and the change from baseline for each of these parameters will also be presented by treatment group and visit as described in Section 6.3.1. In addition, the number and percentage of subjects with orthostatic symptoms will be presented by treatment group and visit.

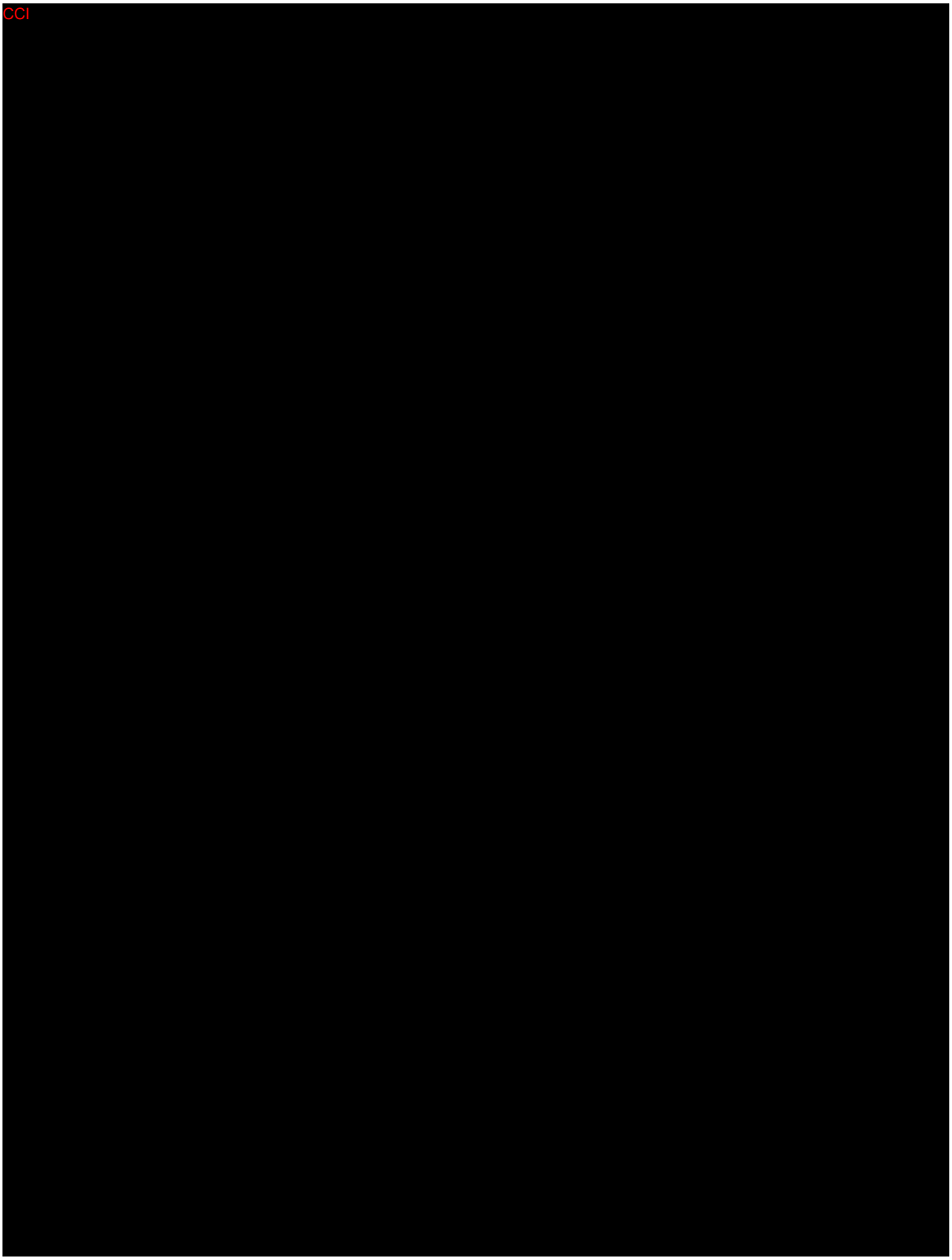
6.3.3.8 S08: UMSARS Part IV

As described in Section 6.3.2.1, the UMSARS is a clinician-rated scale comprised of four parts. The Part IV (Global Disability Scale) score will be assessed as a secondary endpoint.

Change from baseline in the UMSARS Part IV score at Week 48 will be analyzed with the MMRM methodology described in Section 6.3.1.

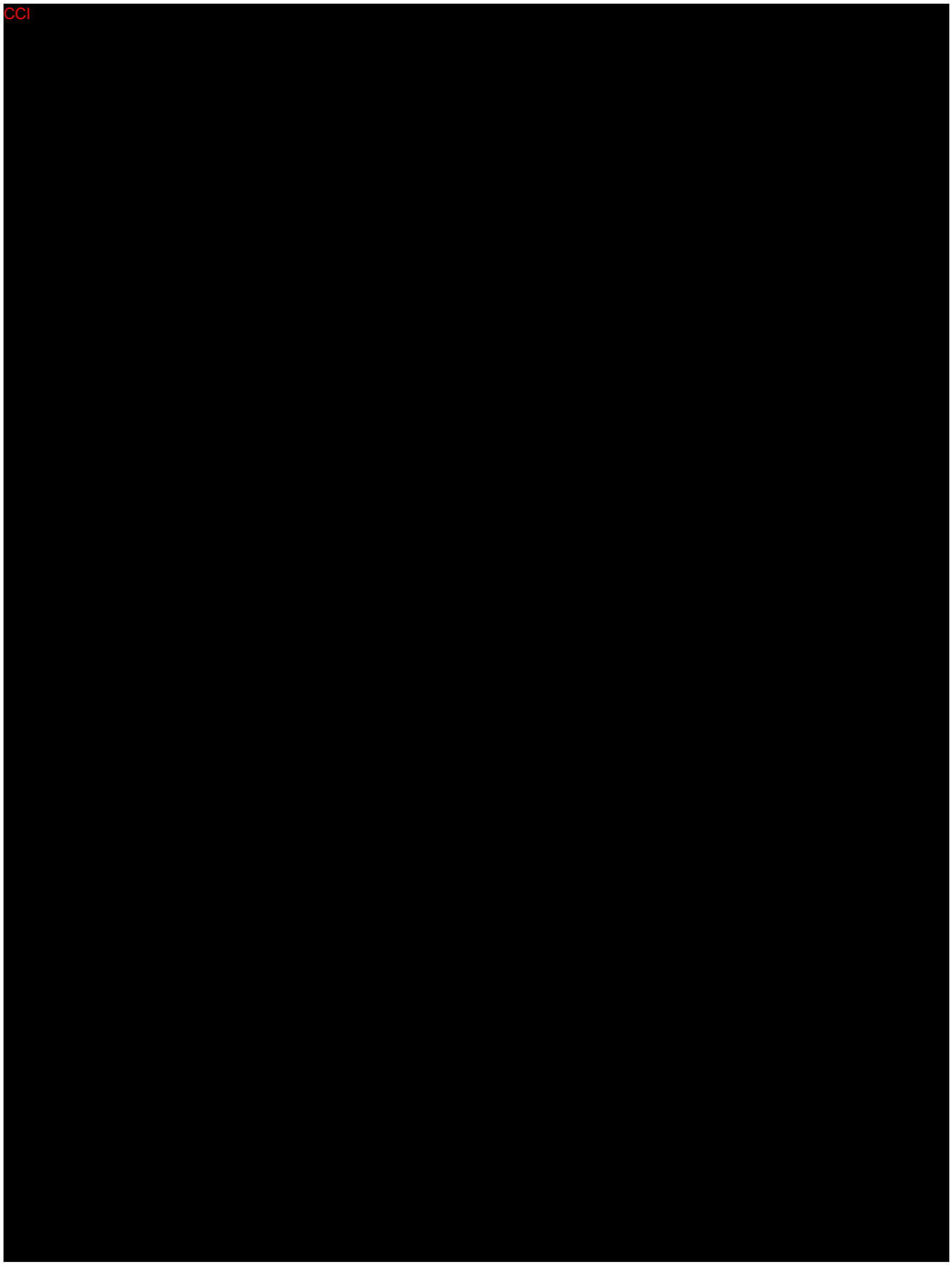
Descriptive statistics for the UMSARS Part IV score and the change from baseline score will also be presented by treatment group and visit as described in Section 6.3.1.

CCI


CCI

CCI

[redacted]


CCI

CCI

[redacted]

CCI

[redacted]

6.4 Safety

Safety and other analyses will be conducted on treated subjects by treatment group. For the DB Phase outputs, treated subjects in the DB Phase will be used; for the OLE Phase outputs, treated subjects in the OLE Phase will be used; and for the combined phase outputs, DB and OLE Verdiperstat treated subjects will be used. All safety and other data will be listed for the entire study with screening, DB Phase, and OLE 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

AEs will be coded using MedDRA and displayed in tables and listings by system organ class (SOC) and preferred term (PT), unless specified otherwise.

Analyses of AEs will be performed for those events that are considered treatment-emergent AEs (TEAEs), where treatment-emergent is defined as any AE that developed, worsened, or became serious on or after first dose of study drug and prior to 30 days after the last dose of study drug.

A TEAE overview without SOC and PT will be presented as the number and percentage of subjects with any of the following TEAEs: TEAE leading to death; serious TEAE; serious TEAE related to study drug; TEAE leading to study drug discontinuation; any TEAE; TEAE related to study drug; and severe AE.

In addition, TEAEs will be tabulated by SOC and PT for all treated subjects for the following endpoints for the DB and OLE phases, separately and combined:

- TEAEs
- TEAEs related to treatment (includes, possibly related and related)
- TEAEs by highest severity (mild, moderate, severe)

- TEAEs related to treatment by highest severity
- Treatment-emergent SAEs
- TEAEs leading to discontinuation of study treatment

In the above tabulations, each subject will contribute only once (i.e., the most related occurrence or the most severe occurrence) to each of the incidence rates in the descriptive analysis, regardless of the number of episodes.

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).

All AEs occurring pre-treatment and during the entire study will be listed. Additional listings will be provided including deaths, SAEs, AEs leading to discontinuation of study drug, and events of special interest (see Section 6.4.2).

CCI

[redacted]

CCI

[redacted]

6.4.3 Laboratory Tests

Laboratory testing will be conducted at Screening, Baseline, and Weeks 2, 4, 8, 12, 24, 36, and 48, Baseline-Extension, Week 4-Extension, Week 8-Extension, Week 12-Extension, Week 24-Extension, Week 36-Extension, Week 48-Extension and/or early discontinuation. Clinical laboratory evaluations include:

Hematology: hemoglobin, hematocrit, platelets, CBC with differential and absolute neutrophil count, RBC count and indices. At Screening and prior to lumbar puncture: INR, PT and partial PT

Serum Chemistry: sodium, potassium, chloride, calcium, magnesium, ALT, AST, LDH, ALP, GGT, phosphorus, bicarbonate, CPK, total protein, albumin, total bilirubin (if greater than 2 mg/dl bilirubin will be fractionated), glucose, creatinine, BUN, uric acid. TSH, free T3 and free T4. Also, anti-TPO antibodies at Baseline, Week 2, Week 4, Week 8, Week 12, Week 24, Week 36, Week 48, Baseline-Extension, Week 4-Extension, Week 8-Extension, Week 12-Extension, Week 24-Extension, Week 36-Extension, Week 48-Extension and whenever TSH is outside of normal range. Additionally, at Screening: total cholesterol, LDL, HDL, triglycerides, folate, HbA1c, P-Amylase or Lipase; Estimated glomerular filtration rate (eGFR).

Urinalysis (collect urine specimen just prior to obtaining weight): macroscopic examination, pH, specific gravity, ketones, nitrites, urobilinogen, leukocyte esterase, protein, creatinine, glucose, and occult blood. If blood, protein, or leukocytes are positive, microscopic examination will be performed on abnormal findings.

Serum pregnancy test will be conducted at Screening, Baseline, and Weeks 4, 12, 24, 36 and 48, Baseline-Extension, Week 4-Extension, Week 12-Extension, Week 24-Extension, Week 36-Extension, Week 48-Extension and early discontinuation. Urine pregnancy tests will be performed (as well as serum) at Baseline and Baseline-Extension (prior to dosing) and at the discretion of the Investigator. FSH is to be conducted on postmenopausal women at Screening to confirm post-menopausal status.

At Screening Only: HBsAg, HCV (reflex HCV-RNA after a positive HCV Ab), HIV antibody detection, and rapid plasma reagin (RPR). Reflex testing will be done for any positive RPR.

At Screening, Baseline and Baseline-Extension Only: Urine Drug Screen for cannabis (medical and recreational), amphetamines (including MDMA/ecstasy), cocaine, barbiturate, PCP, benzodiazepines, tricyclic antidepressants, and/or opiates. Reflex confirmatory testing will be conducted on all positive urine drug screen samples.

At Screening Only: Urine alcohol level.

For those participating in the optional CSF sub-study, analysis conducted by local lab of cell counts (white blood cells and red blood cells [with reflexed differential if either of the counts is abnormal], CSF protein and CSF glucose.

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

On-treatment laboratory abnormalities are those with an assessment date after the date/time of first dose of study drug and within 30 days after the last dose of study drug. For the treatment phase, treatment-emergent laboratory abnormalities will be assessed from the date of first dose of study drug until 30 days after the last dose of study drug.

Clinically significant laboratory abnormalities will be identified as Grade 3 to 4 laboratory test results according to CTCAE version 5.0 if available; otherwise according to DAIDS version 2.1 criteria.

Laboratory Test Change from Baseline

The observed value and change from baseline will be summarized for each continuous laboratory parameter in both conventional and SI units.

Laboratory Test Abnormalities

The number and percentage of treated subjects with at least one on-treatment laboratory assessment will be summarized by treatment group (regardless of baseline) for Grade 0, Grade 1 to 2, Grade 3 to 4, Grade 3, and Grade 4. Note that Grade 3 to 4 abnormalities are considered clinically significant.

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

Liver Function Test Elevations

For the liver function tests, the shift from baseline will be presented by visit and to the maximum observed abnormality for treated subjects in the Randomization and Extension phases. Shift tables will only include treated subjects with an assessment for the specific test of interest at baseline and on treatment.

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

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

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

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

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

- ≤ULN
- >ULN to ≤1.5x ULN
- >1.5x ULN to ≤ 2.0x ULN
- >2x ULN

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

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

For T3 and T4, the following categories will be used to summarize the shift from baseline:

- <LLN
- ≥LLN

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

- <4 mU/L
- 4-10 mU/L
- >10 mU/L

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 DB Phase and DB and OLE Verdiperstat treated subjects.

Additional listings will be presented for the following:

- All abnormal laboratory values considered potentially clinically significant (Grade 3 or 4) for the DB and OLE 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 and Physical Measurements

Vital signs include systolic blood pressure, diastolic blood pressure, heart rate and temperature.

Physical measurements include height, weight, and BMI.

The observed value and change from baseline in vital signs and physical measurements will be summarized at each visit for the double-blind and OLE phases, separately.

A subject listing of vital signs and physical measurements will be provided for enrolled subjects.

6.4.5 Electrocardiogram

Descriptive statistics for ECG interval data (e.g., RR, QRS, PR, QT, QTcF, QTcB), and ventricular heart rate will be reported by treatment group and visit for the double-blind and OLE phases, separately.

The shift from baseline for ECG interpretation will be tabulated by visit for the double-blind and OLE phases, separately and combined. The shift table will include only subjects with a baseline assessment and at least one on-treatment assessment.

In addition, the number and percentage of subjects with at least one post-treatment QTcF > 450 ms, >480 ms, and >500 ms will be summarized by treatment group and visit for the double-blind and OLE phases, separately and combined. Similar number and percentages will be presented for subjects with at least one post-baseline QTcF change from baseline ≥ 30 ms to < 60 ms and those with at least one change from baseline ≥ 60 ms.

A by-subject listing of ECG results will be provided for enrolled subjects.

6.4.6 Sheehan-Suicidality Tracking Scale

The S-STS is a prospective, clinician-administered rating scale that contains 16 questions to track both treatment-emergent suicidal ideation and behaviors. The S-STS will be completed on a paper form at the site at Screening, Baseline and each study visit, including Week 48, Baseline-Extension, Week 2-Extension, Week 12-Extension, Week 24-Extension, Week 36-Extension, Week 48-Extension, or early withdrawal. At Screening, the recall period for completing the S-STS is 6 months prior; at all other visits, the recall period for completing the S-STS is since the last visit (two separate versions will be provided). Subjects who have an S-STS score > 0 should be evaluated by the Investigator. If the Investigator determines that a subject is at risk of suicide or self-harm, appropriate measures to ensure the subject's safety and obtain mental health evaluation must be implemented and the Investigator should evaluate if the subject should be discontinued from the study. In the event the subject is unavailable, the S-STS clinician-administered rating scale will be completed that contains 6 yes/no questions.

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

- Ideation subscale score: Sum of scores (0 – 4) for Questions 2 – 11

- Behavior subscale score: Sum of scores (0 – 4) for Questions 1a, (highest of 12 or any row of 16), (highest of 14 or any row of 15), 17, and 20 based on the scoring used for the 2017 version of scale used in this study. In addition, a Behavior subscale score based on the 2019 version of the scoring will also be calculated where the score for a 'Yes' item 20 is 56 and the scores for 18, 19, 21 and 22 are used from the prior visit if applicable.

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

The self-reported S-STS ideation subscale, behavior subscale, and total score will be summarized as the change from baseline (i.e., <-1, -1, no change, 1, >1) at each visit and the maximum score observed for the double-blind and OLE phases, separately.

All S-STS data will be listed.

6.4.7 Subjects Identified for Narratives

A safety narrative will be prepared for each subject who received at least one dose of Verdiperstat 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 Verdiperstat.

These select events are described in the current version (v1.0) of the Biohaven Safety Narrative Scope for BHV-3241 (Verdiperstat). 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 screened subjects with the select events as described above.

6.4.8 COVID-19 Related Summaries

Analyses are based on the COVID-19 Visit Impact eCRF. At each visit (scheduled or unscheduled) that occur after 01MAR2020, sites complete the COVID-19 Visit Impact eCRF.

COVID-19 visit impact status is based on the response to the lead question "Was this visit impacted by COVID-19 related issues? (yes or no)". If the response to the lead question is "yes", then responses to the following questions are provided:

COVID-19 visit impact type: What was the impact? (check all that apply)

- (1) Missed visit – no assessments done
- (2) In-person visit at site (check all that apply)
    - Not all assessments completed *
    - Scheduled visit occurring earlier or delayed relative to protocol specified schedule *
- (3) Remote visit (indicate one)
    - Virtual visit (video/telemedicine) *
    - Telephone contact *

Subcategories marked with "*" are visit impact characteristics.

COVID-19 visit impact relationship: How was the impact related to COVID-19?

- Subject diagnosed with COVID-19 or quarantined due to COVID-19
- Site closed or access restricted due to COVID-19
- Site open but subject unwilling or unable to come to the site due to COVID-19
- Other

COVID-19 visit impact will be tabulated by as-treated treatment group and overall for the safety analysis set for the double-blind and OLE phases, separately. The number and percentage of subjects in the categories defined by COVID-19 visit impact type, visit impact characteristics, and visit impact relationship will be tabulated. Percentages are based on subjects with ≥ 1 visit with non-missing COVID-19 visit impact.

A by-subject listing of COVID-19 visit impact is provided for the enrolled analysis set. The listing displays COVID-19 visit date, study day derived from the COVID-19 visit date, treatment day derived from the COVID-19 visit date, visit impact status, visit type (scheduled or unscheduled), visit impact type, visit impact characteristics, visit impact relationship, and premature study termination (yes; no; not applicable, subject continuing).

7 CONVENTIONS

7.1 Analysis Periods

Analysis periods are defined as follows:

- Screening phase: to include all assessments on or before the first day of study drug.
- Pre-treatment: measurement date/time on or before the study drug first dose date/time. This period is used to derive baseline values.
- On-treatment in the double-blind phase: to include all assessments after the study drug first dose date/time in the double-blind phase and up to the first day of OLE phase study drug or (last day of DB phase study drug + 7 days for efficacy or 30 days for safety), whichever is earliest, for subjects entering OLE or through the double-blind phase study drug last dose date/time + 7 days to assess efficacy endpoints and + 30 days to assess safety endpoints on treatment.
- On-treatment in the OLE phase: to include all assessments after the open label study drug first dose date/time through the study drug last dose date/time + 7 days to assess efficacy endpoints and + 30 days to assess safety endpoints on treatment
- On-treatment in the study: to include any Verdiperstat treatment (double-blind or OLE phase), and all assessments after the first dose date/time up to the last dose date/time + 7 days to assess efficacy endpoints and + 30 days to assess safety endpoints on treatment, excluding the number of days with a dosing break between the double-blind and OLE phases (applicable if dosing break >30 days).

7.2 Analysis Visit Windows

Study days are calculated from the IWRS randomization date as follows:

- Measurement date – randomization date + 1, if measurement date ≥ randomization date
- Measurement date – randomization date, if measurement date < randomization date.

Treatment days are calculated from the study drug start date as follows:

- Measurement date – study drug start date + 1, if measurement date ≥ study drug start date
- Measurement date – study drug start date, if measurement date < study drug start date.

For post-baseline visits, the protocol specified window is ± 3 days during the double-blind phase (except at Week 48 or Day 337 with window ± 7 days) and ± 7 days during the OLE phase of the study; however, analysis windows will be adjusted to include all data. Refer to Table 5 for details on the analysis visit windows.

Table 5: Analysis Visit Windows

Evaluation	Protocol-Specified Day	Protocol-Specified Window	Analysis-Specified Interval
Double-Blind Phase			
Screening	Day -42 to -1		
Baseline	Day 1		
Week 2	Day 14	Day 11-17	Day 2-22
Week 4	Day 29	Day 26-32	Day 23-43
Week 8	Day 57	Day 54-60	Day 44-71
Week 12	Day 85	Day 82-88	Day 72-127
Week 24	Day 169	Day 166-172	Day 128-211
Week 36	Day 253	Day 250-256	Day 212-295
Week 48	Day 337	Day 334-340	Day 296-425
Open-Label Extension Phase			
Baseline Ext	Day 1		
Week 2 Ext	Day 14	Day 7-21	Day 2-22
Week 4 Ext	Day 29	Day 22-36	Day 23-43
Week 8 Ext	Day 57	Day 50-64	Day 44-71
Week 12 Ext	Day 85	Day 78-92	Day 72-127
Week 24 Ext	Day 169	Day 162-176	Day 128-211
Week 36 Ext	Day 253	Day 246-260	Day 212-295
Week 48 Ext	Day 337	Day 331-344	≥ Day 296

Actual dates and times will be used for PK analysis rather than nominal days and times.

If a subject has more than one record within an analysis window, the record closest to the protocol-specified day will be used in the analysis. In the case of a tie, the earlier record will be used.

Note that in some cases, endpoints are not measured at the specific visits listed in the above table (refer to Appendix I: Schedule of Assessments and Events – Randomized Double-blind Phase and Appendix II: Schedule of Assessments and Events – Open-Label Extension Phase for the schedule of assessments). For example:

- S-STS is assessed at Screening, Baseline, Weeks 2, 4, 12, 24, 36, 48, Baseline Ext, Weeks 2 Ext, 12 Ext, 24 Ext, 36 Ext and 48 Ext; therefore, the analysis-specified window for Week 2 Ext and 12 Ext are study Day 2-71 Ext and Day 72-127 Ext, respectively.

- Laboratory tests are assessed at Screening, Baseline, Weeks 2, 4, 8, 12, 24, 36, 48, Baseline Ext, Weeks 4 Ext, 8 Ext, 12 Ext, 24 Ext, 36 Ext and 48 Ext; therefore, the analysis-specified window for Week 4 Ext is study Day 2-43 Ext.

- UMSARS is assessed at Screening, Baseline, Weeks 4, 12, 24, 36, 48, Baseline Ext, Weeks 12 Ext, 24 Ext, 36 Ext and 48 Ext; therefore, the analysis-specified window for Week 4 Ext and Week 12 Ext are study Day 2-43 and Day 44-127 Ext. The analysis-specified window for the sensitivity analysis for Week 48 will be Day 296-340.

CCI [redacted]

8 CONTENT OF REPORTS

The final CSR is planned after the last planned analysis.

All analyses described in this SAP are included.

9 APPENDICES

9.1 Appendix I: Schedule of Assessments and Events – Randomized Double-blind Phase

Visit	Screening	Baseline (Pre-dose)	2[a]	3	3-1	4	5	6	7[9]	
Day (Visit procedures conducted within a +/- 3 day window, unless otherwise specified)	Maximum of 42 days (-42 to -1)	1	14	29	57	85	169	253	337 (+/- 7 days)	
Week		0	2	4	8	12	24	36	48 (Early D/C)[9]	Week 48 Phone Call (if applicable)[10]
Eligibility Assessments										
Informed Consent (subject and caregiver)	X									
Inclusion/Exclusion Criteria	X	X								
Medical History	X									
Demographic Data	X									
Concomitant Medication review	X	X	X	X		X	X	X	X	X
Safety Assessments										
Physical Examination*	X	X	X	X		X	X	X	X	
Physical Measurements[1]	X	X							X	
Vital Signs[2]	X	X	X	X		X	X	X	X	
ECG (12-lead)	X	X	X	X		X	X	X	X	
Laboratory Tests	X	X	X	X	X	X	X	X	X	
Pregnancy Test[3]	X	X		X		X	X	X	X	
Sheehan Suicidality Tracking Scale	X	X	X	X		X	X	X	X	
Adverse Event Reporting										
Monitor for Nonserious AEs		X	X	X		X	X	X	X	X
Monitor for SAEs	X	X	X	X		X	X	X	X	X

Visit	Screening	Baseline (Pre-dose)	2[a]	3	3-1	4	5	6	7[9]	
Day (Visit procedures conducted within a +/- 3 day window, unless otherwise specified)	Maximum of 42 days (-42 to -1)	1	14	29	57	85	169	253	337 (+/- 7 days)	
Week		0	2	4	8	12	24	36	48 (Early D/C)[9]	Week 48 Phone Call (if applicable)[10]
Efficacy Assessments										
Unified MSA Rating Scale Parts I-IV*	X	X		X		X	X	X	X	
Clinical Global Impression of Severity (CGI-S)		X		X		X	X	X	X	
Clinical Global Impression of Improvement (CGI-I)				X		X	X	X	X	
Patient Global Impression of Severity (PGI-S)		X		X		X	X	X	X	
MSA Quality of Life		X					X		X	
CCI [redacted]										
Time-to-event measures of disease progression		X		X		X	X	X	X	X[10]
Other Assessments										
CCI [redacted]										
Lumbar Puncture[5]		X							X	
Pharmacokinetic Assessments										
Blood PK Sampling[6]		X		X		X	X	X	X	
CCI [redacted]										

Visit	Screening	Baseline (Pre-dose)	2[a]	3	3-1	4	5	6	7[9]	
Day (Visit procedures conducted within a +/- 3 day window, unless otherwise specified)	Maximum of 42 days (-42 to -1)	1	14	29	57	85	169	253	337 (+/- 7 days)	
Week		0	2	4	8	12	24	36	48 (Early D/C)[9]	Week 48 Phone Call (if applicable)[10]
CCI [redacted]										
Pharmacogenetic Assessment										
Blood Pharmacogenetic Sample		X (or a later visit)								
Clinical Drug Supplies										
Randomize		X								
Dispense Study Drug		X		X		X	X	X		
Study Drug Compliance Assessment[8]		X	X[a]	X		X	X	X	X	
Under regular circumstances, visits are to be completed in-person, unless otherwise specified. Some visits can be conducted remotely, as needed, due to COVID-19 restrictions. See Protocol Section 4.3.1, COVID-19 Public Health Emergency, for further details and requirements regarding conduct of remote visits due to COVID-19.										
[a] The Week 2 assessment should generally be conducted at the study site. Per Investigator judgement, additional evaluations/procedures may be conducted at this timepoint as appropriate.										
*Including neurologic exam as a component of general physical exam and UMSARS										
[1] Height at Screening and Week 48; weight at Screening, Baseline, and Week 48.										
[2] Subjects should be seated and resting for approximately 5 minutes prior to vital sign measurements - temperature, blood pressure, and heart rate. Tilt table may be used, if available.										
[3] Serum pregnancy test will be conducted in WOCBP (see section 5.5) at all scheduled visits. Urine pregnancy test will be performed on WOCBP (in addition to serum) at Baseline (within 24 hours prior to first dose).										
CCI [redacted]										

CCI
[6] One PK sample should be collected pre-dose and one PK sample should be collected post-dose. The time of the last BHV-3241/placebo dose prior to the sample collection and time of the PK sample collection and time of most recent meal should be reported on the CRF. At Weeks 4, 12, 24, 36 and 48, the morning dose of study drug should be held and administered in the clinic/office during the study visit, so that one pre-dose (trough) and one post-dose PK sample can be collected. The post-dose sample should be collected within a time window of approximately 2 to 4 hours. Collection of PK samples is not required at the Early Discontinuation visit.
CCI
[8] Subjects should take the first dose of investigational product (IP) while in the office/clinic on the day of the Baseline visit and stay at the clinic for ~4 hours for observation. From start to end of Week 1, subjects will ingest either 300 mg QD of BHV-3241 or matching placebo QD. From start to end of Week 2, subjects will ingest either 300 mg BID of BHV-3241 or matching placebo BID. Starting with Week 3 and for the remainder of the double-blind portion of the study, subjects will ingest either 600 mg BID of BHV-3241 or matching placebo BID.
[9] The Week 48 assessments/procedures should be conducted as an in person visit at the study site. Every effort should be made to conduct the Week 48 visit and maintain a +/- 7 day window. However, due to the COVID-19 public health emergency, if absolutely necessary, in order to minimize potential risks to study participant safety and to comply with governmental and local institutional guidance (e.g., study site has a policy that an in person clinical research visit must be delayed), the Week 48 visit window may be expanded up to an additional 12 weeks (maximum treatment duration of 60 weeks). However, every attempt should be made to conduct the visit as close as possible to the date that the visit is due. Under these circumstances, the Investigator should discuss the specific circumstances of these cases with the sponsor medical monitor (or designee), who must approve the request to extend the treatment duration, prior to any modification of the visit window. If the Week 48 visit window is modified, participants should have safety assessments evaluated remotely (e.g., via phone) and safety labs conducted (at a local lab) at the time of the scheduled Week 48 visit. Subjects who discontinue from study medication prior to Week 48 will have an Early Discontinuation visit. Depending on the reason for the early discontinuation (e.g., declining patient status), some procedures, for example, PK samples, PD samples, lumbar puncture, may not be conducted at this visit. Conduct of all procedures, if possible, is encouraged.
[10] For subjects who discontinue study medication early (prior to Week 48), the site should contact the subject by phone call at approximately Week 48 to collect information on clinical status (including Time to Event measures of disease progression, and if applicable, AEs, SAEs).

9.2 Appendix II: Schedule of Assessments and Events – Open-Label Extension Phase

Visit	Baseline Ext[1] (Pre-dose)	Abbreviated Drug Dispensation Visit[2]	2 Ext (phone)	3 Ext[7] (can be remote)	4 Ext[7] (can be remote)	5 Ext	6 Ext	7 Ext	8 Ext[8]
Day (Visit procedures conducted within a +/- 7 day window, unless otherwise specified)	1		14	29	57	85	169	253	337
Week	0		2 Ext	4 Ext	8 Ext	12 Ext	24 Ext	36 Ext	48 Ext /Early Discontinuation[8]
Eligibility Assessments									
Informed Consent (subject)	X	X							
Medical History	X								
Concomitant Medication review	X	X				X	X	X	X
Safety Assessments									
Physical Examination[3]	X					X	X	X	X
Physical Measurements[4]	X								X
Vital Signs[5]	X	X				X	X	X	X
ECG (12-lead)	X			X if visit is conducted in person	X if visit is conducted in person	X	X	X	X
Laboratory Tests	X			X	X	X	X	X	X
Pregnancy Test[6]	X			X	X	X	X	X	X
Sheehan Suicidality Tracking Scale	X		X			X	X	X	X
Adverse Event Reporting									
Monitor for Nonserious AEs and SAEs	X	X	X	X	X	X	X	X	X

Visit	Baseline Ext[1] (Pre-dose)	Abbreviated Drug Dispensation Visit[2]	2 Ext (phone)	3 Ext[7] (can be remote)	4 Ext[7] (can be remote)	5 Ext	6 Ext	7 Ext	8 Ext[8]
Day (Visit procedures conducted within a +/- 7 day window, unless otherwise specified)	1		14	29	57	85	169	253	337
Week	0		2 Ext	4 Ext	8 Ext	12 Ext	24 Ext	36 Ext	48 Ext /Early Discontinuation[8]
Efficacy Assessments									
Unified MSA Rating Scale Parts I-IV[3] (if visit is conducted remotely [by phone], only conduct Part I)	X					X	X	X	X
MSA Quality of Life	X					X	X	X	X
Time-to-event measures of disease progression	X					X	X	X	X
Dispense Study Drug	X	X				X	X	X	
Study Drug Compliance Assessment	X		X			X	X	X	X

Under regular circumstances, visits are to be completed in-person, unless otherwise specified. Some visits can be conducted remotely, as needed, due to COVID-19 restrictions. See Protocol Section 4.3.1, COVID-19 Public Health Emergency, for further details and requirements regarding conduct of remote visits due to COVID-19.

[1] Baseline visit for the OLE Phase is only required if there is an extended break (≥4 weeks) in dosing between the double-blind and OLE Phases. Otherwise, the double-blind Week 48 visit will take the place of the Baseline-Extension visit. Subjects should take the first dose of IP while in the office/clinic on the day of the Baseline-Extension visit and stay at the clinic for ~ 4 hours for observation. From start to end of Week 1-Extension, subjects will ingest 300 mg QD of BHV-3241. From start to end of Week 2-Extension, subjects will ingest 300 mg BID of BHV-3241. Starting with Week 3-Extension and for the remainder of the OLE, subjects will ingest 600 mg BID of BHV-3241.

[2] Abbreviated visit only needed for patients who completed dosing in the double-blind phase <4 weeks prior to entering the OLE Phase. The double-blind Week 48 Visit serves as the Baseline-Extension visit. This visit is primarily to dispense study drug. Subjects should take the first dose of IP while in the office/clinic on the day of the Baseline-Extension visit and stay at the clinic for ~4 hours for observation. From start to end of Week 1, subjects will ingest 300 mg QD of BHV-3241. From start to end of Week 2, subjects will ingest 300 mg BID of BHV-3241. Starting with Week 3 and for the remainder of the OLE, subjects will ingest 600 mg BID of BHV-3241.

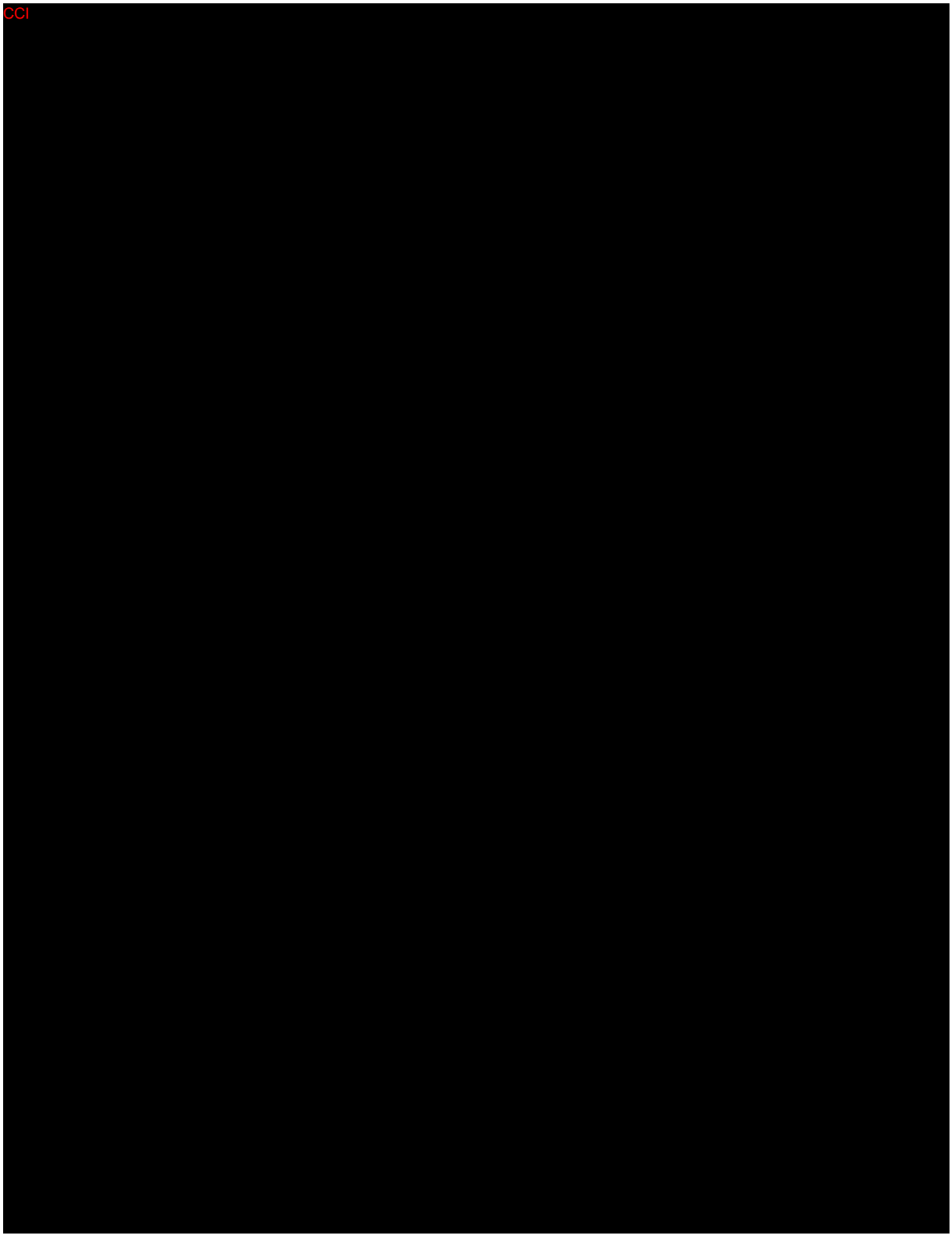
[3] Including neurologic exam as a component of general physical exam and UMSARS.

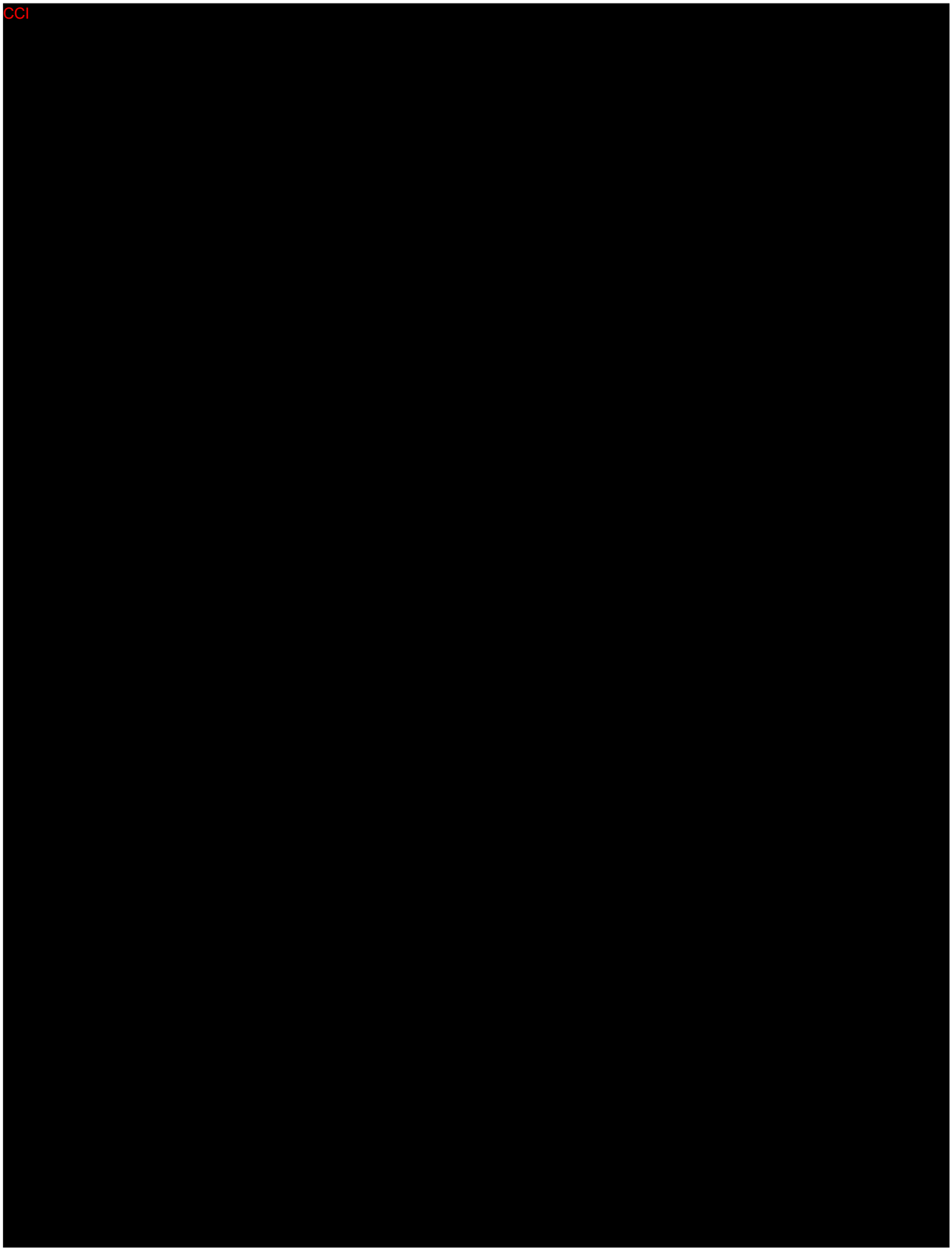
[4] Height and weight.

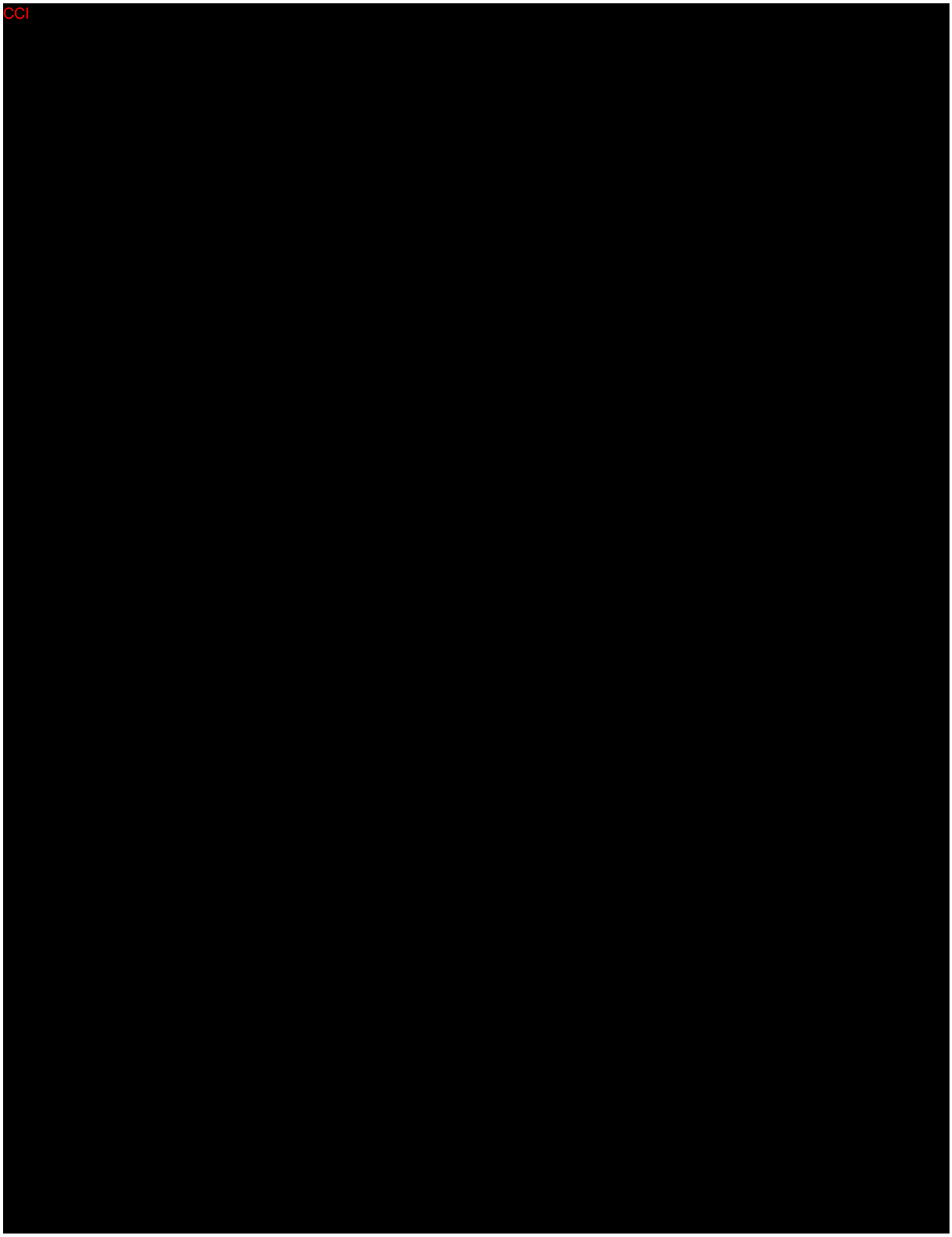
[5] Subjects should be seated and resting for approximately 5 minutes prior to vital sign measurements - temperature, blood pressure, and heart rate. Tilt table may be used, if available.

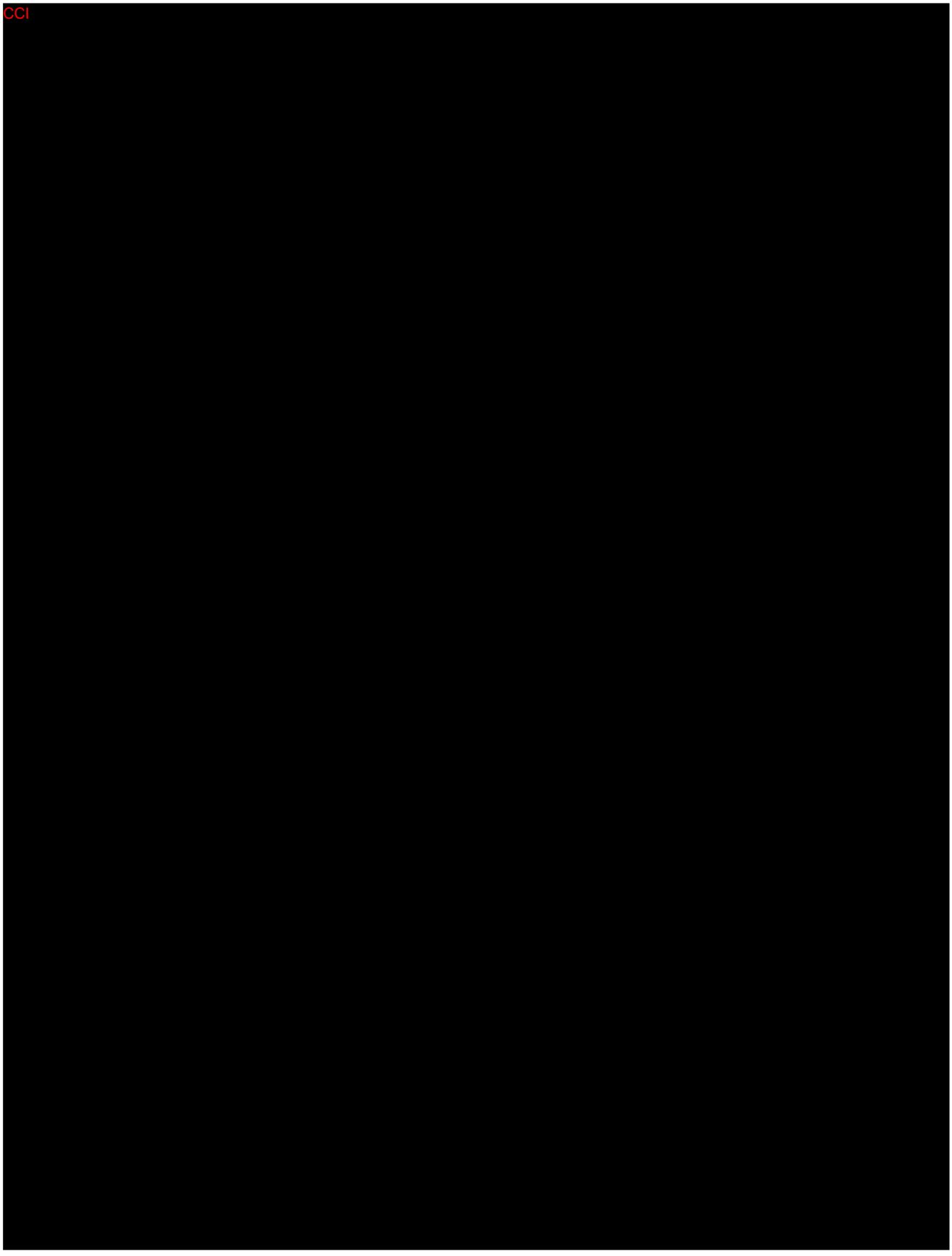
[6] Serum pregnancy test will be conducted in WOCBP (see section 5.5) at all scheduled visits. Urine pregnancy test will be performed on WOCBP (in addition to serum) at Baseline-Extension (within 24 hours prior to first dose of open-label BHV-3241).

[7] Week 4-Extension and Week 8-Extension are primarily to collect lab tests and, if conducted in the clinic/study site, to conduct an ECG. If the patient attends the visit at the clinic/study site, an assessment of adverse events, conduct of an ECG and collection of samples for lab analyses will be completed. If a participant prefers to not attend the visit at the clinic/study site, the lab samples can be collected/conducted at a local laboratory/location and a telephone call to confirm these were collected, as well to assess any adverse events, will take place.
[8] Subjects who discontinue from study medication prior to Week 48-Extension will have an Early Discontinuation visit


CCI


CCI


CCI


CCI

10 REFERENCES

1. Gilman S, Wenning GK, Low PA, et al. Second consensus statement on the diagnosis of multiple system atrophy. Neurology 2008;71:670-6.

2. Levin J, Maass S, Schuberth M, et al. The PROMESA-protocol: progression rate of multiple system atrophy under EGCG supplementation as anti-aggregation-approach. J Neural Transm (Vienna) 2016;123:439-45.

3. Carpenter JR, Roger JH, Kenward MG. Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. J Biopharm Stat 2013;23:1352-71.