TITLE: A Phase 2 Randomized Double-Blind Placebo-Controlled Trial to Evaluate the Efficacy and Safety of BHV-4157 in Patients With Mild to Moderate Alzheimer's Disease

NCT NUMBER: NCT03605667

DATE: 3-December-2020

ADCS and Biohaven Pharmaceuticals

Protocol BHV4157-203

A Phase 2 Randomized Double-Blind Placebo-Controlled Trial to Evaluate the Efficacy and Safety of BHV-4157 in Patients with Mild to Moderate Alzheimer's Disease
Statistical Analysis Plan

Version 3.0
Date: 3 December 2020

SIGNATURE PAGE

Protocol Title:	A Phase 2 Randomized Double-Blind Placebo-Controlled Trial to Evaluate the Efficacy and Safety of BHV-4157 in Patients with Mild to Moderate Alzheimer's Disease
Sponsor:	Biohaven Pharmaceuticals, Inc.
Protocol Number:	BHV4157-203
Document Version/Date:	V3.0 3 December 2020
Author:	PPD PPD PPD UCSD, for ADCS. Signature:_____ Date:_____

Sponsor Approval

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

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

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

Sponsor Signatories:

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

Project Director/ Principal Investigator Signatory

Howard Feldman, MD, Director, Alzheimer's Disease Cooperative Study	Signature:_____ Date:_____

TABLE OF CONTENTS

List of Tables (not Appendix Tables)

REVISION HISTORY

Version	Description of Change
1.0	Original version based on Protocol Version 3.0, Amendment 2
2.0	Revised version based on Protocol Version 6.0. Amendment 5. Revisions primarily include changes related to the potential impact of COVID-19 pandemic crisis and the addition of an open label extension phase to the study. These changes include: • Details on the summarization of data obtained from the CRF collecting information on the impact of COVID-19 pandemic crisis on study visits • Modifications on the analysis plans due to COVID-19 pandemic crisis including an expanded analysis window for the Week 48 to allow for delayed in clinic assessments and clarification on the assessments to be included in the primary analysis • Additional sensitivity analyses included to assess the impact of COVID-19 pandemic crisis on the primary and secondary endpoints as well as key safety domains. • Revised proration method for calculating total score when some subscales are missing on ADAS-cog • Specification on list of subjects for safety narratives • Details on the summarization of data collected in the new Open Label Extension phase • Additional edits and clarification on wording
3.0	Revised version including the following revisions, clarifications and corrections. • Clarification of mITT definition that criteria need only be met by at least one of the co-primary endpoints for a subject to be included. • Updating Meddra version to 23.0 • Clarified the handling of missing APOE e4 status in subgroup definition. • Revised determination of baseline for CDR assessments (Table 1)

	• Clarified in section 4.3.3 that site is also included as a stratification factor in models. • Specified baseline MMSE status subgroup analyses for the co-primary and secondary endpoints. • Revised statistical analysis of MRI parameters to be first focused on assessments in the Week 48 analysis window. Specified additional analyses as exploratory. • Corrected references to sections in section 4.10.2 • Clarified calculation of ADAS-cog score when some performance subtest scores are missing. • Eliminated Summary of AEs "Associated with Study Treatment" • Clarified handling of multiple lab assessments with in a window for by week summaries.

ABBREVIATIONS

Abbreviation	Definition
AE	Adverse event
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
BQL	Below limit of quantification
BUN	Blood urine nitrogen
CI	Confidence interval
CPK	Creatine phosphokinase
CRF	Case Report Form
CSR	Clinical study report
DILI	Drug induced liver injury
ECG	Electrocardiogram
HDL	High-density lipoprotein
GGT	Gamma-glutamyl transferase
LDH	Lactate dehydrogenase
LDL	Low-density lipoprotein
mITT	Modified Intent to Treat
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed model for repeated measures
MMSE	Mini Mental State Exam
PID	Participant Identification Number
PK	Pharmacokinetic
PP	Per protocol set
PT	Preferred term
QD	Quaque die (once daily)
S-STS	Sheehan-Suicidality Tracking Scale
SAE	Serious adverse event
SAP	Statistical analysis plan
SD	Standard deviation
SDS	Sheehan Disability Scale
SOC	System organ class
ULN	Upper limit of normal
WHO Drug	World Health Organization drug reference dictionary

1 INTRODUCTION AND OBJECTIVES OF ANALYSIS

1.1 Introduction

This document presents the statistical analysis plan (SAP) for Biohaven Pharmaceuticals, Protocol BHV4157-203: A Phase 2 Randomized Double-Blind Placebo-Controlled Trial to Evaluate the Efficacy and Safety of BHV-4157 in Patients with Mild to Moderate Alzheimer's Disease.

This SAP is based on Version 6.0, Amendment 5, of the T2 Protocol dated September 12, 2020. It contains the analysis details and methodology to answer the study objectives, including planned summary tables, by-subject listings, and figures, which will provide the basis for the results section of the clinical study report (CSR). Operational aspects related to collection and timing of planned clinical assessments are not repeated in this SAP unless relevant to the planned analyses.

1.2 Objectives of Statistical Analysis

1.2.1 Primary Objective

- To evaluate the efficacy of BHV-4157 as measured by ADAS-Cog 11 and CDR-Sum of Boxes (co-primary endpoints)

1.2.2 Secondary Objectives

- To evaluate the efficacy, safety and tolerability of BHV-4157
    - The efficacy of BHV-4157 will be assessed by the following measures:
        - Volumetric MRI (Quarc bilateral hippocampal volume)
        - Neuropsychiatric Inventory (NPI)
        - Alzheimer's Disease Cooperative Study (ADCS)-Activities of Daily Living (ADCS-ADL)
        - Mini-Mental State Examination (MMSE)
- The safety and tolerability of BHV-4157 will be assessed by:
    - Mortality rates
    - Serious adverse event rates
    - Adverse events
    - Clinical safety laboratories
    - Vital signs

- Physical examinations
- ECGs

- Use of concomitant medications

1.2.3 Exploratory Objectives

- To evaluate the efficacy of BHV-4157 as assessed by the Montreal Cognitive Assessment (MoCA)

- To evaluate the efficacy of BHV-4157 on performance using the AD Composite Score (ADCOMS)

- To evaluate volumetric MRI (bilateral lateral ventricles and whole brain volume)

- To evaluate the efficacy of BHV-4157 on performance on a cognitive composite outcome computed using measures from the NACC Neuropsychological Test Battery [Craft Story 21 Recall (Immediate & Delayed); Benson Complex Figure (Copy & Delayed Recall); Multilingual Naming Test; Letter & Category Fluency; Trail Making Test A & B; Number Span Forward and Backward]

- To analyze the pharmacokinetics of BHV-4157 and riluzole in plasma

- To evaluate treatment response by Apo E genotype

- To assess CSF, serum and plasma biomarkers (Aß42, Aß42/40 ratio, total tau, p-tau-181, neurogranin, NfL, YKL-40, VILIP, SNAP-25, sTREM2, GFAP) at screening, week 24 and week 48 of the double-blind phase. CSF will be assessed in a subset of the study population (estimated n=50 active, n=50 placebo)

Exploratory Open-Label Extension Phase objective is to evaluate the safety and tolerability of BHV-4157 as measured by the change in safety and tolerability measures including: (1) adverse events; (2) clinical laboratory tests; (3) vital signs; (4) physical examinations; (5) ECGs. Additional exploratory endpoint of evaluating long-term effects on clinical outcome measures.

2 STUDY DESIGN

2.1 Synopsis of Study Design

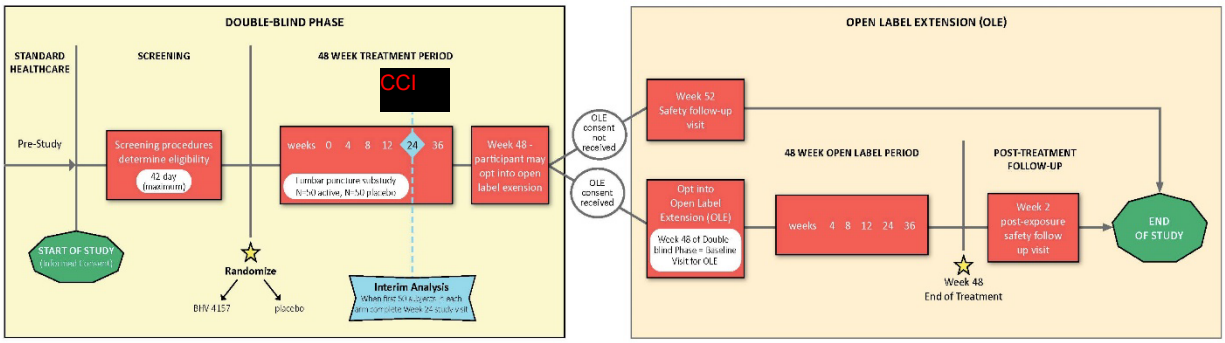
This is a phase 2 multi-center, randomized, double-blind, placebo-controlled, parallel group study in participants with mild to moderate Alzheimer's disease. Approximately 336 participants will be randomized 1:1 to one of two groups: 280 mg of BHV-4157 or placebo. Treatment duration is 48 weeks (12 months). There is a screening period of up to 42 days and a 4-week post-treatment observation period. An interim analysis for futility will be conducted when a sentinel cohort consisting of the first 100 randomized participants has had opportunity to receive 24 weeks of study treatment or longer.

In addition, subjects who are completing or have completed the double-blind treatment phase may be offered the opportunity to enroll in an Open-Label Extension phase, in which subjects will receive open-label treatment with BHV-4157 for approximately 48 weeks. Enrollment into the Open-Label Extension phase for all subjects is contingent on the PI's judgement that open-label treatment offers an acceptable risk-benefit profile for each individual. All subjects who enter the Open-Label Extension phase will need to sign a new informed consent form. All subjects entering the Open-Label Extension phase will start with one capsule of BHV-4157 (140 mg) per day and titrate to two capsules per day after week two (280 mg).

The primary estimand is the effect relative to placebo of treatment as actually taken for 48 weeks, in the protocol-defined modified intent-to-treat study population.

Figure 1 illustrates the study schematic.

Figure 1: Study Schematic



2.2 Randomization Methodology

After completion of all screening evaluations, eligible participants will be randomized in a 1:1 ratio to one of 2 groups: treatment with 280 mg BHV-4157 or placebo. A randomization schedule will be generated and incorporated into the Electronic Data Capture system (EDC) and the randomization ID will be assigned when the site randomizes the participant. A centralized eligibility confirmation procedure will be applied for each participant. A stratified permuted block randomization procedure will be used. The stratification factors will be Screening Visit MMSE [14 to 19; 20 to 24] and Site.

2.3 Unblinding

Blinding is critical to the integrity of this clinical study. However, in the event of a medical emergency in which knowledge of the investigational product is essential for management of the participant, the blind for that participant may be broken after consultation with the ADCS Medical Core.

Unblinding will be managed via the ADCS Unblinded Informatics team. As part of ongoing safety monitoring, an ADCS Unblinded Informatics team member and a bioanalytical scientist designee may be unblinded before data are more generally unblinded after the double-blinded phase of the study has been completed and the database has been locked.

In order to minimize unnecessary analysis of placebo blood PK samples, the bioanalytical scientist will be unblinded to treatment prior to unblinding for the primary endpoint. Results of the blood concentration assay will be kept secure until database lock and unblinding for the primary endpoint.

The interim analysis will be conducted on a locked unblinded data set by the unblinded safety ADCS biostatistician, who is not involved in the final study analysis and who is firewalled from the study team (see section 4.4.1). As part of the interim analysis review, an independent data monitoring committee will be unblinded (see section 4.4.1 for additional detail). Except as noted above, other members of the BHV and ADCS research teams will remain blinded.

Once all participants have completed the double-blind phase of the study, the database will be cleaned, locked, and the data will be analyzed in an unblinded manner.

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

2.4 Efficacy, Safety, and Other Variables

2.4.1 Co-Primary Endpoints

- There are two primary efficacy endpoints:
    - Within-participant change in ADAS-Cog 11 from baseline to week 48 of the double-blind phase, compared between the treatment group and the placebo group
    - Within-participant change in CDR-Sum of Boxes from baseline to week 48 of the double-blind phase, compared between the treatment group and the placebo group

2.4.2 Secondary Endpoints

- The efficacy of BHV-4157 will be assessed by the within-participant changes from baseline to week 48 of the double-blind phase, compared between the treatment group and the placebo group, on the following:
    - Volumetric MRI (Quarc bilateral hippocampal volume)
    - Neuropsychiatric Inventory (NPI) scores
    - Alzheimer's Disease Cooperative Study-Activities of Daily Living (ADCS-ADL) scores
    - Mini-Mental State Examination (MMSE)
- Safety and tolerability as assessed by:
    - Serious adverse events (SAEs) including deaths
    - Adverse events (AEs) leading to discontinuation

- AEs
- AEs judged to be related to study medication
- Clinical safety labs
- Vital signs
- Physical examinations
- ECGs
    - Use of concomitant medications

2.4.3 Exploratory Endpoints

The following will be assessed during the double-blind phase :

- Efficacy of BHV-4157 assessed by the within-participant changes from baseline to week 48, compared between the treatment group and the placebo group, on Montreal Cognitive Assessment (MoCA) scores
- Efficacy of BHV-4157 assessed by the within-participant changes from baseline to week 48, compared between the treatment group and the placebo group, on performance using the AD Composite Score (ADCOMS), a composite outcome computed using 4 ADAS-Cog subscale items, 2 MMSE items, and the CDR Sum of Boxes (Wang et al., 2016)
- Efficacy of BHV-4157 assessed by the within-participant changes from baseline to week 48, compared between the treatment group and the placebo group, on volumetric MRI (bilateral lateral ventricles and whole brain volume)
- Efficacy of BHV-4157 assessed by the within-participant changes from baseline to week 48, compared between the treatment group and the placebo group, on a cognitive composite outcome computed using measures from the NACC Neuropsychological Test Battery [Craft Story 21 Recall (Immediate & Delayed); Benson Complex Figure (Copy & Delayed Recall); Multilingual Naming Test; Letter & Category Fluency; Trail Making Test A & B; Number Span Forward and Backward]
- Pharmacokinetics of BHV-4157 and riluzole in plasma
- Treatment response by APOE genotype
- CSF, serum and plasma biomarker panel (Aß42, Aß42/40 ratio, total tau, p-tau-181, neurogranin, NfL, YKL-40, VILIP, SNAP-25, sTREM2, GFAP) will be assessed at screening, week 24, and week 48. CSF will be assessed in a subset of the study population (estimated n=50 active, n=50 placebo)
- Open Label Exploratory Endpoints The change in safety and tolerability measures including: (1) adverse events; (2) clinical laboratory tests; (3) vital signs; (4) physical examinations; (5)

ECGs. Additional exploratory endpoint of evaluating long-term effects on clinical outcome measures.

3 PARTICIPANT POPULATIONS

3.1 Population Definitions

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

- Enrolled Participants: Participants who signed an informed consent form and were assigned a Participant Identification number (PID)

- Randomized Participants: Enrolled participants who received a randomized treatment assignment. This group forms the ITT population

- Treated Participants: Enrolled and randomized participants who received at least 1 dose of blinded study therapy (BHV 4157 or placebo). This will be the primary analysis population used for safety summaries

- The modified Intent-to-Treat (mITT) population will include all randomized participants who took at least one dose of study medication, who have a baseline assessment of at least one of the co-primary endpoints, and who have at least one efficacy evaluation visit during the double-blind phase following baseline. This will be the primary analysis population used for the efficacy analyses.

- Treated Participants in Open Label Extension: Participants who received at least 1 dose of open label study therapy (BHV 4157) in the Extension Phase. This will be the primary analysis population used for safety summaries focusing on the Extension Phase

- The modified Intent-to-Treat (mITT) population for the Extension Phase will include all randomized participants who took at least one dose of open label study therapy (BHV4157) in the Extension Phase, who have a baseline assessment for the extension phase, and who have at least one efficacy evaluation during the extension phase. This will be the primary analysis population used for the efficacy summaries in the extension phase

- Treated Participants in All Troriluzole Exposure. Participants who received at least 1 dose of Troriluzole in either the Randomization Phase or the Extension Phase. This will be the primary analysis population used for safety summaries focusing on All Troriluzole exposure

The primary population for all efficacy analyses is the mITT population (subjects included in treatment group as randomized). The Treated Participants will be used for analyses of safety endpoints (subject included in treatment group as received).

3.2 Protocol Deviations

Any significant event that does not comply with the inclusion/exclusion criteria, study conduct (e.g., inadequate informed consent, unreported SAEs), or study procedures (e.g., use of

prohibited medications as defined by the protocol; improper breaking of the blind) will be documented as a protocol deviation.

The sponsor, or designee, will be responsible for producing the final protocol deviation file (formatted as a Microsoft Excel file), which will be finalized prior to database lock. Classification of deviations from the protocol as minor or major will be decided on a case-by-case basis by the medical monitor without knowledge of the treatment assigned and before the database lock (Data Blind Review). This file will include site, participant ID, deviation date, deviation type, status (major vs. minor), and a description of the protocol deviation.

All protocol deviations will be presented in a tabulation and data listing. Both the summary tabulation and the listing will include information regarding whether the deviation was potentially related to the COVID-19 global pandemic.

4 STATISTICAL METHODS

4.1 Sample Size Justification and Power

4.2 Overall Study Power

Using approximately 336 randomized participants (with a 1:1 assignment approximately 168 per arm), this study has at least 80% power to reject the null hypothesis of no efficacy in favor of the alternative that BHV4157 is efficacious, given the assumption of a difference in the mean change from baseline to 12 months between the BHV-4157 arm and placebo arm of 2.64 points on the ADAS-Cog 11 and 0.95 points in CDR-SOB, and under a range of interim effect sizes which give 93.7% probability of proceeding at the interim analysis.

This calculation assumes no more than a 26% dropout rate, which would leave 336 X 0.74 = 248 subjects available at the final analysis, or 124 per arm. The calculation also assumes a standard deviation of 6 points on the change in ADAS-Cog 11 in each arm at both 6 months and 12 months, and a standard deviation of 2.3 points on the change in in CDR-SOB at 12 months; in addition a correlation of 45% between 12-month-change on the co-primary ADAS-Cog11 and CDR-SOB endpoints is assumed. These assumptions are supported by published retrospective data from historical ADCS studies[14]. The computations use a 2-sided 2-sample t-test at 5% alpha to approximate power.

Briefly, the statistical reasoning is as follows:

1. P(pass overall) = P(pass final | pass interim) P(pass interim) ≥ P(pass final) P(pass interim)

2. P(pass final) = P(pass final ADAS) P(pass final CDR | pass final ADAS)

3. Numerical integration shows that P(pass final CDR | pass ADAS) is about 3% greater than P(pass final CDR), given the assumed 45% correlation between these endpoints.

Hence, if P(pass interim) = 0.937, P(pass final ADAS) = 0.92, and P(pass final CDR) =0.90, then P(pass overall) ≥ 0.92 x (1.03 x 0.90) x 0.937 = 0.80.

With 124 subjects per arm, the unconditional power for the ADAS-COG endpoint is 92%, if the mean difference in change is 2.64 points with standard deviation of 6 points within each arm. At this sample size, the unconditional power for the CDR-SOB endpoint is 90%, if the mean difference is 0.95 points with standard deviation of 2.3 points within each arm. Hence, if the interim analysis has power 0.937 or greater to pass, the overall power is 80% or more following the statistical reasoning given above.

For perspective, we note that the assumed difference of 2.64 points on ADAS-Cog is 68% of the decline observed in the placebo arm of 3.9 points in studies of mild to moderate AD[14]. The assumed difference of 0.95 points on CDR-SOB is 59% of the decline observed in the placebo arm of 1.6 points[14].

4.2.1 CCI

4.3 General Statistical Methods and Data Handling

4.3.1 General Methods

Overview

The primary study hypothesis is that treatment with BHV-4157 will result in a reduction in total within-participant change from baseline on the ADAS-Cog 11 score and also on the CDR Sum of Boxes, relative to the placebo group at week 48 in the mITT population. The co-primary endpoints at the final analysis will be tested using a hierarchical gatekeeper strategy: if ADAS-Cog 11 is significant at the 5% level, then CDR-SOB will also be tested at 5% significance level. If ADAS-cog is not significant at the 5% level, then CDR-SOB will not be included in the primary analysis, and will be presented for descriptive purposes only. The primary objective of the trial is met only when both co-primary endpoints are significant. If both co-primary endpoints are significant, the remainder of the secondary endpoints will be tested at overall 5% significance level using a Holm-Bonferroni step down test. If the co-primary endpoints are not significant, the secondary endpoints will be presented for descriptive purposes only. This will preserve alpha at 5% overall for all endpoints tested. Both the interim and final analysis will use a mixed effects repeated measures model. The secondary endpoints will compare mean changes from baseline to week 48 using a similar analysis strategy.

4.3.1.1 Estimands, Intercurrent Events, and Missing Data

This discussion follows the framework as in Ratitch et al.[10].

Summary of primary estimand:

The effect relative to placebo of treatment as actually taken for 48 weeks during the double-blind phase, in the protocol-defined mITT study population.

1. Objective

   An estimate of efficacy which would potentially be supportive of registration; thus similar to an ITT, “as taken” estimand.

2. Estimand

   A. The population is the mITT population as defined by the protocol’s inclusion and exclusion criteria.

   B. Efficacy is measured by the change from baseline to 48 weeks in the co-primary endpoints (ADAS-cog and CDR-SOB)

   C. The treatment evaluated is the randomized treatment as it was assigned to study subjects. All types of intercurrent events are incorporated using a treatment policy strategy, which evaluates the randomized treatment as taken including missed or modified doses, drug discontinuation, and concurrent treatments. Study dropout including those potentially related to the 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. Deaths are not expected during the 48 week double-blind phase follow-up in the enrolled subject population.

D. The mean change from baseline between treatment and control arm on the 2 endpoints will be compared at 48 weeks between treatment and placebo arms. A significant difference in both endpoints at familywise error rate of 5% will be required for success.

3. Primary estimator and missing data.

   The estimator is a MMRM using all available assessments. Under the assumption that data are MAR conditional on model covariates and observed assessments, this analysis estimates the treatment as taken in the mITT study population.

   Note that MAR assumes that subjects will follow the same trajectory after dropping out of the study as while on the study, conditional on the model covariates. However, if subjects tend to preferentially drop out after discontinuation of study drug, their post-dropout trajectory (off treatment) may differ from their pre-dropout trajectory (on treatment). In this case the MAR assumption may not hold. A sensitivity analyses is proposed to assess this scenario.

4. Sensitivity analyses and missing data.

   MMRM, using multiple imputation from the control arm to complete assessments missing after discontinuation of study drug. This analysis assumes subjects who discontinue medication and are no longer assessed immediately become similar to subjects who never took any medication, and so provides a lower bound on efficacy, again under the MAR assumption that the time of stopping study medication depends only on past history and covariates

4.3.1.2 Data Handling

All output will be incorporated into Microsoft Excel or Word files, sorted and labeled according to the International Conference on Harmonization (ICH) recommendations, and formatted to the appropriate page size(s).

Tabulations will be produced by randomized treatment group and overall, unless otherwise specified.

Categorical variables will be tabulated with counts and percentages. Continuous variables will be summarized with univariate statistics (e.g., n, mean, standard deviation (SD), median, minimum, and maximum). The minimum and maximum will be presented with the same precision as the data, the mean and percentiles will be presented with the precision of the data + 1 decimal place, and the SD will be presented with the precision of the data + 2 decimal places.

Unless otherwise specified, all statistical tests will be 2-sided with alpha = 0.05. Tests will be declared statistically significant if the calculated p-value is ≤ 0.05. P-values < 0.0001 will be presented as "<0.0001". Otherwise, p-values will be presented to 4 decimal places.

Change from baseline is calculated by subtracting the baseline value from the observed post-baseline value at any subsequent visit. Baseline (specific to the randomized phase or to the open-label phase) will be defined as last available assessment on or before the 1st day the participant receives study medication for that phase.

Tabulations of the following endpoints will present the number of unique participants 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 participant.

Unique Participant ID's will take the form "BHV4157-203-site#-participant#". By-participant listings will display unique PID and "(Age/Sex/Race)" stacked together in the same column using the following conventions:

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

A footnote will describe race abbreviations, e.g., "Race: A = Asian, B = Black or African American, I = American Indian or Alaska Native, M = Multiple, N = Native Hawaiian or Other Pacific Islander, W = White". Participants who report more than one race will be counted only once in the "Multiple" category. Missing age, sex, or race will be displayed as a single blank space.

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

4.3.2 Computing Environment

All statistical analyses will be performed using SAS statistical software (Version 9.4) or R (Version 4.0.0), as specified. Medical history and AEs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA, Version 23.) Concomitant medications will be coded with WHO Drug (Global March 2019) using the WHODrug Insight tool. Laboratory abnormalities will be classified as clinically significant if they are coded as Grade 3 or 4 according to the numeric laboratory test criteria in Common Technical Criteria for Adverse Events (CTCAE) Version 5.0 (2017). For laboratory tests not included in CTCAE Version 5.0, abnormalities will be coded according to Division of AIDS (DAIDS) Table for Grading the Severity of Adult and Pediatric Adverse Events Corrected Version 2.1 (2017).

4.3.3 Adjustments for Covariates and Stratification

The analysis of the primary endpoints and other efficacy endpoints will be analyzed by comparing the change in efficacy outcome at week 48 between treatment groups using a mixed model with repeated measures (MMRM). Fixed effects in the model are APOE status (e4 carrier vs e4 noncarrier), MMSE (mild vs moderate) status, site, baseline value of the endpoint, treatment group (active vs. placebo), visit, visit x baseline and visit x treatment group interaction. Visit is treated as a categorical variable (see section 4.2.8 for definition of visit windows to be use). Subject will be included as random effects in the model. Sites with only one subject will be pooled together.

A stratified permuted block randomization procedure will be used. The stratification factors will be Screening Visit MMSE [mild: 20 to 24; moderate: 14 to 19] and Site.

4.3.4 Multiple Comparisons/Multiplicity

Type 1 errors will be controlled for the co-primary and multiple secondary efficacy endpoints by testing these endpoints with a gate-keeping procedure at a family wise two-sided alpha level of 0.05. The co-primary endpoints, ADAS-Cog 11 and CDR-SOB, will be tested, first testing the ADAS-cog and then, if significant, testing the CDR-SOB. The primary objective of the trial is met only when both co-primary endpoints are significant. If the co-primary endpoints are significant, then the key secondary efficacy endpoints will be tested using the Holm-Bonferroni step down procedure, at family-wise significance level 5%.

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

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

4.3.5 Subpopulations

The subgroups of interest for this study are:

- Sex
- APOE (e4 vs no or missing)
- Mild vs moderate (mild AD:MMSE 20 to 24; moderate AD: MMSE 14 to 19)
- Participation during Covid-19 (opportunity to complete randomization phase before March 13, 2020 vs on or after March 13, 2020)

Except where indicated otherwise, descriptive summaries will be provided for the co-primary endpoints for each subgroup.

4.3.6 Withdrawals, Dropouts, and Loss to Follow-up

Participants who withdraw from the study will not be replaced.

4.3.7 Missing, Unused, and Spurious Data

Unless otherwise noted, efficacy analyses will be based on observed data only. No missing data will be imputed. Imputation will be done for sensitivity analyses only.

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

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

The following general rules will be used:

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

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

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

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

4.3.8 Visit Windows

The protocol-specified in-clinic visit windows are ± 7 days during the Randomization Phase, however, analysis windows will be continuous to include all data. Refer to Appendix 1 for details on the expected evaluation intervals for the details on the schedule of assessments during the Randomization and Extension Phases. The data portal contains a report where all visits and their acceptable windows will be projected at the baseline visit. The baseline visit for the randomization phase is expected to be the date of randomization and first dose of randomized treatment.

On-Treatment Analysis Windows

Table 1: Analysis Windows for Efficacy Parameters: ADAS-cog and CDR in the 48 week Randomization Phase

Evaluation	Protocol-Specified Day	Analysis-Specified Interval
Randomization Phase		
Week 12	Day 84	Day 2-126
Week 24	Day 168	Day 127-210
Week 36	Day 252	Day 211-294
Week 48	Day 336	Day 295-420

Note: Baseline assesment will be defined as the last available assessment on or before first day of randomization phase study drug [For CDR the window for baseline assessment will be extended to include assessments up to and including Day 2 of study drug as one subject completed the CDR after starting study treatment. This assessment was felt to still be a valid baseline assessment due to the limited treatment at the time of assessment. The Week 12 window will be adjusted accordingly to Day 3-126.]. Days based on days from first day of randomization phase study drug.

Table 2: Analysis Windows for All Other Efficacy Parameters Including NPI, ADCS-ADL, MMSE, MoCA, Volumetric MRI, CSF Biomarkers and Blood Biomarkers and NACC measures in the 48 week Randomization Phase.

Evaluation	Protocol-Specified Day	Analysis-Specified Interval
Randomization Phase		
Week 24	Day 168	Day 2-252
Week 48	Day 336	Day 253-420

Note: Baseline assesment will be defined as the last available assessment on or before first day of randomization phase study drug. Days based on days from first day of randomization phase study drug.

Table 3: Analysis Windows for Safety Parameters: Sheehan Suicidal Tracking, Vital Signs, Lab Assessments, PK and ECG

Evaluation	Protocol-Specified Day	Analysis-Specified Interval
Randomization Phase		
Week 4	Day 28	Day 2-42
Week 8	Day 56	Day 43-70
Week 12	Day 84	Day 71-98
Week 16	Day 112	Day 99-126
Week 20	Day 140	Day 127-154
Week 24	Day 168	Day 155-182
Week 28	Day 196	Day 183-210
Week 32	Day 224	Day 211-238
Week 36	Day 252	Day 239-266
Week 40	Day 280	Day 267-294
Week 44	Day 308	Day 295-322
Week 48	Day 336	Day 323-420

Note: Baseline assesment will be defined as the last available assessment on or before first day of randomization phase study drug. Days based on days from first day of randomization phase study drug.

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

Table 4: Analysis Windows for All efficacy assessments and the Volumetric MRI, in the 48 week Extension Phase

Evaluation	Protocol-Specified Day	Analysis-Specified Interval
Extension Phase		
Week 24	Day 168	Day 2-252
Week 48	Day 336	Day 253-420

Note: Baseline assesment will be defined as the last available assessment on or before first day of extension phase study drug as long as it is within 28 days of the first day of extension phase medication. Days based on days from first day of extension phase study drug.

Table 5: Analysis Windows for Safety Parameters: Sheehan Suicidal Tracking, Vital Signs, Lab Assessments, and ECG in the 48 week Extension Phase

Evaluation	Protocol-Specified Day	Analysis-Specified Interval
Extension Phase		
Week 4	Day 28	Day 2-42
Week 8	Day 56	Day 43-70
Week 12	Day 84	Day 71-126
Week 24	Day 168	Day 127-210
Week 36	Day 252	Day 211-294
Week 48	Day 336	Day 295-420

Note: Baseline assesment will be defined as the last available assessment on or before first day of extension phase study drug as long as it is within 28 days of the first day of extension phase medication. Days based on days from first day of extension phase study drug.

4.4 Analysis Periods and Study Phases

Analysis periods and study phases are defined as follows:

- Screening Phase: will include all assessments on or before the first day of study drug.

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

- On-Treatment in Randomization Phase (Efficacy Assessments): except where indicated, "on-treatment" in the Randomization Phase will include all assessments after first day of study drug and up to the first day of dosing of the Open Label Extension Phase study drug (for subjects entering extension) or up to the last day of Randomization Phase study drug + 3 days, whichever comes first.

- On-Treatment in Open Label Extension Phase (Safety Assessments): except where indicated, "on-treatment" in the Open Label Extension Phase will include all assessments after first day of open label study drug and up to the last day of Open Label Extension Phase study drug + 30 days. For AEs, “on-treatment” in the Open Label Extension Phase will include first day of dosing of the Open Label Extension Phase study drug up to the last day of Open Label Extension Phase study drug + 30 days.

- On-Treatment in Open Label Extension Phase (Efficacy Assessments): will include all efficacy assessments after the first day of dosing of OL Extension Phase study drug to the

last day of study drug +3 days. Baseline for the extension will be the last available assessment on or before the first day of dosing of OL Extension Phase medication.

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

4.5 Planned Analyses

DSMB analyses of safety data will be performed quarterly throughout the trial period by an unblinded safety ADCS biostatistician who is not involved in the final study analysis. The quarterly analyses include an Executive Summary and a Study Summary (extracted from the protocol). These summaries are followed by tables and figures detailing elements of the current data base related to safety, specifically: Screening and Enrollment, Enrollment by MMSE, Enrollment by Month, Enrollment by Site, Inclusion and Exclusion, Screen Failure by Inclusion Criteria, Screen Failure by Exclusion Criteria, Early Discontinuation, Protocol Deviations, Demographics and Baseline Outcomes, Vital Signs, Safety, Adverse Events with Onset Prior to Randomization, Adverse Events with Onset Following Randomization, Adverse Events by Severity, AE by relatedness, Adverse Events summarized by MedDRA SOC–PT, Serious Adverse Events, and Safety Labs.

4.5.1 CCI

CCI

[redacted]

4.5.2 Primary Analysis After Double-blind Randomization Phase and Final Analysis After OL Extension Phase

The primary efficacy (including the primary, secondary and exploratory endpoints of the Randomization Phase) and safety analysis will be conducted after the last subject completes their Week 48 visit, or discontinues from the Randomization Phase. The study will be unblinded and the primary analyses will include all data from the double-blind Randomization Phase of the study, but will not include summaries of efficacy and safety data accumulated from the Extension Phase.

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

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

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

4.6 Participant Disposition

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

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

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

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

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

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

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

    - Summarize at any time in double-blind phase

    - Summarize by week (analysis window)

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

    - Summarize at any time in double-blind phase

    - Summarize by study week (analysis window)

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

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

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

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

4.7 Demographic and Baseline Characteristics

Tabulations of demographic and baseline characteristics, including age, sex, racial group, smoking habits, level of education, disease and medical history, and current standard of care treatment, will be made for all treated participants. A separate set of tabulations will be made for participants enrolled but not randomized. Demographic information will be summarized by treatment group and for all treatment groups combined.

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

- Subjects completing or discontinuing the double-blind randomization phase before March 13, 2020 (the reference date for the start of the COVID-19 pandemic based on the date National Emergency declared)

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

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

4.8 Efficacy Evaluation

Unless otherwise noted, all efficacy analyses will be conducted using all available in-clinic assessments from the mITT population during the double blind phase. Remote assessments during the double blind phase will be included in sensitivity analyses. All efficacy data will be included in listings by participant, treatment group, visit and phase (as applicable).

4.8.1 ADAS-Cog

The ADAS-Cog is a rater-administered scale with eight performance subtests: Word Recall; Commands; Constructional Praxis; Naming; Ideational Praxis; Orientation; Word Recognition; Remembering Word Recognition Test Instructions[11]. Three examiner ratings, Spoken Language, Language Comprehension, and Word Finding Difficulty, also are obtained. The test is scored in terms of errors, with higher scores reflecting poorer performance and greater impairment. Scores can range from 0 (best) to 70 (worst).

Please refer to Appendix 2 regarding prorating subtest scores when there are partially missing data.

4.8.2 CDR Sum of Boxes

The CDR sum of boxes is a numeric scale used to quantify the severity of symptoms of dementia and it assesses a participant's cognitive and functional performance in six areas: memory, orientation, judgment & problem solving, community affairs, home & hobbies, and personal care. Scores in each of these are combined to obtain a composite score ranging from 0 (best) to 18 (worst). If any item is missing, the assessment is considered missing.

The CDR sum of boxes will be assessed at visits conducted in clinic or via remote communication. Only in-clinic assessments will be considered in the primary analysis; sensitivity analysis will include similar analysis including remote assessments.

4.8.3 Analysis of the Co-Primary Outcome Measures

The primary analyses use the within-participant change in ADAS-Cog 11 and in CDR-SOB as the outcome in two separate similar mixed effects repeated measures models, including all available in-clinic outcomes during the double blind phase in the mITT population. Fixed effects in the models are APOE status (e4 carrier vs e4 noncarrier), MMSE (mild vs. moderate) status, site, baseline ADAS-Cog or CDR-SOB score, respectively, treatment group (active vs placebo), visit, visit x baseline score interaction, and visit x treatment group interaction. Visit is treated as a categorical variable. Sites with only one subject will be pooled together.

Descriptive statistics showing the timing of the primary outcome assessment will be presented by treatment arm as well as break-outs by baseline characteristics (e.g. APOE status, Mild/Moderate status based on MMSE, age group, and sex).

A random effect will be included for participant. The within-participant covariance will be unstructured. If the model does not converge, MMSE status will be removed from the model. If the model still does not converge, the following structures for the within-subject covariance will be fit, sequentially, until the structure is found that results in convergence of the model: Huynh-Feldt, Toeplitz, Autoregressive (1), and Compound Symmetry. Error degrees of freedom will be calculated using the Kenward-Roger approximation if an unstructured covariance is used; otherwise, a sandwich estimator will be utilized to estimate the covariance structure and degrees of freedom will be calculated using the between-within method[13]. For the ADAS-Cog 11and CDR-SOB, each primary endpoint will be tested using model-adjusted least squares means at the week 48 visit. Point estimates, standard errors, two-sided 95% confidence intervals, and p-values will be presented.

The two co-primary endpoints will be tested for significance using a hierarchical gatekeeper strategy which preserves a 5% family-wise type I error rate. If a significant improvement in ADAS-Cog 11 is found with BHV-4157 relative to placebo at 5% level, then a similar analysis to the above will be conducted comparing CDR-SOB between the BHV-4157 and placebo arms. If a significant difference is found in CDR-SOB at 5% level, then both ADAS-Cog 11 and CDR-SOB are declared significant at 5% joint significance level, using this hierarchical gatekeeper analysis strategy. If there is no significant difference in ADAS-Cog 11, then CDR-SOB cannot be found to have a significant difference. A significant difference on both co-primary endpoints (at the 5% level) must be found in order to meet the primary trial objective.

The following sensitivity analysis will be conducted to support the principal analysis, following the same methods as above.

1. Imputation for missing data. All available in-clinic data will be used. Multiple imputations will be used to fill in any missing assessments. The imputation model will be a MMRM model. The parameters of the imputation model will use data from the placebo arm only - a jump to reference approach as described by Ayela, Lipkovich, Molenberghs & Mallinckrodt[1] and Carpenter, Roger & Kenward[2]. This model will be used as the imputation model for both the treated arm and the placebo arm, The imputation model will follow that specified in the primary analysis plan, however incorporating the additional baseline covariates of age, ADAS-COG, CDR-SOB, ADCS-ADL, and the time-dependent COVID indicator specified below. Proper imputation from this model will be used, following Rubin[12] and Tsiatis[16], in that the mth imputation is based on a new parameter $\hat{\beta}m$, drawn from a $N(\hat{\beta},\hat{\Sigma})$ distribution, where $\hat{\beta}$ and $\hat{\Sigma}$ are the maximum likelihood estimates from the imputation model. The augmented data (observed data, completed by the imputed data) will then be analyzed as specified for the primary analysis. This analysis will provide a lower bound on the estimated efficacy, as described in section 4.2.1.1.

   Inference will use the rules for multiple imputation as outlined in Rubin[12]. 20 imputations will be used. This analyses will be implemented in PROC MI in SAS with the MONOTONE option using predictive mean matching (option REGPMM) with

K=20, or in a similar manner in R. The outcome variable will be imputed in a monotone sequential manner.

2. Investigation of the impact of COVID-19 will be conducted by the following separate analyses:

   i) A fixed effect to indicate an assessment occurring after March 13 (during the COVID pandemic) will be added to the model ("COVID indicator"), as well as the interaction of the COVID indicator x time. If the model with the interaction term fails to converge, the interaction term will be omitted.

   ii) A fixed effect to indicate the timing of the Week 48 assessment will be included in the primary analysis model. The fixed effect will be subject-based and include 2 levels. The subjects either discontinuing early or completing their in-clinic assessment within the originally specified week 48 window (Days 323-372) will be in one level of the effect and subjects with their week 48 assessment in the expanded portion of the window (days 373-420) in the another level.

   iii) Remote assessments will be included in the model, with the model as specified for the primary analysis.

   iv) Remote assessments will be included in the model, with the model as specified for the sensitivity analysis which includes the COVID indicator.

Additional supplemental analyses will run the primary model for each of the co-primaries but only include in-clinic "on-treatment" assessments (i.e. done within 3 days of the final dose of double-blind medication – see Section 4.3). These analysis will assess a more hypothetical estimand which focus on the effect of the study medication as intended.

Subgroup Analyses

Except where indicated, descriptive statistics for the two co-primary endpoints ADAS-Cog 11 score and CDR-SOB score, and their change from baseline, will be tabulated by visit, separately, for the subgroups defined by:

- Sex
- Mild vs Moderate MMSE status
- APOE e4 status
- Participation during Covid-19 (opportunity to complete randomization phase before March 13, 2020 vs on or after March 13, 2020)

As a further exploratory analysis, for Mild vs Moderate MMSE baseline status subgroups an MMRM analysis will be utilized on each subgroup using the model specification as described above.

In addition, descriptive statistics will also be presented by visit in the Open Label Extension Phase, by original randomized group (BHV-4157 and Placebo to BHV4157).

4.8.4 Secondary Efficacy Endpoints

4.8.4.1 Volumetric MRI

Bilateral hippocampal volume (as well as the exploratory endpoints of bilateral lateral ventricles and whole brain volume) will be measured as volumetric MRI. MRI imaging of the brain will be performed in order to measure brain atrophy over time. Results from vMRI studies suggest that the patterns of atrophy in AD can reliably be detected and tracked across time. Atrophy of the medial temporal lobe, including hippocampus and entorhinal cortex, has long been described in vMRI studies of AD. Hippocampal volume derived from MRI correlates with histological hippocampal volume and degree of neuronal loss and AD pathology, and entorhinal cortical thickness change appears to be an early and sensitive indicator of neurodegeneration associated with AD. Longitudinal MRI measures of regional and whole brain volumetric change provide a valuable complement to cognitive measures in that they are not influenced by temporary symptomatic improvements, and they provide an early index of the study drug's ability to reach the target organ and have an effect on AD-related atrophy.

The analysis of structural change will be performed through Quarc nonlinear registration (see Holland et al., 2011[6]) comparing each participant's follow-up scan to the initial baseline scan. For the interim analysis, the week 24 scan will be registered to the baseline scan, and for the final analysis, the latest timepoint scan will be registered to the baseline scan. The procedure provides percent deformation within regions of interest (ROIs) obtained through segmentation of the baseline scan. The registration of serially-acquired brain volumes yields a deformation field that represents the volumetric shift of internal structure required for optimal overlaying of the two scans. The anatomically-based segmentation is overlaid on the smoothed voxel-wise deformation field and an average percent deformation is created for each ROI. Visual QC of registration is performed and values from scans passing QC are returned to the statistical team for analysis.

This treatment effect will be mean difference in change in volumetric MRI, compared between the BVH-4157 and placebo groups.

Subjects who have a baseline MRI measure and at least one MRI evaluation on-treatment following baseline will form the mITT population for this analysis. The number of images which are not quantifiable due to scan alignment failure will be summarized by study arm. The percent deformation in volumetric MRI will first be analyzed using the mITT population via an analysis of covariance (ANCOVA) with baseline MMSE as a covariate based only on subjects with an assessment in the Week 48 analysis window as the main analysis. As an exploratory analysis, an additional analysis also using the same ANCOVA model, however including all subjects with a post-baseline assessment, will be conducted to compare between BHV-4157 and placebo based on subject's last post baseline MRI.

Finally, although only the last assessed post-baseline MRI deformation will be available at time of database lock, the MRI processing team will remain blind and will process the remaining post-baseline MRIs. A subsequent analysis utilizing all on-treatment post-baseline assessments will be done based on an MMRM analysis with the model including fixed effect terms for baseline MMSE score, visit, visit x baseline MMSE score interaction, and visit x treatment group interaction. Visit is treated as a categorical variable. A random effect will be included for

participant. The within-participant covariance will be unstructured. A similar approach to calculating degrees of freedom and convergence issues will be taken as done for the primary endpoint.

As a further exploratory analysis, for Mild vs Moderate MMSE baseline status subgroups an analysis will be utilized on each subgroup using the model specification as described above.

In addition, descriptive statistics will also be presented at endpoint in the Open Label Extension Phase, by original randomized group (BHV-4157 and Placebo to BHV4157).

4.8.4.2 ADCS-ADL

The ADCS-ADL scale is a validated tool for assessing instrumental and basic activities of daily living based on a 23-item structured interview of the study partner. The scale has a range of 0 to 78, with lower scores indicating greater impairment[5].

This treatment effect will be summarized as the difference in change from baseline in the ADCS-ADL total score between the BVH-4157 and placebo groups.

The change from baseline ADCS-ADL total score will be analyzed using the mITT set via a MMRM analysis model using the same approach as in the primary analysis.

As a further exploratory analysis, for Mild vs Moderate MMSE baseline status subgroups an MMRM analysis will be utilized on each subgroup using the model specification as described above.

When computing the total score, if 25% or fewer of the questions are missing within a question category, the score for each missing question will be prorated using the average score of the completed questions, categorized by basic ADL and IADL:

> Basic ADL (questions 1-6) will be prorated using the average score of completed questions 1-6 (no more than 1 of these question can be prorated).
>
> Instrumental ADL (questions 7-23) will be prorated using the average score of completed questions 7-23 (no more than 4 of these questions can be prorated).

If more than 25% of questions with in a subgroup are missing, the total score will be missing.

The ADCS-ADL will be assessed at visits conducted in clinic or via remote communication. Only in-clinic assessments will be considered in the primary analysis; sensitivity analysis will include similar analysis including remote assessments.

4.8.4.3 NPI Total Scores

The NPI is a validated, reliable, multi-item instrument to assess psychopathology in AD based on interview with the study partner[3]. The NPI evaluates both the frequency and severity of 12 neuropsychiatric features, including delusions, hallucinations, agitation/aggression, dysphoria, anxiety, euphoria, apathy, disinhibition, irritability/lability, and aberrant motor behavior, as well

as evaluates sleep and appetite/eating disorders. Frequency assessments range from 1 (occasionally, less than once per week) to 4 (very frequently, once or more per day or continuously) as well as severity (1 = mild, 2 = moderate, 3 = severe). The overall score and the score for each subscale are the product of severity and frequency. This treatment effect will be summarized as the difference in change from baseline in the NPI total score between the BVH-4157 and placebo groups.

As for the primary endpoint, the change from baseline in NPI total score will be analyzed using the mITT set via a MMRM analysis model using the same approach as in the primary analysis.

As a further exploratory analysis, for Mild vs Moderate MMSE baseline status subgroups an MMRM analysis will be utilized on each subgroup using the model specification as described above.

In the event that an NPI is partially, but not completely administered, individual NPI scores will be imputed within the scoring algorithm using the following guidelines:

- If there are 4 or fewer items with missing data on an NPI (severity and/or frequency), and if the participant has a previous or subsequent complete NPI, the scores for each missing item will be used from the assessment completed on the date closest to the assessment with missing data (either the previous NPI or subsequent NPI) in order to calculate a total score.

- If 5 or more questions are missing, the entire assessment will be considered as missing.

The NPI will be assessed at visits conducted in clinic or via remote communication. Only in-clinic assessments will be considered in the primary analysis; sensitivity analysis will include a similar analysis including remote assessments.

4.8.4.4 MMSE Scores

The MMSE is a validated, brief, frequently used screening instrument for Alzheimer's disease drug studies[4]. The MMSE scale evaluates orientation, memory, attention, concentration, naming, repetition, comprehension, and ability to create a sentence and to copy two overlapping pentagons. The MMSE is scored as the number of correctly completed items with a lower score indicative of poorer performance and greater cognitive impairment. The total score ranges from 0 (worse) to 30 (perfect performance).

When computing the total score, if three or fewer of the 30 items on the MMSE are missing, the last non-missing observation for those items will be carried forward in order to calculate a total score. If four or more items are missing, the MMSE Total Score will be missing.

The change from baseline in MMSE Total Score will be analyzed using the mITT set via a MMRM analysis model, similar to the primary endpoint.

As a further exploratory analysis, for Mild vs Moderate MMSE status subgroups an MMRM analysis will be utilized on each subgroup using the model specification as described above.

4.8.5 Exploratory Efficacy Endpoints

4.8.5.1 MoCA Scores

The MoCA is a brief mental status screening instrument that evaluates orientation, memory, executive functions, visuospatial ability, language, abstraction, attention, and naming[8].

When computing the total score, if three or fewer of the 30 items on the MoCA are missing, the last non-missing observation for those items will be carried forward to compute a total score. If four or more items are missing, the MoCA Total Score will be missing.

The change from baseline in MoCA Total Score will be analyzed using the mITT set via a MMRM analysis model.

4.8.5.2 Composite Neuropsychological Test Battery Score

A cognitive composite outcome measure will be computed using measures from the NACC Neuropsychological Test Battery[18]. The composite outcome will comprise the overall average of the Z-scores from each of the following NACC subtest measures: Craft Story 21 total immediate Recall; Craft Story 21 total delayed recall; Benson Figure copy score; Benson Figure delayed recall; Multilingual Naming Test total correct; Letter Fluency total score; Category Fluency total score; Trail Making Test A time to completion; Trail Making Test B time to completion; Number Span Forward total correct; and Number Span Backward total correct. Z-scores for each participant on each measure will be calculated using the overall average of all participants at baseline as reference.

The total composite neuropsychological test battery score will be calculated as the overall average Z-score from each of the completed subtests. If more than 2 NACC Neuropsychological Test Battery subtest measures are missing, the total composite neuropsychological test battery score will be missing. Please refer to Appendix 3 regarding prorating subtest scores when there are partially (i.e., item-level) missing data.

The change from baseline in composite neuropsychological test battery score will be analyzed using the mITT set via a mixed model as above. However since there are no repeated measures on participants, there will not be a participant-level random effect. This analysis is exploratory for potential publication purposes only.

In addition, descriptive statistics will also be presented by visit in the Open Label Extension Phase, by original randomized group (BHV-4157 and Placebo to BHV4157).

4.8.5.3 ADCOMS

The AD Composite Score (ADCOMS) is a weighted composite outcome optimized for progression in an MCI stage of AD, and comprised of scores from 4 ADAS-Cog subscales (Delayed Word Recall, Orientation, Word Recognition, and Word Finding Difficulty), 2 MMSE items (Orientation to Time, and Intersecting Pentagons), and the CDR Sum of Boxes (Wang et al., 2016). Because Delayed Word Recall is not assessed within this trial, the ADCOMS cannot be calculated using the methods and weights described by Wang and colleagues (2016) and this

composite is not optimized for the mild-to-moderate population included in our study. However, exploratory analyses may be run using historical data from a similar mild-to-moderate population to combine and weight elements of the ADAS-Cog and MMSE with the CDR-SB to develop a composite score using similar methodology. Change from baseline to Week 48 on this composite outcome will be analyzed using the mITT set via an MMRM analysis model to assess impact on AD progression.

4.8.5.4 CSF, Serum and Plasma Biomarker Panel

CSF, serum and plasma biomarker panel (including, for example, Aß42, Aß42/40 ratio, total tau, p-tau-181, neurogranin, NfL, YKL-40, VILIP, SNAP-25, sTREM2, GFAP) will be assessed at screening, week 24 and week 48. CSF will be collected and analyzed in a subset of the study population (estimated n=50 active, n=50 placebo) at screening, week 24 and week 48.

The change from baseline in the response through week 48 will be analyzed using the mITT population via a Mixed Model for Repeated Measures (MMRM) analysis model, similar to the primary analysis. Fixed effects in the models are APOE status (e4 carrier vs e4 noncarrier), baseline score, treatment group (active vs placebo), visit, visit x baseline score, and visit x treatment group interaction. If appropriate, the response variable will be log transformed. Subjects who have a baseline measure and at least one evaluation on-treatment following baseline will form the study population for this analysis.

Summary statistics will be provided by arm. A p-value and 95% confidence interval for the estimated effect will be included in the summary.

4.9 Pharmacokinetic Evaluations

All PK analyses will be conducted using the treated population.

A PK sample will be collected at baseline, weeks 4, 8, 12, 24 and 48 of the Randomization Phase. Additionally, PK samples should be drawn if there are any SAEs that could possibly be drug related or severe AEs that could be drug related. Date and time of doses on the day of visits and day prior will be collected in the case report forms along with the time of last meal.

Participants who are able to schedule a morning visit for baseline, weeks 2, 4, 8, 12, 24 and 48 can be instructed to hold their dose of study drug that morning until after a PK trough sample is obtained, if possible and appropriate.

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

Individual plasma concentration data will be displayed in listings for the Randomization Phase.

4.10 Safety and Other Analyses

Safety and other exploratory analyses will be conducted on the Treated Participants Population for the respective phase. Separate summaries will be provide for Randomization and Extension

Phases. In addition, where indicated safety summaries for the combined all BHV-4157 exposure will be provided.

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

Continuous values will be summarized by descriptive statistics (mean, median, SD, minimum, maximum, and number of participants). Mean line plots over time may be displayed if applicable with separate lines for each treatment.

Categorical variables will be summarized using the number and percentage of participants in each category for each treatment group and overall. The denominators for calculating the percentages will be based on the number of participants with non-missing assessments at a particular visit for the safety population.

In order to assess the impact of the COVID-19 pandemic crisis on safety results, select safety summaries will be repeated separating subjects by whether they were active (on-treatment) in the double-blind Randomization phase on or after March 13, 2020 or only active prior to this date.

4.10.1 Extent of Exposure and Compliance to Study Treatment

Participants will receive placebo (QD) or BHV-4157 (140 mg QD followed by 140 mg BID) during the Randomization Phase. For the Extension Phase subjects will receive open label BHV-4157 (140 mg QD followed by 140 mg BID).

For the Randomization Phase, the extent of participant exposure (including missed dose days) will be quantified as the number of days on study drug (placebo or BHV-4157) and measured from the time the participant received the first dose until the time the participant received the last dose (i.e., total days on randomized study medication = last day of double-blind randomized study medication – day 1 of double blind randomized study medication + 1).

For the Extension Phase, the extent of participant exposure (including missed dose days) will be quantified as the number of days on study drug (open label BHV-4157) and measured from the time the participant received the first dose in the Extension Phase until the time the participant received the last dose (i.e., total days on open label medication = last day of open label study medication – day 1 of open label study medication + 1).

For combined exposure of BHV-4157 either the randomization and the extension phase, the extent of participant exposure (including missed dose days) will be quantified as the number of days on study drug (BHV-4157) and measured from the time the participant received the first dose BHV-4157 until the time the participant received the last dose of BHV-4157 regardless of phase (i.e., total days on randomized study medication = last day of BHV-4157 study medication – day 1 of BHV-4157 study medication + 1).

Extent of participant exposure for a given phase or overall for BHV4157 will also be calculated only including days where number of capsules taken was >0 (total days on study medication).

For the Randomization Phase, the number of participants on-treatment and their average daily dose (including minimum and maximum) will be summarized by 14-day intervals for the first 28 days and then by month (28-day intervals) as well as overall for the randomization phase. For participants on placebo the number of capsules will be summarized.

For the Randomization Phase, the extent of study medication exposure will also be summarized by a Kaplan-Meier plot presenting proportion of participants still on Randomization Phase study medication (including days of missed dose).

For the Extension Phase and the entire study, the number of subjects on-treatment and their average daily dose (including minimum and maximum) will be summarized by month (30-day intervals) for the as well as overall for the Extension Phase and the entire study (BHV-4157 exposure).

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

- Randomization Phase % compliance = total days on study medication (i.e. number of days where number of capsules taken > 0) during the Randomization Phase / (last dose date in Randomization Phase– first dose date + 1) × 100

- Extension Phase % compliance = total days on BHV-4157 in Extension Phase / (last dose date – first dose date in Extension Phase + 1) × 100

- Overall % compliance = total days on BHV-4157 during the study / [(last BHV-4157 dose date – first BHV-4157 dose date + 1) – (number of days with a dosing break in dosing between the Randomization and Extension Phases)]

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

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

4.10.2 Adverse Events

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

Analyses of AEs will be performed for those events that are considered treatment-emergent (TEAE), where treatment-emergent is defined as events that first occurred or worsened on or after first dose of study medication. See section 4.4 for the detailed definition of “on-treatment” treatment-emergent AEs (TEAEs) and see section 4.3.7 for additional details when onset date is partial.

The number and percentage of participants with the following AEs will be summarized by treatment group and overall for the Randomization and Extension Phases, separately, as well as just overall for subjects after at least one dose of BHV-4157 in any phase.

- TEAEs
- TEAEs Related, Probably Related or Possibly Related to Treatment (An AE is considered to be related to study drug if the relationship is either definitely, probably or possibly related.)
- Treatment-emergent SAEs
- TEAEs leading to discontinuation of study treatment
- TEAEs by highest severity (mild, moderate, severe)
- TEAEs Related, Probably Related or Possibly Related to Treatment by highest severity (mild, moderate, severe)
- TEAEs indicating potential Interstitial Lung Disease (using Standardized MedDRA Queries (SMQ) Interstitial Lung Disease including Eosinphilic Pneumonia and Hypersensitivity Pneumonitis)

In the above tabulations, each participant will contribute only once for a given preferred term (i.e., the most related occurrence or the most severe occurrence) to each of the incidence rates in the descriptive analysis, regardless of the number of episodes. No formal hypothesis-testing analysis of AEs incidence rates will be performed.

All AEs occurring pre-treatment, during the Randomization and Open Label Extension Phases, and throughout the entire study will be listed. Additional listings will be provided including all deaths, SAEs, and AEs leading to withdrawal of study drug.

Listings will indicate whether subject was active during COVID-19 pandemic crisis and the date of first visit impacted.

4.10.3 Laboratory Data

Clinical laboratory evaluations include:

- Hematology: hemoglobin, hematocrit, platelets, complete blood count with differential and absolute neutrophil count
- Serum Chemistry: sodium, potassium, chloride, calcium, ALT, AST, LDH, alkaline phosphatase, GGT, phosphorous, bicarbonate, CPK, total protein, albumin, direct bilirubin, indirect bilirubin, total bilirubin, glucose, creatinine, BUN, uric acid, total cholesterol, LDL, HDL, triglycerides
- Urinalysis: pH, specific gravity, protein, glucose, ketones, urobilinogen, bilirubin, blood

Clinical laboratory values will be expressed using conventional (US) and standard international (SI) units. In the event of repeat values within the same analysis visit, if any measurement has abnormal results, that measurement will be used for presentation in by-visit tables. If none of the measurements are abnormal, or all of them are abnormal, the worst measurement in the analysis

visit interval will be used for presentation in by-visit tables. All measurements will be presented in listings and considered for evaluation of potential drug- induced liver injury (DILI) or abnormalities.

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

On-treatment laboratory abnormalities are those with an assessment date after the date/time of first dose of study drug and within 30 days after the last dose of study drug. For the Randomization Phase, treatment-emergent laboratory abnormalities will be assessed from the date of first dose of study drug until: 1) the first day of the Extension Phase or 2) if the subject did not continue into the Extension Phase, 30 days after the last dose of study drug. For the Extension Phase, treatment-emergent lab abnormalities will be assessed from the first dose during the Extension Phase until 30 days after the last dose of study drug. (See section 4.3)

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

Clinical laboratory test results will be graded according to CTCAE version 5.0 if criteria for test available, otherwise according to DAIDS version 2.1 criteria, if available for test. The number and percent of participants with at least one on-treatment lab assessment by highest grade will be summarized for each treatment (regardless of baseline) for the Randomization and Extension Phases, separately as well as just overall for subjects after at least one dose of BHV-4157 in any phase. In addition, the shift from baseline to worst on-treatment lab (based on either normal limits or grading, where available) will be tabulated by visit and the overall minimum and maximum observed for treated subjects in the Randomization and Extension Phases, separately, as well as just overall for subjects after at least one dose of BHV-4157 in any phase. The shift tables will include only participants with a baseline assessment and at least one on-treatment assessment.

For the liver function tests, AST, ALT, Alk Phos, and BILI the shift from baseline will also be presented by visit and to the maximum observed abnormality for the Randomization Phase and Extension Phases, separately, as well as just overall for subjects after at least one dose of BHV-4157 in any phase. Baseline in the shift tables will be defined as per Section 4.2.8; the last available assessment on or before the first day of study treatment in that phase. Shift tables will only include treated participants with an assessment for the specific test of interest at baseline and on treatment. The following categories will be used to summarize the shift from baseline based on the upper limit of normal (ULN) range for ALT and AST:

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

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

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

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

- <=ULN
- >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 Total Bilirubin BILI:

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

An evaluation of drug-induced serious hepatotoxicity (eDISH) scatter plot will display the maximum TBL ratio of value to ULN on the y-axis versus the maximum ALT ratio of value to ULN on the x-axis, where the maxima is not necessarily concurrent, for treated subjects in the randomization phase and DB and OL BHV-4157 treated subjects. Both axes will be on the log10 scale. Ratios < 0.1 x ULN will be set to 0.1. Sample sizes in the legend will represent subjects with paired ratios. A horizontal reference line will be placed at 2 x ULN, and a vertical reference line will be placed at 3 x ULN. The lower left quadrant will be labeled “Normal Range”, the upper left quadrant will be labeled “Hyperbilirubinemia”, the lower right quadrant will be labeled “Temple’s Corollary”, and the upper right quadrant will be labeled “Possible Hy’s Law Range.”

All laboratory data will be presented in data listings. Additional listings will be presented for all abnormal laboratory values considered potentially clinically significant (Grade 3 or 4) for the Randomization and Extension Phases. Participants 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 will also be presented in a listing. Note that these abnormalities do not need to occur on concurrent visits.

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

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

4.10.4 Physical Examinations

Physical examination findings and number of participants are summarized as the count and percentage of participants by eCRF pre-defined categories at last visit. Results of Physical Examination will be included as a listing for each arm and overall.

4.10.5 Vital Signs and Physical Measurements

The observed value and change from baseline in vital signs and physical measurements will be summarized for each arm at each visit for the Randomization and Extension Phases, separately. The summary will be based on the "on-treatment" data after the date/time of first dose of study drug and within 30 days after the last dose of study drug (see section 4.3 for detailed definition of "on-treatment").

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

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

The number and percent of subjects meeting these criteria will be summarized for the Randomization and Extension Phase, separately as well as the overall BHV-4157 treated exposure. A subject listing will provide for each vital sign measure abnormality criteria listing any subject meeting the criteria with a complete list of all of their values that particular vital sign measure with the specific assessments that meet the criteria indicated.

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

4.10.6 Electrocardiogram

Descriptive statistics for ECG interval data (e.g., QRS, PR, QT, QTcF), and ventricular heart rate will also be reported by visit for the Randomization and Extension Phases, separately. The summary will be based on the data after the date/time of first dose of study drug and within 30 days after the last dose of study drug (see section 4.3 for detailed definition of "on-treatment").

In addition, the number and percentages of subjects with at least one post-treatment QTcF > 450 ms, >480 ms, and >500 ms will be summarized for the Randomization and Extension Phase, separately, as well as the overall BHV-4157 treated exposure. Similar number and percentages will be presented for subjects with at least one post-baseline QTcF Change from baseline >= 30 ms to < 60ms and those with at least one change from baseline >= 60 ms. A subject listing will provide each QTcF sign measure, specifying any subject meeting the criteria with a complete list of their QTcF assessments with the ones meeting the criteria indicated.

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

4.10.7 Concomitant Medications

Concomitant medications will be coded with WHODrug (Global March 2019) using the WHODrug Insight tool. Each codable concomitant medication will be assigned five values: Drug Name, Drug Record Number, SeqNum1, SeqNum 2, and Anatomic Therapeutic Class (ATC) code. Results will be tabulated by treatment group, ATC code, and Drug Name during the Randomization and Extension Phases, separately. Note that all concomitant medications with a given Drug Record Number will be assigned the same ATC code.

Unless the start date of the medication is after the last study drug dose date, or the end date of the medication is prior to the start date of the study drug, the medication will be considered 'concomitant.' Partial onset and dates need to be considered in the determination of the medication being concomitant or not.

4.10.8 Sheehan-Suicidality Tracking Scale (S-STS)

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

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

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

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

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

The self-reported S-STS ideation subscale, behavior subscale, and total score will be summarized as the change from baseline (i.e., <-1, -1, no change, 1, >1) at each visit and at Maximum score in the Randomization and Extension Phases, presented separately. Both

versions of the Total score will be summarized. All S-STS assessments will be included in summaries regardless of being completed at study site or remote.

4.10.9 Subjects Identified for Narratives

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

- All deaths on-treatment and post-treatment through the end of the study

- SAEs on-treatment, which includes up to 30 days after the last dose of study drug; SAEs that occur > 30 days (i.e., during the follow-up period) will be included per the clinical judgment of the Biohaven medical monitor

- All premature discontinuations of study drug due to AEs (either identified through "action taken" or "end of treatment status")

- The following on-treatment events of special interest:
  - Neutropenia based on laboratory results and defined as minimum absolute neutrophil count < 500 per mm3
  - LFT abnormalities:
    - ALT or AST > 3x ULN
    - ALT or AST > 3x ULN, and serum total bilirubin > 2x ULN

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

These select events are described in the current version (v3) of the Biohaven Safety Narrative Scope for BHV-4157 (troriluzole). Because select events may be subject to change, updates to the list of events or selection algorithms after database lock may be described in a Note to File (NTF) rather than amending the SAP.

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

5 REFERENCES

1. Ayela B, Lipkovich I, Molenberghs G, Mallinckrodt CH. (2014). A multiple imputation-based approach to sensitivity analyses and effectiveness assessments in longitudinal clinical trials. *Journal of Biopharmaceutical Statistics*, *24*(2). 211–228.

2. Carpenter J, Roger J, Kenward M. Analysis of longitudinal trials with missing data: a framework for relevant, accessible assumptions, and inference via multiple imputation. *Journal of Biopharmaceutical Statistics,* 23 2013: 1352–1371.

3. Cummings, J. L. (1997). The Neuropsychiatric Inventory: assessing psychopathology in dementia patients. *Neurology, 48*(5 Suppl 6), S10-16.

4. Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res, 12*(3), 189-198.

5. Galasko D1, Bennett D, Sano M, Ernesto C, Thomas R, Grundman M, Ferris S. An inventory to assess activities of daily living for clinical trials in Alzheimer's disease. The Alzheimer's Disease Cooperative Study. *Alzheimer Dis Assoc Disord,* 1997;11 Suppl 2:S33-9.
6. Holland, D., McEvoy, L. K., Dale, A. M., & Alzheimer's Dis, N. (2012). Unbiased comparison of sample size estimates from longitudinal structural measures in ADNI. *Human Brain Mapping, 33*(11), 2586-2602. doi:10.1002/hbm.21386

7. Mallinckrodt, C., Molenberghs, G., and Rathmann, S. (Jan-Feb 2017). Choosing Estimands in Clinical Trials with Missing Data. *Pharmaceutical Statistics*, *16*(1): 29-36. doi: https://doi.org/10.1002/pst.1765|
8. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, Cummings JL, Chertkow H. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*(4), 695-9.

9. Qiu Y, Jacobs DM, Messer K, Salmon DP, Feldman HH (2019). Cognitive heterogeneity in probable Alzheimer disease: Clinical and neuropathologic features. *Neurology, 22*: 10-212.

10. Ratitch, B., Bell, J., Mallinckrodt, C., Bartlett, JW., Goel, N., Molenberghs, G., O' Kelly, M., Singh, P., and Lipkovich, I (April 2019). Choosing Estimands in Clinical Trials: Putting the ICH E9(R1) Into Practice. *Therapeutic Innovation & Regulatory Science*, 4. doi: 10.1177/2168479019838827.

11. Rosen, W. G., Mohs, R. C., & Davis, K. L. (1984). A new rating scale for Alzheimer's disease. *Am J Psychiatry, 141*(11), 1356-1364. doi:10.1176/ajp.141.11.1356

12. Rubin, D. B. (1987). Multiple Imputation for Nonresponse in Surveys. Wiley and Sons, Harvard.

13. Schluchter, M. D., and Elashoff, J. D. (1990). “Small-Sample Adjustments to Tests with Unbalanced Repeated Measures Assuming Several Covariance Structures.” *Journal of Statistical Computation and Simulation* 37:69–87.

14. Thomas RG, Albert M, Petersen RC, Aisen PS. Longitudinal decline in mild-to-moderate Alzheimer's disease: Analyses of placebo data from clinical trials. *Alzheimers Dement*, *12*(5), 598-603. doi: 10.1016/j.jalz.2016.01.002. Epub 2016 Feb 23.
15. Turner, R. S., Thomas, R. G., Craft, S., van Dyck, C. H., Mintzer, J., Reynolds, B. A., . . . Alzheimer's Dis Cooperative, S. (2015). A randomized, double-blind, placebo-controlled trial of resveratrol for Alzheimer disease. *Neurology, 85*(16), 1383-1391. doi:10.1212/wnl.0000000000002035

16. Tsiatis, A. A. *Semiparametric theory and missing data*. Springer, 2006.

17. Wang J, Logovinsky V, Hendrix SB, Stanworth SH, Perdomo C, Xu L, Dhadda S, Do I, Rabe M, Luthman J, Cummings J. ADCOMS: a composite clinical outcome for prodromal Alzheimer's disease trials. *Journal Neuroly, Neurosurgy and Psychiatry*. 2016 Sep 1;87(9):993-9.
18. Weintraub S, Besser L, Dodge HH, Teylan M, Ferris S, Goldstein FC, Giordani B, Kramer J, Loewenstein D, Marson D, Mungas D. Version 3 of the Alzheimer Disease Centers’ neuropsychological test battery in the Uniform Data Set (UDS). Alzheimer disease and associated disorders. 2018 Jan;32(1):10.

19. Xiong et al (2017). Linear Combinations of Multiple Outcome Measures to Improve the Power of Efficacy Analysis ---Application to Clinical Trials on Early Stage Alzheimer Disease. *Journal of Biostatistics and Epidemiology, 1*(1): 36-58.

APPENDIX

APPENDIX 1 - Study Plan and Procedures at Each Visit

Table 1: Schedule of Assessments and Events – Double-Blind Phase

Visit Number	1	2	3	4	5	6	7	8	9 or Early Term	10	
Study Visit Time Point	Screening (within 42 days prior to baseline)	Baseline (Week 0)	Week 2 (±3 days)	Week 4 (±7 days)	Week 8 (±7 days)	Week 12 (±7 days)	Week 24 (±7 days)	Week 36 (±7 days)	Week 48[1] (±7 days)	Follow-Up Week 52 (±7 days)	Unscheduled Visit[2]
Informed Consent	X										
Eligibility Review	X	X									
Randomization[3]		X									
Med History/Demographics	X										
Modified Hachinski Ischemic Scale	X										
Height[4] & Weight	X	X		X	X	X	X	X	X	X	
Physical Examination	X	X		X	X	X	X	X	X	X	
Neurological Examination	X										
Vital Signs[5]	X	X		X	X	X	X	X	X	X	X
Concomitant Medication	X	X	X	X	X	X	X	X	X	X	
Adverse Events[6]	X	X	X	X	X	X	X	X	X	X	X
12-lead ECG (resting)	X					X	X		X	X	
Clinical Safety Blood Tests[7]	X	X		X	X	X	X	X[8]	X	X	
B12 (blood test)	X										
Urinalysis[9]	X	X		X		X	X		X		
HIV, HBsAg, HCV antibody	X										
ApoE4 (blood test)	X										
Blood Collection for Biomarker Banking		X					X		X		
Blood Collection for Pharmacokinetics[10]		X		X	X	X	X		X		
Visit Number	1	2	3	4	5	6	7	8	9 or Early Term	10	

Study Visit Time Point	Screening (within 42 days prior to baseline)	Baseline (Week 0)	Week 2 (±3 days)	Week 4 (±7 days)	Week 8 (±7 days)	Week 12 (±7 days)	Week 24 (±7 days)	Week 36 (±7 days)	Week 48 (±7 days)	Follow-Up week 52 (±7 days)	Unscheduled Visit
Volumetric MRI[11]	X						X		X		
Lumbar Puncture (LP)[11,12]	X						X		X		
CSF Biomarkers[13]	X						X		X		
Post-LP Safety Telephone[14]	X						X		X		
Sheehan Suicidality Tracking	X	X		X	X	X	X	X	X	X	
ADAS-Cog 11		X				X	X	X	X		
CDR-SOB		X				X	X	X	X		
NPI		X					X		X		
ADCS-ADL		X					X		X		
Craft Story 21 Recall		X							X		
Benson Complex Figure		X							X		
MINT		X							X		
Letter & Category Fluency		X							X		
Trail Making Test A & B		X							X		
Number Span Test		X							X		
MMSE	X	X					X		X		
MoCA		X					X		X		
Research Satisfaction Survey		X				X	X	X	X		
Dispense Study Drug[15]		X		X	X	X	X	X			
Study Drug Instruction Phone Call[16]			X								
Study Drug Accountability				X	X	X	X	X	X		
Treatment Blinding Questionnaire[17]									X		

Every effort should be made to conduct the Week 48 visit and maintain the +/- 7 day window. However, due to concerns related to the COVID-19 pandemic, the Week 48 visit window may be modified beyond the +/- 7 day window, in order to minimize any potential risks to study participant safety and to comply with governmental and local institutional guidance (e.g., study site has a policy that a clinical research visit must be delayed). Under these circumstances, the visit window may be extended (up to a maximum treatment duration of 60 weeks), but every attempt should be made to conduct the visit as close to the date the visit is due as possible. If the visit window is modified, participants should be evaluated remotely (e.g., via phone) at the time of the scheduled Week 48 visit to perform and document appropriate safety assessments. Study medication may be sent to the participant via tracked and certified courier. For any such cases, the investigator should discuss the specific circumstances of each case with the sponsor medical monitor (or designee) who must approve the request prior to any modification of the visit window: PPD Cell PPD PPD

1 The following must be collected at any unscheduled visit: vital signs, prior and concomitant medications, adverse events, and protocol deviations. All other unscheduled visit procedures are subject to PI discretion or request from the Medical Monitor.

2 Randomization must occur at the baseline visit after eligibility is confirmed.
3 Height is done at screening only.
4 Vital signs include sitting blood pressure, pulse, temperature, respiration rate, and weight.
5 The SAE reporting period starts at the screening visit (i.e., when the participant or LAR signs consent). The reporting period for non-serious AEs starts at the baseline visit at the time of receiving the first dose of study drug. The end of the reporting period for both SAEs and AEs is 30 days after the study drug has been discontinued.
6 Clinical Safety Laboratory Tests will be performed by a Central Laboratory. Assessment includes the following: Hematology (hemoglobin, hematocrit, platelets, RBC, WBC, differential count, and absolute neutrophil count), Chemistry (sodium, potassium, chloride, calcium, ALT, AST, LDH, alkaline phosphatase, GGT, phosphorus, bicarbonate, CPK, total protein, albumin, indirect bilirubin, direct bilirubin, total bilirubin, glucose, creatinine, BUN, uric acid, total cholesterol, LDL, HDL, triglycerides, folate).
7 Performing additional safety labs at Week 36 is now strongly encouraged, given the COVID-19 pandemic and unanticipated issues related to collecting lab samples that may potentially arise.
8 Urinalysis to include: pH, specific gravity, protein, glucose, ketones, urobilinogen, bilirubin, blood
9 A blood sample for pharmacokinetic (PK) measures in plasma should be collected at baseline, week 4, week 8, week 12, week 24, and week 48. Participants should take their dose at their routine time on the days of these visits. Date and time of last dose should be collected in case report forms for entry into the EDC system. Participants who are able scheduled for a morning visit should be instructed to hold their dose of study drug that morning until after a PK trough sample is obtained, if possible and appropriate in the judgment of the investigator. Blood samples for PK in plasma should be drawn if there are any SAEs that could possibly be drug-related or severe AEs that could be drug-related. In addition, for participants that are undergoing lumbar puncture, PK samples should be drawn at the time of lumbar puncture. Date and time of doses on the day of lumbar puncture and day prior should be collected in case report forms, for entry into the EDC system.
10 All vMRIs must be performed per imaging protocol and must use the same scanner throughout study.
    - **Screening vMRI:** if a participant has **not** had an MRI performed within 6 months of screening (i.e., within 6 months from the date of informed consent), then an MRI must be performed as part of the screening requirements for this study, per the imaging protocol, and should be one of the last screening procedures performed to determine final eligibility in order to prevent participants from undergoing unnecessary MRIs. If a participant **has** had an MRI within 6 months of screening (i.e., within 6 months from the date of informed consent) **but** the MRI does not follow the study-specific imaging protocol, that MRI can be used to help determine eligibility; however, another MRI must be performed per the imaging protocol, and must occur as close to, and prior to, the baseline visit, after all other eligibility criteria have been confirmed.
    - **Week 24 and week 48 vMRI:** the protocol window for vMRI at week 24 and week 48 is 14 days before and up to14 days after the clinic visit. If participant is terminating early at 36 week or after, obtain vMRI. If MRI is performed on the same day as a lumbar puncture at week 48, the vMRI must be conducted before the lumbar puncture. Otherwise, at least a 3-day window between vMRI and the lumbar puncture is required.
11 It is anticipated that lumbar puncture will be performed on an estimated 100 study participants as part of a sub-study (separate informed consent required).
12 Only for the CSF sub-study. Visit windows for CSF are: up to 14 days prior to first dose of study medication and up to 14 days prior to weeks 24 and 48. If a participant in the CSF sub-study is terminating early, they do not need to undergo lumbar puncture.
13 Post lumbar puncture safety follow up telephone call must occur 1 to 3 days after the lumbar puncture is performed.
14 First dose of study medication will be given in the clinic at the baseline clinic visit after completion of the baseline visit procedures.
15 Participants will receive a phone call during Week 2 from site study personnel in order to assess study drug tolerability. If tolerable, then the participant will be instructed to proceed with dose escalation.
16 Treatment blinding questionnaire to be administered to site PI and Raters.

Table 2: Schedule of Assessments and Events – Open-Label Extension Phase

Visit Number	1	2	3	4	5	6	7	8	9	10	
Study Visit Timepoint	Baseline[1]	Abbrev-iated Drug Dispensation Visit[2]	Ext Week 2 (+/- 7 days)	Ext Week 4 (+/- 7 days)	Ext Week 8[3] (+/- 7 days)	Ext Week 12 (+/- 7 days)	Ext Week 24 (+/- 7 days)	Ext Week 36 (+/- 7 days)	Ext Week 48 or Early Term (+/- 7 days)	Week2 Post last dose (+/- 7 days)	Unsched-uled Visit[4] (+/- 7 days)
Informed Consent	X	X									
Medical History	X										
Adverse Event Assessment	X	X	X	X	X[3]	X	X	X	X	X	
Telephone Check-in[5] *Includes AE assessment and concomitant medication review*			X		X[3]						
Laboratory Assessments[6]	X			X	X[3]	X	X	X	X		
Urinalysis	X						X		X		
Physical Exam	X								X		
Physical Measurement - Weight	X						X		X		
Vital Signs	X	X		X		X	X	X	X	X	X
12-Lead ECG	X					X	X		X		
Concomitant Medication Review	X	X	X	X	X	X	X	X	X	X	X
Sheehan Suicidality Tracking Scale (S-STS)	X	X		X		X	X	X	X	X	
Dispense Study Drug	X	X		X		X	X	X			
Volumetric MRI[7]	X								X		
ADAS-Cog 11	X						X		X		
CDR-SOB	X						X		X		
Study Drug Accountability				X		X	X	X	X		
Blood Collection for Biomarker Banking							X		X		
Craft Story 21 Recall	X						X		X		
Letter & Category Fluency	X						X		X		
Number Span Test	X						X		X		
MoCA	X						X		X		

1. Baseline visit for the Extension Phase is only required if there is an extended break (≥4 weeks) in dosing between the Randomization and Extension Phases.
2. For participants who will have been off of the study drug for less than or equal to 4 weeks, the subject should come into the investigative site for an Abbreviated Drug Dispensation Visit (also "Abbreviated OLE Baseline Visit"), at which time any major medical or medications changes will be reviewed prior to dispensing riluzole. In this case, most procedures from the Week 48 visit will also serve as the Open-Label Extension Phase Baseline visit
3. Extension Week 8 is primarily to collect Liver Function tests. If a participant prefers not to come into the office, these can be collected at a local laboratory, with a telephone call confirming these were collected as well as a review of concomitant medications and an assessment of any adverse events. If the participant comes into the center, an assessment of adverse events and Liver function tests will be completed. Although all other visits can be conducted remotely as needed due to COVID-19 restrictions, under regular circumstances visits are to be completed in-person. See Protocol Section 8.1.12, Open-Label Extension Phase, for further details and requirements regarding remote visit completion due to COVID-19.
4. The following must be collected at any unscheduled visit: vital signs, prior and concomitant medications, adverse events, and protocol deviations. All other unscheduled visit procedures are subject to PI discretion or request from the Medical Monitor.
5. Telephone calls to subjects will be made between visits during Extension Phase (Weeks 2 and 8) to monitor subject condition, review concomitant medications and assess any adverse events.
6. Clinical Safety Laboratory Tests will be performed by a Central Laboratory. Assessment includes the following: Hematology (hemoglobin, hematocrit, platelets, RBC, WBC, differential count, and absolute neutrophil count), Chemistry (sodium, potassium, chloride, calcium, ALT, AST, LDH, alkaline phosphatase, GGT, phosphorus, bicarbonate, CPK, total protein, albumin, indirect bilirubin, direct bilirubin, total bilirubin, glucose, creatinine, BUN, uric acid, folate).
7. Volumetric MRI is not required if the participant's Double-Blind Week 48 MRI scan occurred less than than 3 months prior to their joining the Open-Label Extension Phase.

Table 2: Schedule of Assessments and Events – Open-Label Extension Phase

Visit Number	1	2	3	4	5	6	7	8	9	10	
Study Visit Timepoint	Baseline[1]	Abbrev-iated Drug Dispensation Visit[2]	Ext Week 2 (+/- 7 days)	Ext Week 4 (+/- 7 days)	Ext Week 8[3] (+/- 7 days)	Ext Week 12 (+/- 7 days)	Ext Week 24 (+/- 7 days)	Ext Week 36 (+/- 7 days)	Ext Week 48 or Early Term (+/- 7 days)	Week2 Post last dose (+/- 7 days)	Unsched-uled Visit[4] (+/- 7 days)
Informed Consent	X	X									
Medical History	X										
Adverse Event Assessment	X	X	X	X	X[3]	X	X	X	X	X	
Telephone Check-in[5] *Includes AE assessment and concomitant medication review*			X		X[3]						
Laboratory Assessments[6]	X			X	X[3]	X	X	X	X		
Urinalysis	X						X		X		
Physical Exam	X								X		
Physical Measurement - Weight	X						X		X		
Vital Signs	X	X		X		X	X	X	X	X	X
12-Lead ECG	X					X	X		X		
Concomitant Medication Review	X	X	X	X	X	X	X	X	X	X	X
Sheehan Suicidality Tracking Scale (S-STS)	X	X		X		X	X	X	X	X	
Dispense Study Drug	X	X		X		X	X	X			
Volumetric MRI[7]	X								X		
ADAS-Cog 11	X						X		X		
CDR-SOB	X						X		X		
Study Drug Accountability				X		X	X	X	X		
Blood Collection for Biomarker Banking							X		X		
Craft Story 21 Recall	X						X		X		
Letter & Category Fluency	X						X		X		
Number Span Test	X						X		X		
MoCA	X						X		X		

8. Baseline visit for the Extension Phase is only required if there is an extended break (≥4 weeks) in dosing between the Randomization and Extension Phases.

9. For participants who will have been off of the study drug for less than or equal to 4 weeks, the subject should come into the investigative site for an Abbreviated Drug Dispensation Visit (also "Abbreviated OLE Baseline Visit"), at which time any major medical or medications changes will be reviewed prior to dispensing riluzole. In this case, most procedures from the Week 48 visit will also serve as the Open-Label Extension Phase Baseline visit
10. Extension Week 8 is primarily to collect Liver Function tests. If a participant prefers not to come into the office, these can be collected at a local laboratory, with a telephone call confirming these were collected as well as a review of concomitant medications and an assessment of any adverse events. If the participant comes into the center, an assessment of adverse events and Liver function tests will be completed. Although all other visits can be conducted remotely as needed due to COVID-19 restrictions, under regular circumstances visits are to be completed in-person. See Protocol Section 8.1.12, Open-Label Extension Phase, for further details and requirements regarding remote visit completion due to COVID-19.
11. The following must be collected at any unscheduled visit: vital signs, prior and concomitant medications, adverse events, and protocol deviations. All other unscheduled visit procedures are subject to PI discretion or request from the Medical Monitor.
12. Telephone calls to subjects will be made between visits during Extension Phase (Weeks 2 and 8) to monitor subject condition, review concomitant medications and assess any adverse events.
13. Clinical Safety Laboratory Tests will be performed by a Central Laboratory. Assessment includes the following: Hematology (hemoglobin, hematocrit, platelets, RBC, WBC, differential count, and absolute neutrophil count), Chemistry (sodium, potassium, chloride, calcium, ALT, AST, LDH, alkaline phosphatase, GGT, phosphorus, bicarbonate, CPK, total protein, albumin, indirect bilirubin, direct bilirubin, total bilirubin, glucose, creatinine, BUN, uric acid, folate).
14. Volumetric MRI is not required if the participant's Double-Blind Week 48 MRI scan occurred less than than 3 months prior to their joining the Open-Label Extension Phase.

APPENDIX 2 - Calculating the ADAS-Cog Total Score when there are missing data

The ADAS-Cog is a rater-administered scale with eight performance subtests: Word Recall; Commands; Constructional Praxis; Naming; Ideational Praxis; Orientation; Word Recognition; Remembering Word Recognition Test Instructions. Three examiner ratings, of Spoken Language, Language Comprehension, and Word Finding Difficulty, also are obtained. The test is scored in terms of errors, with higher scores reflecting poorer performance and greater impairment. Scores can range from 0 (best) to 70 (worst).

If a participant is unable to complete any of the eight performance subtests, the rater must code the reason for incompletion:

- Not done for reasons other than physical or cognitive
- Participant refused
- - Participant unable to complete for physical reasons
- - Participant unable to complete for cognitive reasons

When total scores for performance subtest measures are missing because the participant is unable to complete the task for cognitive reasons, the worst possible score for the measure will be assigned.

Because Remembering Test Instructions is scored solely based on performance on Word Recognition, if Word Recognition was not completed for any reason, then Remembering Test Instructions must also be scored as not completed.

<u>For a given individual performance subtest:</u>

If 25% or fewer of the items within a subtest are missing, scores for the missing items will be prorated as the average of the completed items on that subtest. This procedure will be employed for a maximum of two performance subtests.

If more than 25% of the items within any performance subtest are missing, the entire subtest is missing.

<u>For the ADAS-Cog Total Score</u>

The total score is the sum of 11 subscales (8 performance subtests and 3 examiner ratings). If more than two subscales (either performance subtest or examiner ratings) are missing or more than 2 performance subtests have partial missing data or if both the Word Recall and Word Recognition subtests are missing, then the ADAS-Cog total score will be missing. Otherwise, the total score will be based on the sum of the scores from the non-missing subscales multiplied (prorated) by the ratio (max ADAS-cog Total Score of 70/max total possible from non-missing subscales).

For example, if the Word Recall performance subtest , which ranges from 0-10 (maximum=10), is missing, and another subtest Commands, which ranges from a score of 0-5 (maximum = 5) is also missing, then the total ADAS-cog score will be calculated as the sum from the 9 non-missing subscales multiplied by the ratio (70/(70 – (10 + 5)) = 70/55 = 1.27, the resulting prorated total score will be rounded up to the nearest integer

APPENDIX 3 - Calculating the composite neuropsychological test battery score when there are missing data

A cognitive composite outcome measure will be computed using measures from the NACC Neuropsychological Test Battery. The composite outcome will comprise the overall average of the Z-scores from each of the following NACC subtest measures: Craft Story 21 total immediate Recall; Craft Story 21 total delayed recall; Benson Figure copy score; Benson Figure delayed recall; Multilingual Naming Test total correct; Letter Fluency total score; Category Fluency total score; Trail Making Test A time to completion; Trail Making Test B time to completion; Number Span Forward total correct; and Number Span Backward total correct. Z-scores for each participant on each measure will be calculated using the overall average of all participants at baseline as reference.

Missing total scores for subtest measures from the NACC Neuropsychological Test Battery are designated as missing for the following reasons:

- · Participant Unable for Cognitive Reasons
- · Participant Unable for Other Reasons (Physical, auditory, etc.)
- · Participant/Study Partner Refused
- · Oversight
- · Other Reason
- · Unknown

When total scores for subtest measures are missing because the participant was unable to complete the task for cognitive reasons, the worst possible score for the measure will be assigned. Total scores that are missing for any other reason will be treated as missing in calculating the composite neuropsychological test battery score.

For partially missing (item-level) missing data, if fewer than 25% of items within a NACC subtest measure are missing, scores for the missing items will be prorated as the average of the completed items on that subtest. This procedure will be employed for a maximum of two NACC subtests. If more than 25% of the items within any NACC subtest are missing, the entire subtest is missing.

The total composite neuropsychological test battery score will be calculated as the overall average Z-score from each of the completed subtests. If more than 2 NACC subtests are missing, the total composite neuropsychological test battery score will be missing.