Janssen Research & Development

Statistical Analysis Plan

A Multicenter, Randomized, Placebo-Controlled, Double-Blind Study to Investigate the Safety, Tolerability, and Clinical Efficacy of Seltorexant (JNJ-42847922) on Behavioral and Psychological Symptoms of Dementia in Patients with Probable Alzheimer's Disease

Protocol 42847922ALZ2001; Phase 2a

JNJ-42847922 (seltorexant)

Status:	Approved
Date:	November 16, 2023
Prepared by:	Janssen Research & Development, LLC
Document No.:	EDMS-RIM-700114; 3.0

Compliance: The study described in this report was performed according to the principles of Good Clinical Practice (GCP).

TABLE OF CONTENTS

VERSION HISTORY

Table – SAP Version History Summary

SAP Version	Approval Date	Change	Rationale
1.0	08/15/22		
2.0	09/07/2023	1.1: Updated CDR cutoff for population from 2 to 1	Change is consistent with protocol amendment
		5.1.1: Updated Visit Window for MMSE	Added baseline and Day 43 windows to be consistent with schedule of activities for MMSE and windows for other parameters
		5.3.1: Clarified the data usage rules regarding ICE strategies for the estimands	Updated rules to remain consistent with other studies in the program
		5.3.2, 5.3.3: Clarified rescue medication ICE	Existing wording did not specify how to handle rescue medication occurring after treatment discontinuation
		5.3.2, 5.5.7, 5.5.10: Removed community dwelling/assisted living status as factors from MMRM/ANCOVA models	Insufficient numbers of participants were stratified to each of the levels to include as a factor in models
		5.3.2, 5.5.7, 6.3: Changed primary models to use per-eDC values of stratification factors, rather than per-IWRS values	Significant mis-stratification based on these stratification factors was observed. Study team agreed that per-eDC values should be used for primary analysis
		5.3.2.1.1: Added that in the case that a structured covariance matrix needs to be used, a robust sandwich variance estimator will be used.	Change to remain consistent with health authority feedback for other studies in the program
		5.3.4: Section added to describe sensitivity analyses related to use of per-eDC vs per-IWRS values of stratification factors	Given that primary analysis was changed to use per-eDC values, a sensitivity analysis based on per-IWRS values was desired
		5.4.1.2.2, 5.5.5.2, 5.5.8.1: Clarified that both SDI summary scores should be used.	Two different SDI summary scores exist. Both are of interest. Existing wording did not specify that both should be included

		5.5.1.2: Remove Change from Baseline in NPI-12 Caregiver Distress Score MMRM analysis	Team agreed that means and mean changes in Caregiver Distress over time were sufficient for exploration of this endpoint
		5.5.5.2: Specified the exact NPI-C analyses to be performed for rebound analysis	Specified sub-totals are of more interest for this study than total score
		5.6.3.1: Added Shift in Metabolic Lab value table	To assist in clinical understanding of treatment
		5.6.3.1, 5.6.3.3: Clarified that only DB Lab and ECG values may be TE.	Clarifying that post-discontinuation of treatment labs would not be considered TE
		5.7.2.1: Actigraphy section updated to specify mobility parameters to be analyzed, and to specify analyses performed to explore correlations	After seeing the parameters included in actigraphy data, the team agreed upon exact parameters to analyze, and analyses to perform
		5.7.4: Added Baseline NPI-12 Sleep Doman Score, Age Group II, and BMI to listing of Subgroups	Updated subgroups of interest based on team consensus
		6.1: Added eCOA abbreviation to abbreviation table	For clarification
		6.3: Removed Prior Medications of Special Interest from Baseline Disease Characteristics table. Added baseline CDR score.	Prior Medications of Special Interest to be included in a separate output. Baseline CDR added to observe impact of protocol amendment
		6.8: Updated several markedly abnormal thresholds for hematology variables to be expressed in Fraction of 1, rather than %. Updated lab test names. Updated Vitamin B12 limits to be expressed in correct units.	Updates made to align table with structure of received data
3.0	11/14/2023	5.4.1.1.1	Clarified the calculation of CMAI-C total score
		5.6.3.1	Team decision to produce shift in lab value table for all chemistry labs. Added categories for metabolic lab shifts

1. INTRODUCTION

This statistical analysis plan (SAP) contains definitions of analysis sets, derived variables, and statistical methods for all planned analyses for study 42847922ALZ2001.

1.1. Objectives and Endpoints

For the following objectives and endpoints, probable AD with clinically significant agitation/aggression is defined as: Diagnostic and Statistical Manual of Mental Disorders-5th Edition (DSM-5) diagnosis of probable Alzheimer's Disease (AD) and criteria of a syndrome diagnosis of agitation based on International Psychogeriatric Association (IPA) consensus clinical and research definition of agitation in cognitive disorders, have an MMSE total score from 10 to 24 (inclusive), and a Clinical Dementia Rating (CDR) of ≥1. Additionally, patients will be characterized by the presence of clinically significant agitation/aggression defined as NPI-12 agitation/aggression (A/A) domain score ≥4 with NPI frequency score ≥2 at entry and baseline with no more than 35% of improvement allowed in NPI-12 A/A score during the screening period.

Double-blind Treatment Phase

Objectives	Endpoints
Efficacy	
Primary	
• To investigate the effect of seltorexant versus placebo on the sum of agitation and aggression domain scores (A+A) of the Neuropsychiatric Inventory - Clinician rating (NPI-C) in participants with probable AD with clinically significant agitation/aggression.	• Change from baseline to Day 43 on NPI-C A+A scores.
Secondary –Efficacy	
To investigate the effect of seltorexant versus placebo on the following in participants with probable AD with clinically significant agitation/aggression:	
• Cohen-Mansfield Agitation Inventory community version (CMAI-C) total score	• Change from baseline to Day 43 on the total CMAI-C score.
• Sleep symptoms as assessed by the Sleep Disorder Inventory (SDI).	• Change from baseline to Day 43 on SDI total score.
Secondary –Pharmacokinetics	
• To investigate the plasma pharmacokinetics (PK) of seltorexant and its M12 metabolite in participants with probable AD with clinically significant agitation/aggression.	• Plasma concentrations of seltorexant and its M12 metabolite at Day 15 and Day 43.

Objectives	Endpoints
Exploratory –Digital	
To investigate the effect of seltorexant versus placebo on the following in participants with probable AD with clinically significant agitation/aggression:	
• Sleep and activity metrics using an actigraphy device.	• Change from baseline on sleep and activity metrics.
• Sleep and mobility using an off-body touchless sensor in a subset of participants, (if feasible and available)	• Change from baseline on sleep and mobility measurements.
Exploratory –Biomarkers	
• To investigate the effect of seltorexant versus placebo in participants with probable AD with clinically significant agitation/aggression on blood biomarkers including but not limited to: o biomarkers of AD, eg, p217+tau, $A\beta_{1-42/40}$ levels and ratio, NfL, and other exploratory biomarkers	• Change from baseline in levels of blood biomarkers of AD
Other Exploratory	
To investigate the effect of seltorexant versus placebo on the following in participants with probable AD with clinically significant agitation/aggression:	
• Variables derived from the NPI-12/NPI-C rating: o The sum of the NPI-C delusion and hallucination domain scores (NPI-C D+H) o Individual NPI-C domain scores for agitation, aggression, dysphoria, delusions, and hallucinations o NPI-12 total score o NPI-12 Caregiver Distress total score o All NPI-12 individual item scores	• Change from baseline over time on the specified NPI-12/NPI-C variables.
• SDI Caregiver Distress score	• Change from baseline over time in SDI caregiver distress
• Short-term rebound effects on behavioral and psychological symptoms of dementia (BPSD) as well as withdrawal symptoms	• Change in Physician Withdrawal Checklist (PWC-20) from End of DB Treatment/EW to first follow up visit. • Change in NPI-12, NPI-C and SDI scores from End of DB Treatment/EW to first follow up visit.
• CMAI-C factor scores	• Change from baseline over time on CMAI-C factor scores.
• Depressive symptoms as assessed by the Geriatric Depression Scale, Informant version (GDSI).	• Change from baseline on GDSI Total score at Day 43.

Objectives	Endpoints
• The association between treatment effect of seltorexant in sleep and treatment effect in agitation/aggression	• Correlation between change in NPI-C A+A and change in SDI total score, and mediation effect on change in NPI-C A+A through change in SDI total score
• Cognitive symptoms	• Change from baseline to Day 43 on Alzheimer's Disease Assessment Scale Cognitive subscale 14-item version (ADAS-Cog-14) composite cognitive score and from screening to Day 43 on the Mini-Mental State Examination (MMSE) total score.
Safety	
To investigate the effect of seltorexant versus placebo on the following in participants with probable AD with clinically significant agitation/aggression:	
• Safety and tolerability of seltorexant	• Change from baseline in Vital signs, Clinical Labs (chemistry, hematology, urinalysis), ECG, C-SSRS, and Treatment Emergent Adverse Events (TEAEs) including AEs of special interest (AESI).

Statistical hypotheses pertaining to the primary endpoint are described in Section 2.

Descriptive analysis of pharmacokinetic data is described in section 5.7.1. Details about further analysis of biomarker and pharmacokinetic data pertaining to the additional exploratory objectives will be described in their respective analysis plans separately.

1.2. Study Design

This is a multicenter, DB randomized, parallel-group, placebo-controlled, 6-week study to assess the efficacy and safety of seltorexant 20mg in participants aged 55 to 85 years, inclusive, with probable AD with clinically significant agitation/aggression.

This study will consist of 3 phases: a screening phase (up to 28 days), a double-blind treatment phase (43 days), and a post-treatment follow-up phase (14 days after DB treatment phase).

Approximately 86 participants with probable AD and clinically significant agitation/aggression will be randomly assigned to receive placebo or seltorexant 20 mg in a 1:1 ratio for 6 weeks in the DB treatment phase. Randomization will be stratified by baseline NPI-12 sleep domain score (<4, ≥4) and community dwelling/assisted living.

Participants who discontinue early from study intervention in the DB treatment phase will be further assessed during additional follow-up visits every 2 weeks until Day 57.

Participants will take their assigned study intervention at home, once daily at bedtime during DB treatment phase.

2. STATISTICAL HYPOTHESES

This study is designed to show that the treatment effect in improving agitation and aggression (as measured by change from baseline on Day 43 in NPI-C A+A score) of seltorexant 20 mg is superior to placebo in participants with probable AD with clinically significant agitation/aggression.

If μ_T is the mean change in NPI-C A+A total score for seltorexant group and μ_P is the mean change in NPI-C A+A total score for the placebo group, then the hypothesis can be written as follows:

H_0: $\mu_T - \mu_P \geq 0$ vs.
H_1: $\mu_T - \mu_P < 0$

Superiority can be concluded if the two-sided p-value for the testing of the hypothesis above is less than 0.20.

3. SAMPLE SIZE DETERMINATION

Assuming a treatment difference of 5 points in change from baseline in NPI-C A+A score between seltorexant 20 mg and placebo, a standard deviation of 10 (effect size of 0.5), and a 15% drop-out rate during the 6-week double-blind phase, 86 participants (randomized in 1:1 ratio to placebo and seltorexant 20 mg) will need to be enrolled in the double-blind phase, in order to achieve 80% power to detect treatment difference, at one-sided 0.1 significance level.

4. POPULATIONS (ANALYSIS SETS) FOR ANALYSIS

Analysis Sets	**Description**
All Enrolled	All participants who sign the ICF
All Randomized	All participants who were randomized in the study.
Full	All randomized participants who take at least 1 dose of study intervention
Follow-up	All randomized participants who provide data in the follow-up visit(s).
Safety	All randomized participants who take at least 1 dose of study intervention

5. STATISTICAL ANALYSES

5.1. General Considerations

5.1.1. Visit Windows

As participants do not always adhere to the protocol visit schedule, the following rules are applied to assign actual visits to analysis visits. Listed below are the visit windows and the target days for each visit. The reference day is Study Day 1. If a participant has 2 or more actual visits in 1 visit window, the visit closest to the target day will be used as the protocol visit for that visit window. The other additional visit(s) will not be used in the summaries or

analyses, but they can be used for determination of clinically important endpoints. If 2 actual visits are equidistant from the target day within a visit window, the later visit is used.

All assignments will be made in chronological order. Once a visit date is assigned to a visit window, it will no longer be used for a later time point except for the endpoint. Listed below (Table 1) are the analysis visit windows and the target days for each visit defined in the protocol.

Table 1– Visit Windows

Parameter	Analysis Phase	Scheduled Visit Number	Time Interval (label on output)	Time Interval (Day)*	Target Time Point (Day)*
NPI-12 SDI C-SSRS	SCR	1	Screening	<1	
	DB	2	Baseline (DB)	<=1	1
		4	Day 15	2 to 22	15
		5	Day 29	23 to 36	29
		6	Day 43	37 to End of DB	43
		DB final visit	Endpoint (DB)	2 to End of DB	
	FU (DB)	7	Day 57	End of DB + 1 to End of FU	10
			Day 15 EW	2 to 22	15
			Day 29 EW	23 to 36	29
			Day 43 EW	37 to 48	43
			Day 57 EW	49 to End of FU	57
NPI-C	DB	2	Baseline (DB)	<=1	1
		4	Day 15	2 to 22	15
		5	Day 29	23 to 36	29
		6	Day 43	37 to End of DB	43
		DB final	Endpoint (DB)	2 to End of DB	
	FU (DB)	7	Day 57	End of DB + 1 to End of FU	10
			Day 15 EW	2 to 22	15
			Day 29 EW	23 to 36	29
			Day 43 EW	37 to 48	43
			Day 57 EW	49 to End of FU	57
CMAI-C ADAS-Cog-14	DB	2	Baseline (DB)	<=1	1
		4	Day 15	2 to 29	15
		6	Day 43	30 to End of DB	43
		DB final	Endpoint (DB)	2 to End of DB	
GDSI	DB	2	Baseline (DB)	<=1	1
		6	Day 43	2 to End of DB	43
		DB final visit	Endpoint (DB)	2 to End of DB	
MMSE	SCR	1	Screening	<1	
	DB	2	Baseline (DB)	<=1	
		6	Day 43	2 to End of DB	43
		DB final visit	Endpoint (DB)	2 to End of DB	
Weight Labs	SCR	1	Screening	<1	
	DB	2	Baseline (DB)	<=1	1
		DB final visit	Endpoint (DB)	2 to End of DB	
Vital Signs	SCR	1	Screening	<1	
	DB	2	Baseline (DB)	<=1	1
		4	Day 15	2 to 29	15
		6	Day 43	30 to End of DB	43

		DB final visit	Endpoint (DB)	2 to End of DB	

Parameter	Analysis Phase	Scheduled Visit Number	Time Interval (label on output)	Time Interval (Day)*	Target Time Point (Day)*
	FU (DB)	7	Day 57	End of DB + 1 to End of FU	10
			Day 15 EW	2 to 29	15
			Day 43 EW	30 to 48	43
			Day 57 EW	49 to End of FU	57
12-lead ECG	SCR	1	Screening	<1	
	DB	2	Baseline (DB)	<=1	1
		6	Day 43	2 to End of DB	43
		DB final visit	Endpoint (DB)	2 to End of DB	
PWC-20	DB	DB final visit	Endpoint (DB)	2 to End of DB	43
	FU (DB)	7	Day 57	End of DB + 1 to End of FU	10
Actigraphy	SCR		Screening	<1	
	DB		Baseline (DB)	-7 to -1	
			Week 1	1 to 7	
			Week 2	8 to 14	
			Week 3	15 to 21	
			Week 4	22 to 28	
			Week 5	29 to 35	
			Week 6	36 to End of DB	

* Relative to the first day of the respective phases for DB and FU (DB) phase; Assignment of visits to the FU (DB) will be made to the assessments that were recorded after double-blind treatment stop date. Participants who discontinue DB prior to Day 35 will have additional follow-up visits performed every 2 weeks until Day 50-57 - the follow-up data from these participants are assigned relative to the first day of DB phase with visit labels ending with EW.

5.1.2. Analysis Phases

Double-Blind Analysis Phase

The analysis reference start date of the double-blind analysis phase is the date of the first dose of double-blind study intervention. The analysis reference end date of the double-blind analysis phase is the maximum of the date of the last visit in the double-blind phase and date of completion or early withdrawal from the double-blind phase. For randomized participants who did not receive any study intervention in the double-blind phase, both analysis reference start and end dates are missing for the double-blind analysis phase.

Assignment of adverse events to double-blind analysis phase is described separately in Section 5.6.2.

Follow-up Phase

Start and end dates for the follow-up phase are only defined for participants who continued into the follow-up phase. The analysis reference start date of the follow-up analysis phase is the day after the reference end date for the double-blind analysis phase. The analysis reference end date of the follow-up analysis phase is the maximum of the last follow-up visit date or the study disposition date.

5.1.3. Imputation of Efficacy

Imputation method for missing data will include the following methods (Table 2).

Table 2: Imputation of Missing Efficacy Data

Imputation	Method/Rule
Multiple Imputation (MI) method	Copy Reference

5.2. Participant Dispositions

The number of screened participants and reason for screen failures will be summarized overall.

The number of participants in the following disposition categories will be summarized throughout the study by intervention group and overall for the all randomized, full and follow-up analysis sets pertinent to each phase:

- Participants randomized
- Participants who received study intervention
- Participants who completed, discontinued and reasons for discontinuation of study intervention during double-blind phase
- Participants who completed, terminated and reasons for termination from the study
- Intercurrent events in the double-blind phase

The distribution of the time to discontinuation of study intervention in double-blind phase will be displayed with Kaplan-Meier curves. A participant who discontinues study intervention at any time will be considered as an event, and the date of study intervention discontinuation will be used in the time to event calculation. A participant who completes study intervention will be censored and the date of last dose of study intervention will serve as the time of censoring.

Listings of participants will be provided for the following categories:

- Participants who discontinued double-blind study intervention
- Participants who were unblinded during the study period
- Participants who were randomized yet did not receive study intervention.

Additional analyses of disposition data for assessing and mitigating the impact of COVID-19 on study outcome are presented in Section 6.9**.**

5.3. Primary Endpoint(s) Analysis

The primary analysis will be based on the full analysis set using the NPI-C A+A score.

5.3.1. Definition

The primary efficacy endpoint is the change in NPI-C A+A score from baseline (DB) to Day 43.

The Neuropsychiatric Inventory-Clinician Rating (NPI-C) (de Medeiros 2010) is an instrument that

expands upon the items and 12 symptom domains of the original NPI. The modified rating methodology of the NPI-C incorporates the expert clinician's impressions to the data provided by the patients and caregivers. For each domain, the domain score will be calculated as the sum of clinician severity ratings (from 0=does not occur to 3=severe) for each item. NPI-C A+A score is the sum of the domain scores from the Agitation and Aggression domains at a given time point, and ranges from 0 to 63. Higher scores represent a more serious condition.

Negative changes in NPI-C A+A represent improvement.

Two estimands are defined for the primary endpoint: The primary estimand is Estimand 1, and the supplementary estimand is Estimand 2.

The sections below describe the primary analyses performed for each primary and supplementary estimand. For treatment policy strategy, data collected after participants experienced intercurrent events will be included in analyses; for hypothetical strategy, data collected after participants experienced an intercurrent event of treatment switch will not be included in analyses. Otherwise, data observed up to and including DB disposition day will be included.

5.3.2. Estimand 1

Estimand 1 will be the **primary estimand**. The analysis will be performed on the full analysis set.

Primary Trial Objective: To evaluate the efficacy of seltorexant 20 mg compared with placebo in improving agitation/aggression symptoms in participants with probable AD with clinically significant agitation/aggression.

Estimand Scientific Question of Interest: What is the benefit in improving agitation/aggression symptoms from seltorexant 20 mg versus placebo in adults and elderly participants with probable AD with clinically significant agitation/aggression based on the change from baseline in NPI-C A+A score if the participants take study intervention as directed?

Estimand 1 is defined by the following 5 components:

Study Intervention:

- Experimental: Seltorexant dose of 20 mg
- Control: Placebo

Population: Participants with probable AD with clinically signification agitation/aggression, as reflected by the inclusion/exclusion criteria.

Variable: Change in NPI-C A+A total score from baseline to Day 43

Summary Measure: Difference in intervention means

Intercurrent events and their corresponding strategies:

Intercurrent Events	Name of Strategy for Addressing Intercurrent Events and Its Description
1. Treatment discontinuation of study drug without switching	Hypothetical strategy: as if the intercurrent event had not occurred
2. Switch of treatment	Hypothetical strategy: see above
3. Allowed rescue medication (including use after treatment discontinuation, but prior to DB disposition)	Treatment policy strategy: all observed values of the endpoint are used regardless of whether or not the participant had experienced this intercurrent event

5.3.2.1. Analysis Methods

5.3.2.1.1. Primary Analysis

Descriptive statistics of the actual values and the change from baseline (DB) to each postbaseline time point will be presented for NPI-C A+A score by intervention group.

NPI-C A+A score will be analyzed by a Mixed-Effect Model for Repeated Measures (MMRM) based on observed case. The fixed terms included in the model will be intervention group, baseline NPI Sleep domain score category (per eDC), time, and time-by-intervention interaction, and baseline (DB) NPI-C A+A total score as a covariate. The within-subject covariance between visits will be estimated via an unstructured variance-covariance matrix. In case of convergence problems, alternative variance-covariance structures will be tried in the following order, with the first structure that converges being used in the analysis: heterogeneous Toeplitz, standard Toeplitz, and AR(1) with separate subject random effect; in addition, a robust sandwich variance estimator will be used to address the potential impact of covariance matrix misspecification on the estimation and testing of treatment effect. The Kenward-Roger method will be used for approximating the denominator degrees of freedom. Comparison between seltorexant and placebo at Day 43 will be performed using the appropriate contrast. Difference in least square means and 2-sided 80% CI will be presented.

Least squares mean changes from baseline(DB) (+/- SE) will be presented graphically over time.

No multiplicity adjustment will be included for efficacy endpoints.

5.3.3. Estimand 2

Estimand 2 will be a **supplementary estimand**. The analysis will be performed on the full analysis set.

All components as described under Estimand 1 (Section 5.3.2) apply to Estimand 2 except for the strategy for addressing intercurrent events.

Intercurrent events and their corresponding strategies:

Intercurrent Events	Name of Strategy for Addressing Intercurrent Events and Its Description
1. Treatment discontinuation of study drug without switching	Treatment policy strategy: all observed values of the endpoint are used regardless of whether or not the participant had experienced this intercurrent event
2. Switch of treatment	Hypothetical strategy: as if the intercurrent event had not occurred
3. Allowed rescue medication (including use after treatment discontinuation, but prior to Day 43)	Treatment policy strategy: all observed values of the endpoint are used regardless of whether or not the participant had experienced this intercurrent event

5.3.3.1. Analysis Methods

5.3.3.1.1. Primary Analysis

The primary analysis for Estimand 2 will include implementation of Copy Reference (CR) MI method.

Step 1 – Multiple Imputation (MI)

If there are participants with a non-monotone missing data pattern, datasets with only monotone missing data patterns will be created first by imputing the intermediate missing values using methods such as the MCMC (Markov Chain Monte Carlo) method. 500 imputations will be performed to create 500 unique datasets which now have monotone missing (ie., missing data after the participant experienced an intercurrent event) data pattern.

Analysis assumptions: MAR is assumed for intermediate missing data. Missing not at random (MNAR) is assumed for monotone missing (i.e., data that is deemed missing after the participant experienced an intercurrent event under hypothetical strategy, or missing data after the occurrence of an intercurrent event under treatment policy) in the experimental intervention group, where efficacy scores are assumed as if participant had always been in the control group. MAR is assumed for missing data in the control group.

Step 2 – Analysis

The same MMRM analysis as described for the primary efficacy analysis (Section 5.3.2.1.1) will be performed for each set of the adjusted fully imputed datasets.

Step 3 – Pooling

Rubin's methodology will be applied to the MMRM results from the 500 imputed datasets to produce final inferences (Rubin, D. 1987).

The efficacy data that is either missing or not used after the intercurrent event at a given timepoint

will be imputed using the imputation model of the control group, i.e., conditional on the data observed or imputed at previous timepoints relative to the mean of the model for the control group.

This approach does not assume a sustained benefit of experimental intervention for the efficacy data that is either missing or not used after the intercurrent event and uses an imputation method that is based on the control group distribution and the estimated correlations between time points in the control group.

Descriptive statistics of the actual values and the change from baseline(DB) to each postbaseline time point will be presented for NPI-C A+A total score by intervention group. Under the treatment policy strategy, data from participants who discontinued the study drug only (intercurrent event 1 under this estimand) and provided additional follow-up data every 2 weeks until Day 43 will be included in this analysis.
Least squares mean changes from baseline(DB) (+/- SE) will be presented graphically over time.

5.3.4. Sensitivity Analysis Related to Mis-stratifications

Sensitivity analyses to evaluate the impact of using per-eDC values of stratification factors rather than per-IWRS values on the primary endpoint will be performed. The analyses described in sections 5.3.2 and 5.3.3 will be performed using the per-IWRS values of the NPI Sleep domain score category instead of the per-eDC values. Difference in least square means and 2-sided 80% CI will be presented.

5.4. Secondary Endpoint(s) Analysis

5.4.1. Secondary Efficacy Endpoint(s)

Similar to the primary endpoint, the primary analysis of the secondary efficacy endpoints will be based on the full analysis set.

5.4.1.1. Cohen-Mansfield Agitation Inventory —Community Version (CMAI-C)

5.4.1.1.1. Definition

The first secondary endpoint is the change in CMAI-C from baseline(DB) to Day 43.

The CMAI-C is a 37-item scale that measures the ability of a drug to reduce overall frequency of agitation symptoms, including aggressive behaviors (Cohen-Mansfield 1991). Individual items are rated by an expert clinician on a scale of 1 to 7 in which a score of 7 represents the most frequent for each item assessed. CMAI-C total score is the sum of the first 36 individual item responses.

Negative changes in CMAI-C total score indicate improvement.

5.4.1.1.2. Analysis Methods

The change from baseline in CMAI-C total score will be analyzed using the same MMRM model as described for the NPI-C A+A score, with baseline CMAI-C total score as the covariate. Comparison between seltorexant and placebo at Day 43 will be performed using the appropriate contrast. Difference in least square means and 2-sided 80% CI will be presented.

Descriptive statistics of the actual values and the change from baseline(DB) to each postbaseline time point will be presented for CMAI-C total score by intervention group.

Least squares mean changes from baseline(DB) (+/- SE) will be presented graphically over time.

5.4.1.2. Sleep Disorder Inventory (SDI)

5.4.1.2.1. Definition

The Sleep Disorders Inventory (SDI) is based on a caregiver interview and an expanded version of one item of the NPI-12 (Tractenberg 2003). It describes the frequency, severity, and caregiver burden of sleep-disturbed behaviors during a period prior to its administration.

The SDI was created by expanding item 11 of the 12-item NPI. According to the basic structure of the NPI-12, the respondent is asked a 'screening' question, a general indication of whether or not symptoms in that particular behavioral area are present. If the screening question is positive, then specific sub-questions are asked. The SDI consists of the seven sub-questions from the NPI-12 sleep disturbance item. Each of the sub-questions was made into a separate question with frequency, severity, and caregiver distress rated by the caregiver with respect to the participant.
Thus, in contrast to a single rating for frequency and severity for all sleep disturbance-related behaviors, which would be incorporated into an overall NPI-12 score, the SDI score is derived after the caregiver rates the frequency and severity of each of the seven separate sleep disturbance symptoms. The SDI Average Total score is calculated by multiplying the average frequency score across all items by the average severity score across all items. The SDI Summed Product score is the calculated by summing the product of frequency and severity scores for each item.

5.4.1.2.2. Analysis Methods

The change from baseline in SDI average total score and SDI summed product score will be analyzed using the same MMRM model as described for the NPI-C A+A score, with baseline SDI Average Total score and baseline SDI Summed Product score, respectively, as the covariate. Comparison between seltorexant and placebo at Day 43 will be performed using the appropriate contrast. Difference in least square means and 2-sided 80% CI will be presented.

Descriptive statistics of the actual values and the change from baseline(DB) to each postbaseline time point will be presented for SDI average total score and SDI summed product score by intervention group.

Least squares mean changes from baseline(DB) (+/- SE) will be presented graphically over time.

5.5. Exploratory Endpoint(s) Analysis

Exploratory analyses will be performed for the endpoints described in the sections below for the full analysis set.

5.5.1. Variables Derived from NPI-12

5.5.1.1. Definition

The NPI-12 is a measure of psychobehavioral disturbances (Cummings 1994; 1997), assessing the

frequency and severity of disturbances in 12 domains, based on a caregiver interview. An answer of "No" on the domain screening question indicates that the patient does not exhibit symptoms in the domain and will result in a domain score of 0. Otherwise, frequency of each domain is rated on a 4-point scale (from 1=rarely to 4=very often) and severity on a 3-point scale (from 1=mild to 3=severe), with the score for each domain being the product of the frequency and severity scores, such that each domain is scored from 1 to 12. The NPI-12 total score is the sum of all individual NPI-12 domain scores.

For each domain, there is also an assessment of caregiver burden. Caregiver distress, based on the response to the question "how emotionally distressing do you find this behavior?", is rated on a 6-point scale (from 0 = not at all to 5 = very severely or extremely), with the total score for caregiver distress ranging from 0 (best) to 60 (worst).

The following variables derived from the NPI-12 will be analyzed:

- NPI-12 total score
- NPI-12 Caregiver Distress total score
- All NPI-12 individual item scores

5.5.1.2. Analysis

The change from baseline in NPI-12 total score will be analyzed using the same MMRM model as described for the NPI-C A+A score, with baseline NPI-12 total score as the covariate. Comparison between seltorexant and placebo at Day 43 will be performed using the appropriate contrast. Difference in least square means and 2-sided 80% CI will be presented.

Descriptive statistics of the actual values and the change from baseline(DB) to each postbaseline time point will be presented for NPI-12 total score by intervention group.

Least squares mean changes from baseline(DB) (+/- SE) will be presented graphically over time.

Descriptive statistics of the actual values and the change from baseline(DB) to each postbaseline time point will be presented for NPI-12 Caregiver Distress total score by intervention group.

Descriptive statistics of the actual values and the change from baseline(DB) to each postbaseline time point will be presented for each NPI-12 individual score by intervention group.

5.5.2. Variables Derived from NPI-C

5.5.2.1. Definition

The following variables derived from the NPI-C will be analyzed:

- The sum of the NPI-C delusion and hallucination domain scores (NPI-C D+H)
- Individual NPI-C domain scores for agitation, aggression, dysphoria, delusions, and hallucinations

5.5.2.2. Analysis

The change from baseline in NPI-C D+H score will be analyzed using the same MMRM model as described for the NPI-C A+A score, with baseline NPI-C D+H score as the covariate. Comparison between seltorexant and placebo at Day 43 will be performed using the appropriate contrast. Difference in least square means and 2-sided 80% CI will be presented.

Descriptive statistics of the actual values and the change from baseline(DB) to each postbaseline time point will be presented for NPI-C D+H total score by intervention group.

Least squares mean changes from baseline(DB) (+/- SE) will be presented graphically over time.

The change from baseline in the specified individual NPI-C domain scores will be analyzed using the same MMRM model as described for the NPI-C A+A score, with the corresponding baseline NPI-C domain score as the covariate.

Descriptive statistics of the actual values and the change from baseline(DB) to each postbaseline time point will be presented for the specified individual NPI-C domain scores by intervention group.

5.5.3. SDI Caregiver Distress Score

5.5.3.1. Definition

Caregiver distress is collected for each item in the SDI, as defined in Section 5.4.1.2.1. The total caregiver distress score is the sum of the distress score for each item.

5.5.3.2. Analysis

Descriptive statistics of the actual values and the change from baseline(DB) to each postbaseline time point will be presented for the SDI Caregiver Distress score by intervention group.

5.5.4. Patient Withdrawal Checklist (20 items; PWC-20)

5.5.4.1. Definition

The PWC-20 is a reliable and sensitive instrument for the assessment of discontinuation symptoms (Rickels K, Garcia-Espana R, Mandos LA, Case GW. Physician Withdrawal Checklist (PWC-20). J Clin Psychopharmacol. 2008;28(4):447-451.). The PWC-20 is a simple and accurate method used to assess potential withdrawal symptoms following cessation of study intervention.

5.5.4.2. Analysis

The proportion of participants with withdrawal symptoms at the end of DB phase and follow-up visits will be presented by intervention group.

In addition, symptoms at follow-up will be compared to the end of therapy visit for double-blind phase and will be summarized using the following categories: new or worsened symptoms, symptoms present and unchanged, no symptoms, and improved.

Additionally, PWC-20 total score will be computed, and summary statistics of the total score at each visit will be provided.

Bar chart of the incidence of withdrawal symptoms (i.e, severity of mild or worse) over time for all 20 items will be provided for each item.

5.5.5. Short-term rebound effects on behavioral and psychological symptoms of dementia (BPSD) as well as withdrawal symptoms

5.5.5.1. Definition

The NPI-12, NPI-C and SDI will be used to monitor changes in symptoms between withdrawal and the first follow up visit.

5.5.5.2. Analysis

The baseline(FU) is defined as the value at the end of DB phase for participants who have entered the follow-up phase.

Descriptive statistics of the actual values and the change from baseline (FU) to the first follow up visit will be presented for the NPI-12 total score, NPI-C A+A score, NPI-C D+H score, NPI-C Dysphoria score, SDI average total score, and SDI summed product score by intervention group.

5.5.6. CMAI-C Factor Scores

5.5.6.1. Definition

CMAI-C factors are derived from the original CMAI-C scale. Each item is classified as a physically nonaggressive behavior (PNAB), physically aggressive behavior (PAGB), verbally nonaggressive behavior (VNAB), or verbally aggressive behavior (VAGB). The CMAI-C factor score for each factor is calculated by summing the individual scores for each item within the factor.

- Physically nonaggressive behaviors (PNAB): General restlessness, repetitious mannerisms, pacing, trying to get to a different place, handling things inappropriately, hiding, and inappropriate dressing or undressing; scale items 14, 15, 16, 17, 18, 19, 21, 22
- Physically aggressive behaviors (PAGB): Hitting, pushing, scratching, grabbing things, and grabbing people (kicking, biting); scale items 10, 13, 20, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35
- Verbally nonaggressive behaviors (VNAB): Negativism, doesn't like anything, constant requests for attention, verbal bossiness, complaining or whining, relevant interruptions, irrelevant interruptions, and repeating sentences; scale items 1, 2, 3, 6, 7, 8, 11
- Verbally aggressive behaviors (VAGB): Screaming, cursing, temper outbursts, and making strange noises; scale items 4, 5, 9, 12, 23

5.5.6.2. Analysis

Descriptive statistics of the actual values and the change from baseline(DB) to each post-baseline time point will be presented for each CMAI-C factor by intervention group.

5.5.7. Geriatric Depression Scale, Informant Version (GDSI)

5.5.7.1. Definition

The GDSI is a 30- or 15-item self-report assessment used to identify depressive symptomatology in the elderly. GDSI is a version developed to be completed based on the informant (a study partner for the purpose of the study) rating of the participant's symptoms. The GDSI consists of the same 15 items as the original GDS 15-item self-report version.

5.5.7.2. Analysis

Descriptive statistics of the patient-reported GDSI total score at baseline(DB) and endpoint (DB), as well as the change from baseline to Endpoint (DB), will be presented by intervention group.

An ANCOVA model will be used to test the difference of change from baseline at endpoint(DB) in GDSI total score between seltorexant and placebo. The model will include intervention and baseline NPI Sleep domain score category (per eDC) as factors, and baseline(DB) GDSI total score value as a covariate. Difference of least square means and 2-sided 80% CI will be presented.

5.5.8. The association between treatment effect of seltorexant in sleep and treatment effect in agitation/aggression

5.5.8.1. Analysis

Correlations between change in NPI-C A+A score and change in SDI average total score and between change in NPI-C A+A score and change in SDI summed product score over time will be explored; Pearson and Spearman correlation coefficients will be calculated. Scatter plot of change in scores of the 2 assessment scales over time will be presented for each intervention group.

A mediation analysis will be performed to examine the mediating role of improvement in sleep assessed by change from baseline in SDI score on change from baseline in NPI-C A+A score provided that both endpoints demonstrate a statistically significant difference between seltorexant and placebo. This analysis will assess the extent to which change in NPI-C A+A score may be mediated by or independent of change in SDI score based on observed data (i.e., data collected at each time point without carrying forward previous values) at Day 43. The analysis will consider both change in NPI-C A+A score and change in SDI score as continuous variables.

- Simulation-based Counterfactual Approach: this analysis introduced by Imai et. al obtains the natural direct effect and natural indirect effect using numerical simulations. The approach uses parametric bootstrapping to construct the point estimate and its uncertainty estimates for direct effect and indirect effect from the bootstrap sampling distribution. For each bootstrapped sample, change from baseline in NPI-C A+A score and change from baseline in SDI score will be analyzed. The ANCOVA analysis of change in NPI-C A+A score from baseline will include factors for treatment (placebo and seltorexant dose groups), change from baseline (Day 1, predose) in SDI score, treatment-by-change from baseline in SDI score

interaction if the interaction is significant, and baseline NPI-C A+A score as a covariate. The ANCOVA analysis of change in SDI score will include factors for treatment (placebo and seltorexant dose groups), and baseline NPI-C A+A score as a covariate. Estimate of the controlled direct effect will be obtained by plugging in the estimated coefficient values and the mean level of change in SDI score into the analytic expressions.

5.5.9. Cognitive Symptoms

5.5.9.1. Definition

The Alzheimer's Disease Assessment Scale Cognitive subscale 14-item version (ADAS-Cog-14) is a rater administered instrument that was designed to assess the severity of dysfunction in cognition characteristic of persons with AD (Rosen 1984). The ADAS-Cog-11 consists of 11 tasks measuring the disturbances of memory, language, praxis, attention, and other cognitive abilities, which are often referred to as the core symptoms of AD.

The modified ADAS-Cog 14-item scale includes all original ADAS-Cog items with the addition of a number cancellation task, Maze task and a delayed free recall task, for a total of 90 points, with higher scores indicative of worse cognitive performance.

The Mini-Mental State Examination (MMSE) is a 30-point questionnaire that is used extensively in clinical and research settings to measure cognitive impairment. It is commonly used in medicine and allied health to screen for dementia. The test is divided into two sections: the first section requires vocal responses and covers orientation, memory, and attention. The second part tests ability to name, follow verbal and written commands, write a sentence spontaneously, and copy a complex polygon similar to a Bender- Gestalt Figure. The score ranges from 0 (minimum score) to 30 (maximum score) and it is calculated by the sum of the sub-items scored 0 (incorrect answer) or 1 (correct answer) (Creavin 2016; Folstein 1975).

5.5.9.2. Analysis

The change from baseline in ADAS-Cog-14 total score will be analyzed using the same MMRM model as described for the NPI-C A+A score, with baseline ADAS-Cog-14 total score as the covariate. Comparison between seltorexant and placebo at Day 43 will be performed using the appropriate contrast. Difference in least square means and 2-sided 80% CI will be presented. Descriptive statistics of the actual values and the change from baseline(DB) to each postbaseline time point will be presented for ADAS-Cog-14 total score by intervention group.

Least squares mean changes from baseline(DB) (+/- SE) will be presented graphically over time.

The change from baseline in MMSE total score will be analyzed using the same ANCOVA model as described for the GDSI score, with baseline MMSE total score as the covariate. Comparison between seltorexant and placebo at Day 43 will be performed using the appropriate contrast. Difference in least square means and 2-sided 80% CI will be presented.

Descriptive statistics of the actual values and the change from baseline(DB) to each postbaseline time point will be presented for MMSE total score by intervention group.

5.5.10. Subgroup Analyses

For the subgroups listed in Section 5.7.4, subgroup analyses using MMRM will be performed for the change in NPI-C A+A total score at Day 43. The fixed terms in the model will be intervention group, baseline NPI-12 sleep domain score category (<4, ≥4) per eDC, time, subgroup, time-by-intervention interaction, intervention-by-subgroup interaction, and time-by-intervention-by-subgroup interaction as factors, and baseline(DB) NPI-C A+A total score as a covariate. Point estimate of the treatment difference and 2-sided 80% CI will be estimated using appropriate contrasts.

The analysis results (difference of LS means and 80% CI) of the different subgroups will be displayed in a forest plot.

Additionally, descriptive statistics of the actual values and the change from baseline(DB) to each postbaseline time point will be presented for NPIC A+A total score by intervention group and subgroup.

Similar subgroup analyses will be performed for the change in CMAI-C score. The results will be displayed in a forest plot.

The analyses will be performed for the Full analysis set.

5.5.11. Change in Care Settings

A listing of participants who have changed their care settings (community dwelling or assisted living) during the DB phase will be provided for the Full analysis set.

5.6. Safety Analyses

All safety analyses will be based on safety analysis set.

For all continuous safety variables, descriptive statistics by intervention group will include the N, mean, standard deviation, median, minimum, and maximum. Categorical variables will be summarized by intervention group using frequency counts and percentages.

5.6.1. Extent of Exposure

The number and percentage of participants who receive study intervention will be summarized.

Descriptive statistics for duration of study intervention (N, mean, SD, median, and range (minimum, maximum)) will be summarized.
Duration of intervention will be summarized in the following duration categories: <=7 days, 8-14 days, 9-21 days, 22-28 days, 29-35 days, 36-42 days, >42 days.

Study intervention duration is defined as (date of last dose of study intervention – date of first dose of study intervention) +1. Number of doses is defined as the total number of dose administrations, which is the total number of days that study intervention was administered to the participant (excluding days off study intervention).

The analysis will be performed on the safety analysis set.

5.6.1.1. Intervention Compliance

Compliance will be summarized descriptively for each study intervention.

The percent compliance will be categorized and the number and percentage of participants in each category will be summarized by intervention group for study agent.

Compliance will be calculated for the study intervention as:

Compliance (%) = (number of days where the expected number of tablets were taken)/(days expected to be dosed (includes days where the study intervention was not taken)) x 100

The analysis will be performed on the safety analysis set.

5.6.2. Adverse Events

The verbatim terms used in the CRF by investigators to identify adverse events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). Any AE occurring at or after the initial administration of study intervention through the day of last dose plus 2 days is considered to be treatment emergent. AEs with onset date > date of last dose of double-blind study intervention plus 2 days will be assigned to follow-up phase. If the event occurs on the day of the initial administration of study intervention (and action taken is not entered as 'Not Applicable' in the database), then the event will be considered as treatment emergent. If the event date is recorded as partial or completely missing, then the event will be considered to be treatment emergent unless it is known to be prior to the first administration of study intervention based on partial onset date or resolution date. All reported adverse events will be included in the analysis. For each adverse event, the number and percentage of participants who experience at least 1 occurrence of the given event will be summarized by planned intervention group.

Summary tables will be provided for treatment-emergent adverse events:

- AEs (all AEs, and AEs with incidence of at least 5% in any treatment group)
- Serious AEs (SAEs)
- AEs leading to discontinuation of study intervention
- AEs by severity
- AEs by relationship to study intervention
- AEs of special interest (See Appendix 6.6 for list of adverse events in each category)

In addition to the summary tables, listings will be provided for participants who had:

- SAEs
- AEs leading to discontinuation of study intervention
- AEs of special interest

A listing of participants who died will be provided.

5.6.3. Additional Safety Assessments

5.6.3.1. Clinical Laboratory Tests

Descriptive statistics will be presented for all chemistry, hematology, and urinalysis laboratory tests at scheduled time points in the double-blind phase.

Change from baseline(DB) to all post-baseline visits in the double-blind phase will be summarized for chemistry, hematology, and urinalysis tests and displayed by study intervention group.

Shift in chemisty lab values from baseline to maximum post-baseline value will be summarized for the double-blind phase and displayed by study intervention group. Shift in Metabolic Lab values specifically from baseline to maximum post-baseline value will be summarized for the double-blind phase and displayed by study intervention group based on the following categories.

Glucose:

- from <100 mg/dL to [≥100 mg/dL - <126 mg/dL] (normal to borderline)
- from [≥100 mg/dL - <126 mg/dL] to ≥126 mg/dL (borderline to high)
- from <100 mg/dL to ≥126 mg/dL (normal to high)

Triglycerides:

- from <150 mg/dL to ≥200 mg/dL (normal to high/very high)
- from <150 mg/dL to ≥500 mg/dL (normal to very high)
- from [≥150 mg/dL - <200 mg/dL] to ≥200 mg/dL (borderline to high/very high)
- from [≥150 mg/dL - <200 mg/dL] to ≥500 mg/dL (borderline to very high)
- from [≥200 mg/dL - <500 mg/dL] to ≥500 mg/dL (high to very high)

Total Cholesterol

- from <200 mg/dL to ≥200 mg/dL (normal to borderline/high)
- from <200 mg/dL to ≥240 mg/dL (normal to high)
- from <200 mg/dL to [≥200 mg/dL - <240 mg/dL] (normal to borderline)
- from [≥200 mg/dL - <240 mg/dL] to ≥240 mg/dL (borderline to high)

HDL Cholesterol

- from ≥40 mg/dL to <40 mg/dL (normal to low)

LDL Cholesterol

- from < 100mg/dL to ≥160 mg/dL (normal to high)

Clinical laboratory test values in the double-blind phase will be considered “treatment-emergent markedly abnormal” (TEMA) using the criteria defined by the sponsor listed in Section 6.8.

- If the postbaseline value is above the upper limit of the markedly abnormal criteria and the baseline(DB) value is equal to or below the upper limit, then the postbaseline marked abnormality will be considered TE. The same applies to the postbaseline value being below the lower limit of the markedly abnormal criteria with the baseline value being equal to or above the lower limit of the markedly abnormal criteria.

- If the baseline(DB) value is missing, a postbaseline marked abnormality will always be considered as TE.

The number and percentage of participants with treatment-emergent markedly abnormal values in the double-blind phase will be presented by study intervention group.

The incidence of participants with treatment-emergent ALT values >3*upper normal limit (ULN) or AST value > 3*ULN will be presented for the double-blind phase. Additionally, incidence of treatment-emergent hepatic toxicity (suspected Hy's Law [U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). [2009]] cases) defined as (ALT values >3*ULN or AST values > 3*ULN) AND total bilirubin values >2*ULN will be presented for the double-blind phase. Similar to the markedly abnormal analysis, only participants with baseline(DB) values that don't meet the criteria or are missing will be eligible for these analyses.

A listing of participants with treatment emergent markedly abnormal laboratory values will be provided. A listing of participants with ALT > 3* ULN or AST values > 3*ULN) and participants with hepatic toxicity (suspected Hy's Law cases) will be provided.

5.6.3.2. Vital Signs

Continuous vital sign parameters including temperature, weight, pulse, blood pressure (systolic and diastolic), and Body Mass Index (BMI) will be summarized at each assessment time point. Body Mass Index will be calculated as weight (kg)/(height (m))^2, at each time point that body weight is measured. The height measurement collected at screening will be used in the calculation. Change from the baseline(DB) will be summarized for the double-blind phase. Descriptive statistics (mean, standard deviation, median, minimum and maximum) will be presented. Abnormality criteria (based on criteria defined below) will be applied to postbaseline values in the double-blind phase. Postbaseline values will be considered treatment-emergent if they meet both value and change criteria in the table below.

For criteria that do not include an increase or decrease from baseline(DB) for the double-blind phase:

- Treatment-emergent will be concluded if the postbaseline value is above the upper limit and the baseline(DB) value is below the upper limit (eg, Normal or Low). The same applies to the postbaseline value being below the lower limit with the baseline value being above the lower limit (eg, Normal or High).
    - If the baseline(DB) value is missing, a postbaseline abnormality will always be considered as TE.

Incidence of treatment-emergent clinically important vital signs during intervention, as defined in Table 4, will be summarized for participants who had a baseline(DB) assessment and at least 1 postbaseline assessment for that vital sign. A listing of participants with treatment-emergent clinically important abnormalities in vital signs will be presented, along with a listing of all vital sign measurements.

Table 4: Clinically Important Abnormalities in Vital Signs

Vital Sign	Abnormal Category	Criteria
Pulse	Abnormally high	[≥100] bpm and with [≥15] bpm increase from baseline
	Abnormally low	[≤50] bpm and with [≥15] bpm decrease from baseline
Systolic blood pressure	Abnormally high	[≥180] mm Hg and with [≥20] mm Hg increase from baseline
	Abnormally low	[≤90] mm Hg and with [≥20] mm Hg decrease from baseline
Diastolic blood pressure	Abnormally high	[≥105] mm Hg and with [≥15] mm Hg increase from baseline
	Abnormally low	[≤50] mm Hg and with [≥15] mm Hg decrease from baseline
Temperature	Abnormally high	[>37.5]ºC
	Abnormally low	[<35.5]ºC
Weight	Abnormally high	increase [≥7%] from baseline
	Abnormally low	decrease [≥7%] from baseline

5.6.3.3. Electrocardiogram

The ECG parameters that will be analyzed are heart rate, PR interval, QRS interval, QT interval, and corrected QT (QTc) interval using the following correction methods: Bazett's formula (QTcB), Fridericia's formula (QTcF).

Bazett's formula: QTcB (msec) = QT (msec) / (RR (msec)/1000)$^{1/2}$; if RR is missing, use QT (msec) * (HR(bpm)/60)$^{1/2}$;

Fridericia's formula: QTcF (msec) = QT (msec) / (RR (msec)/1000)$^{1/3}$; if RR is missing, use QT (msec) * (HR(bpm)/60)$^{1/3}$;

Baseline ECG is defined as the average of all ECG results collected up to and including the day of the first dose of study intervention.

The number and percentage of participants with QTc interval increases from average baseline to the maximum postbaseline value will be summarized at each time point for double-blind phase. Refer to the following Table 5 for summary categories.

Table 5: Criteria for Abnormal QTc Values and Changes From Baseline

QTc value (msec)	Normal QTc	≤450 for male, ≤470 for female
		>450 to ≤480 for male, >470 to ≤480 for female
		>480 to ≤500
		>500
Clinically significant QTc (msec)	No	≤500
	Yes	>500
QTc change from baseline (msec)	No concern	≤30
	Concern	>30-60
	Clear concern	>60

Descriptive statistics of ECG parameters and change from average baseline will be summarized at each scheduled time point for double-blind phase.

Abnormality criteria (based on criteria defined in Table 6 below) will be applied to average baseline and postbaseline values.

Postbaseline abnormalities will be compared with their corresponding average baseline result:

- Treatment-emergent will be concluded if the postbaseline value in the double-blind phase is above the upper limit and the average baseline value is below the upper limit (eg, Normal or Low). The same applies to the postbaseline value being below the lower limit with the average baseline value being above the lower limit (eg, Normal or High).
- If the average baseline value is missing, a postbaseline abnormality in the double-blind phase will always be considered as treatment-emergent.

The number and percentage of participants with treatment-emergent ECG values outside predefined limits will be presented by study intervention over time for DB phase of the study:

Table 6: Abnormal Limits for ECG Parameters

ECG Parameter	Outside of normal limit if ...	
	Abnormally low	Abnormally high
Heart Rate (bpm)	≤ 50 bpm	≥100 bpm
PR interval (msec)	≤ 120 msec	≥200 msec
QRS interval (msec)	≤ 60 msec	≥120 msec
QT interval (msec)	≤ 200 msec	≥500 msec
RR interval (msec)	≤ 600 msec	≥1200 msec

The interpretation of the ECGs as determined by a central reader will be displayed by the number and percentage of participants meeting the normality criteria. The interpretation will be summarized over time.

A listing of treatment emergent clinically relevant ECG abnormalities will also be provided.

5.6.3.4. Other Safety Parameters

5.6.3.4.1. Columbia Suicide Severity Rating Scale (C-SSRS)

The C-SSRS is a low-burden measure of the spectrum of suicidal ideation and behavior that was developed to assess severity and track suicidal events through any intervention (Posner K, Oquendo MA, Gould M, Stanley B, Davies M. Columbia Classification Algorithm of Suicide Assessment (C-CASA): classification of suicidal events in the FDA's pediatric suicidal risk analysis of antidepressants. Am J Psychiatry. 2007;164:1035-1043.). It is a clinical interview providing a summary of both suicidal ideation and behavior that can be administered during any evaluation or risk assessment to identify the level and type of suicidality present. The C-SSRS has been used frequently in clinical studies, and is a validated, standard measure for suicidal ideation assessment. Using the C-SSRS, the outcomes will be categorized using the scoring for the 11 categories:

Suicidal Ideation (1-5)	
1	Wish to be dead
2	Non-specific active suicidal thoughts
3	Active suicidal ideation with any methods (not plan) without intent to act
4	Active suicidal ideation with some intent to act, without specific plan
5	Active suicidal ideation with specific plan and intent
Suicidal Behavior (6-10)	
6	Preparatory acts or behavior

7	Aborted attempt
8	Interrupted attempt
9	Actual attempt
10	Suicide
Non-suicidal self-injurious behavior (11)	
11	Non-suicidal self-injurious behavior

At each time point, an event of suicidal ideation or behavior will be assigned a score of 1 to 10 based on the maximum response for the C-SSRS at that visit. If no events qualify for a score of 1 to 10, a score of 0 will be assigned (0="no suicidal ideation or behavior that can be assessed on the basis of C-SSRS"). A participant with an event of non-suicidal self-injurious behavior only will not be considered as having suicidal ideation or behavior; therefore, a score of 0 will be assigned. However, an additional score of 11 will be assigned to summarize any participants with an event of non-suicidal self-injurious behavior.

Shifts from baseline(DB) to the maximum score pertaining to suicidal ideation or suicidal behavior (i.e., scores 1 to 10) will be summarized by intervention for DB phase.

The maximum score (of scores 0 to 10) assigned to each participant will be grouped into one of three broad categories: No suicidal ideation or behavior (0), Suicidal ideation (1-5), Suicidal behavior (6-10). Shifts from baseline(DB) to the maximum category will be summarized by intervention for DB phase.

A frequency distribution of the scores for the 11 categories (0 to 10) will be provided by study intervention at each time point for DB phase. In addition, the number and proportion of participants with non-suicidal self-injurious behavior (a score of 11) will be provided by study intervention at each time point for DB phase.

5.7. Other Analyses

5.7.1. Pharmacokinetics

PK analyses will be addressed in a separate plan.

5.7.2. Actigraphy

Objective sleep parameters are measured using Actigraphy.

The device, Phillips Actiwatch Spectrum®, is a wristwatch actigraphy system that provides accurate and objective activity, sleep, wake, and light-exposure data. Subjects will wear the device continuously during the screening and double-blind phases. Data will be downloaded and transmitted electronically to a study data file during study visits.

5.7.2.1. Analysis

The following parameters will be used for the analysis:

Parameter	Description
Total sleep time (min)	Number of scored sleep epochs, in minutes.

Parameter	Description
Number of wake bouts	A wake bout is an epoch or continuous epochs scored as awake
Sleep efficiency (%)	Scored total sleep time divided by total time in bed minus total invalid time, multiplied by 100.
Sleep onset latency (min)	Time between the start of a given rest interval and the following Sleep Start Time, in minutes.
Wake after sleep onset (min)	Number of epochs of the given sleep Interval scored as WAKE by the actigraphy software.
24 hour mean motor activity (counts/min)	The average of all valid physical activity counts for all epochs in a 24 hour period divided by the epoch length in minutes.
Daytime mean motor activity (counts/min)	The average of all valid physical activity counts for all epochs scored as ACTIVE divided by the epoch length in minutes.
Notes: An epoch is 30 to 60 seconds.	

The continuously collected data will be grouped into weekly intervals in the double-blind phase. Descriptive statistics of the mean weekly data (calculated based on non-missing daily data) and the change from baseline to each postbaseline time point in the double-blind phase will be presented by intervention group for the parameters listed in the table.

The change from baseline(DB) in each of the parameters will be analyzed using the same MMRM model as described for the primary endpoint with the respective baseline(DB) value as the covariate. The correlation between actigraphy parameters and the primary efficacy endpoint, SDI summary scores, and CMAI-C scores will be explored. Pearson and Spearman correlation coefficients will be calculated. Scatter plot of change in scores of the 2 assessment scales over time will be presented for each intervention group.

5.7.3. Biomarkers and Off-Body Sensor Data

Analysis of biomarker data will be discussed in a separate SAP. Analysis of off-body sensor data will be included in a separate plan.

5.7.4. Definition of Subgroups

Subgroup analyses of the primary endpoint will be performed for the full analysis set for the following subgroups:

Subgroup	Definition
Sex	Male Female Other
Age Group	Adult (<65 years) Elderly (≥ 65 years)
Age Group II	<65 years 65-<75 years ≥75 years
BMI	underweight (<18.5 kg/m2) normal (18.5-<25 kg/m2) overweight (25-<30 kg/m2) obese (≥30 kg/m2)
Baseline NPI-12 sleep domain score (per eDC)	<4 ≥4

6. SUPPORTING DOCUMENTATION

6.1. Appendix 1 List of Abbreviations

AD	Alzheimer's Disease
ADAS-Cog-14	Alzheimer's Disease Assessment Scale Cognitive subscale 14-item version
AE	adverse event
AESI	adverse event(s) of special interest
ALT/SGPT	alanine aminotransferase
ANCOVA	analysis of covariance
AST/SGOT	aspartate aminotransferase
ATC	anatomic and therapeutic class
BMI	body mass index
BPSD	Behavioral and Psychological Symptoms of Dementia
C-SSRS	Columbia Suicide Severity Rating Scale
CDR	Clinical Depression Rating
CI	confidence interval
CIR	copy increment from reference
CMAI-C	Cohen-Mansfield Agitation Inventory community version
CR	copy reference
CRF	case report form
CV	coefficient of variation
DB	double-blind
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, 5th edition
ECG	electrocardiogram
eCOA	Electronic clinical outcome assessment
eDC	Electronic data capture
EW	Early Withdrawal
GDSI	Geriatric Depression Scale, Informant version
IPA	International Psychogeriatric Association
IQ	interquartile
IS	insomnia symptoms
IWRS	interactive web response system
LS	least squares
MAR	missing at random
MCMC	Markov Chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
MI	multiple imputation
MMRM	mixed model for repeated measures
MMSE	Mini-Mental State Exam
MNAR	missing not at random
NPI-12	Neuropsychiatric Inventory – 12 item
NPI-12 A/A	Neuropsychiatric Inventory – 12 item Agitation/Aggression Domain
NPI-C	Neuropsychiatric Inventory – Clinician Rating
NPI-C A+A	Neuropsychiatric Inventory – Clinician Rating Agitation + Aggression Domains
NPI-C D+H	Neuropsychiatric Inventory – Clinician Rating Delusion + Hallucination Domains
PK	pharmacokinetic(s)
PWC-20	Physician Withdrawal Checklist-20
SAE	serious adverse event
SAP	Statistical Analysis Plan
SD	standard deviation
SDI	Sleep Disorder Inventory
SE	standard error
SNRI	selective norepinephrine reuptake inhibitors

SSRI	selective serotonin reuptake inhibitors
TE	Treatment Emergent
TEAE	Treatment Emergent Adverse Event
TEMA	Treatment Emergent Markedly Abnormal
ULN	Upper limit normal

6.2. Appendix 2 Changes to Protocol-Planned Analyses

6.3. Appendix 3 Demographics and Baseline Characteristics

The number of participants in each analysis set will be summarized and listed by intervention group and overall. In addition, the distribution of participants by site ID will be presented unless otherwise noted.

Table 7 presents a list of the demographic variables that will be summarized by intervention group and overall for the full analysis set.

Table 7: Demographic Variables

Continuous Variables	**Summary Type**
Age (years)	Descriptive statistics (N, mean, standard deviation [SD], median and range [minimum and maximum]).
Weight (kg)	
Height (cm)	
Body Mass Index (BMI) (kg/m^2)	
Categorical Variables	Frequency distribution with the number and percentage of participants in each category.
Age Group (Adult [<65 years], Elderly [≥65 years])	
Age Group II (55-64 years, 65-74 years, and ≥75 years)	
Sex (male, female, unknown, undifferentiated)	
Race[a] (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or other Pacific Islander, White, Multiple, Not Reported)	
Ethnicity (Hispanic or Latino, not Hispanic or Latino, Not Reported, Unknown)	
BMI (underweight <18.5 kg/m^2, normal 18.5-<25 kg/m^2, overweight 25-<30 kg/m^2, obese ≥30 kg/m^2)	

[a]If multiple race categories are indicated, the Race is recorded as 'Multiple'

Table 8 presents a list of the baseline disease characteristics variables that will be summarized by intervention group and overall for the full analysis set.

Table 8: Baseline Disease Characteristics

Continuous Variables	**Summary Type**
Baseline NPI-C A+A score	Descriptive statistics (N, mean, standard deviation [SD], median and range [minimum and maximum]).
Baseline NPI-12 total score	
Baseline NPI-12 A/A domain score	
Baseline NPI-12 sleep domain score per eCOA	
Baseline CMAI-C total score	
Baseline MMSE	
Categorical Variables	
Family history of diagnosed Alzheimer's disease or other cause of dementia	Frequency distribution with the number and percentage of participants in each category.
Baseline NPI-12 sleep domain score category per eDC (<4, ≥4)	
Community dwelling/assisted living per eDC	
Baseline MMSE category (10 to 20, 21 to 24)	
Baseline CDR score (1, 2-3)	

6.4. Appendix 4 Protocol Deviations

In general, the following list of major protocol deviations may have the potential to impact participants' rights, safety or well-being, or the integrity and/or result of the clinical study. Participants with major protocol deviations will be identified prior to database lock and the participants with major protocol deviations will be summarized by category for safety analysis sets.

- Developed withdrawal criteria but not withdrawn
- Entered but did not satisfy criteria
- Received a disallowed concomitant treatment
- Received wrong treatment or incorrect dose
- Other

Number of participants not meeting inclusion criteria, or meeting exclusion criteria will be summarized by study intervention group for the safety analysis set.

Additional analyses of protocol deviations for assessing and mitigating the impact of COVID-19 on study outcome are presented in Section 6.9.

6.5. Appendix 5 Prior and Concomitant Medications

Prior medications and concomitant medications will be coded using the World Health Organization Drug Dictionary (WHO-DD). Prior medications are defined as any therapy used before the day of first dose (partial or complete) of study intervention. Concomitant medications are defined as any therapy used on or after the same day as the first dose of study intervention, including those that started before and continue on after the first dose of study intervention.

Summaries of concomitant medications will be presented by ATC level 2 and ATC level 4 terms and base preferred term for the double-blind and follow-up phases for safety and follow-up analysis sets, respectively. The proportion of participants who receive each concomitant medication will be summarized as well as the proportion of participants who receive at least 1 concomitant medication. In addition, concomitant medications of special interest will be summarized. See Appendix 7 Medications of Special Interest for list of categories of medications of special interest.

Prior medications of special interest will be summarized by ATC level 2 and ATC level 4 terms and base preferred term.

6.6. Appendix 6 Adverse Events of Special Interest

AE Special Interest Category	**Event Type**
Somnolence	Somnolence
Cataplexy	Cataplexy
Sleep paralysis	Sleep paralysis
Complex, sleep-related behaviors/parasomnias	Confusional arousal
	Somnambulism
	Sleep terror
	Bruxism
	Sleep sex
	Sleep-related eating disorder
	Sleep behaviour disorder
	Catathrenia
Fall	Fall
Motor vehicle accident	Motor vehicle accident

6.7. Appendix 7 Medications of Special Interest

Categories for medications of special interest are defined as follows:

Medications of Special Interest Category
Antidepressants
Benzodiazepines
Hypnotic/sedative including z-drugs
Antipsychotics
AChE inhibitors and memantine
Aducanumab
Non-Pharmacological Therapy

6.8. Appendix 8 Criteria for Treatment-emergent Markedly Abnormal Laboratory Values

Laboratory Parameter	Unit	Low	High
Clinical Chemistry			
Albumin	g/dL	2.4	6.0
Albumin	g/L	24	60
Alkaline phosphatase	U/L	N/A	250
Alanine transaminase (SGPT)	U/L	N/A	200
Aspartate transaminase (SGOT)	U/L	N/A	250
Bicarbonate	mEq/L	15.1	34.9
Bicarbonate	mmol/L	15.1	34.9
Bilirubin (direct)	mg/dL	N/A	3.0
Bilirubin (direct)	µmol/L	N/A	51.3
Bilirubin (total)	mg/dL	N/A	3.0
Bilirubin (total)	µmol/L	N/A	51.3
Blood urea nitrogen	mg/dL	N/A	50
Blood urea nitrogen	mmol/L	N/A	17.9
Calcium	mg/dL	6	12
Calcium	mmol/L	1.497	2.994
Chloride	mEq/L or mmol/L	94	112
Cholesterol	mg/dL	N/A	300
Cholesterol	mmol/L	N/A	7.758
Creatine kinase	U/L	N/A	990
Creatinine	mg/dL	N/A	3
Creatinine	µmol/L	N/A	265.2
Gamma glutamyl transferase	U/L	N/A	300
Glucose Plasma	mg/dL	40	300
Glucose Plasma	mmol/L	2.204	16.653
High-density lipoprotein cholesterol (HDL)	mg/dL	35	N/A
High-density lipoprotein cholesterol (HDL)	mmol/L	0.905	N/A
Lactic acid dehydrogenase (LDH)	U/L	N/A	500
Low-density lipoprotein cholesterol (LDL)	mg/dL	N/A	160
Low-density lipoprotein cholesterol (LDL)	mmol/L	N/A	4.1376
Phosphate	mg/dL	2.0	6.0
Phosphate	mmol/L	0.6458	1.9374
Potassium	mmol/L	3.0	5.8
Sodium	mEq/L	125	155
Sodium	mmol/L	125	155
Total protein	g/L	50	N/A
Total protein	g/dL	5	N/A
Triglycerides	mg/dL	N/A	500
Triglycerides	mmol/L	N/A	5.645
Uric acid	mg/dL	1.5	10
Uric acid	µmol/L	89.22	594.8
Hematology			
Hematocrit - female	%	28	50
- male	%	24	55
Hemoglobin	g/dL	8	19
Hemoglobin	g/L	80	190
Hemoglobin A1c	%	4	8
Neutrophils/Leukocytes	Fraction of 1	0.3	0.9
Monocytes/Leukocytes	Fraction of 1	N/A	0.2
Eosinophils/Leukocytes	Fraction of 1	N/A	0.1
Basophils/Leukocytes	Fraction of 1	N/A	0.06
Lymphocytes/Leukocytes	Fraction of 1	0.1	0.6
MCV	fL	40	140

MCH	pg	22	38
Reticulocytes/Erythrocytes	Fraction of 1	0.005	0.028
Platelet count	10^9/L; giga/L	100	600
Red blood cell (RBC) count - female	10^{12}/L; tera/L	3.0	N/A
- male	10^{12}/L; tera/L	3.0	N/A
White blood cell (WBC) count	10^9/L; giga/L	2.5	15.0
Urinalysis			
Urine pH		N/A	8.5
Urine specific gravity		< 1.001	> 1.035
Other			
Folate	ng/mL	4	4.8
B12	pmol/L	111	812

Note: Values should be flagged as markedly abnormally low if the value is less than the value indicated in the "Low" column. Likewise, values should be flagged as markedly abnormally high if the value is greater than the value indicated in the "High" column.
Note: The same limits apply to both males and females unless gender is indicated.
N/A = Not applicable.

6.9. Appendix 9 Analyses for Assessing and Mitigating the Impact of COVID- 19 on Study Outcome

The following measures will be taken to handle the impact of COVID-19 on study outcome:

1. Listing and summary of treatment discontinuations and study discontinuations including reasons due to COVID-19 will be presented.

2. Protocol deviations related to COVID-19 including missing visits and remote visits due to COVID will be summarized; corresponding listing will be provided.

7. REFERENCES

1. Cohen-Mansfield J. Instruction Manual for the Cohen-Mansfield Agitation Inventory (CMAI). 1991.
2. Creavin ST, et al. Mini-Mental State Examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations. Cochrane Database of Systematic Reviews 2016, Issue 1. Art. No.: CD011145. https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD011145.pub2/full.
3. Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J. The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia. Neurology. 1994;44(12):2308-2314.
4. Cummings JL. The Neuropsychology in dementia patients. Neurology. 1997;48(5 Suppl 6):S10-S16.
5. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975;12(3):189-198
6. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. Psychological Methods; 2010; Vol. 15, No. 4, pp. 309–334.
7. Posner K, Oquendo MA, Gould M, Stanley B, Davies M. Columbia Classification Algorithm of Suicide Assessment (C-CASA): classification of suicidal events in the FDA's pediatric suicidal risk analysis of antidepressants. Am J Psychiatry. 2007;164:1035-1043.
8. Rickels K, Garcia-Espana R, Mandos LA, Case GW. Physician Withdrawal Checklist (PWC-20). J Clin Psychopharmacol.2008;28(4):447-451.
9. Rosen WG, Mohs RC, Da mer's disease. Am J Psychiatry. 1984;141(11):1356-1364.
10. Rubin, D. (1987), Multiple Imputation for Nonresponse in Surveys, Wiley, New York, XXIX+258 pp.
11. Tractenberg RE, Singer CM, Cummings JL, Thal LJ. The Sleep Disorders Inventory: an instrument for studies of sleep disturbance in persons with Alzheimer's disease. J Sleep Res. 2003a;12(4):331-337. doi: 10.1046/j.0962-1105.2003.00374.x. PMID: 14633245; PMCID: PMC4418643.
12. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). [2009] Guidance for industry drug induced liver injury: premarketing clinical evaluation. Available: http://www.fda.gov/downloads/Drugs/.../Guidances/UCM174090.pdf.