Janssen Research & Development Statistical Analysis Plan

A Multicenter, Double-Blind, Randomized, Parallel-Group, Placebo-Controlled, Study to Evaluate the Efficacy and Safety of Seltorexant 20 mg as Adjunctive Therapy to Antidepressants in Adult and Elderly Patients with Major Depressive Disorder with Insomnia Symptoms Who Have Responded Inadequately to Antidepressant Therapy

Protocol 42847922MDD3002; Phase 3

JNJ-42847922 (seltorexant)

Status: Approved
Date: 10 August 2022
Prepared by: Janssen Research & Development,
LLC Document No: EDMS-RIM-121671, 2.0

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		Not Applicable	Initial release
1.1		Fixed Formatting Errors Section 5.1: Update visit names for follow-up visits. Minor edits to the analysis windows Section 5.3: Specified the intercurrent event "switch of add-on treatment and/or underlying antidepressant" separately to three distinct events to improve clarity. Clarified the summary measure and added rules for handling multiple intercurrent events at the same time. Clarified data to be used under each estimand Section 5.3.2 Added a supportive analysis excluding Ukrainian subjects ongoing in DB phase as of Feb. 24, 2022 Section 5.4.2.2: Updated analyses to be consistent with not imputing percent reduction from baseline(DB) to Day 43 in MADRS total score Section 5.5.1.2: Clarified that sQUALwill be analyzed using the ANCOVA on actual change (not on ranked data). Section 5.5.6.2 Changed analysis method for PGI-S from ANCOVA on ranked data to MMRM on actual data. In addition, similar change applied to PGI-C in section 5.5.7.2 and CGI-S in Section 5.5.8.2 Section 5.6: Updated safety analyses to be based on planned intervention received Section 5.6.1: Updated compliance calculation for study intervention. Added rule for handling missing	Made the analysis plan consistent with the protocol objectives.

		data of the last dose of study intervention Section 5.6.3: Clarified definition of Treatment emergent markedly abnormal value. Clarified that only participants with baseline (DB) values that don't meet criteria or if baseline value is missing will be eligible for analyses of hepatic toxicity. Added RR Interval to ECG table Section 5.6.3.4: Clarified scoring rule for C-SSRS. Clarified the visit to which follow-up symptoms will be compared for PWC. Section 5.7.3 Updated region categorization Appendix 3: Added HDRS-17 to baseline disease characteristic table Appendix 5: Added summary of antidepressants taken during the DB phase Appendix 6: Updated preferred term based on MEDRA 24.1. Appendix 9: Updated units for Hemoglobin A1c, removed LDH Appendix 10: Specified that if number of non-site visits is small, then sensitivity analysis may not be performed	

1. INTRODUCTION

This statistical analysis plan (SAP) contains definitions of analysis sets, derived variables, and statistical methods for all planned analyses for study 42847922MDD3002.

1.1. Objectives and Endpoints

For the following objectives and endpoints, major depressive disorder with insomnia symptoms (MDDIS) is defined as: Diagnostic and Statistical Manual of Mental Disorders-5th Edition (DSM-

5) diagnosis of Major Depressive Disorder (MDD) and a) moderate to severe IS by a patient version Insomnia Severity Index (ISI) total score of ≥15 at the end of screening and b) a positive response for IS (MDD symptoms Item 4) on the Structured Clinical Interview for DSM-5 Axis I Disorders Clinical Trials Version (SCID-CT). In addition, the clinician version ISI total score ≥15 is also required since this version was used in the Phase 2 program.

Double-blind Treatment Phase

Objectives	Endpoints
Efficacy	
Primary	
• To assess the efficacy of seltorexant 20 mg compared with placebo as adjunctive therapy to an antidepressant in improving depressive symptoms in participants with MDDIS who have had an inadequate response to current antidepressant therapy with an SSRI or SNRI	• Change from baseline to Day 43 in the MADRS total score in participants with MDDIS.
Key Secondary	
• MDD symptoms other than insomnia symptoms	• Change from baseline to Day 43 in the MADRS without sleep item (MADRS-WOSI) total score.
• To assess the efficacy of seltorexant compared with placebo in participants with MDDIS, as adjunctive therapy to an antidepressant on patient-reported assessment of sleep outcomes	• Change from baseline to Day 43 in sleep disturbance using the Patient Reported Outcome Measurement Information System-Sleep Disturbance (PROMIS-SD) Short Form (8a) T-score in participants with MDDIS.
Secondary	
To assess the efficacy of seltorexant compared with placebo as adjunctive therapy to an antidepressant in participants with MDDIS on the following:	
• Core symptoms of depression	• Change from baseline to Day 43 in the MADRS-6 total score.

• Response of depressive symptoms	• Proportion of responders on depressive symptoms scale, defined as a ≥50% improvement in MADRS total score from baseline to Day 43.
• Patient-reported symptoms of depression	• Change from baseline to Day 43 in Patient Health Questionnaire, 9-Item (PHQ-9)
Exploratory	
To assess the efficacy of seltorexant compared with placebo as adjunctive therapy to an antidepressant on the following:	
• Patient-reported sleep diary	• Change from baseline to Day 43 in subjective sleep parameters as measured by the Consensus Sleep Diary (CSD).
• Remission of depressive symptoms	• Proportion of participants with remission of depressive symptoms, defined as a MADRS total score ≤12 at Day 43.
• Patient-reported health-related quality of life	• Change from baseline to Day 43 in health-related quality of life and health status, as assessed by the European Quality of Life, 5 Dimension, 5-Level (EQ-5D-5L) questionnaire.
• Patient-reported global functioning (work/school, social, and family life)	• Change from baseline to Day 43 in the Sheehan Disability Scale (SDS) total score.
• Patient-reported insomnia symptoms	• Change from baseline to Day 43 in the patient-reported ISI total score.
• Patient-reported assessment of sleep outcomes	• Change from baseline over time in sleep symptoms using the Patient Global Impression of Severity (PGI-S).
• Patient-reported assessment of change in depressive symptoms	• Change from baseline over time in depressive symptoms using the Patient Global Impression of Change (PGI-C).
• Clinical symptom severity of depression	• Change from baseline over time in the Clinical Global Impression-Severity (CGI-S) score.
• Patient-reported rumination symptoms	• Change from baseline to Day 43 in the RRS total score.
Safety	

• To assess the safety and tolerability of seltorexant as adjunctive therapy to an antidepressant in participants with MDD in the short-term compared with placebo	• Adverse events (AEs) including AEs of special interest (AESI) • Vital signs • Suicidality assessment using the C-SSRS • Withdrawal symptoms assessment using the Physician Withdrawal Checklist 20-items (PWC-20) • Laboratory values and ECG • Patient-reported sexual functioning using Arizona Sexual Experiences Scale (ASEX)

Additional exploratory objectives:

- To identify diagnostic biomarkers and to investigate changes in MDD-related biomarkers (eg, HPA axis, metabolic function, and biomarkers of immune system activation) in relation to clinical response on depression and IS upon adjunctive treatment with seltorexant.
- To identify genetic (eg, CYP genes) and other factors that may influence the PK, safety, or tolerability of seltorexant.
- To assess the plasma exposure of seltorexant and its M12 metabolite along with alpha-1-acid glycoprotein levels in participants with MDDIS when used as adjunctive treatment.

Statistical hypotheses pertaining to the primary and the key secondary endpoints are described in Section 2.

Details about the 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 with seltorexant 20 mg study to assess the efficacy and safety of seltorexant as an adjunctive therapy in adults (18 to 64 years, inclusive) and elderly (65 to 74 years, inclusive) participants with MDDIS, who have had an inadequate response to current antidepressant therapy with a SSRI/SNRI.

The study will consist of 3 phases: a screening phase (up to 30 days), a DB treatment phase (43 days), and a posttreatment follow-up phase (7 to 14 days after double-blind treatment phase).

Approximately 386 participants with MDDIS 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 country, age group (adults [<65years] versus elderly [≥65 years]), baseline rumination level (RRS total score <54, ≥54), and by baseline MADRS group (MADRS total score <24, ≥24).

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 50-57, if they do not withdraw the consent.

Participants will take their assigned study intervention at home, once daily at bedtime during DB treatment phase. Participants will continue to take their baseline SSRI/SNRI antidepressant (at the same dose, without change, and at approximately the same time of the day as prior to entering the study) during the study starting at screening, and through the DB treatment phase and the follow- up phase.

2. STATISTICAL HYPOTHESES

This study is designed to show that the intervention treatment effect in improving depressive symptoms (as measured by change from baseline on Day 43 in MADRS total score) of seltorexant as an adjunctive MDD treatment is superior to placebo in participants with MDDIS.

If μ_T is the mean change in MADRS total score for seltorexant group and μ_P is the mean change in MADRS 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 0.05.

3. SAMPLE SIZE DETERMINATION

Approximately 386 participants (randomized in 1:1 ratio to placebo and seltorexant 20 mg) are planned to be enrolled in the DB treatment phase. The enrollment is targeted to achieve approximately 374 participants eligible to be included in the full analysis set 1 (FAS1). Assuming a treatment difference of 4.4 points in change from baseline in MADRS total score between seltorexant and placebo, a standard deviation of 12, and a 1-sided significance level of 0.025 (equivalently 2-sided 0.05), this sample size (ie.,374 in FAS1) will provide approximately 90% power in a pairwise comparison between seltorexant and placebo in the primary efficacy analysis, accounting for a drop-out rate of approximately 15%. The assumed treatment difference and standard deviation used in this calculation are based on Phase 2 (42847922MDD2001) study results, as well as on clinical judgment.

An interim analysis (IA) for futility is planned when a combined total of 276 participants with MDDIS have completed the DB phase in 42847922MDD3002 and another phase 3 study, 42847922MDD3001. Details about the interim analysis are presented in a separate IA SAP.

4. POPULATIONS (ANALYSIS SETS) FOR ANALYSIS

Analysis Sets	Description
Enrolled	All participants who sign the ICF
Randomized	The randomized analysis set includes all participants who were randomized in the study.
Full Analysis Set 1 (FAS1)	The full analysis set 1 (FAS1) includes all randomized participants who received at least 1 dose of study intervention in the DB phase and had baseline MADRS total score ≥24 (per IWRS).
Full Analysis Set 2 (FAS2)	The full analysis set 2 (FAS2) includes all randomized participants who received at least 1 dose of study intervention in the DB phase.
Safety	The safety analysis set includes all randomized participants who received at least 1 dose of study intervention in the DB phase.
Follow-up Analysis Set	The follow-up analysis set includes all randomized participants who entered the follow-up phase.

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: Visit Windows

Parameter	Analysis Phase	Scheduled Visit Number	Time Interval (label on output)	Time Interval (Day)*	Target Time Point (Day)*
MADRS PROMIS-SD CGI-S PGI-S PHQ-9	DB	3	Baseline (DB)	<=1	1
		6	Day 15	2 to 22	15
		7	Day 29	23 to 36	29
		8	Day 43	37 to End of DB	43
		DB final	Endpoint (DB)	2 to End of DB	
	FU (DB)	10	Day 50-57	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 50-57 EW	49 to End of DB	54

Parameter	Analysis Phase	Scheduled Visit Number	Time Interval (label on output)	Time Interval (Day)*	Target Time Point (Day)*
CSD**	DB	3	Average Baseline	<=1	1
		8	Day 43	2 to End of DB	43
		DB final	Endpoint (DB)	2 to End of DB	
	FU (DB)		Day 50-57	End of DB + 1 to End of FU	10
ISI (patient)	SCR	1	Screening 1	<= -7	-19
		2	Screening 2	-6 to -1	-3
	DB	3	Baseline (DB)	<=1	1
		8	Day 43	2 to End of DB	43
		DB final visit	Endpoint (DB)	2 to End of DB	
SDS RRS ASEX	DB	3	Baseline (DB)	<=1	1
		8	Day 43	2 to End of DB	43
		DB final visit	Endpoint (DB)	2 to End of DB	
PGI-C	DB	6	Day 15	2 to 22	15
		7	Day 29	23 to 36	29
		8	Day 43	37 to End of DB	43
		DB final	Endpoint (DB)	2 to End of DB	
	FU (DB)	10	Day 50-57	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 50-57 EW	49 to End of DB	54
EQ-5D-5L	DB	3	Baseline (DB)	<=1	1
		8	Day 43	2 to End of DB	43
		DB final	Endpoint (DB)	2 to End of DB	
ECG Lab	SCR	1	Screening	<1	
	DB	3	Average Predose	<=1	1
		8	Day 43	2 to End of DB	43
		DB final	Endpoint (DB)	2 to End of DB	
Vital Signs Weight	SCR	1	Screening	<1	
	DB	3	Baseline (DB)	<=1	1
		6	Day 15	2 to 22	15
		7	Day 29	23 to 36	29
		8	Day 43	37 to End of DB	43
		DB final	Endpoint (DB)	2 to End of DB	
	FU (DB)	10	Day 50-57	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 50-57 EW	49 to End of DB	54
C-SSRS	SCR	1	Screening	<1	
	DB	3	Baseline (DB)	<=1	1
		5	Day 8	2 to 11	8
		6	Day 15	12 to 22	15
		7	Day 29	23 to 36	29
		8	Day 43	37 to End of DB	43
		DB final	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)	10	Day 50-57	End of DB + 1 to End of FU	10
			Day 8 EW	2 to 11	8
			Day 15 EW	12 to 22	15
			Day 29 EW	23 to 36	29
			Day 43 EW	37 to 48	43
			Day 50-57 EW	49 to End of DB	54
PWC	DB	8	Endpoint (DB)	2 to End of DB	43
	FU (DB)	9	Follow-up 1 (DB)	1 to 5	1
		10	Follow-up 2 (DB)	6 to End of FU	10

) 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)*
MADRS PROMIS-SD CGI-S PGI-S PHQ-9	DB	3	Baseline (DB)	<=1	1
		6	Day 15	2 to 22	15
		7	Day 29	23 to 36	29
		8	Day 43	37 to End of DB	43
		DB final	Endpoint (DB)	2 to End of DB	
	FU (DB)	10	Day 50-57	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 50-57 EW	49 to End of DB	54
CSD**	DB	3	Average Baseline	<=1	1
		8	Day 43	2 to End of DB	43
		DB final	Endpoint (DB)	2 to End of DB	
	FU (DB)		Day 50-57	End of DB + 1 to End of FU	10
ISI (patient)	SCR	1	Screening 1	<= -7	-19
		2	Screening 2	-6 to -1	-3
	DB	3	Baseline (DB)	<=1	1
		8	Day 43	2 to End of DB	43
		DB final visit	Endpoint (DB)	2 to End of DB	
SDS RRS ASEX	DB	3	Baseline (DB)	<=1	1
		8	Day 43	2 to End of DB	43
		DB final visit	Endpoint (DB)	2 to End of DB	
PGI-C	DB	6	Day 15	2 to 22	15
		7	Day 29	23 to 36	29
		8	Day 43	37 to End of DB	43
		DB final	Endpoint (DB)	2 to End of DB	
	FU (DB)	10	Day 50-57	End of DB + 1 to End of FU	10
			Day 15 EW	2 to 22	15

Parameter	Analysis Phase	Scheduled Visit Number	Time Interval (label on output)	Time Interval (Day)*	Target Time Point (Day)*
			Day 29 EW	23 to 36	29
			Day 43 EW	37 to 48	43
			Day 50-57 EW	49 to End of DB	54
EQ-5D-5L	DB	3	Baseline (DB)	<=1	1
		8	Day 43	2 to End of DB	43
		DB final	Endpoint (DB)	2 to End of DB	
ECG Lab	SCR	1	Screening	<1	
	DB	3	Average Predose	<=1	1
		8	Day 43	2 to End of DB	43
		DB final	Endpoint (DB)	2 to End of DB	
Vital Signs Weight	SCR	1	Screening	<1	
	DB	3	Baseline (DB)	<=1	1
		6	Day 15	2 to 22	15
		7	Day 29	23 to 36	29
		8	Day 43	37 to End of DB	43
		DB final	Endpoint (DB)	2 to End of DB	
	FU (DB)	10	Day 50-57	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 50-57 EW	49 to End of DB	54
C-SSRS	SCR	1	Screening	<1	
	DB	3	Baseline (DB)	<=1	1
		5	Day 8	2 to 11	8
		6	Day 15	12 to 22	15
		7	Day 29	23 to 36	29
		8	Day 43	37 to End of DB	43
		DB final	Endpoint (DB)	2 to End of DB	
	FU (DB)	10	Day 50-57	End of DB + 1 to End of FU	10
			Day 8 EW	2 to 11	8
			Day 15 EW	12 to 22	15
			Day 29 EW	23 to 36	29
			Day 43 EW	37 to 48	43
			Day 50-57 EW	49 to End of DB	54
PWC	DB	8	Endpoint (DB)	2 to End of DB	43
	FU (DB)	9	Follow-up 1 (DB)	1 to 5	1
		10	Follow-up 2 (DB)	6 to End of FU	10

*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.
** The analysis windows for CSD pertains to the average of the assessments that were performed 7 days prior to and including the study visit.

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 disposition date at follow-up.

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	1) Copy Reference 2) Copy Increment from Reference 3) Delta Adjustment
Non-Responder	Participants with missing values will be imputed as non-responders for analyses that use imputed data.
Non-Remitter	Participants with missing values will be imputed as non-remitters for analyses that use imputed data.

Imputation of total scores will be performed for the following efficacy scales as shown in Table 3 below. If the number of items with missing scores is greater than the maximum number of items presented in the table, the total score will be missing.

For the remaining efficacy assessments which require adding multiple item scores, the total score will be missing if any item score is missing.

Table 3: Imputation of Total Score for Efficacy Scales

Efficacy Scale	Total Number of Items	Maximum Number of Items That Can Be Missing	Formula for Total Score
MADRS	10	1	Sum of non-missing item scores * (10 / number of non-missing items)

MADRS-WOSI	9	1	Sum of non-missing item scores * (9 / number of non-missing items)
PROMIS-SD	8	4	Sum of non-missing item scores * (8 / number of non-missing items). If the result is a fraction, the imputed total score will be rounded up to the nearest whole number.

5.1.4. Pooling Algorithm

If low enrollment were to occur in a country, it may be combined with other countries in the similar geographic region for analysis.

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, FAS1, FAS2, safety 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 follow-up phase of 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 during 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 Appendix 10.

5.3. Primary Endpoint(s) Analysis

The primary analysis will be based on full analysis set 1 (FAS1) using the MADRS total score collected during the double-blind phase for submissions other than the EU dossier. For the EU dossier, the primary analysis set will be based on full analysis set 2 (FAS2).

5.3.1. Definition

The primary efficacy endpoint is the change in MADRS total score from baseline (DB) to Day 43.

The MADRS is a clinician-administered scale designed to measure depression severity and detects changes due to antidepressant intervention (Montgomery 1979). The scale consists of10 items, each of which is scored from 0 (item not present or normal) to 6 (severe or continuous presence of the symptoms). Higher scores represent a more severe condition. The MADRS evaluates reported sadness, apparent sadness, inner tension, reduced sleep, reduced appetite, concentration

difficulties, lassitude, inability to feel, pessimistic thoughts, and suicidal thoughts. The test exhibits high inter-rater reliability. The typical recall period for the MADRS is 7 days.

The MADRS total score is the sum of scores from individual question items at a given time point, and ranges from 0 to 60. Higher scores represent a more severe condition. Imputation of total score is presented in Section 5.1.3.

Negative changes in MADRS total score indicate improvement.

Five estimands are defined for the primary endpoint: Estimand 1, Estimand 2a using the FAS1 analysis set, and Estimand 2, Estimand 2b, and Estimand 1a using the FAS2 analysis set. With the exception of the European Union (EU) dossier, the primary estimand is Estimand 1, and the supplementary estimands are Estimand 1a, Estimand 2, Estimand 2a and Estimand 2b. For the EU dossier, the primary estimand is Estimand 2, and the supplementary estimands are Estimand 1, Estimand 1a, Estimand 2a and Estimand 2b.

The sections below describe the primary and sensitivity analyses performed for each primary and supplementary estimand.

5.3.2. Estimand 1

Estimand 1 will be the **primary estimand** for submissions other than the EU dossier, and a **supplementary estimand** for the EU dossier. The analysis will be performed on FAS1.

Primary Trial Objective: To evaluate the efficacy of seltorexant 20 mg compared with placebo as adjunctive therapy to an antidepressant in improving depressive symptoms in participants with MDDIS who have had an inadequate response to current antidepressant therapy with a SSRI or SNRI.

Estimand Scientific Question of Interest: What is the antidepressant benefit from seltorexant 20 mg versus placebo as adjunctive treatment to a SSRI or SNRI in adults and elderly participants with

MDDIS, who have had an inadequate response to treatment with a SSRI/SNRI based on the change from baseline in MADRS total 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 as an adjunctive treatment to SSRI or SNRI
- Control: Placebo as an adjunctive treatment to SSRI or SNRI

Population:

Participants with MDDIS who have had an inadequate response to current antidepressant therapy with a SSRI/SNRI, as reflected by the inclusion/exclusion criteria, and who have a baseline MADRS total score ≥24.

Variable: Change in MADRS 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 add-on study drug only	Hypothetical strategy: as if the intercurrent event had not occurred
2. Treatment discontinuation of both underlying antidepressant and add-on study drug	Hypothetical strategy: see above
3. Switch of add-on study drug	Hypothetical strategy: see above
4. Switch of underlying antidepressant	Hypothetical strategy: see above
5. Switch of add-on study drug and underlying antidepressant	Hypothetical strategy: see above

The population component requires a criterion of having baseline MADRS ≥24. The reason is that the study uses HDRS with remote rating for inclusion/exclusion during screening (as an effort to control potential rater inflation), however, at Day 1 (baseline), site rating of MADRS is needed as the primary efficacy evaluation. In order to avoid the situation when HDRS rating has non- negligible discrepancy with MADRS rating, the population for the primary estimand is pre- planned to exclude participants with low MADRS (i.e., <24) at baseline since these participants do not represent the population of

interest. This approach was discussed and agreed upon at the End of Phase 2 meeting with FDA (February 2020).

The hypothetical strategy attempts to estimate the treatment effect in the hypothetical scenario as if the intercurrent events had not occurred. As such, any data observed after an intercurrent event of switch of treatment is excluded. Otherwise, data observed up to and including DB disposition day will be included.

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 in the double-blind phase will be presented for MADRS total score by intervention group.

MADRS total 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, country, age group (adults [<65 years] and elderly [≥65 years]), baseline(DB) rumination level (RRS total score <54, ≥54), time, and time-by-intervention interaction, and baseline(DB) MADRS 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. 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 95% CI will be presented.

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

A supportive analysis excluding Ukrainian participants ongoing in Double-blind phase on February 24, 2022 will be conducted using the same MMRM model used for the primary analysis. Comparison between seltorexant and placebo at Day 43 will be performed using the appropriate contrast. Difference in least square means and 2-sided 95% CI will be presented.

Justification of MMRM: Based on data from the short-term phase 2b study MDD2001, it was observed that within each treatment group, there was no obvious difference in MADRS change trajectories between completers and drop-outs, and for each drop-out reason, MADRS trajectories do not follow a specific pattern. In addition, in the 20mg group (the dose being studied in phase 3 program), those who dropped out generally showed in improvement prior to discontinuation. Based on these observations, it is considered reasonable to assume missing at random (MAR) as the missingness mechanism.

MAR assumption cannot be tested vs. MNAR using observed data, however, simulation findings (Siddiqui 2009). indicate that in the presence of a mixture of the three missing mechanisms (MCAR,

MAR, MNAR) with differential dropout rates between treatment groups, the MMRM approach is able to re-estimate the true treatment difference consistently with a negligible bias and control Type I error rate. Therefore, the primary analysis in this phase 3 study will be MMRM, with sensitivity analyses (tipping point) (Section 5.3.2.1.2) to stress test the efficacy findings under MNAR assumption. This approach was discussed and agreed upon at the End of Phase 2 meeting with FDA (February 2020)

5.3.2.1.2. Sensitivity Analysis

The sensitivity analysis for Estimand 1 will include the delta adjustment multiple imputation method which will be implemented on the MADRS total score. This method will employ the following 3 steps:

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: Missing at random (MAR) is assumed for intermediate missing data (i.e., missing data between non-missing observations). Monotone missing data will first be imputed by MAR-based MI regression, and the imputed scores in the experimental intervention group will be adjusted to be worse than the other participants in same group with non-missing data as discussed below.

The imputed values will be adjusted by adding δ_C to the imputed values for participants randomized to the control group and adding δ_A to the imputed values for participants randomized to the experimental intervention group. Delta-adjusted fully imputed datasets will be generated for different combinations of δ_C and δ_A values as defined below:

- $\delta_C = 0$ and $\delta_A = 0$ *to* Δ^* in increments of 1 (experimental group-only adjustment analysis)

- Adding positive values results in higher (worse) scores. Δ^* represents the adjustments leading to the 'tipping point', so the smallest delta adjustments values at which conclusions change from *favorable* (i.e. statistically significant: 2-sided p-value ≥0.05) to *unfavorable* (fail to reject the null hypothesis of no intervention difference).

- Imputed values in both the experimental and the control groups will be adjusted using a range of delta values, and delta-adjusted fully imputed datasets will be generated for each combination.

These methods will be applied to all data not used after the intercurrent events under hypothetical strategy. In addition, another version of these methods will be implemented, where the delta adjustments will be applied to all data not used after the intercurrent events under hypothetical strategy except for those caused by discontinuation reasons due to COVID-19, and for discontinuation reasons not related to study intervention including Lost to Follow-Up, Withdrawal by Subject or Other.

Step 2 – Analysis

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) Between-group comparisons to placebo at Day 43 (e.g. 2-sided p-values, point estimates for intervention difference) will be displayed graphically for each considered δ_A, up to the 'tipping point' adjustment. For analysis of the imputed datasets using a range of delta-adjustments on both experimental and control groups, a tipping point two-way map will be generated.

5.3.3. Estimand 2

Estimand 2 will be the **primary estimand** for EU dossier, and a **supplementary estimand** for submissions other than the EU dossier. The analysis will be performed on FAS2.

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

Population:

Participants with MDDIS who have had an inadequate response to current antidepressant therapy with a SSRI/SNRI, as reflected by the inclusion/exclusion criteria (i.e., regardless of the baseline MADRS total score).

Intercurrent events and their corresponding strategies:

Intercurrent Events	Name of Strategy for Addressing Intercurrent Events and Its Description
1. Treatment discontinuation of add-on study drug only	Treatment policy strategy: all observed values of the endpoint are used regardless of whether or not the participant had experienced this intercurrent event
2. Treatment discontinuation of both underlying antidepressant and add-on study drug	Hypothetical strategy: as if the intercurrent event had not occurred

3. Switch of add-on study drug	Hypothetical strategy: as if the participant had discontinued treatment instead of switching
4. Switch of underlying antidepressant	Hypothetical strategy: as if the participant had discontinued treatment instead of switching
5. Switch of add-on study drug and underlying antidepressant	Hypothetical strategy: as if the participant had discontinued treatment instead of switching

For participants experiencing multiple intercurrent events at the same time, intercurrents in categories 3-5 will override the rest. For example, if a patient starts a new add-on study drug the day after taking the last dose of their previous add-on study medication, this intercurrent event will be considered a switch of add-on study drug, and not a discontinuation. However, if more than one day passes between the last dose of their previous add-on medication and the first dose of the new add-on medication, this situation will be considered a discontinuation of add-on study drug intercurrent event followed by a switch of add-on study drug intercurrent event.

The treatment policy strategy attempts to estimate the treatment effect irrespective of the intercurrent event. Therefore, follow-up data will be used in the analysis for patients who discontinue add-on study drug only. The hypothetical strategy attempts to estimate the treatment effect in the hypothetical scenario as if the intercurrent events had not occurred. As such, any data observed after an intercurrent event of switch of treatment is excluded. Otherwise, data observed up to and including DB disposition day will be included.

5.3.3.1. Analysis Methods

Once monotone missing datasets are created after imputing intermediate missing data using the MCMC method as described in Section 5.3.2.1.2, the monotone missing data will be imputed using the following 2 methods: (i) Copy Reference (CR) MI method for the primary analysis (ii) Copy Increment from Reference (CIR) MI method as a sensitivity analysis. Each method has a different set of assumptions as described in the sections below.

Steps 2 and 3 from Section 5.3.2.1.2 will then be implemented: performing MMRM analysis on each imputed dataset, and pooling of the results from the imputed datasets.

5.3.3.1.1. Primary Analysis

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

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.

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 MADRS total score by intervention group. Under the treatment policy strategy, data from participants who discontinued the add-on 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.

5.3.3.1.2. Sensitivity Analysis

The sensitivity analysis for Estimand 2 will include implementation of Copy Increment from Reference (CIR) MI method.

Analysis assumptions: MAR is assumed for intermediate missing data. MNAR will be 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 it will be assumed that the efficacy data that is not missing or not used after the intercurrent event in these participants immediately adopt a distribution with predicted mean values at future visits where change in mean from visit to visit is similar to those in the control group. MAR is assumed for missing data in the control group.

This approach assumes that the participant's efficacy under the experimental intervention versus the control group is maintained as at the last time for which the efficacy scores are available.

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

5.3.4. Estimand 1a

Estimand 1a will be a **supplementary estimand** for all submissions. The analysis will be performed on FAS2.

Estimand 1a is similar to Estimand 1 (Section 5.3.2) in all but the population component. The population under this estimand includes participants with MDDIS who have had an inadequate response to current antidepressant therapy with a SSRI/SNRI, as reflected by the inclusion/exclusion criteria (i.e., regardless of the baseline MADRS total score).

5.3.4.1. Analysis Methods

5.3.4.1.1. Primary Analysis

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

The primary analysis under Estimand 1a is the same as the primary analysis under Estimand 1 (Section 5.3.2.1.1) where an MMRM analysis will be performed.

5.3.5. Estimand 2a

Estimand 2a will be a **supplementary estimand** for all submissions. The analysis will be performed on FAS1.

Estimand 2a is similar to Estimand 2 (Section 5.3.3) in all but the population component. The population under this estimand includes participants with MDDIS who have had an inadequate response to current antidepressant therapy with a SSRI/SNRI, as reflected by the inclusion/exclusion criteria, and who have a baseline MADRS total score ≥24.

5.3.5.1. Analysis Methods

5.3.5.1.1. Primary Analysis

Descriptive statistics of the actual values and the change from baseline(DB) to each postbaseline time point will be presented for MADRS total score by intervention group. Under the treatment policy strategy, data from participants who discontinued the add-on study drug only (intercurrent event 1 under this estimand) and provided additional data every 2 weeks until Day 43 will be included in this analysis.

The primary analysis under Estimand 2a is the same as the primary analysis under Estimand 2 (Section 5.3.3.1.1) where CR MI will be implemented.

5.3.6. Estimand 2b

Estimand 2b will be a **supplementary estimand** for all submissions. The analysis will be performed on FAS2.

Estimand 2b is similar to Estimand 2 (Section 5.3.3) in all but the strategy for the intercurrent event of treatment discontinuation of both underlying antidepressant and add-on study drug; under this estimand, treatment policy strategy will be implemented – under this strategy all observed values of the endpoint will be used regardless of whether or not the participant had experienced this intercurrent event. The strategy for the other 2 intercurrent events will be the same as that of Estimand 2.

5.3.6.1. Analysis Methods

5.3.6.1.1. Primary Analysis

Descriptive statistics of the actual values and the change from baseline(DB) to each postbaseline time point will be presented for MADRS total score by intervention group. Under the treatment policy strategy, data from participants who discontinued the add-on study drug only, and from those who discontinued both underlying antidepressant and add-on study drug (intercurrent events 1 and 2 under this estimand) and provided additional follow-up data every 2 weeks until Day 43 will be included in this analysis.

The primary analysis under Estimand 2b is the same as the primary analysis under Estimand 2 (Section 5.3.3.1.1) where CR MI will be implemented.

Additional analyses of the primary endpoint for assessing and mitigating the impact of COVID- 19 on study outcome are presented in Appendix 10**.**

5.4. Secondary Endpoint(s) Analysis

5.4.1. Key Secondary Endpoint(s)

Similar to the primary endpoint, the primary analysis of the key secondary endpoints will be based on full analysis set 1 (FAS1) for submissions other than the EU dossier. For the EU dossier, the primary analysis will be based on full analysis set 2 (FAS2).

5.4.1.1. MADRS – WOSI (Without Sleep Item)

5.4.1.1.1. Definition

The first key secondary endpoint is the change in MADRS-WOSI from baseline(DB) to Day 43.

MADRS-WOSI is defined as the full MADRS without the sleep item (Montgomery 1979).

The MADRS-WOSI total score is the sum of scores of the remaining 9 items at a given time point, ranging from 0 to 54. Higher scores represent a more severe condition. Imputation of total score is presented in Section 5.1.3.

Negative changes in MADRS-WOSI total score indicate improvement.

5.4.1.1.2. Estimand(s)

The same estimands as defined for the primary endpoint in Section 5.3 are defined for MADRS-WOSI, with the variable being the following: Change in MADRS-WOSI total score from baseline to Day 43.

5.4.1.1.3. Analysis Methods

Same analyses as for the estimands under the primary endpoint, and the corresponding sensitivity and supplementary analyses will be implemented for change in MADRS-WOSI total score. The MMRM model will include baseline(DB) MADRS-WOSI total score as the covariate.

Descriptive statistics of the actual values and the change from baseline(DB) to each postbaseline time point in the double-blind phase will be presented for MADRS-WOSI total score by intervention group. Under the treatment policy strategy for Estimand 2, Estimand 2a and Estimand 2b, MADRS-WOSI data from participants who discontinued the add-on study intervention only (intercurrent event 1 under these estimands), and from participants who discontinued both underlying antidepressant and add-on study drug (intercurrent event 2 under Estimand 2b) and provided additional data every 2 weeks until Day 43 will be included in the analyses corresponding to each estimand.

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

The fixed sequence testing procedure will be applied to control the familywise error rate (FWER) at two-sided 0.05 level accounting for multiplicity due to the primary and key secondary efficacy endpoints. The fixed sequence testing procedure will first test the primary endpoint at two-sided 0.05 level. If the hypothesis corresponding to the primary endpoint is rejected, then the MADRS-WOSI will be tested at the same alpha level. If the hypothesis corresponding to the primary endpoint is not rejected, then the testing procedure will stop.

5.4.1.2. PROMIS-SD

5.4.1.2.1. Definition

The second key secondary endpoint is the change in PROMIS-SD T-score from baseline(DB) to Day 43.

Developed under a National Institutes of Health (NIH) initiative, the Patient Reported Outcomes Measurement Information System captures self-reported, qualitative health aspects in the domains of physical, mental, and social health (Yu L 2011).

The PROMIS-SD (Short Form 8a) subscale consists of a static 8-item questionnaire. The PROMIS-SD instruments assess self-reported perceptions of sleep quality, sleep depth, and restoration associated with sleep. This includes perceived difficulties and concerns with getting to sleep or staying asleep, as well as perceptions of the adequacy of and satisfaction with sleep. Sleep Disturbance does not focus on symptoms of specific sleep disorders, nor does it provide subjective estimates of sleep quantities (total amount of sleep, time to fall asleep, amount of wakefulness during sleep). The Sleep Disturbance short form is universal rather than disease-specific. It assesses sleep disturbance over the past 7 days.

Each question has five response options ranging in value from one to five. To find the total raw score for a short form with all questions answered, sum the values of the response to each question. For the

8-item form, the lowest possible raw score is 8; the highest possible raw score is 40. Lower scores indicate less sleep disturbance. The "direction" of the responses is not the same for all questions, i.e., sometimes a response of "not at all" indicates more sleep disturbance and sometimes a response of "not at all" indicates less sleep disturbance.

"My sleep quality was" ranges from 5=very poor to 1=very good "My sleep was refreshing" ranges from 5=not at all to 1=very much

"I had a problem with my sleep" ranges from 1=not at all to 5=very much "I had difficulty falling asleep" ranges from 1=not at all to 5=very much "My sleep was restless" ranges from 1=not at all to 5=very much

"I tried hard to get to sleep" ranges from 1=not at all to 5=very much

"I worried about not being able to fall asleep" ranges from 1=not at all to 5=very much "I was satisfied with my sleep" ranges from 5=not at all to 1=very much

A score can be approximated if a participant skips a question. Imputation of total score is presented in Section 5.1.3.

The total raw score or pro-rated raw score will be converted into a T-score for each participant based on the table in Appendix 6.8. The T-score rescales the raw score into a standardized score with a mean of 50 and an SD of 10.

5.4.1.2.2. Estimand(s)

The same estimands as defined for the primary endpoint in Section 5.3 are defined for PROMIS- SD, with the variable being the following: Change in PROMIS-SD T-score from baseline to Day 43.

5.4.1.2.3. Analysis Methods

Same analyses as for the estimands under the primary endpoint, and the corresponding sensitivity and supplementary analyses will be implemented for change in PROMIS-SD T-score. The MMRM model will include baseline(DB) PROMIS-SD T-score as the covariate.

Descriptive statistics of the actual values and the change from baseline(DB) to each postbaseline time point in the double-blind phase will be presented for PROMIS-SD raw score and the T-score by intervention group. Under the treatment policy strategy for Estimand 2, Estimand 2a and Estimand 2b, PROMIS-SD T-score data from participants who discontinued the add-on study drug only (intercurrent event 1 under these estimands), and from participants who discontinued both underlying antidepressant and add-on study drug (intercurrent event 2 under Estimand 2b) and provided additional follow-up data every 2 weeks until Day 43 will be included in the analyses corresponding to each estimand.

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

Similar to the first key secondary endpoint, a fixed sequence testing procedure will be applied to control the FWER at two-sided 0.05 level accounting for multiplicity due to the primary and key secondary efficacy endpoints. If the hypothesis corresponding to MADRS-WOSI is rejected, then the PROMIS-SD T-score key will be tested at the same alpha level of 2-sided 0.05. If the hypothesis corresponding to the MADRS-WOSI is not rejected, then the testing procedure will stop.

Additional analyses of the key secondary endpoint for assessing and mitigating the impact of COVID-19 on study outcome are presented in Appendix 10**.**

5.4.2. Other Secondary Endpoint(s)

The following analyses will be performed on the FAS1 and FAS2 populations.

5.4.2.1. MADRS-6

5.4.2.1.1. Definition

The 6-item MADRS is a clinician-administered scale designed to measure the core symptoms of depression severity and detects changes due to antidepressant intervention (Montgomery 1979). It is a subset of MADRS (10-item). The MADRS scale is a validated, reliable scale and acceptable to regulatory health authorities as a primary scale to determine efficacy in major depression.

The MADRS-6 subscale score is the sum of scores for the following MADRS items: reported sadness, apparent sadness, inner tension, lassitude, inability to feel, and pessimistic thoughts.

Negative changes in MADRS-6 total score indicate improvement.

5.4.2.1.2. Analysis

Descriptive statistics of the actual values and the change from baseline(DB) to each postbaseline time point in the double-blind phase will be presented for MADRS-6 total score by intervention group.

The change from baseline(DB) in MADRS-6 total score at Day 43 will be analyzed using the same MMRM model as described for the primary endpoint with baseline(DB) MADRS-6 total score as the covariate.

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

5.4.2.2. Response Based on MADRS Total Score

5.4.2.2.1. Definition

A participant is defined as a responder at a given time point if the percent improvement from baseline(DB) in MADRS is ≥50% at that time point (i.e., percent change ≤-50%). Participants who do not meet such criterion will be considered as non-responders. Imputation of missing response status is presented in Section 5.1.3.

5.4.2.2.2. Analysis

The number and percentage of participants who achieve a response will be summarized at each time point during the double-blind phase by intervention group. The analysis will be performed on observed as well as imputed data (participants with missing values will be imputed as non- responders).

Response rates over time will be plotted.

The cumulative response rate, defined as the percentage of participants experiencing at least a given value of percent reduction from baseline(DB) to Day 43 in MADRS total score, will be presented graphically, for observed data.

5.4.2.3. Patient Health Questionnaire - 9 Item (PHQ-9)

5.4.2.3.1. Definition

The 9-item Patient Health Questionnaire - 9 Item (PHQ-9) scale scores each of the 9 symptom domains of the DSM-5 MDD criteria and it has been used both as a screening tool and a measure of response to intervention for depression. Each item is rated on a 4-point scale (0=not at all, 1=several days, 2=more than half the days, and 3=nearly every day). The participant's item responses are summed to provide a total score (range of 0 to 27), with higher scores indicating greater severity of depressive symptoms. The recall period is 2 weeks.

Negative changes in PHQ-9 total score indicate improvement.

5.4.2.3.2. Analysis

Descriptive statistics of the actual values and the change from baseline(DB) to each postbaseline time point in the double-blind phase will be presented for PHQ-9 total score by intervention group.

The change from baseline(DB) in PHQ-9 total score at Day 43 will be analyzed using the same MMRM model as described for the primary endpoint with baseline(DB) PHQ-9 total score as the covariate.

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

5.5. Tertiary/Exploratory Endpoint(s) Analysis

Exploratory analyses will be performed for the endpoints described in the sections below for FAS1 and FAS2 analysis sets.

5.5.1. Consensus Sleep Diary

5.5.1.1. Definition

Subjective sleep parameters are measured by means of Consensus Sleep Diary (CSD).

The CSD (Carney 2012). is a prospective measure that involves daily completion within an hour of getting out of bed. The CSD includes: (1) time into bed, (2) time of sleep attempt, (3) time it took to fall asleep, (4) number of awakenings during the main sleep period, (5) total estimated duration of awakenings, (6) time of final awakening, (7) rise time, (8) subjective report of sleep quality (very poor, poor, fair, good, very good), and (9) comment (e.g., comments such as "I had a cold last night") (this item is not collected in this study). Of these nine items, the first seven are used to derive estimates of other sleep indices. The following parameters will be analyzed:

Consensus Sleep Diary – Sleep Indices	
Self-reported sleep onset latency (sSOL)	Recorded time to fall asleep (item 3)
Subjective wake after sleep onset (sWASO)	Total duration of the nighttime awakenings (item 5)
Subjective total sleep time (sTST)	Time of final awakening (item 6) – time into bed (item 1) – (sSOL + sWASO)
Subjective number of nighttime awakenings (s-nNAW)	Number of awakenings during the sleep period (item 4)
Subjective (sQUAL)	Subjective report of quality of sleep (item 8)

5.5.1.2. Analysis

As the CSD data will be collected for up to 7 days prior to each scheduled visit, the data analyzed will be the average of the data pertaining to each visit. The average of the ordinal data pertaining to sQUAL (0 = Very Poor to 4 = Very Good) will be rounded to the nearest integer.

Descriptive statistics of the average actual values and the change from average baseline(DB) to each postbaseline time point in the double-blind phase will be presented by intervention group for sSOL, sTST, sWASO, s-nNAW, and sQUAL.

An ANCOVA model will be used to test the difference of change from average baseline(DB) at endpoint(DB) in the continuous endpoints (sSOL, sTST, sWASO, and s-nNAW) and sQUAL between seltorexant and placebo. The model will include intervention, country, age group, and

baseline(DB) rumination level as factors, and the respective average baseline(DB) value as a covariate. Difference of least square means and 2-sided 95% CI will be presented.

5.5.2. Remission Based on MADRS Total Score

5.5.2.1. Definition

A participant is defined as a remitter at a given time point if the MADRS total score is [≤12] at that time point. Participants who do not meet such criterion will be considered as non-remitters. Imputation of missing remission status is presented in Section 5.1.3.

5.5.2.2. Analysis

The number and percentage of participants who achieve remission will be summarized at each time point during the double-blind phase by intervention group. The analysis will be performed on observed as well as imputed data (participants with missing values will be imputed as non- remitters).

Remission rates over time will be plotted.

5.5.3. European Quality of Life (EuroQol) Group, 5 Dimension, 5-Level questionnaire (EQ-5D-5L)

5.5.3.1. Definition

The EQ-5D-5L is a standardized instrument for use as a measure of health outcome, primarily designed for self-completion by respondents. It is a descriptive system comprised of the following 5 dimensions: Mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each of the 5 dimensions is divided into 5 levels of perceived problems (Level 1 indicating no problem, Level 2 indicating slight problems, Level 3 indicating moderate problems, Level 4 indicating severe problems, and Level 5 indicating extreme problems) (EuroQol Group. About EQ-5D.2014).

Participants select an answer for each of the 5 dimensions considering the response that best matches their health “today”. Individual scores from the 5 dimensions will be used to obtain a weighted health status index as shown below: (i) Scores from each dimension will be combined to obtain a 5L profile score: eg, a score of 1 for each dimension will give a 5L profile score of 11111. Dimension scores will be combined in the following order: Mobility, Self-Care, Usual Activities, Pain/Discomfort, Anxiety/Depression. (ii) The value set of the Health Status Index (HSI) for various values of 5L profile scores is published for Canada in the following website: https://www.ncbi.nlm.nih.gov/pubmed/26492214/. (iii) The Canadian value set will be used to get the HSI values for all the countries participating in the study.

A sum score, with a possible range of 0 to 100, is derived as follows: (sum of the scores from the 5 dimensions minus 5) * 5. Negative changes in the sum score indicate improvement.

5.5.3.2. Analysis

Descriptive statistics of the actual values and the change from baseline(DB) to each postbaseline time point in the double-blind phase will be presented by intervention group for weighted EQ-5D health status index, the EQ-VAS, and the sum score.

Individual dimension responses will also be summarized at each visit with frequency counts and percentage of participants by intervention group.

5.5.4. Sheehan Disability Scale (SDS)

5.5.4.1. Definition

The SDS, a patient-reported outcome measure, is a 5-item questionnaire which has been widely used and accepted for assessment of functional impairment and associated disability (Leon 1996). The first 3 items assess disruption of (1) work/school work, (2) social life/leisure activities, and (3) family life/home responsibilities using a 0-10 rating scale. The total score is the sum of the first 3 items, ranging from 0 to 30. Higher scores indicate greater impairment. It also

has one item on days lost from school or work and one item on days when underproductive. Negative changes in SDS total score indicate improvement.

5.5.4.2. Analysis

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

An ANCOVA model will be used to test the difference of change from baseline at endpoint(DB) in SDS total score between seltorexant and placebo. The model will include intervention, country, age group, and baseline(DB) rumination level as factors, and baseline(DB) SDS total score value as a covariate. Difference of least square means and 2-sided 95% CI will be presented.

In addition, descriptive statistics of the actual values and the change from baseline, as well as a frequency distribution, will be presented by intervention group for each of the 5 SDS individual item scores at each time point. The percentage of participants who have not worked/studied at all during the past 7 days will be summarized at each time point during the double-blind phase by intervention group.

5.5.5. Insomnia Severity Index (ISI) (Patient Version)

5.5.5.1. Definition

The ISI is a 7-item questionnaire assessing the nature, severity, and impact of insomnia. The dimensions evaluated are: sleep onset, sleep maintenance, early morning awakening problems; sleep dissatisfaction; interference of sleep problem with daily functioning; noticeability of sleep problems by others; and distress caused by the sleep difficulties. A 5-point Likert scale (0-4) is used to rate

each item. The total score is the sum of the 7 item scores, ranging from 0 to 28. The total score is interpreted as follows: absence of insomnia (0-7); sub-threshold insomnia (8-14); moderate insomnia (15-21); and severe insomnia (22-28).

Negative changes in ISI total score indicate improvement.

5.5.5.2. Analysis

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

The change from baseline(DB) in ISI total score at endpoint (DB) will be analyzed using the same ANCOVA model as described for SDS total score (Section 5.5.4.2) with baseline(DB) ISI total score as the covariate. Difference of least square means and 2-sided 95% CI will be presented.

In addition, the shift in ISI score category from baseline to endpoint (DB) will be presented. A frequency distribution of the ISI individual item scores at baseline(DB) and endpoint(DB) will be provided by intervention group.

5.5.6. Patient Global Impression of Severity (PGI-S) for Insomnia

5.5.6.1. Definition

The PGI-S for sleep symptoms consists of 2 patient-reported items to capture the participant's perceived severity of difficulty falling and staying asleep as well as the problem of not feeling rested the next day. Both items have the recall period of the past 7 days. The 2 dimensions of PGI- S are measured on a 5-point scale: 1=no difficulty of falling asleep or staying asleep (for *severity of difficulty falling and staying asleep*), or I did not have this problem (for *problem of not feeling rested the next day*); 2=mild; 3=moderate; 4=severe; 5=very severe.

Negative changes in PGI-S dimension scores indicate improvement.

5.5.6.2. Analysis

Descriptive statistics of the actual values and the change from baseline(DB) will be presented by intervention group at each postbaseline timepoint in the double-blind phase for observed case data for each dimension.

The change from baseline(DB) in PGI-S dimension score at each postbaseline timepoint will be analyzed using an MMRM model with baseline(DB) PGI-S dimension score as a covariate.

A frequency distribution of the PGI-S dimension scores over time will be provided by intervention group. The frequencies of PGI-S scores will be plotted at baseline(DB) and endpoint(DB) for the 2 dimensions.

5.5.7. Patient Global Impression of Change (PGI-C) for Depression

5.5.7.1. Definition

The PGI-C for depression is a single item that will capture the participant's perceptions of improvement or deterioration in depression symptoms compared with when the participant started the study intervention. The change of measured on a 7-point scale: 1=much better; 2=somewhat better; 3=a little better; 4=about the same; 5=a little worse; 6=somewhat worse; 7=much worse.

Higher scores indicate deterioration of symptoms.

5.5.7.2. Analysis

Descriptive statistics of the actual values at each postbaseline time point in the double-blind phase will be presented by intervention group for observed case data.

An ANOVA model will be used to test the difference in PGI-C score at each postbaseline time point in the double-blind phase between seltorexant and placebo. The model will include the PGI-C score as the dependent variable, intervention, country, age group, baseline(DB) rumination level as factors.

A frequency distribution of the PGI-C over time will be provided by intervention group.

5.5.8. Clinical Global Impression – Severity (CGI-S)

5.5.8.1. Definition

The CGI-S (Depression) provides an overall clinician-determined summary measure of the severity of the participant's illness that takes into account all available information, including knowledge of the participant's history, psychosocial circumstances, symptoms, behavior, and the impact of the symptoms on the participant's ability to function (Guy 1976). The CGI-S evaluates the severity of psychopathology on a scale of 1 to 7. Considering total clinical experience with the depression population, a participant is assessed on severity of illness at the time of rating according to: 0=not assessed; 1=normal, not at all ill; 2=borderline mentally ill; 3=mildly ill; 4=moderately ill; 5=markedly ill; 6=severely ill; 7=among the most extremely ill patients.

A score of 0 indicates that the participant was not assessed and will be treated as missing. The score will be summarized as recorded.

Negative changes in CGI-S score indicate improvement.

5.5.8.2. Analysis

Descriptive statistics of the actual values and the change from baseline(DB) will be presented at each post-baseline timepoint in the double-blind phase by intervention group for observed case data.

The change from baseline(DB) in CGI-S score at each postbaseline timepoint will be analyzed using an MMRM model with baseline(DB) CGI-S score as a covariate.

A frequency distribution of the CGI-S scores over time will be provided by intervention group. The frequencies of CGI-S scores will be plotted at baseline and endpoint(DB).

5.5.9. Ruminative Response Scale (RRS)

5.5.9.1. Definition

The RRS assesses rumination as the process of "compulsively focused attention on the symptoms of one's distress, and on its possible causes and consequences, as opposed to its solutions" as established by Nolen-Hoeksema in 1998. The 22 items of the RRS measure aspects of rumination, brooding and reflective pondering (Nolen-Hoeksema 2000). A 4-point Likert scale (1=almost never, 2=sometimes, 3=often, 4=almost always) is used to rate each item.

The RRS total score is the sum of the 22 item scores, ranging from 22 to 88. A higher total score indicates higher degree of rumination.

Negative changes in RRS total score indicate improvement.

5.5.9.2. Analysis

Descriptive statistics of the actual values and the change from baseline(DB) to endpoint(DB) will be presented for RRS total score by intervention group.

The change from baseline(DB) in RRS total score at endpoint (DB) will be analyzed using the ANCOVA model with intervention, country, age group as factors and baseline(DB) RRS total score as a covariate. Difference of least square means and 2-sided 95% CI will be presented.

5.5.10. Correlation between MADRS total score and PROMIS-SD T-score

Correlation between change in MADRS total score and change in PROMIS-SD T-score at Week 6 will be explored; Pearson and Spearman correlation coefficients will be calculated. Scatter plot of change in scores of the 2 assessment scales at Week 6 will be presented for each intervention group.

5.5.11. Subgroup Analyses

For the subgroups listed in Section 5.7.3, subgroup analyses using MMRM will be performed for the change in MADRS total score at Day 43. The fixed terms in the model will be intervention group, country, age group (adults, elderly), baseline (DB) RRS level (RRS total score <54, ≥54), time, subgroup, time-by-intervention interaction, intervention-by-subgroup interaction, and time- by-intervention-by-subgroup interaction, and baseline(DB) MADRS total score as a covariate. Point estimate of the treatment difference and 2-sided 95% CI will be estimated using appropriate contrasts. The terms in the model will be adjusted for the subgroups of region and age group II (18-34 years, 35-54 years, 55-64 years, ≥ 65 years). Country will not be included in the model where the subgroup of interest is Region and age group will not be included where the subgroup of interest is Age group II.

The analysis results (difference of LS means and 95% CI) of the different subgroups will be displayed in a forest plot.

Similar analyses will be performed for the change in PROMIS-SD T-score and the results will be displayed in a forest plot. Sub-group analyses for the change in MADRS-WOSI total score may be performed if this assessment shows a meaningful difference from MADRS total score.

The analyses will be performed for the FAS1 and FAS2 analysis sets.

5.5.12. Analysis of Efficacy data in the Follow-up Phases

Descriptive statistics of the actual values and the change from the baseline (DB) for the data collected in the follow-up phase will be presented for the following efficacy data:

MADRS total score, MADRS-WOSI score, MADRS-6 score, PROMIS-D raw and T-scores, individual domains of CSD, PGI-S (severity of difficulty falling and staying asleep, and problem of not feeling rested the next day), PGI-C, CGI-S.

The analyses will be performed for the follow-up analysis set for the follow-up phase.

5.6. Safety Analyses

All safety analyses will be based on safety analysis set based on planned intervention, unless otherwise specified.

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 and of the background antidepressant 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, 15-21 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. If the date of last dose of study intervention is missing, it will be imputed as the day prior to the double-blind disposition date for calculating exposure duration and compliance. 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).

Background antidepressant duration is defined as (minimum of end of double blind phase and end of antidepressant use – maximum of start of double blind phase and start of antidepressant use) +1.

The analysis will be performed on FAS1, FAS2 and 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 and for the background antidepressant.

Compliance will be calculated for the study intervention as:

Compliance (%) = (number of days when participant took the expected number of tablets)/(days expected to be dosed (includes days where the antidepressant was not taken) x 100)

Compliance will be calculated for the background antidepressant as:

Compliance (%) = (number of days actually dosed)/(days expected to be dosed (includes days where the antidepressant was not taken) x 100

The analysis will be performed on FAS1, FAS2 and 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 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.

Clinical laboratory test values will be considered "treatment-emergent markedly abnormal" (TEMA) using the criteria defined by the sponsor listed in Appendix 6.9.

- 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 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 criteria or if baseline value is missing will be eligible for these analyses.

A listing of participants with 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.

In addition, the number of subjects with the following shifts in chemistry laboratory values from baseline(DB) to the maximum postbaseline time point will be presented for the double-blind phase:

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).

5.6.3.2. Vital Signs and Physical Examination Findings

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. 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]°C
	Abnormally low	[<35.5]°C
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 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 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 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 2007). 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.

5.6.3.4.2. Arizona Sexual Experiences Scale (ASEX)

The ASEX is a patient-reported 5-item rating scale that quantifies sex drive, arousal, vaginal lubrication/penile erection, ability to reach orgasm, and satisfaction from orgasm. The scale has shown satisfactory reliability and validity (McGahuey 2000).

Each of the 5 items is rated on a 6-point Likert scale, ranging from 1 to 6. The total score is the sum of the 5 items. Higher scores indicate more sexual dysfunction. If any item of the scale is missing at a visit, the total score for that scale at that visit will be left blank.

For each of the 5 items, a frequency distribution will be provided by intervention group and gender at each time point. In addition, for each of the 5 items, a frequency distribution will be provided by intervention group, combining the responses for the genders. For this analysis, "vaginal lubrication/penile erection" (question 3) will be summarized as one question. The ASEX total score at each time point and the change from baseline(DB) will be summarized with descriptive statistics by intervention group.

The number and percentage of participants who have ASEX total score 19 or greater, or a score of 5 or greater on any item, or a score of 4 or greater on any 3 items, reflecting sexual dysfunction, will be summarized at each time point by intervention group for double-blind phase.

5.6.3.4.3. Physician Withdrawal Checklist (20 items; PWC-20)

The PWC-20 is a reliable and sensitive instrument for the assessment of discontinuation symptoms (Rickels 2008). The PWC-20 is a simple and accurate method used to assess potential withdrawal symptoms following cessation of study intervention.

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 endpoint (DB) 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.

5.6.3.4.4. Menstrual Cycle Tracking

Menstrual cycle in premenopausal women will be tracked during the double-blind and follow-up phases. Participants who are experiencing menses during the screening evaluation belonging to the safety analysis set will be included in the analysis.

Descriptive summary statistics (n, mean, SD, median, minimum, and maximum) of the duration of menses and the length of menstrual cycle will be provided for the duration of the double-blind phase. This analysis will be repeated for the following 2 sub-groups: women who use contraception, and those who don't use contraception. Duration of menses in the follow-up phase will be summarized for the follow-up analysis set.

5.7. Other Analyses

5.7.1. Pharmacokinetics

PK analyses will be performed on the safety analysis set.

Descriptive statistics (N, mean, SD, median, range, CV (%) and IQ range) will be used to summarize plasma concentrations at each sampling time point and for each PK parameter of Active Study Intervention.

5.7.2. Biomarkers

Analysis of biomarker data will be discussed in a separate SAP.

5.7.3. Definition of Subgroups

Subgroup analyses of the primary and key secondary endpoints will be performed for the FAS1 and FAS2 analysis sets for the following subgroups:

Subgroup	**Definition**
Sex	Male Female Other
Race	Asian Black White Other
Ethnicity	Hispanic or Latino Not Hispanic or Latino

Country	Argentina Denmark Finland Korea Malaysia Poland Chile Ukraine USA Slovakia
Region	European Union North America Rest of World
Age Group	Adult (<65 years) Elderly (≥ 65 years)
Age Group II	18-34 years 35-54 years 55-64 years ≥ 65 years
Baseline 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
Number of major depressive episodes	< 3 ≥ 3
Family history of psychiatric disorders	Yes No Unknown
Number of failed prior antidepressants	1 2 3 or more
Current antidepressant type	SSRI SNRI
Prior use of benzodiazepines	Yes No
Baseline RRS per IWRS	<54 ≥ 54

5.8. Interim Analyses

One interim analysis of the primary endpoint is planned for this study. The aim of the IA is to make recommendations to continue the study, or stop the study for futility. Details about the analysis will be discussed in an IA SAP.

5.8.1. Independent Data Monitoring Committee (IDMC)

An independent data monitoring committee will monitor the SAEs on a bi-monthly basis. Additionally, the IDMC will meet periodically to review safety data and meet once to evaluate the unblinded efficacy data at the IA and make recommendation on whether to declare futility. Details about the membership of the IDMC and its roles and responsibilities are discussed in a separate IDMC charter.

6. SUPPORTING DOCUMENTATION

6.1. Appendix 1 List of Abbreviations

AE	adverse event
AESI	adverse event(s) of special interest
ALT/SGPT	alanine aminotransferase
ANCOVA	analysis of covariance
ASEX	Arizona Sexual Maturity Scale
AST/SGOT	aspartate aminotransferase
ATC	anatomic and therapeutic class
BMI	body mass index
C-SSRS	Columbia Suicide Severity Rating Scale
CGI-S	Clinical Global Impression-Severity
CI	confidence interval
CIR	copy increment from reference
CR	copy reference
CRF	case report form
CSD	consensus sleep diary
CV	coefficient of variation
DB	double-blind
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, 5th edition
ECG	electrocardiogram
EQ-5D-5L	European Quality of Life, 5-Dimension, 5-Level
EQ-VAS	European Quality of Life-Visual analog Scale
EU	European Union
FAS1	full analysis set 1
FAS2	full analysis set 2
FWER	familywise error rate
HSI	health status index
ICH	International Conference on Harmonisation
IDMC	Independent Data Monitoring Committee
IQ	interquartile
IS	insomnia symptoms
ISI	Insomnia Sleep Index
IWRS	interactive web response system
LS	least squares
MADRS	Montgomery-Asberg Depression Rating Scale
MADRS-WOSI	Montgomery-Asberg Depression Rating Scale-Without Sleep Item
MAR	missing at random
MCMC	Markov Chain Monte Carlo
MDD	major depressive disorder
MDDIS	major depressive disorder with insomnia symptoms
MedDRA	Medical Dictionary for Regulatory Activities
MI	multiple imputation
MMRM	mixed model for repeated measures
MNAR	missing not at random
NIH	National Institutes of Health
PGI-C	Physician Global Impression-Change
PGI-S	Physician Global Impression-Severity
PHQ-9	Patient Health Questionnaire-9 Item

PK	pharmacokinetic(s)
PWC-20	Physician Withdrawal Checklist-20
RRS	Ruminative Response Scale
s-nNAW	subjective number of nighttime awakenings

SAE	serious adverse event
SAP	Statistical Analysis Plan
SCID-CT	Structured Clinical Interview for DSM-5-Clinical Trial Version
SD	standard deviation
SDS	Sheehan Disability Scale
SE	standard error
SNRI	selective norepinephrine reuptake inhibitors
sQUAL	subjective quality of sleep
SSRI	selective serotonin reuptake inhibitors
sSOL	subjective sleep onset latency
sTST	subjective total sleep time
sWASO	subjective wake after sleep onset

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 region, country, and 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 FAS1 and FAS2 analysis sets.

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 (18-34 years, 35-54 years, 55-64 years, and ≥65 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)	
Country	
Region	

[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 FAS1 and FAS2 analysis sets.

Table 8: Baseline Disease Characteristics

Continuous Variables	**Summary Type**
Age (years) when diagnosed with MDD	Descriptive statistics (N, mean, standard deviation [SD], median and range [minimum and maximum]).
Duration (weeks) of current depressive episode	
Baseline MADRS total score	
Baseline CGI-S score	
Baseline patient-rated ISI total score per eDC	
Baseline clinician-rated ISI total score per eDC	
Baseline RRS total score per eDC	
Baseline HDRS-17 total score	
Categorical Variables	Frequency distribution with the number and percentage of participants in each category.
Baseline CGI-S (Depression) score (Normal (not at all ill), Borderline mentally ill, Mildly ill, Moderately ill, Markedly ill, Severely ill, Among the most extremely ill patients)	
Baseline RRS per IWRS (54, ≥54)	
Baseline RRS per eDC (54, ≥54)	

Categorical Variables	Summary Type
Current antidepressant type (SSRI, SNRI)	Frequency distribution with the number and percentage of participants in each category.
Number of failed prior antidepressants (1, 2, 3 or more)	
SCID-CT DSM-5 specifiers for MDD	
Prior medication use of benzodiazepines	
Number of major depressive episodes to date, including current episode (<3, ≥	
Family history of alcohol abuse (Yes, No)	
Family history of anxiety disorder (Yes, No)	
Family history of bipolar disorder (Yes, No)	
Family history of depression (Yes, No)	
Family history of schizophrenia (Yes, No)	
Family history of substance abuse (Yes, No)	

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 Appendix 10.

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 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. Antidepressant medications taken during the double-blind phase will be summarized. 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 antidepressant medications, prior medications other than antidepressants, and prior medications of special interest will be summarized by ATC level 2 and level 4 terms and base preferred term.

6.6. Appendix 6 Adverse Events of Special Interest

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

6.8. Appendix 8 Conversion of Raw Score to T-Score for PROMIS-SD

Sleep Disturbance 8a *Short Form Conversion Table*		
Raw Score	**T-Score**	**SE***
8	28.9	4.8
9	33.1	3.7
10	35.9	3.3
11	38.0	3.0
12	39.8	2.9
13	41.4	2.8
14	42.9	2.7
15	44.2	2.7
16	45.5	2.6
17	46.7	2.6
18	47.9	2.6
19	49.0	2.6
20	50.1	2.5
21	51.2	2.5
22	52.2	2.5
23	53.3	2.5
24	54.3	2.5
25	55.3	2.5
26	56.3	2.5
27	57.3	2.5
28	58.3	2.5
29	59.4	2.5
30	60.4	2.5
31	61.5	2.5
32	62.6	2.5
33	63.7	2.6
34	64.8	2.6
35	66.1	2.7
36	67.5	2.8
37	69.0	3.0
38	70.8	3.2
39	73.0	3.5
40	76.5	4.4

*SE= Standard Error on T-score metric

Adult version

6.9. Appendix 9 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
Hemoglobin A1c	Fraction of 1	0.04	0.08
High-density lipoprotein cholesterol (HDL)	mg/dL	35	N/A
High-density lipoprotein cholesterol (HDL)	mmol/L	0.905	N/A
Low-density lipoprotein cholesterol (LDL)	mg/dL	89	160
Low-density lipoprotein cholesterol (LDL)	mmol/L	2.3015	4.1376
Phosphate	mg/dL	2.2	8.1
Phosphate	mmol/L	0.71038	2.61549
Potassium	mmol/L	3.0	5.8
Sodium	mEq/L	125	155
Sodium	mmol/L	125	155
Total protein	g/L	50	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	%	0.28	0.50
- male	%	0.24	0.55
Hemoglobin	g/dL	8	19
Hemoglobin	g/L	80	190
Neutrophils	%	30	90
Monocytes	%	N/A	20
Eosinophils	%	N/A	10

Basophils	%	N/A	6
Lymphocytes	%	10	60
Reticulocytes	%	0.5	1.5
Platelet count	10^9/L; giga/L	100	600
Red blood cell (RBC) count - female	10^{12}/L; tera/L	3.0	5.5
- male	10^{12}/L; tera/L	3.0	6.4
White blood cell (WBC) count	10^9/L; giga/L	2.5	15.0
	Y		

Laboratory Parameter	Unit	Low	High
Urinalysis			
Urine pH		N/A	6.5
Urine specific gravity		< 1.001	> 1.035

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.

6.10. Appendix 10 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 discontinuation and study discontinuation including reasons due to COVID-19 will be presented.

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

3. Additional analyses to assess the potential impact of COVID-19 in the primary and key secondary endpoints.

    a. Mean summaries over time by types of collection (remote, home, site, etc).

    b. Sensitivity analyses will be provided by including types of collection in the primary models for MADRS, MADRS-WOSI, and PROMIS-SD. If the number of non-site visits is small, then the sensitivity analysis may not be conducted as the results would not be expected to differ from the main analyses.

4. Sensitivity analysis for Estimand 1: Additional tipping analysis will be performed for Estimand 1 where the delta adjustments will be applied to all missing data under hypothetical strategy including those caused by discontinuation reasons due to COVID-19; delta adjustments will not be applied to missing data due to discontinuation reasons not related to study intervention including Lost to Follow-Up, Withdrawal by Subject or Other.

7. REFERENCES

Carney CE, Buysse DJ, Ancoli-Israel S, et al. The Consensus sleep diary: Standardizing prospective sleep self-monitoring. Sleep (2012);35(2):287-302.

EuroQol Group. About EQ-5D. http://www.euroqol.org/about-eq-5d.html (Accessed 18 Jun 2014).

EuroQol Group. EQ-5D-5L User Guide: basic information on how to use the EQ-5D-5L instrument. Version 2.0; Oct 2013. http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/UserGuide_EQ-5D-5L_v2.0_October_2013.pdf (Accessed 6 Aug 2014).

Guy W. CGI: Clinical Global Impressions. In: ECDEU Assessment Manual for Psychopharmacology. Rockville, MD: US Department of Health and Human Services; printed 1976 (reprinted 1991):217-222.

Leon AC, Olfson M, Portera L, Farber L, Sheehan DV. Assessing psychiatric impairment in primary care with the Sheehan Disability Scale. Int J Psychiatry Med. (1997);27(2):93-105.

McGahuey CA, Gelenberg AJ, Laukes CA, et al. The Arizona Sexual Experience Scale (ASEX): reliability and validity. J Sex Marital Ther. (2000);26(1):25-40.

Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry. (1979);134:382-389.

Nolen-Hoeksema S. The role of rumination in depressive disorders and mixed anxiety/depressive symptoms. J Abnorm Psychol. (2000);109(3):504-511.

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.

Rickels K, Garcia-Espana R, Mandos LA, Case GW. Physician Withdrawal Checklist (PWC-20). J Clin Psychopharmacol. (2008);28(4):447-451.

Rubin, D. (1987), Multiple Imputation for Nonresponse in Surveys, Wiley, New York, XXIX+258 pp.

Sheehan DV, Harnett-Sheehan K, Raj BA. The measurement of disability. Int Clin Psychopharmacol. (1996) Jun;11 Suppl 3:89-95.

Siddiqui, Ohidul, HM James Hung, and Robert O'Neill. "MMRM vs. LOCF: a comprehensive comparison based on simulation study and 25 NDA datasets." Journal of biopharmaceutical statistics 19.2 (2009): 227- 246.

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.

Yu L, Buysse DJ, Germain A, et al. Development of short forms from the PROMIS™ sleep disturbance and sleep-related impairment item banks. Behav Sleep Med. (2011);10(1):6-24.