Janssen Research & Development *

Statistical Analysis Plan
Amendment 2

A 6-Month, Multicenter, Double-Blind, Randomized, Flexible-Dose, Parallel-Group Study to Compare the Efficacy, Safety, and Tolerability of JNJ-42847922 versus Quetiapine Extended-Release as Adjunctive Therapy to Antidepressants in Adult Subjects With Major Depressive Disorder Who Have Responded Inadequately to Antidepressant Therapy

Protocol 42847922MDD2002; Phase 2

JNJ-42847922 (Selective Orexin-2 Receptor Antagonist)

Status: Approved
Date: 2 July 2019
Prepared by: Janssen Research & Development, LLC
Document No.: EDMS-153916699 3.0

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

TABLE OF CONTENTS

AMENDMENT HISTORY

SAP Version	Issue Date
Original SAP	31 August 2018
Amendment 1	24 May 2019
Amendment 2	3 July 2019

Amendments below are listed beginning with the most recent amendment.

Amendment 1 (24 May 2019)

The overall reason for the amendment: removed analysis windows for salivary cortisol; modified subgroups; added efficacy parameters MADRS-6 and MADRS-WOSI; added conversion rules for lab values; added analysis for plasma concentrations and removed biomarker analysis.

Applicable Section(s)	Description of Change (s)
Section 2.3	Removed analysis windows for salivary cortisol from Table 1 (this is moved to a separate biomarker SAP)
Section 2.6	Additional subgroups added: Baseline MADRS total score, Class of antidepressant study medication, and Final dose. Modifications made to subgroup categorizations of number of antidepressants used and number of major depressive episodes.
Section 4.6	Clarified that prior medications will be summarized by base preferred term.
Section 5	Added MADRS-6 Total score and MADRS-WOSI Total score to the efficacy variable table. Added section 5.3.14 and 5.4.2
Section 5.2.3	Possible sensitivity analysis added excluding a site with GCP issues
Section 5.3.1.2	Added BOCF definition and analysis for MADRS Added subgroup analysis by mode dose
Section 5.4.1	Consensus Sleep Diary moved to exploratory efficacy endpoint Modification made to average calculations for subjective sleep parameters
Section 6.2	Added conversion rules for chemistry laboratory values
Section 6.5.4	Added PWC-20 total score
Section 7	Added that Plasma concentrations for JNJ-42847922, M12, and M16 will be summarized using descriptive statistics
Section 8	Biomarker analysis removed to be defined in a separate document
Attachment 3	Updated conversion factor for Triglycerides

Amendment 2 (3 July 2019)

The overall reason for the amendment: added sensitivity analyses, clarification of subgroup analysis methods, minor error fixes.

Applicable Section(s)	Description of Change (s)
Section 2.3	Revised visit window definition for PWC
Section 5.2.3	Added an analysis for time to treatment-related discontinuation, and Kaplan-Meier curves for time to discontinuation by Week 12.
Section 5.2.3.1	Clarification on methods for subgroup analysis
Section 5.3.3.2	Added a summary of time to sustained response by modal dose
Section 5.4.1.2	Added an ANCOVA analysis on ranks for CSD parameters sFRESH and sQUAL
Section 5.4.2.2	Removed MADRS-WOSI ANCOVA analysis

ABBREVIATIONS

AE	adverse event
ANCOVA	analysis of covariance
ASEX	Arizona Sexual Experiences Scale
BMI	body mass index
BOCF	baseline observation carried forward
CGI-S	Clinical Global Impression-Severity
CI	confidence interval
CRF	case report form
CSD-M	Consensus Sleep Diary-Morning Administration
CSR	Clinical Study Report
C-SSRS	Columbia Suicide Severity Rating Scale
DSM-5	Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition
ECG	electrocardiogram
ESRS-A	Extrapyramidal Symptom Rating Scale-Abbreviated
FT_4	free thyroxine
HAM-A	Hamilton Anxiety Rating Scale
HbA1c	hemoglobin A1c
HOMA	Homeostatic Assessment
HPA	hypothalamic-pituitary-adrenal
HRUQ	Healthcare Resource Use Questionnaire
HVLT-R	Hopkins Verbal Learning Test-Revised
IP	investigational products
ISI	Insomnia Severity Index
IWRS	interactive web response system
LOCF	last observation carried forward
LS	least-squares
MADRS	Montgomery-Asberg Depression Rating Scale
MDD	major depressive disorder
MedDRA	Medical Dictionary for Regulatory Activities
MGH-ATRQ	Massachusetts General Hospital-Antidepressant Treatment Response Questionnaire
MMRM	mixed model for repeated measures
MRU	medical resource utilization
PGI-S	Patient Global Impression-Severity
PK	pharmacokinetic(s)
PRO	patient-reported outcome(s)
PROMIS-SD	Patient Reported Outcomes Measurement Information System-Sleep Disturbance
PROMIS-SRI	Patient Reported Outcomes Measurement Information System-Sleep Related Impairment
PWC	Physician Withdrawal Checklist
QLDS	Quality of Life in Depression Scale
SAE	serious adverse event
SAP	Statistical Analysis Plan
SCID-CT	Structured Clinical Interview for DSM-5 Axis I Disorders– Clinical Trials Version
SD	standard deviation
SDMT	Symbol Digit Modalities Test
SIGH-A	structured interview guide for the Hamilton Anxiety Scale
SMDDS	Symptoms of Major Depressive Disorder Scale
SNRI	serotonin-norepinephrine reuptake inhibitor
SSRI	selective serotonin reuptake inhibitor
TEAE	treatment-emergent adverse event
TEMA	treatment-emergent markedly abnormal
TMT-Part B	Trail Making Test-Part B
TSH	thyroid-stimulating hormone
TST	Total Sleep Time
US	United States
WASO	Wake after sleep onset
WHO-DD	World Health Organization Drug Dictionary
WOSI	Without Sleep Item
XR	extended-release

1. INTRODUCTION

This statistical analysis plan (SAP) contains definitions of analysis sets, derived variables, and statistical methods for all planned analyses for the clinical study report (CSR) for study 42847922MDD2002.

This SAP does not include planned analyses on biomarkers or pharmacogenomics data or on pharmacokinetic (PK) and exposure/response analyses, which will be specified as appropriate in separate documents.

1.1. Trial Objectives

Objectives	Endpoints
Primary	
• To assess, in subjects with major depressive disorder (MDD) who have had an inadequate response to current antidepressant therapy with a selective serotonin reuptake inhibitor (SSRI) or serotonin-norepinephrine reuptake inhibitor (SNRI), the efficacy of flexibly dosed JNJ-42847922 (20 mg or 40 mg) compared to flexibly dosed quetiapine extended-release (XR) (150 mg or 300 mg) as adjunctive therapy to an antidepressant drug in delaying time to all-cause discontinuation of study drug over a 6-month (24-week) treatment period.	• Time to all-cause discontinuation of study drug.
Secondary	
• To assess the efficacy of JNJ-42847922 compared to quetiapine XR in the adjunctive treatment of MDD based on:	
– Achieving and sustaining remission of depressive symptoms	– The proportion of subjects achieving (at Week 12) and sustaining remission at Weeks 18 and 24. Remission is defined as Montgomery-Asberg Depression Rating Scale (MADRS) total score ≤12.
– Achieving and sustaining response of depressive symptoms	– The proportion of subjects achieving (at Week 12) and sustaining response at Weeks 18 and 24. Response is defined as ≥50% improvement from baseline MADRS total score.
– Improving depressive symptoms in subjects with significant insomnia versus those without significant insomnia	– Change from baseline to Weeks 12, 18, and 24 in the MADRS total score in subjects with significant insomnia (baseline Insomnia Severity Index [ISI] score ≥15) versus those without significant insomnia (baseline ISI

Objectives	Endpoints
	score <15).
– Improving anxiety symptoms	– Change from baseline to Weeks 12, 18, and 24 in the 14-item Hamilton Anxiety Rating Scale (HAM-A) total score.
• To assess the effect of JNJ-42847922 compared to quetiapine XR as adjunctive therapy for MDD on metabolic parameters during long-term (6 months) treatment, including:	
– Weight	• The proportion of subjects with weight gain ≥7% of baseline body weight at end-of-study assessment.
– Lipids	Triglycerides • The proportion of subjects with shifts in triglycerides from normal to high (<150 mg/dL at baseline to ≥200 mg/dL at any post-baseline assessment). • The proportion of subjects with shifts in triglycerides from borderline to high (≥150 and <200 mg/dL at baseline to ≥200 mg/dL at any post-baseline assessment). • The proportion of subjects with shifts in triglycerides from normal to very high (<150 mg/dL at baseline to ≥500 mg/dL at any post-baseline assessment). • The proportion of subjects with shifts in triglycerides from borderline to very high (≥150 mg/dL and <200 mg/dL at baseline to ≥500 mg/dL at any post-baseline assessment). • The proportion of subjects with shifts in triglycerides from high to very high (≥200 mg/dL and <500 mg/dL at baseline to ≥500 mg/dL at any post-baseline assessment).
– Blood Glucose	Fasting glucose • The proportion of subjects with shifts in fasting blood glucose from normal to borderline (<100 mg/dL at baseline to ≥100 and <126 mg/dL at any post-baseline assessment). • The proportion of subjects with shifts in fasting blood glucose from borderline to high (≥100 to <126 mg/dL at baseline to

Objectives	Endpoints
	≥126 mg/dL at any post-baseline assessment). • The proportion of subjects with shifts in fasting blood glucose from normal to high (<100 mg/dL at baseline to ≥126 mg/dL at any post-baseline assessment).
• To assess the efficacy of JNJ-42847922 compared to quetiapine XR as adjunctive therapy for MDD in the improvement of global depressive symptom severity.	• Change from baseline to Weeks 12 and 24 in the Clinical Global Impression-Severity (CGI-S) score.
• To characterize subject perceptions of global MDD severity.	• Change from baseline to Weeks 12 and 24 in the Patient Global Impression-Severity (PGI-S) score.
• To assess the effect of JNJ-42847922 compared to quetiapine XR as adjunctive therapy for MDD on the subject's assessment of quality of life.	• Change from baseline to Weeks 12 and 24 in Quality of Life in Depression Scale (QLDS).
• To assess the effect of JNJ-42847922 compared to quetiapine XR as adjunctive therapy for MDD on the subject's assessment of sleep impairment and daytime problems due to lack of sleep.	• Change from baseline to Weeks 12 and 24 in Patient Reported Outcomes Measurement Information System-Sleep Disturbance (PROMIS-SD Short Form 8a). • Change from baseline to Weeks 12 and 24 in Patient Reported Outcomes Measurement Information System-Sleep-Related Impairment (PROMIS-SRI Short Form 8a).
• To assess the effect of JNJ-42847922 compared to quetiapine XR as adjunctive therapy for MDD on the subject's assessment of reduction of depressive symptoms.	• Change from baseline to Weeks 12 and 24 in Symptoms of Major Depressive Disorder Scale (SMDDS)
• To assess the effect of JNJ-42847922 compared to quetiapine XR as adjunctive therapy for MDD on cognitive function.	• Change from baseline to Weeks 6, 12, and 24 in Symbol Digit Modalities Test (SDMT). • Change from baseline to Weeks 6, 12, and 24 in Trail Making Test - Part B (TMT-Part B). • Change from baseline to Weeks 6, 12, and 24 in Hopkins Verbal Learning Test-Revised (HVLT-R).
CCI [redacted]	• Change from baseline to Weeks 6 and 24 in salivary cortisol levels as measured at home upon awakening and at home during the evening.

Objectives	Endpoints
• To assess the safety and tolerability of JNJ-42847922 compared with quetiapine XR as adjunctive treatment in subjects with MDD.	• Safety assessments including: – Adverse events (AEs). – Proportion of all serious adverse events (SAEs) and events of special interest (e.g., parasomnias). – Vital signs, physical examinations, electrocardiogram (ECG), and laboratory parameters. – Sexual functioning using the Arizona Sexual Experiences Scale (ASEX). – Extrapyramidal symptoms assessed by the Extrapyramidal Symptom Rating Scale-Abbreviated (ESRS-A). – Suicidality assessed using the Columbia Suicide Severity Rating Scale (C-SSRS). – Potential withdrawal effects assessed by the Physician Withdrawal Checklist (PWC).
Exploratory	
• To collect medical resource utilization (MRU) data that may be used in future economic modeling.	• Healthcare Resource Use Questionnaire (HRUQ).
• To assess the effect of JNJ-42847922 compared with quetiapine XR as adjunctive therapy for MDD on self-reported measures of sleep.	• Change from baseline in subjective sleep parameters as measured by the Consensus Sleep Diary-Morning Administration (CSD-M) after the first 6 weeks of treatment (using the average of 3 morning measurements in Week 6). • Change from baseline in subjective sleep parameters as measured by the CSD-M after 3 to 6 months of treatment (using the average of 3 morning measurements in Weeks 12, 18, and 24, respectively).
CCI [redacted]	• Correlation between baseline biomarkers and clinical outcome at Week 6. • Correlation between change in baseline biomarker at Weeks 6 and 24 and clinical outcome.

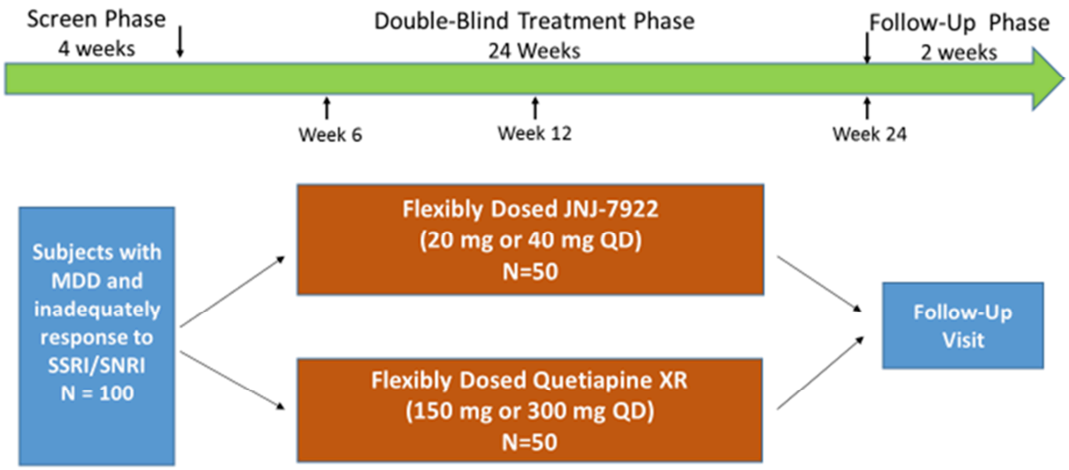
1.2. Trial Design

This is a multicenter, randomized, double-blind, active-controlled, parallel-group, flexible-dose, 6-month study in adult subjects with MDD who have an inadequate response to current antidepressant therapy with an SSRI/SNRI.

The study will consist of 3 phases: an up to 4-week screening phase, a 6-month double-blind treatment phase, and a 2-week follow-up phase culminating in a follow-up/end-of-study visit. A total of 100 subjects will be randomized in a 1:1 ratio to receive either flexibly dosed JNJ-42847922 (20 or 40 mg) or flexibly dosed quetiapine XR (150 or 300 mg) as adjunctive therapy to their current SSRI or SNRI. Subjects will continue to take their baseline SSRI/SNRI antidepressant (at the same dose, without change, every day and at approximately the same time as prior to entering the study) throughout the screening, double-blind, and follow-up phases.

A schematic overview of the study design is provided in Figure 1.

Figure 1: Schematic Overview of the Study Design



1.3. Statistical Hypotheses for Trial Objectives

The primary efficacy endpoint is the time to all-cause discontinuation of study drug. The null hypothesis to be tested to address the primary objective of the study is that there is no difference between JNJ-42847922 and quetiapine XR, as adjunctive treatment to an SSRI/SNRI, in the time to all-cause discontinuation of study drug, based on the primary efficacy endpoint.

1.4. Sample Size Justification

The primary purpose of the study is to investigate potential differentiating features between JNJ-42847922 and quetiapine XR. Although the sample size was not chosen based on a specific

set of endpoints, for evaluating the time to discontinuation of study drug in the JNJ-42847922 group versus the quetiapine XR group, a one-sided log-rank test with an overall sample size of 100 subjects (50 in the JNJ-42847922 group and 50 in the quetiapine XR group) provides approximately 92% power at a 1-sided 0.10 significance level to detect a hazard ratio of 0.415, assuming the proportion of subjects who discontinue from study drug in the quetiapine XR group is 50%. The planned sample size of the study is consistent with the nature of a Phase 2 exploratory study.

1.5. Randomization and Blinding

Central randomization will be implemented in this study. Subjects will be randomly assigned to 1 of 2 treatment groups (JNJ-42847922 or quetiapine XR) based on a computer-generated randomization schedule prepared before the study by or under the supervision of the sponsor. The randomization will be balanced by using randomly permuted blocks and will be stratified by insomnia status (significant insomnia symptoms [ISI score ≥15] at baseline versus no significant insomnia symptoms [ISI score <15] at baseline).

The interactive web response system (IWRS) will assign a unique treatment code, which will dictate the treatment assignment and matching study drug kit for the subject. The requestor must use his or her own user identification and personal identification number when contacting the IWRS, and will then give the relevant subject details to uniquely identify the subject.

To maintain the study blind, the study drug container will have a label containing the study name, study drug number, and reference number. The label will not identify the study drug in the container. However, if it is necessary for a subject's safety, the study blind may be broken and the identity of the study drug ascertained. The study drugs will be identical in appearance and will be packaged in identical containers.

Under normal circumstances, the blind should not be broken until all subjects have completed the study and the database is finalized. Otherwise, the blind should be broken only if specific emergency treatment/course of action would be dictated by knowing the treatment status of the subject. In such cases, the investigator may in an emergency determine the identity of the treatment by contacting the IWRS.

2. GENERAL ANALYSIS DEFINITIONS

2.1. Study Reference Start and End Dates

The overall reference start date for the study is defined as the date of the first dose of double-blind medication (the date is missing for screened subjects who did not receive a dose of double-blind medication). The overall reference end date for the study is the end of trial date including the last follow-up visit, i.e., the study reference end date is the maximum of the date of the last visit in the double-blind treatment phase, or date of the last visit in the follow-up phase, or date of disposition in the double-blind phase (Trial Disposition for Double-blind Phase case report form [CRF] page), or date of disposition in the follow-up phase (Disposition at Follow-up CRF page).

2.2. Analysis Phases

There are 3 analysis phases defined in this study: Screening, Double-blind, and Follow-up (post double-blind). Each analysis phase has its own analysis reference start date.

Screening

The screening phase begins on the date informed consent is obtained and ends 1 day prior to the date of the first dose of study agent in the double-blind treatment phase. The screening phase end date is left missing for those subjects who did not receive study agent.

Double-blind Phase

The analysis reference start date of the double-blind analysis phase is the date of the first dose of double-blind medication. The analysis reference end date of the double-blind analysis phase (except for Adverse Events) 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 (Trial Disposition for Double-blind Phase CRF page). For Adverse Events, the analysis reference end date of the double-blind analysis phase is the date of the last dose of study drug plus 2 days. For randomized subjects who did not receive any medication in the double-blind phase, both analysis reference start and end dates are missing for the double-blind analysis phase.

Follow-up Phase

Start and end dates for the follow-up phase are only defined for subjects 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 (Disposition at Follow-up CRF page).

2.3. Visit Windows

As subjects 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 (which is the first day the study agent was taken in the double-blind phase). If a subject has 2 or more actual visits in one 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 end point (DB). Listed below (Table 1) are the 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/Time)
MADRS, Vital Signs, C-SSRS	Screening	1	Screening	<1	-28 to -1
	DB	2	Baseline	≤1	1
		4	Week 2	2 to 21	14
		5	Week 4	22 to 35	28
		6	Week 6	36 to 63	42
		7	Week 12	64 to 105	84
		8	Week 18	106 to 147	126
		9	Week 24	148 to end of DB	168
		DB last visit	End point (DB)	2 to end of DB	
	Follow-up	10	Follow-up	End of DB + 1 to end of FU	182
CGI-S, PGI-S	DB	2	Baseline	≤1	1
		4	Week 2	2 to 21	14
		5	Week 4	22 to 35	28
		6	Week 6	36 to 63	42
		7	Week 12	64 to 105	84
		8	Week 18	106 to 147	126
		9	Week 24	148 to end of DB	168
		DB last visit	End point (DB)	2 to end of DB	
	Follow-up	10	Follow-up	End of DB + 1 to end of FU	182
HAM-A	DB	2	Baseline	≤1	1
		6	Week 6	2 to 63	42
		7	Week 12	64 to 105	84
		8	Week 18	106 to 147	126
		9	Week 24	148 to end of DB	168
		DB last visit	End point (DB)	2 to end of DB	
	Follow-up	10	Follow-up	End of DB + 1 to end of FU	182
SDMT, TMT-Part B, HVLT-R	DB	2	Baseline	≤1	1
		6	Week 6	2 to 63	42
		7	Week 12	64 to 126	84
		9	Week 24	127 to end of DB	168
		DB last visit	End point (DB)	2 to end of DB	
QLDS, SMDDS	DB	2	Baseline	≤1	1
		4	Week 2	2 to 21	14
		5	Week 4	22 to 35	28
		6	Week 6	36 to 63	42
		7	Week 12	64 to 105	84
		8	Week 18	106 to 147	126
		9	Week 24	148 to end of DB	168
		DB last visit	End point (DB)	2 to end of DB	
PROMIS-SD, PROMIS-SRI	DB	2	Baseline	≤1	1
		4	Week 2	2 to 28	14
		6	Week 6	29 to 63	42
		7	Week 12	64 to 105	84
		8	Week 18	106 to 147	126
		9	Week 24	148 to end of DB	168
		DB last visit	End point (DB)	2 to end of DB	
CSD-M	DB	Prior to Visit 2	Average Predose	≤1	-2, -1, 1
		Prior to Visit 6	Week 6 (Average)	2 to 63	40, 41, 42

Parameter	Analysis Phase	Scheduled Visit Number	Time Interval (label on output)	Time Interval (Day)*	Target Time Point (Day/Time)
		Prior to Visit 7	Week 12 (Average)	64 to 105	82, 83, 84
		Prior to Visit 8	Week 18 (Average)	106 to 147	124, 125, 126
		Prior to Visit 9	Week 24 (Average)	148 to end of DB	166, 167, 168
		DB last visit	End point (DB)	2 to end of DB	
HRUQ	DB	2	Baseline	≤1	1
		6	Week 6	2 to 63	42
		7	Week 12	64 to 105	84
		8	Week 18	106 to 147	126
		9	Week 24	148 to end of DB	168
		DB last visit	End point (DB)	2 to end of DB	
	Follow-up	10	Follow-up	End of DB + 1 to end of FU	182
Physical Exam	Screening	1	Screening	<1	-28 to -1
	DB	1	Baseline	≤1	-28 to -1
		9	Week 24	2 to end of DB	168
		DB last visit	End point (DB)	2 to end of DB	
ISI	Screening	1	Screening	<1	-28 to -1
	DB	2	Baseline	≤1	1
Serum Chemistry, Urinalysis, Hematology, Lipid Panel, Weight, BMI, Waist Circumference	Screening	1	Screening	<1	-28 to -1
	DB	2	Baseline	≤1	1
		6	Week 6	2 to 63	42
		7	Week 12	64 to 126	84
		9	Week 24	127 to end of DB	168
		DB last visit	End point (DB)	2 to end of DB	
HbA1c, TSH, FT_4	Screening	1	Screening	<1	-28 to -1
	DB	1	Baseline	≤1	-28 to -1
		7	Week 12	2 to 126	84
		9	Week 24	127 to end of DB	168
		DB last visit	End point (DB)	2 to end of DB	
ECG	DB	1	Screening	<1	-28 to -1
		1, 2	Predose	≤1	-28 to 1
		1, 2	Average Predose	≤1	1
		6	Week 6	2 to 105	42
		9	Week 24	106 to end of DB	168
		DB last visit	End point (DB)	2 to end of DB	
ASEX	DB	2	Baseline	≤1	1
		7	Week 12	2 to 126	84
		9	Week 24	127 to end of DB	168
		DB last visit	End point (DB)	2 to end of DB	
ESRS-A	DB	4	Week 2	2 to 21	14
		5	Week 4	22 to 35	28
		6	Week 6	36 to 63	42
		7	Week 12	64 to 126	84
		9	Week 24	127 to end of DB	168
		DB last visit	End point (DB)	2 to end of DB	
PWC[a]	DB	10	End point (DB)	2 to end of DB	

Parameter	Analysis Phase	Scheduled Visit Number	Time Interval (label on output)	Time Interval (Day)*	Target Time Point (Day/Time)
	Follow-up	10	Follow-up	End of DB + 1 to end of FU	182

*Relative to Study Day 1; DB=double-blind; FU=follow-up
[a] PWC was collected at one timepoint. For subjects who discontinued the study, PWC was collected at time of withdrawal. Otherwise, it was collected at the follow-up visit.

2.4. Pooling Algorithm for Analysis Centers

Subjects will be enrolled at sites in the United States (US). Actual site enrollment rates will be monitored to avoid gross imbalances across centers. Sites will not be pooled for analyses.

2.5. Analysis Sets

Subjects will be classified into the following analysis sets: all randomized, full analysis set, and safety analysis set.

2.5.1. All Randomized Analysis Set

The all randomized analysis set includes all subjects who were randomized in the study (i.e., subjects who reported a randomization date, or were assigned a randomization number) regardless of whether or not study agent was received.

2.5.2. Efficacy Analysis Set

2.5.2.1. Full Analysis Set

The full analysis set includes all randomized subjects who received at least 1 dose of study agent except subjects who received potentially defective investigational products (IP) and were subsequently discontinued early because of the IP issue.

2.5.3. Safety Analysis Set

The safety analysis set includes all randomized subjects who received at least 1 dose of study agent.

2.6. Definition of Subgroups

Subgroup	Definition
Baseline ISI score	• ≥15 • <15 using the randomization stratification factor as entered into IWRS; this will also be used in the analyses as the baseline insomnia status
Baseline MADRS total score	• ≤ median • > median
MDD subtype based on SCID-CT	• With anxious distress • Without anxious distress based on the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition (DSM-5) Axis I Disorders – Clinical Trials Version (SCID-CT)
Sex	• Male

Subgroup	Definition
	• Female • Undifferentiated
Age Group	• 18-34 • 35-54 • 55-64 • 65-70
Sex and Age Group	• Adults (18-64) male, Elderly (≥65) male, Adults (18-64) female, Elderly (≥65) female
MDD with Anxiety based on HAMA-A	• Yes, defined as baseline HAM-A total score ≥18 • No, defined as baseline HAM-A total score <18
Class of antidepressant study medication	• SNRI • SSRI
Final dose	• Seltorexant 20mg • Seltorexant 40mg • Quetiapine XR 150mg • Quetiapine XR 300mg
Number of antidepressants of sufficient dose and with inadequate response taken for at least 4 weeks during the current episode	• 1 • 2 or more based on the Massachusetts General Hospital-Antidepressant Treatment Response Questionnaire (MGH-ATRQ)
Number of major depressive episodes in lifetime, including current episode	• ≤3 • >3
Prior medication use of adjunctive antipsychotics	• Yes • No
Prior medication use of benzodiazepines	• Yes • No
Baseline body mass index (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
Race	• categories as collected on eCRF
Ethnicity	• categories as collected on eCRF

2.7. Study Day and Relative Day

Study Day 1 or Day 1 refers to the start of the first study agent administration. All efficacy and safety assessments at all visits will be assigned a day relative to this date.

Study day or relative day for a visit is defined as:

- Visit date - (date of Study Day 1) +1, if visit date is ≥date of Day 1
- Visit date - Date of Day 1, if visit date <date of Day 1

There is no 'Day 0'.

2.8. Baseline and End point (DB)

Baseline is defined as the last observation prior to the start of the first study agent administration, with the exception of the average predose ECG measurement and the average predose CSD-M measurement, which is defined as the average of all predose ECG results collected up to and

including the day of the first dose of study agent and the average of all predose CSD-M measurements collected up to and including the day of the first dose of study agent, respectively.

End point (DB) is defined as the last available postbaseline result within the double-blind phase. Unscheduled visit results are included in this definition and will be considered as the end point value if the unscheduled visit result is the last postbaseline result available within the double-blind phase.

2.9. Imputation Rules for Missing AE Date/Time of Onset/Resolution

Partial AE onset dates will be imputed as follows:

- If the onset date of an AE is missing day only, it will be set to:
  - First day of the month that the AE occurred, if month/year of the onset of AE is different than the month/year of the study agent start
  - The day of study agent start, if the month/year of the onset of AE is the same as month/year of the study agent start date and month/year of the AE resolution date is later
  - The day of study agent start or day of AE resolution date, whichever is earliest, if month/year of the onset of AE and month/year of the study agent start date and month/year of the AE resolution date are same.
- If the onset date of an AE is missing both day and month, it will be set to the earliest of:
  - January 1 of the year of onset, as long as this date is on or after the study agent start date
  - Month and day of the study agent start date, if this date is the same year that the AE occurred
  - Last day of the year if the year of the AE onset is prior to the year of the study agent start date
  - The AE resolution date.
- Completely missing onset dates will not be imputed.

Partial AE resolution dates not marked as ongoing will be imputed as follows:

- If the resolution date of an AE is missing day only, it will be set to the earliest of the last day of the month of occurrence of resolution or the day of the date of death, if death occurred in that month.
- If the resolution date of an AE is missing both day and month, it will be set to the earliest of December 31 of the year or the day and month of the date of death, if death occurred in that year.
- Completely missing resolution dates will not be imputed.

AE onset/resolution dates with missing times will be imputed as follows:

- A missing time of onset of an AE will be set to:

- 00:01 as long as the onset date is different from the study agent start date
- The time of the study agent start if this is the same day the AE occurred.

- The missing time of resolution of an adverse event will be set to 23:59.

If a missing time is associated with a partial or missing date, the date will be imputed first prior to imputing the time.

2.10. Imputation Rules for Missing Prior/Concomitant Medication Dates

2.10.1. Prior Medications

Prior medications or therapy are those taken by subjects before the start of dosing of first study agent. Medications will be classified as prior if the medication start date is complete and prior to the date of first dose of study agent or the medication end date is complete and prior to the date of first dose of study agent.

If the medication start day is missing, and the month and year of the start date are not missing, then if:

- The month and year of the start date of medication is earlier than the month and year of the initial study agent administration; or

- The CRF indicates the medication was taken prior (prior medication flag=Yes) and the month and year of the start date of medication is the same as the month and year of the initial study agent administration

then the medication will be considered prior.

If the medication start month and day are missing, and the year of the start date is not missing, then if:

- The year of the start date of medication is earlier than the year of the initial study agent administration; or

- The CRF indicates the medication was taken prior when the year of the start date of medication is the same as the year of the initial study agent administration

then it will also be considered prior.

If the medication start date is completely missing, and the CRF indicates it was taken prior, it will also be considered prior.

2.10.2. Concomitant Medications Taken During the Double-blind Phase

Concomitant medications taken during the double-blind phase are those that started on the same day as the first dose or after the start of dosing or those continuing from predose (prior medication flag=Yes) and the CRF indicates the medication is ongoing or the medication stop

date is on or after the first dose of study agent. Medications that start after the analysis reference end date of the double-blind phase are not considered concomitant medications taken during the double-blind phase.

If the medication start date is missing the day, but the month and year are complete, then if the month and year of the start date are on or prior to the month and year of the analysis reference end date of the double-blind phase and:

- The CRF indicates that the medication is ongoing, or
- The stop date of the medication is missing, or
- The stop date of the medication is on or after the initial study agent administration, or
- The stop date month and year (day is missing) of the medication is on or after the initial study agent administration, or
- The stop date year (day and month are missing) of the medication is on or after the initial study agent administration

then the medication is classified as concomitant during the double-blind phase.

If the medication start date is missing the month and day, but the year is complete, then if the year of the start date is the same as or prior to the year of the analysis reference end date of the double-blind phase and:

- The CRF indicates that the medication is ongoing, or
- The stop date of the medication is missing, or
- The stop date of the medication is on or after the initial study agent administration, or
- The stop date month and year (day is missing) of the medication is on or after the initial study agent administration, or
- The stop date year (day and month are missing) of the medication is on or after the initial study agent administration

then the medication is classified as concomitant during the double-blind phase.

If the medication start date is completely missing then if:

- The CRF indicates that the medication is ongoing, or
- The stop date of the medication is missing, or
- The stop date of the medication is on or after the initial study agent administration, or

- The stop date month and year (day is missing) of the medication is on or after the initial study agent administration, or

- The stop date year (day and month are missing) of the medication is on or after the initial study agent administration

then the medication is classified as concomitant during the double-blind phase.

For concomitant medications categorized as "taken during DB" based on the rules above, the duration during the DB phase will be calculated as: Minimum of medication end date and the DB End date - Maximum of the medication start date and the DB start date + 1. If there are partial start/end dates, the imputation rules for the duration during the DB phase are as follows:

If the month of the start date is not missing but the day is, then the duration is imputed as:

> Minimum of the medication end date and the DB End Date - Maximum of the first day of the month and the DB Start Date + 1;

If the year of the start date is not missing but the month is, then the duration is imputed as:

> Minimum of the medication end date and the DB End Date - Maximum of the first day of the year and the DB Start Date + 1;

If the month of the end date is not missing but the day is, then the duration is imputed as:

> Minimum of the last day of the month and the DB End Date - Maximum of the medication start date and the DB Start date + 1;

If the year of the end date is not missing but the month is, then the duration is imputed as:

Minimum of the last day of the year and the DB End Date - Maximum of the medication start date and the DB Start date + 1.

2.10.3. Concomitant Medications Taken During the Follow-up Phase

Follow-up concomitant medications are those that started after the analysis reference end date of the double-blind phase.

3. INTERIM ANALYSIS AND DATA MONITORING COMMITTEE REVIEW

There is no planned interim analysis for this study.

4. SUBJECT INFORMATION

The number of subjects in each analysis set will be summarized and listed by treatment group and overall. In addition, the distribution of subjects by site ID will be presented for the full analysis set.

4.1. Demographics and Baseline Characteristics

Table 2 presents a list of the demographic variables that will be summarized by treatment group and overall for the full analysis set.

Table 2: Demographic Variables

Continuous Variables:	**Summary Type**
Age (years), calculated based on date of informed consent	Descriptive statistics (N, mean, standard deviation [SD], median and range [minimum and maximum]).
Weight (kg)	
Height (cm)	
Body Mass Index (BMI) (kg/m^2) calculated as Weight (kg)/[Height (m)]^2	
Waist circumference (cm)	
Categorical Variables	
Age (18-34 years, 35-54 years, 55-64 years, 65-70 years)	Frequency distribution with the number and percentage of subjects in each category.
Sex (male, female, undifferentiated)	
Race[a] (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or other Pacific Islander, White, Not Reported, Multiple)	
Ethnicity (Hispanic or Latino, not Hispanic or Latino)	
BMI (underweight <18.5 kg/m^2, normal 18.5-<25 kg/m^2, overweight 25-<30 kg/m^2, obese ≥30 kg/m^2)	

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

Table 3 presents a list of the psychiatric history variables that will be summarized by treatment group and overall for the full analysis set.

Table 3: Psychiatric History Variables

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 PGI-S score	
Baseline ISI score per eDC	
Categorical Variables	
Current antidepressant type (SSRI, SNRI)	Frequency distribution with the number and percentage of subjects in each category.
Antidepressant treatment history (number of medications with inadequate response taken for at least 4 weeks during the current episode as obtained in the MGH-ATRQ)	
SCID-CT DSM-5 specifiers for MDD (anxious distress, mixed features, melancholic features, atypical features, peripartum onset, seasonal pattern)	
MDD with Anxiety based on HAMA-A (Yes=defined as baseline HAM-A total score ≥18, No=defined as baseline HAM-A total score <18)	
Prior medication use of adjunctive antipsychotics (Yes, No)	
Prior medication use of benzodiazepines (Yes, No)	
Baseline CGI-S score (1=normal [not at all ill]; 2=borderline ill; 3=mildly ill; 4=moderately ill; 5=markedly ill; 6=severely ill; 7=among the most extremely ill patients)	
Baseline PGI-S score (1=none, 2=mild, 3=moderate, 4=severe)	
Baseline ISI score per IWRS (≥15, <15)	

Baseline ISI score per eDC (≥15, <15) (if different from IWRS)	
Number of major depressive episodes (1, 2, ≥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)	

By-subject listings of the demographic and baseline characteristics as well as general medical history, psychiatric history and tobacco use history will be provided.

4.2. Disposition Information

The number of screen failures will be summarized overall.

The number of subjects in the following disposition categories will be summarized throughout the study by treatment group and overall:

- Subjects randomized
- Subjects receiving study agent
- Subjects who completed the double-blind phase
- Subjects who discontinued from the double-blind phase
    - Reasons for discontinuation from the double-blind phase
- Subjects completing the follow-up phase

Listings of subjects will be provided for the following categories:

- All randomized subjects
- Subjects who discontinued from the double-blind phase
- Subjects who discontinued from the follow-up phase
- Subjects who were unblinded during the study period
- Subjects who were randomized yet did not receive study agent.

These summaries and listings will be provided for the all randomized analysis set.

4.3. Treatment Compliance

4.3.1. Treatment Compliance With Study Agent

Study agent compliance will be summarized descriptively.

Compliance for each subject will be calculated based on the percent of the scheduled number of capsules of study agent actually taken within the double-blind phase. Study agent compliance will be calculated as follows:

Study agent compliance (%) = (actual number of capsules taken/total number of capsules supposed to be taken) x100.

The number of capsules supposed to have been taken will be calculated as the duration of treatment within the phase (i.e., date of last dose of study agent – date of first dose of study agent + 1) multiplied by 2 (since each "dose" of study agent consists of 2 capsules).

Descriptive statistics on the percent (%) compliance will be summarized by treatment group for the safety analysis set. In addition, percent compliance will be categorized as <60%, 60%-<80%, 80%-100%, >100% and the number and percentage of subjects in each category will be summarized.

4.3.2. Treatment Compliance With Baseline SSRI/SNRI Antidepressant

Compliance with baseline SSRI/SNRI antidepressant will be summarized descriptively.

Compliance for each subject will be calculated based on the percent of the scheduled number of days that the baseline SSRI/SNRI antidepressant was actually taken within the double-blind phase. It is defined as:

Compliance with baseline SSRI/SNRI (%) = (number of days baseline SSRI/SNRI antidepressant taken / number of days in double-blind phase)*100.

The number of days in the double-blind phase will be calculated as the duration of double-blind treatment phase (i.e., completion or discontinuation date of the double-blind phase – date of first dose of study drug + 1).

Descriptive statistics on the percent compliance will be summarized by treatment group for the safety analysis set. In addition, percent compliance will be categorized as <60%, 60%-<80%, 80%-100%, >100% and the number and percentage of subjects in each category will be summarized.

4.4. Extent of Exposure

Total duration of exposure (including days off drug) is defined as (date of last dose of study agent – date of first dose of study agent) + 1. Number of doses is defined as the total number of dose administrations. Taking more than 2 tablets within 4 hours is considered as overdose.

Descriptive statistics (N, mean, SD, median, minimum, and maximum) for total duration of exposure (including days off drug) and for number of doses will be presented by treatment group for the safety analysis set.

The mode daily dose will be calculated for each subject, based on days on drug. The lowest mode dose is selected in case of ties. The frequency of mode daily dose will be presented by treatment group.

The mean daily dose of study agent for a subject is calculated as the sum of total daily dose during the treatment phase divided by the total number of days exposed. The final dose is the last non-zero dose received during the double-blind phase. The calculation of mean daily dose and final dose will exclude days off study drug. Descriptive statistics (N, mean, SD, median, minimum and maximum) for mean daily dose and final dose will be presented by treatment group.

A by-subject listing of study drug administration will be provided.

4.5. Protocol Deviations

In general, the following list of major protocol deviations may have the potential to impact subjects' rights, safety or well-being, or the integrity and/or result of the clinical study. Subjects with major protocol deviations will be identified prior to database lock and the number and percentage of subjects with major protocol deviations will be summarized by category using the full analysis set.

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

More categories may be included depending on the nature of the protocol deviation. A subject may be counted in more than one deviation category.

A by-subject listing showing the specific major protocol deviations will also be provided.

A summary of the number of subjects in the safety analysis set not meeting each inclusion/exclusion criterion will be presented.

4.6. Prior and Concomitant Medications

Prior 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 agent. Concomitant medications are defined as any therapy used on or after the same day as the first dose of study agent, including those that started before and continue after the first dose of study agent.

If the medication start date is recorded as partial or completely missing, then the medication will be considered to be concomitant unless it is known to be prior to the first administration of study drug based on partial start date or stop date or the CRF indicates that the medication was taken

prior (prior medication flag=Yes) (see Section 2.10 for detailed classification of prior and concomitant medications).

Prior medications will be summarized by treatment group and standardized medication name for the safety analysis set. The proportion of subjects who receive each prior medication will be summarized as well as the proportion of subjects who receive at least one prior medication. In addition, the number and percent of subjects who receive antidepressant medications prior to the study will be summarized.

Summaries of concomitant medications (other than antidepressant medications) will be presented by treatment group and base preferred term for the safety analysis set, for those medications used during the double-blind phase and for those used during the follow-up phase separately. Definitions for concomitant medications used during the double-blind phase and those used during the follow-up phase are provided in Section 2.10.2 and 2.10.3, respectively. The proportion of subjects who receive each concomitant medication will be summarized as well as the proportion of subjects who receive at least one concomitant medication.

In addition, summary tables of the antidepressant concomitant medications received during the study will be presented by treatment group, for those medications used during the double-blind phase and for those used during the follow-up phase separately.

A by-subject listing of all prior and concomitant medication will also be provided.

5. EFFICACY

All efficacy analyses will be based on the full analysis set.

The efficacy variables for this study are listed in Table 4.

Table 4: Efficacy Variables

Efficacy Variable		Endpoint
Time to all-cause discontinuation of study drug	• Time to all-cause discontinuation of study drug	Primary
MADRS	• Change from baseline to Weeks 2, 4, 6, 12, 18, and 24 in the MADRS total score	Secondary
	• The proportion of subjects achieving (at Week 12) and sustaining remission at Weeks 18 and 24. Remission is defined as MADRS total score ≤12.	Secondary
	• The proportion of subjects achieving (at Week 12) and sustaining response at Weeks 18 and 24. Response is defined as ≥50% improvement from baseline MADRS total score.	Secondary

Efficacy Variable		Endpoint
	• Change from baseline to Weeks 12, 18, and 24 in the MADRS total score in subjects with baseline ISI score ≥15 versus subjects with baseline ISI score <15	Secondary
	• Change from baseline to Weeks 2, 4, 6, 12, 18, and 24 in the MADRS-6 total score	Secondary
	• Change from baseline to Weeks 2, 4, 6, 12, 18, and 24 in the MADRS-WOSI total score	Exploratory
HAM-A	• Change from baseline to Weeks 6, 12, 18, and 24 in the HAM-A total score	Secondary
CGI-S	• Change from baseline to Weeks 2, 4, 6, 12, 18 and 24	Secondary
PGI-S	• Change from baseline to Weeks 2, 3, 6, 12, 18 and 24	Secondary
QLDS	• Change from baseline to Weeks 2, 4, 6, 12, 18 and 24	Secondary
PROMIS-SD	• Change from baseline to Weeks 2, 6, 12, 18 and 24	Secondary
PROMIS-SRI	• Change from baseline to Weeks 2, 6, 12, 18 and 24	Secondary
SMDDS	• Change from baseline to Weeks 2, 4, 6, 12, 18 and 24	Secondary
SDMT	• Change from baseline to Weeks 6, 12, and 24	Secondary
TMT-Part B	• Change from baseline to Weeks 6, 12, and 24	Secondary
HVLT-R	• Change from baseline to Weeks 6, 12, and 24	Secondary
CSD-M	• Change from baseline in subjective sleep parameters after the first 6 weeks of treatment	Exploratory
	• Change from baseline in subjective sleep parameters after 3 to 6 months of treatment (using the average of up to 7 morning measurements in Weeks 12, 18, and 24, respectively)	Exploratory

5.1. Analysis Specifications

5.1.1. Level of Significance

The primary efficacy endpoint will be evaluated at a two-sided significance level of 0.20. For all other efficacy endpoints, no multiplicity adjustment will be done and point estimates and confidence intervals will be presented.

5.1.2. Data Handling Rules

As sensitivity analyses, the change in MADRS total score and HAM-A total score will be analyzed using an analysis of covariance (ANCOVA) model, using last observation carried forward (LOCF) data. These analyses are further described in Sections 5.3.1.2 and 5.3.4.2.

For the analyses of remission of depressive symptoms based on MADRS total score (Section 5.3.2) and response of depressive symptoms based on MADRS total score (Section 5.3.3), subjects with missing values will be imputed as non-responders/non-remitters.

For the graphical presentations of the cumulative response rates, defined as the percentage of subjects experiencing at least a given value of percent reduction from baseline to Week 24 for MADRS total score (Section 5.3.3.2), both observed and LOCF data will be presented.

5.1.3. Imputation Methods for Missing Items

For MADRS, imputation of the total score when there are missing items is described in Section 5.3.1.1. Likewise, for PROMIS-SD and PROMIS-SRI, imputation of the total score when there are missing items is described in Sections 5.3.8.1 and 5.3.9.1, respectively. For QLDS, imputation of the total score when there are missing items is described in Section 5.3.7.1. For all other scales where multiple items are summed to create a total, if any item of the scale is missing at a visit, the total score for that scale at that visit will be left blank.

5.2. Primary Efficacy Endpoint

5.2.1. Definition

The primary efficacy endpoint, time to all-cause discontinuation of study drug, is defined as the number of days from the first dose of study drug to the last dose of study drug. Subjects who complete double-blind treatment are not considered to have discontinued.

5.2.2. Estimand

The primary estimand, the main clinical quantity of interest to be estimated in the study, is defined by the following 4 components:

Population: subjects with MDD who have had an inadequate response to current antidepressant therapy with an SSRI/SNRI, as defined through appropriate inclusion/exclusion criteria to reflect the targeted patient population;

Variable: time to all-cause discontinuation of study drug (see Section 5.2.1);

Intervention Effect: the effect of the initially randomized treatment together with the oral SSRI/SNRI antidepressant;

Summary Measure: the difference in survival distributions.

The primary analysis will be based on the full analysis set.

5.2.3. Analysis Methods

The primary efficacy endpoint, time to all-cause discontinuation of study drug, will be estimated by the Kaplan-Meier method and summarized (number of discontinuations, number of censored subjects, median, 25th and 75th percentile, if estimable) by treatment group.

A stratified log-rank test (stratified for baseline insomnia status) will be used to test the hypothesis that there is no difference between the JNJ-42847922 group and quetiapine XR in the probability of discontinuing study drug, using a two-sided 0.20 significance level.

As a sensitivity analysis, similar to the stratified log-rank test above, a log-rank test without adjusting for stratification factor will be carried out.

As an additional sensitivity analysis, the log-rank analysis may be repeated excluding PPD site due to possible GCP issues.

The estimate of the hazard ratio and its 80% confidence interval will be provided based on the Cox proportional hazards model with treatment and baseline insomnia status as the factors. In addition, a similar model with treatment as the only factor will be carried out.

The distribution of the time to discontinuation of study agent will be displayed with Kaplan-Meier curves, for the full analysis set. Subjects who discontinued the double-blind phase at any time due to any reason will be considered an 'Event' and their date of last dose of study agent based on the exposure dataset will be used in the time to event calculation. Subjects who complete the double-blind phase will be censored and the date of last dose of study agent based on the exposure dataset will serve as the time of censoring. Kaplan-Meier curves will also be produced for the Week 12 timepoint (Day 84). Subjects who do not experience an event prior Week 12 will be censored at day 84.

In addition to time to all-cause discontinuation, the stratified log-rank analysis described above will be repeated for time to treatment-related discontinuation. Examples of potentially non-treatment related discontinuations include subjects relocating or traveling out of the area and unable to return for study visits.

5.2.3.1. Subgroup Analyses

Descriptive statistics (number of discontinuations, number of censored subjects, median, 25th and 75th percentile, if estimable) will be provided for the primary efficacy endpoint (time to all-cause discontinuation of study drug) by the subgroups identified in Section 2.6. Subgroup analysis will be performed using the Cox proportional hazards model. The model will include terms for treatment, subgroup and treatment-by-subgroup. One subgroup at a time will be included in the model. A forest plot with the corresponding hazard ratios and 80% CI will be presented for the subgroups. If there are insufficient numbers in the subgroups to draw meaningful conclusions, the subgroup analysis will not be performed. The Kaplan-Meier plots will be provided for the subgroups with baseline ISI $\geq$15 and baseline ISI <15.

5.3. Secondary Efficacy Endpoints

5.3.1. Change in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score

5.3.1.1. Definition

A secondary efficacy endpoint is the change in MADRS total score from baseline to Weeks 12, 18, and 24. The MADRS will be performed by independent, centralized remote raters during the study. The MADRS is a clinician-administered scale designed to measure depression severity and consists of 10 items, each of which is scored from 0 (item not present or normal) to 6 (severe or continuous presence of the symptoms). The MADRS evaluates apparent sadness, reported sadness, inner tension, reduced sleep, reduced appetite, concentration difficulties, lassitude, inability to feel, pessimistic thoughts, and suicidal thoughts. The recall period for the MADRS is 7 days.

A total score (0 to 60) is calculated by adding the scores of all 10 items. Higher scores represent a more severe condition. Imputation of the total score will be performed only when 1 item score is missing. If 2 or more items are missing, the total score will be left missing. The total score will be imputed by calculating the sum of the scores of the non-missing items and multiplying it by the ratio of the maximum possible number of items (i.e., 10) to the number of non-missing items (i.e., 9).

The MADRS change from baseline at Week 24 is calculated as (MADRS total score at Week 24 – Baseline MADRS total score). Change from baseline to other time points is calculated similarly. Negative changes in MADRS total score indicate improvement.

5.3.1.2. Analysis Methods

Descriptive statistics of the actual values on Baseline, Week 2, Week 4, Week 6, Week 12, Week 18, Week 24, and at follow-up and the change from baseline to each postbaseline time point will be presented for MADRS total score by treatment group. A frequency distribution of the MADRS individual item scores at each time point will be provided by treatment group.

MMRM

A mixed model for repeated measures (MMRM) analysis of the observed data from the JNJ-42847922 and quetiapine XR groups will be used to adjust for important covariates. The MMRM will include time (scheduled Week), treatment (JNJ-42847922 and quetiapine XR), baseline insomnia status (present/absent), and treatment-by-time interaction as factors and baseline MADRS total score as a covariate. Data from the follow-up phase will not be used. An unstructured variance-covariance matrix will be used for observations clustered by subject. 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. Subjects in the full analysis set who do not have complete data will still contribute to the estimates at each time point, but will have less weight in the analysis than those subjects with complete data. The

Kenward-Roger method will be used for approximating the denominator degrees of freedom. Based on the MMRM, the comparison between JNJ-42847922 and quetiapine XR will be performed using the appropriate contrasts directly from the MMRM analysis, using estimates at each postbaseline time point. A 80% confidence interval (CI) for the difference in least-squares (LS) means will be calculated based on the contrast test statistic.

As a secondary analysis, age and gender will be added to the MMRM model.

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

ANCOVA

As a sensitivity analysis, the change in MADRS total score from baseline to each postbaseline time point in the double-blind phase will be analyzed using an ANCOVA model, using LOCF and BOCF data. The ANCOVA model will include factors for treatment, baseline insomnia status (present/absent), and baseline MADRS total score as a continuous covariate. A 80% CI for the difference in LS Means will be calculated based on the contrast test statistic.

The last postbaseline observation during the double-blind phase will be carried forward as the "End Point (DB)" for that phase. Besides the end point assessment, the LOCF values will be created for intermediate postbaseline time points as well. These imputed time points will be labeled 'WEEK X LOCF'.

For example, if a subject has a visit on Day 125 for the Week 18 visit and then a final visit on Day 130, the visit on Day 125 will be slotted to "Week 18" and "Week 18 LOCF" because Day 125 is closer to the target Day 126 than Day 130. The Day 130 visit will be used as "Week 24 LOCF" and "End Point (DB)".

For subjects who discontinue treatment early, impute the missing MADRS scores post-discontinuation as the baseline value, and label as "Week xx BOCF" or "End Point (DB) BOCF".

As a secondary analysis, age and gender will be added to the ANCOVA model.

SUBGROUP ANALYSIS

Baseline ISI Score

The ISI has 7 questions, each rated on a 5-point Likert scale ranging from 0 to 4. The total score is the sum of each individual item and can range from 0 to 28 (28 = most severe insomnia).

Subjects will be dichotomized based on their baseline ISI score (≥15 versus <15) to evaluate whether there are differences in these subgroups (i.e., evaluating change from baseline to each postbaseline time point in the MADRS total score in subjects with ISI score ≥15 versus subjects

with ISI score <15). Descriptive statistics of the change from baseline to each postbaseline time point in the MADRS total score will be presented by treatment group for these subgroups.

Mode Dose

Descriptive statistics of the change from baseline to each postbaseline time point in the MADRS total score will be presented by treatment group and mode dose as defined in Section 4.4.

5.3.2. Achieving and Sustaining Remission of Depressive Symptoms

5.3.2.1. Definition

A secondary efficacy endpoint is the proportion of subjects achieving (at Week 12) and sustaining remission at Weeks 18 and 24. Remission is defined as MADRS total score ≤12. A subject must have a MADRS total score ≤12 at all 3 time points (Week 12, Week 18, and Week 24) to be defined as a sustained remitter. Subjects who do not meet such criterion will be considered as non-sustained remitters. The calculation of the MADRS total score is described in Section 5.3.1.1.

5.3.2.2. Analysis Methods

The number and percentage of subjects who achieve and sustain remission will be summarized by treatment group. In addition, the number and percentage of subjects who have a MADRS total score ≤12 at each postbaseline time point will be presented by treatment group.

The point estimate and 2-sided 80% confidence interval will be provided for the relative risk of achieving and sustaining remission using a Mantel-Haenszel test controlling for baseline insomnia status (present/absent).

5.3.3. Achieving and Sustaining Response of Depressive Symptoms

5.3.3.1. Definition

A secondary efficacy endpoint is the proportion of subjects achieving (at Week 12) and sustaining response at Weeks 18 and 24. Response is defined as ≥50% improvement from baseline MADRS total score. A subject must have a ≥50% improvement from baseline MADRS total score at all 3 time points (Week 12, Week 18, and Week 24) to be defined as a sustained responder (yes=1). Subjects who do not meet such criterion will be considered as non-sustained responders and will be assigned a value of 0 (i.e., no). The calculation of the MADRS total score is described in Section 5.3.1.1. The percentage change from baseline for MADRS total score is calculated as 100*(MADRS total score at Week X – Baseline MADRS total score)/(Baseline MADRS total score). Negative percent changes in MADRS total score indicate improvement (e.g., percent change < -50% indicates improvement >50%).

5.3.3.2. Analysis Methods

The number and percentage of subjects who achieve and sustain a response will be summarized by treatment group. In addition, the number and percentage of subjects who have a ≥50%

improvement from baseline MADRS total score at each postbaseline time point will be presented by treatment group.

The point estimate and 2-sided 80% confidence interval will be provided for the relative sustained response using a Mantel-Haenszel test controlling for baseline insomnia status (present/absent).

The cumulative response rate, defined as the percentage of subjects experiencing at least a given value of percent reduction from baseline to Week 24 in MADRS total score, will be presented graphically, for both observed and LOCF data.

The cumulative distribution function of the time to sustained response will be estimated by the Kaplan-Meier method. Time to sustained response will be summarized (number of sustained responders, number of censored subjects, median, 25th and 75th percentile, if estimable) by treatment group. Sustained response is defined as the first occurrence of response that is maintained through the Week 24 assessment. Subjects who discontinue early are not considered to have sustained response. A stratified log rank test (stratified for baseline insomnia status) will be used to test the hypothesis that there is no difference between JNJ-42847922 and quetiapine XR in the probability of achieving sustained response. In addition, time to sustained response will be summarized by modal dose.

5.3.4. Change in Hamilton Anxiety Rating Scale (HAM-A)

5.3.4.1. Definition

A secondary efficacy endpoint is the change from baseline to Weeks 12, 18, and 24 in the HAM-A total score. The original 14-item clinician-administered HAM-A scale assesses the severity of different anxiety-related symptoms. It is a commonly used, validated scale in clinical trials assessing anxiety. However, the original HAM-A lacks instructions for administration and clear anchor points for the assignment of severity ratings. For this reason, the structured interview guide version (SIGH-A) will be used in this study.

Each of the 14-items in the scale is scored on a 5-point scale, ranging from 0 (not present) to 4 (very severe, symptom is incapacitating). A total score (0 to 56) is calculated by adding the scores of all 14 items, where 0-13 indicates normal range, 14-17 indicates mild severity, 18–24 mild to moderate severity, 25–30 moderate to severe, and ≥31 severe. Higher scores represent a more severe condition.

The HAM-A change from baseline at Week 24 is calculated as (HAM-A total score at Week 24 – Baseline HAM-A total score). Change from baseline to other time points is calculated similarly. Negative changes in HAM-A total score indicate improvement.

The somatic factor score of HAM-A is defined as the sum of items 7 to 13; and the psychic factor score of HAM-A is defined as the sum of items 1-6 and 14.

5.3.4.2. Analysis Methods

Descriptive statistics of the actual values on Baseline, Week 6, Week 12, Week 18, Week 24, and at follow-up and the change from baseline to each postbaseline time point will be presented for HAM-A total score, and HAM-A somatic and psychic factor scores by treatment group.

The change from baseline in HAM-A total score will be analyzed using the same MMRM as described in Section 5.3.1.2 for MADRS total score, with the continuous covariate "baseline MADRS total score" changed to "baseline HAM-A total score".

As a sensitivity analysis, the change in HAM-A total score from baseline to each postbaseline time point will be analyzed using an ANCOVA model, using LOCF data. The derivation of the LOCF values is described in Section 5.3.1.2. The ANCOVA model will include factors for treatment, baseline insomnia status (present/absent), and baseline HAM-A total score as a continuous covariate. A 80% CI for the difference in LS Means will be calculated based on the contrast test statistic.

In addition, a frequency distribution over time of the HAM-A total score categories and HAM-A individual item scores at each time point will be provided by treatment group.

The shift in HAM-A total score categories (normal, mild, mild to moderate, moderate to severe, severe) from baseline to each time point will be presented by treatment group.

The correlations between HAM-A total score and MADRS total score will be evaluated using Pearson and Spearman correlation coefficients at each time point. These correlations will look at the correlations of the change from baseline. Scatter plots will also be presented.

5.3.5. Clinical Global Impression-Severity (CGI-S)

5.3.5.1. Definition

A secondary efficacy endpoint is the change from baseline to Weeks 12 and 24 in the CGI-S score. The CGI-S provides an overall clinician-determined summary measure of the severity of the subject's illness that takes into account all available information, including knowledge of the subject's history, psychosocial circumstances, symptoms, behavior, and the impact of the symptoms on the subject's ability to function. The CGI-S evaluates the severity of psychopathology on a scale of 1 to 7. Considering total clinical experience with the depression population, a subject is assessed on severity of illness at the time of rating according to: 1=normal (not at all ill); 2=borderline ill; 3=mildly ill; 4=moderately ill; 5=markedly ill; 6=severely ill; 7=among the most extremely ill patients. The CGI-S permits a global evaluation of the subject's condition at a given time.

The CGI-S change from baseline at Week 24 is calculated as (CGI-S score at Week 24 – Baseline CGI-S score). Change from baseline to other time points is calculated similarly. Negative changes in CGI-S score indicate improvement.

5.3.5.2. Analysis Methods

A frequency distribution over time of the CGI-S scores at Baseline, Week 2, Week 4, Week 6, Week 12, Week 18, Week 24, and at follow-up will be provided by treatment group. In addition, descriptive statistics of the actual values and the change from baseline will be presented by treatment group for observed case data.

The analysis of the change in CGI-S score from baseline to each postbaseline time point will be performed using an ANCOVA model on the ranks of the change in score with treatment and baseline insomnia status as factors, and unranked baseline CGI-S score as a covariate.

In addition, the correlation between CGI-S score and HAM-A total score and MADRS total score, separately, will be evaluated using Spearman correlation coefficients. These correlations will look at the correlations of the actual values at each time point, as well as the change from baseline. Scatter plots will also be presented.

5.3.6. Patient Global Impression-Severity (PGI-S)

5.3.6.1. Definition

A secondary efficacy endpoint is the change from baseline to Weeks 12 and 24 in the PGI-S score. The PGI-S is a self-report scale to measure severity of illness (0=not depressed, 1=mildly depressed, 2=moderately depressed, 3=very depressed, 4=extremely depressed). Considering all aspects of depression, subjects will rate their severity on the PGI-S.

The PGI-S change from baseline at Week 24 is calculated as (PGI-S score at Week 24 – Baseline PGI-S score). Change from baseline to other time points is calculated similarly. Negative changes in PGI-S score indicate improvement.

5.3.6.2. Analysis Methods

A frequency distribution over time of the PGI-S scores at Baseline, Week 2, Week 4, Week 6, Week 12, Week 18, Week 24, and at follow-up will be provided by treatment group. In addition, descriptive statistics of the actual values and the change from baseline will be presented by treatment group for observed case data.

The analysis of the change in PGI-S score from baseline to each postbaseline time point will be performed using an ANCOVA model on the ranks of the change in score with treatment and baseline insomnia status as factors, and unranked baseline PGI-S score as a covariate.

5.3.7. Quality of Life in Depression Scale (QLDS)

5.3.7.1. Definition

A secondary efficacy endpoint is the change from baseline to Weeks 12 and 24 in the QLDS score. The QLDS is a disease specific patient-reported outcome (PRO) designed to assess health related quality of life in patients with MDD. The instrument has a recall period of "at the present time", contains 34-items with "true"/"not true" response options, and takes approximately 5 to 10 minutes to complete.

Each statement on the QLDS is given a score of "1" or "0". A score of "1" is indicative of adverse quality of life. All item scores are summed to give a total score that ranges from 0 (good quality of life) to 34 (very poor quality of life).

As the measure contains positive as well as negative items, item scores are not always in the same order - that is, not all 'true' responses are allocated a score of "1". The scores for each item can be found in Attachment 1.

Imputation of the total score will be performed only when between 1 to 6 items are missing responses. If 7 or more items are missing, the total score will be left missing. The total score will be imputed as follows:

$$T = \frac{x}{34 - m} \times 34$$

where *T* is the final total score, *x* is the item summation score, and *m* is the number of missing items.

5.3.7.2. Analysis Methods

Descriptive statistics of the actual values on Baseline, Week 2, Week 4, Week 6, Week 12, Week 18, and Week 24 and the change from baseline to each postbaseline time point will be presented for QLDS total score by treatment group.

The change from baseline in QLDS total score will be analyzed using the same MMRM as described in Section 5.3.1.2 for MADRS total score, with the continuous covariate "baseline MADRS total score" changed to "baseline QLDS total score".

Frequency distributions of the QLDS individual items will be provided at each assessment time point by treatment group.

5.3.8. Patient Reported Outcome Measurement Information System-Sleep Disturbance (PROMIS-SD)

5.3.8.1. Definition

A secondary efficacy endpoint is the change from baseline to Weeks 12 and 24 in the PROMIS-SD. The PROMIS-SD Short Form subscale consists of a static 8-item questionnaire. Using a recall period of the past 7 days, it assesses the concepts of sleep initiation (2 items), quality of sleep (3 items), early morning feelings (2 items) and worrying about sleep (1 item).

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. Note that 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. However, for the 8-item form, at least 4 items must have been answered in order to calculate a score. After confirming that enough responses were provided, sum the response scores from the items that were answered. Multiply this sum by the total number of items in the short form. Finally, divide by the number of items that were answered. For example, if a respondent answered 5 of 8 questions and answered all items with the second lowest response option (2), you would sum all responses (10), multiply by the number of items in the short form (8) and divide by the number of items that were answered (5). Here (10x8)/5=16. If the result is a fraction, round up to the nearest whole number.

The formula is:

(Raw sum x number of items on the short form) / Number of items that were actually answered

This is a pro-rated raw score.

The raw score (i.e., the total raw score or pro-rated raw score) can be converted into a T-score for each participant based on the table in Attachment 2. The T-score rescales the raw score into a standardized score with a mean of 50 and an SD of 10.

The change in the PROMIS-SD score from baseline to each time point in the double-blind phase will be calculated, for both the raw score and the T-score.

5.3.8.2. Analysis Methods

Descriptive statistics of the actual values on Baseline, Week 2, Week 6, Week 12, Week 18, and Week 24 and the change from baseline to each postbaseline time point will be presented for PROMIS-SD raw score and the T-score by treatment group.

The change from baseline to each postbaseline time point in the PROMIS-SD raw score and the T-score will be analyzed using the same MMRM as described in Section 5.3.1.2 for MADRS total score, with the covariate "baseline MADRS total score" changed to "baseline PROMIS-SD raw score" or "baseline PROMIS-SD T-score", respectively.

Frequency distributions of the PROMIS-SD individual items will be provided at each assessment time point by treatment group.

5.3.9. Patient Reported Outcome Measurement Information System- Sleep Related Impairment (PROMIS-SRI)

5.3.9.1. Definition

A secondary efficacy endpoint is the change from baseline to Weeks 12 and 24 in the PROMIS-SRI. The PROMIS-SRI Short Form subscale consists of a static 8 item questionnaire. The PROMIS adult Sleep-Related Impairment item bank focuses on self-reported perceptions of alertness, sleepiness, and tiredness during usual waking hours, and the perceived functional impairments during wakefulness associated with sleep problems or impaired alertness. Though Sleep-Related Impairment does not directly assess cognitive, affective, or performance impairment, it does measure waking alertness, sleepiness, and function within the context of overall sleep-wake function. The Sleep-Related Impairment short form is universal rather than disease-specific. It assesses sleep-related impairment 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 related impairment. Note that the "direction" of the responses is not the same for all questions, i.e., sometimes a response of "not at all" indicates more sleep related impairment and sometimes a response of "not at all" indicates less sleep related impairment.

"I had a hard time getting things done because I was sleepy" ranges from 1=not at all to 5=very much
"I felt alert when I woke up" ranges from 5=not at all to 1=very much
"I felt tired" ranges from 1=not at all to 5=very much
"I had problems during the day because of poor sleep" ranges from 1=not at all to 5=very much
"I had a hard time concentrating because of poor sleep" ranges from 1=not at all to 5=very much
"I felt irritable because of poor sleep" ranges from 1=not at all to 5=very much
"I was sleepy during the daytime" ranges from 1=not at all to 5=very much
"I had trouble staying awake during the day" ranges from 1=not at all to 5=very much

A score can be approximated if a participant skips a question. However, for the 8-item form, at least 4 items must have been answered in order to calculate a score. After confirming that enough responses were provided, sum the response scores from the items that were answered. Multiply this sum by the total number of items in the short form. Finally, divide by the number of items that were answered. For example, if a respondent answered 5 of 8 questions and answered all items with the second lowest response option (2), you would sum all responses (10), multiply by the number of items in the short form (8) and divide by the number of items that were answered (5). Here (10x8)/5=16. If the result is a fraction, round up to the nearest whole number.

The formula is:

(Raw sum x number of items on the short form) / Number of items that were actually answered

This is a pro-rated raw score.

The total raw score or pro-rated score can be converted into a T-score for each participant based on the table in Attachment 3. The T-score rescales the raw score into a standardized score with a mean of 50 and an SD of 10.

The change in the PROMIS-SRI score from baseline to each time point in the double-blind phase will be calculated, for both the raw score and the T-score.

5.3.9.2. Analysis Methods

Descriptive statistics of the actual values on Baseline, Week 2, Week 6, Week 12, Week 18, and Week 24 and the change from baseline to each postbaseline time point will be presented for PROMIS-SRI raw score and the T-score by treatment group.

The change from baseline to each postbaseline time point in the PROMIS-SRI raw score and the T-score will be analyzed using the same MMRM as described in Section 5.3.1.2 for MADRS total score, with the covariate "baseline MADRS total score" changed to "baseline PROMIS-SRI raw score" or "baseline PROMIS-SRI T-score", respectively.

Frequency distributions of the PROMIS-SRI individual items will be provided at each assessment time point by treatment group.

5.3.10. Symptoms of Major Depressive Disorder Scale (SMDDS)

5.3.10.1. Definition

A secondary efficacy endpoint is the change from baseline to Weeks 12 and 24 in the SMDDS. The SMDDS assesses patient-reported symptoms associated with MDD. The target population includes adults (aged 18 and older) with a clinical diagnosis of MDD who are being treated in an ambulatory setting. The SMDDS has been developed in a patient sample including both males and females, varying levels of age, race, education, marital status, and severity. This 16-item instrument has a 7-day recall period, and subjects respond to each question using a rating scale between 0 ("Not at all" or "Never") to 4 ("Extremely" or "Always"). Before summing the items to create a total score, item 11 ("how often did you have a poor appetite") and item 12 ("how often did you over eat") are combined into a single score by selecting the highest severity on either item. The total score is then created by summing the responses on the 15 items. The total score ranges from 0 to 60 with a higher score indicating more severe depressive symptomatology.

5.3.10.2. Analysis Methods

Descriptive statistics of the actual values on Baseline, Week 2, Week 4, Week 6, Week 12, Week 18, and Week 24 and the change from baseline to each postbaseline time point will be presented for SMDDS total score by treatment group.

The change from baseline in SMDDS total score will be analyzed using the same MMRM as described in Section 5.3.1.2 for MADRS total score, with the continuous covariate "baseline MADRS total score" changed to "baseline SMDDS total score".

Frequency distributions of the SMDDS individual items will be provided at each assessment time point by treatment group.

5.3.11. Symbol Digit Modalities Test (SDMT)

5.3.11.1. Definition

A secondary efficacy endpoint is the change from baseline to Weeks 6, 12, and 24 in the SDMT. The SDMT is a widely used, paper-and-pencil assessment of complex scanning and visual tracking, requiring elements of attention, visuoperceptual processing, working memory, and cognitive/psychomotor speed. The test is viewed as a robust screening test for adult neuropsychological impairment and is sensitive to impairments in cognitive function associated with MDD. The SDMT measures the time to pair abstract symbols with specific numbers. The test includes a coding key consisting of 9 abstract symbols, each paired with a number ranging from 1 to 9. Following the key, the subject is presented with randomly ordered symbols and is required to write the number corresponding to each symbol as fast as possible. The number of correct substitutions within 90 seconds is recorded.

5.3.11.2. Analysis Methods

Descriptive statistics of the actual values on Baseline, Week 6, Week 12, and Week 24 and the change from baseline to each postbaseline time point will be presented for SDMT score by treatment group.

The change from baseline in SDMT score will be analyzed using the same MMRM as described in Section 5.3.1.2 for MADRS total score, with the continuous covariate "baseline MADRS total score" changed to "baseline SDMT score".

5.3.12. Trail Making Test – Part B (TMT–Part B)

5.3.12.1. Definition

A secondary efficacy endpoint is the change from baseline to Weeks 6, 12, and 24 in the TMT-Part B. The TMT-Part B measures divided attention and executive function (tracking and sequencing). The subject is instructed to draw a line to connect a set of 25 consecutively numbered and lettered circles, alternating sequentially between numbers and letters (i.e., 1 A 2 B). The subject is instructed to work as quickly as possible while still maintaining accuracy. The TMT-Part B is sensitive to cognitive decline associated with MDD.

5.3.12.2. Analysis Methods

Descriptive statistics of the actual values on Baseline, Week 6, Week 12, and Week 24 and the change from baseline to each postbaseline time point will be presented for TMT-Part B score by treatment group.

The change from baseline in TMT-Part B score will be analyzed using the same MMRM as described in Section 5.3.1.2 for MADRS total score, with the continuous covariate "baseline MADRS total score" changed to "baseline TMT-Part B score".

5.3.13. Hopkins Verbal Learning Test-Revised (HVLT-R)

5.3.13.1. Definition

A secondary efficacy endpoint is the change from baseline to Weeks 6, 12, and 24 in the HVLT-R. The HVLT-R, a measure of verbal learning and memory, is a 12-item word list recall test. Administration includes 3 learning trials, a delayed recall (20-minute) trial, and a 24-word recognition list (including 12 target and 12 foil words). The test administrator reads instructions and word lists aloud, and records words recalled/recognized by the subject. Scores include learning, delayed recall, and recognition. The HVLT-R is a well-validated and widely used measure of verbal episodic memory.

5.3.13.2. Analysis Methods

Descriptive statistics of the actual values on Baseline, Week 6, Week 12, and Week 24 and the change from baseline to each postbaseline time point will be presented by treatment group for the following: total recall (sum of total correct responses for Trials 1, 2, and 3), delayed recall (number of correct responses for Trial 4), total number of true-positive errors, and recognition discrimination index (total number of true positives minus total number of false positives).

5.3.14. MADRS-6

5.3.14.1. Definition

MADRS-6 is the depression subscale of the full MADRS, including the following 6 items: Apparent Sadness, Reported Sadness, Inner tension, Lassitude, Inability to feel, Pessimistic thoughts.

A total score (0 to 36) is calculated by adding the scores of all 6 items. Higher scores represent a more severe condition. Imputation of the total score will be performed only when 1 item score is missing. If 2 or more items are missing, the total score will be left missing. The total score will be imputed by calculating the sum of the scores of the non-missing items and multiplying it by the ratio of the maximum possible number of items (i.e., 6) to the number of non-missing items (i.e., 5).

The MADRS-6 change from baseline at Week 24 is calculated as (MADRS-6 total score at Week 24 – Baseline MADRS-6 total score). Negative changes in MADRS-6 total score indicate improvement.

5.3.14.2. Analysis Methods

Descriptive statistics of the actual values on Baseline, Week 2, Week 4, Week 6, Week 12, Week 18, Week 24 and at follow-up and the change from baseline to each postbaseline time point will be presented for MADRS-6 total score by treatment group.

The change from baseline over time in MADRS-6 total score will be analyzed using the same MMRM as described in Section 5.3.1.2, with the continuous covariate "baseline MADRS total score" changed to "baseline MADRS-6 total score".

As a sensitivity analysis, the change in MADRS-6 total score from baseline to each postbaseline time point will be analyzed using an ANCOVA model, using LOCF data. The derivation of the LOCF values is described in Section 5.3.1.2. The ANCOVA model will include factors for treatment, baseline insomnia status (present/absent), and baseline MADRS-6 total score as a continuous covariate. A 80% CI for the difference in LS Means and p-value will be calculated based on the contrast test statistic.

5.4. Exploratory Efficacy Endpoints

5.4.1. Consensus Sleep Diary – Morning Administration (CSD-M)

5.4.1.1. Definition

An exploratory efficacy endpoint is the change from baseline in subjective sleep parameters as measured by the CSD-M after the first 6 weeks of treatment and after 3 to 6 months of treatment. Subjects will be asked to provide answers to questions to determine their subjective experience of sleep by recording their answers in a daily sleep diary (CSD-M). The CSD-M is the only sleep diary developed with rigorous methodology for patient reported outcome development, including employing user/focus group feedback and expert feedback to establish construct validity. It has undergone psychometric testing and its content validity has been confirmed by patient focus groups. The parameters recorded include:

- self-reported sleep onset latency (sSOL)
- subjective Total Sleep Time (sTST)
- subjective Wake After Sleep Onset (sWASO)
- subjective number of nighttime awakenings (s-nNAW)
- subjective quality of sleep (sQUAL)
- subjective refreshed feeling on waking (sFRESH)

Subjects will complete the CSD-M at home after their normal morning routine and preferably within an hour of waking for the day. The CSD-M will be completed for 3 mornings leading up to and including the targeted study visit on Study Day 1, Week 6, Week 12, Week 18, and Week 24.

The definitions and computations for deriving the subjective sleep parameters of sSOL, sWASO, sTST, and s-nNAW are shown in Table 5.

Table 5: Calculation of Subjective Sleep Parameters

Parameter	Definition	Computation
sSOL (minutes)	How many minutes it takes to fall asleep, starting from the moment of intention to fall asleep	Derived from CSD-M, Question 3: "How long did it take you to fall asleep?" with the following instructions: "Beginning at the time you wrote in question 2 (i.e., 'What time did you try to go to sleep?'), how long did it take you to fall asleep."
sWASO (minutes)	Total amount of time awake during the night, excluding sSOL (how many minutes it takes to fall asleep) and amount of awake time between the final awakening and the time of getting out of bed	Derived from CSD-M, Question 5: "In total, how long did these awakenings last?" with the following instructions: "What was the total time you were awake between the time you first fell asleep and your final awakening."
sTST (minutes)	Actual time slept	Derived from CSD-M, Question 8: "In total, how long did you sleep?" with the following instructions: "This should just be your best estimate, based on when you went to bed and woke up, how long it took you to fall asleep, and how long you were awake. You do not need to calculate this by adding and subtracting; just give your best estimate."
s-nNAW (number of times)	Number of awakenings, excluding the final awakening before the final arising	Derived from CSD-M, Question 4: "How many times did you wake up, not counting your final awakening?" with the following instructions: "How many times did you wake up between the time you first fell asleep and your final awakening?"

Source: Buysse et al, 2006.[1]

For sSOL, sTST, sWASO, s-nNAW, the average of up to 7 morning measurements will be used as the measurement for that time point. If only one measurement is provided, it will be used as the average.

The parameter sQUAL is derived from CSD-M, Question 9: "How would you rate the quality of your sleep?", with the instructions "Sleep Quality is your sense of whether your sleep was good or poor." There are 5 possible responses: "Very poor", "Poor", "Fair", "Good", and "Very good".

The parameter sFRESH is derived from CSD-M, Question 10: "How rested or refreshed did you feel when you woke-up for the day?", with the instructions "This refers to how you felt after you were done sleeping for the night, during the first few minutes that you were awake." There are 5 possible responses: "Not at all rested", "Slightly rested", "Somewhat rested", "Well-rested", and "Very well-rested".

For sQUAL and sFRESH, in order to create a single "average" response, the categorical responses will be mapped to a 0-4 scale and an average will be calculated. This average will also be mapped back to the closest categorical response. For example, if sQUAL is answered "Very poor", "Poor" and "Good", these would be mapped to scores of 0, 1, and 3, respectively, and the average would be calculated as 1.33. As this is closest to 1, the average response would be "Poor".

5.4.1.2. Analysis Methods

Descriptive statistics of the actual values on Average predose, Week 6 (Average), Week 12 (Average), Week 18 (Average), and Week 24 (Average) and the change from Average predose to each postbaseline time point will be presented by treatment group for sSOL, sTST, sWASO, and s-nNAW.

The change from Average predose for sSOL, sTST, sWASO, s-nNAW, sFRESH and sQUAL will be analyzed using the same MMRM as described in Section 5.3.1.2 for MADRS total score, with the continuous covariate "baseline MADRS total score" changed to the corresponding baseline score for sSOL, sTST, sWASO, s-nNAW, sFRESH, and sQUAL respectively. For sFRESH and sQUAL, an additional analysis of the change from average predose to each postbaseline time point will be performed using an ANCOVA model on the ranks of the change in score with treatment and baseline insomnia status as factors, and unranked average predose score as a covariate.

In addition, for sFRESH and sQUAL, a frequency distribution of the average categorical response will be provided for Average predose, Week 6 (Average), Week 12 (Average), Week 18 (Average), and Week 24 (Average). In addition, the shift in average categorical response from Average predose to each of the postbaseline time points will be presented.

Descriptive statistics of the actual values on Average predose, Week 6 (Average), Week 12 (Average), Week 18 (Average), and Week 24 (Average) and the change from Average predose to each postbaseline time point will be presented by treatment group for average number of naps, the average total duration of naps (in minutes), the average number of drinks containing alcohol, and the average number of caffeinated drinks.

5.4.2. MADRS-WOSI

5.4.2.1. Definition

MADRS-WOSI is defined as the full MADRS with the item Reduced Sleep excluded.

A total score (0 to 54) is calculated by adding the scores of all 9 items. Higher scores represent a more severe condition. Imputation of the total score will be performed only when 1 item score is missing. If 2 or more items are missing, the total score will be left missing. The total score will be imputed by calculating the sum of the scores of the non-missing items and multiplying it by the ratio of the maximum possible number of items (i.e., 9) to the number of non-missing items (i.e., 8).

The MADRS-WOSI change from baseline at Week 24 is calculated as (MADRS-WOSI total score at Week 24 – Baseline MADRS-WOSI total score). Negative changes in MADRS-WOSI total score indicate improvement.

5.4.2.2. Analysis Methods

Descriptive statistics of the actual values on Baseline, Week 2, Week 4, Week 6, Week 12, Week 18, Week 24 and at follow-up and the change from baseline to each postbaseline time point will be presented for MADRS-WOSI total score by treatment group.

The change from baseline over time in MADRS- WOSI total score will be analyzed using the same MMRM as described in Section 5.3.1.2, with the continuous covariate "baseline MADRS total score" changed to "baseline MADRS- WOSI total score".

6. SAFETY

All safety analyses will be based on the safety analysis set based on actual treatment received, unless otherwise specified.

6.1. Adverse Events

The verbatim terms used in the CRF by investigators to identify AEs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). Any AE occurring at or after the initial administration of study agent through the day of last dose plus 2 days is considered to be treatment-emergent. If the AE occurs on the day of the initial administration of study agent, and either AE time or time of administration are missing, then the AE will be assumed to be 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 agent based on partial onset date or resolution date.

All reported treatment-emergent adverse events (TEAEs) will be included in the analysis. For each TEAE, the number and percentage of subjects who experience at least 1 occurrence of the given event will be summarized by system organ class, preferred term, and treatment group (and overall).

Summary tables will be provided for:

- TEAEs
- TEAEs occurring in ≥5% of subjects in either treatment group
- Treatment-emergent serious AEs (SAEs)
- TEAEs leading to discontinuation of study agent
- TEAEs by severity
- TEAEs by most recent treatment dose
- AEs by relationship to study agent
- AEs leading to dose interruption (overall and by most recent treatment dose)
- AEs leading to dose modification (increase or decrease)

For the summaries of TEAEs by severity/relationship to study drug, the observation with the most severe occurrence/closest relationship to study agent will be chosen if there is more than one incident of the same TEAE for the subject.

In addition, a summary table will be generated showing the number and percentage of subjects who experience at least 1 occurrence of a given TEAE by system organ class, preferred term, lowest level term, and treatment group.

A summary of all somnolence-related TEAEs (MedDRA preferred terms: somnolence, hypersomnia, and sedation; regardless of start time) will be presented.

Adverse events of special interest are cataplexy, sleep paralysis, and complex, sleep-related behaviors (parasomnias). Subjects with TEAEs of special interest will be presented separately, by preferred term and treatment group. The AEs to be included in the summary of TEAEs of special interest are marked as such on the AE CRF.

A summary table of non-TEAEs occurring on or after the date of last dose of study drug plus 3 days until the overall reference end date for the study will also be provided.

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

- Died
- Had SAEs
- Had AEs leading to discontinuation of study agent
- Had somnolence-related TEAEs
- Had suicide-related TEAEs
- Had TEAEs of special interest

6.2. Clinical Laboratory Tests

Descriptive statistics will be presented for thyroid-stimulating hormone (TSH) and free thyroxine (FT_4), as well as all chemistry (including the lipid panel), hematology (including hemoglobin A1c [HbA1c]), and urinalysis (pH and specific gravity) laboratory parameters at each scheduled time point. In addition, change from baseline to all postbaseline time points will be summarized for chemistry, hematology, urinalysis (pH and specific gravity), TSH, and FT_4 parameters and displayed by treatment group.

The number and percentage of subjects with treatment-emergent postbaseline markedly abnormal postbaseline values will be presented by treatment group. Clinical laboratory test values will be considered "treatment-emergent markedly abnormal" (TEMA) using the criteria defined by the sponsor listed in Attachment 4. The identification of TEMA laboratory values is based on the postbaseline value being out of range while the baseline value is either missing or within the range given in Attachment 4. If postbaseline laboratory results are above the upper limit and the baseline value is below the lower limit, then the postbaseline abnormality will also be considered TEMA. The same applies to the postbaseline value being below the lower limit with the baseline value being above the upper limit.

In addition, the number of subjects with the following shifts in chemistry laboratory values from baseline to the maximum postbaseline time point will be presented:

- Glucose:
    - from <100 mg/dL to ≥126 mg/dL (normal to high)
    - from <100 mg/dL to [≥100 mg/dL - <126 mg/dL] (normal to borderline high)
    - from [≥100 mg/dL - <126 mg/dL] to ≥126 mg/dL (borderline high 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 high to high/very high)
    - from [≥150 mg/dL - <200 mg/dL] to ≥500 mg/dL (borderline high 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/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 high)
    - from [≥200 mg/dL - <240 mg/dL] to ≥240 mg/dL (borderline high to high)
- HDL Cholesterol: from ≥40 mg/dL to <40 mg/dL (normal to low).

The following conversion rules will be used: Glucose 1 mg/dL=0.05551 mmol/L; Triglycerides 1 mg/dL=0.01129 mmol/L; Total Cholesterol, HDL Cholesterol 1 mg/dL=0.02586 mmol/L.

The incidence of subjects with treatment-emergent ALT values >3*upper normal limit (ULN), AST > 3*ULN, or total bilirubin >2*ULN will be presented for the double-blind phase. Additionally, incidence of treatment-emergent hepatic toxicity (suspected Hy's Law[4] 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 subjects with baseline (ALT values ≤3*ULN or AST ≤3*ULN) (AND baseline total bilirubin values ≤2*ULN for hepatic toxicity) (or if baseline value is missing) will be eligible for these analyses.

A listing of subjects with markedly abnormal laboratory values will be provided.

6.2.1. Homeostatic Assessment (HOMA) Modeling

Insulin resistance and beta-cell function based upon fasting glucose and insulin using the homeostatic assessment (HOMA) [3] model will be assessed. Two variables, HOMA IR (insulin resistance) and HOMA-%B (beta-cell function) will be derived. The relationship between glucose and insulin secretion, mathematically approximated using a simple nonlinear solution, is given below:

$$HOMA\ Insulin\ Resistance\ (IR) = \frac{FI}{22.5e^{-lnFG}}$$

$$HOMA\ Beta\ Function\ (B) = \frac{20\ x\ FI}{[FG - 3.5]}$$

where FG = fasting glucose (mmol/L); FI = fasting insulin (mU/L); Insulin:
1 µIU/mL = 6.945 pmol/L. HOMA IR and HOMA-%B will not be derived if FG is <=3.5.

The descriptive statistics for HOMA-IR and HOMA-%B at baseline (DB) and end point (DB) for the safety analysis set will include the following:

Geometric mean (GM) = exp(mean(logs));

GM mean ± 1 SD = (exp(mean(logs) - 1 SD(logs)), exp(mean(logs) + 1 SD(logs)));

where logs indicates the natural logarithm of the HOMA values.

6.3. Vital Signs and Physical Examination Findings

Continuous vital sign variables including weight, waist circumference, temperature, supine and standing pulse, supine and standing blood pressure (systolic and diastolic), and BMI will be summarized at each assessment time point by treatment group. BMI 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. Changes from baseline to each postbaseline time point will be summarized. Descriptive statistics (N, mean, SD, median, minimum and maximum) will be presented.

Incidence of treatment-emergent clinically important abnormalities in vital signs during the double-blind phase, as defined in Table 6, will be summarized for subjects who had at least one postbaseline assessment for that vital sign. If the baseline value is missing, the postbaseline value will be compared against the abnormally low/abnormally high criteria. Vital sign assessments collected during follow-up will not be used for this summary.

Table 6: Clinically Important Abnormalities in Vital Signs

Vital Sign Parameter	Postbaseline value outside of normal limit if:

	Abnormally low	**Abnormally high**
Pulse (bpm)	A decrease from baseline of ≥15 to a value ≤50	An increase from baseline of ≥15 to a value ≥100
Systolic BP (mmHg)	A decrease from baseline of ≥20 to a value ≤90	An increase from baseline of ≥20 to a value ≥180
Diastolic BP (mmHg)	A decrease from baseline of ≥15 to a value ≤50	An increase from baseline of ≥15 to a value ≥105
Body weight (kg)	A decrease from baseline of ≥7%	An increase from baseline of ≥7%
Body temperature (°C)	<35.5°C	>37.5°C

BP = blood pressure

A listing of subjects with treatment-emergent clinically important abnormalities in vital signs will be presented.

Orthostatic hypotension is defined as an absolute decrease in systolic (>20 mm Hg) or diastolic (>10 mm Hg) blood pressure after standing for at least 1 minute relative to supine position with an increase in pulse rate of >15 beats per minute (Table 7). The number and percentage of subjects who experience treatment-emergent orthostatic hypotension outside of predefined limits at any time during the double-blind phase and for whom the orthostatic hypotension was not present at baseline will be tabulated. Vital sign assessments collected during follow-up will not be used for this summary.

Table 7: Abnormal Limits for Orthostatic Hypotension Parameters (Changes in Vital Signs in Standing Relative to Supine Position)

Vital Sign	**Outside of normal limit if difference (standing minus supine)**
(1) Pulse (bpm)	> 15 bpm
(2a) Systolic blood pressure (mm Hg) (SBP)	< -20 mm Hg
(2b) Diastolic blood pressure (mm Hg) (DBP)	< -10 mm Hg

Note: Orthostatic hypotension requires that conditions (1) and [(2a) or (2b)] are met.

For subjects who are unable to stand and have the vital signs measured in a sitting or supine position instead of the standing position, the difference between standing and supine values will remain missing.

A listing of subjects with treatment-emergent orthostatic hypotension will be presented.

In addition, a by-subject listing of the physical examination data will be presented.

6.4. Electrocardiogram

Twelve-lead ECGs will be recorded in a supine position so that the different ECG intervals (RR, PR, QRS, QT) can be measured. ECGs will be assessed at screening, Baseline, Week 6, and

Week 24. The ECGs will be read by a central reader. The ECG parameters that will be analyzed are heart rate, PR interval, RR interval, QRS interval, QT interval, and QTc using the following correction methods: Bazett's formula (QTcB) and Fridericia's formula (QTcF).

Bazett's formula: $QTcB\ (msec) = QT\ (msec) * (HR(bpm)/60)^{1/2}$

Fridericia's formula: $QTcF\ (msec) = QT\ (msec) * (HR(bpm)/60)^{1/3}$

Descriptive statistics for observed values and changes from average predose will be presented by treatment group for the above ECG parameters at each scheduled time point. Average predose ECG is defined as the average of all ECG results collected up to and including the day of the first dose of study drug.

The identification of treatment-emergent abnormal ECG values is based on the post-baseline value (a value occurring after the first study drug administration) being out of range while the average predose value is either missing or within the limits given in Table 8. If post-baseline ECG results are above the upper limits (abnormally high) and the average predose value is below the lower limits (abnormally low), then the post-baseline abnormality will also be considered treatment-emergent. The same applies to the post-baseline value being below the lower limits (abnormally low) with the average predose value being above the upper limits (abnormally high). The number and percentage of subjects with treatment-emergent ECG values outside the pre-defined normal limits defined below will be presented by treatment group.

Table 8: 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

In addition, the number and percentage of subjects within each of the categories defined below will be presented for the average predose and the maximum postbaseline value during the DB phase by treatment group. The maximum postbaseline value during the double-blind period will be computed for each ECG parameter using data from both scheduled and unscheduled visits.

Categories to assess QT prolongation:

QTc Interval:

- Normal QTc (≤450 msec for male, ≤470 msec for female)
- QTc (>450 to ≤480 msec for male, >470 to ≤480 msec for female)
- QTc (>480 to ≤500 msec)

- QTc (>500 msec)

Clinically significant QTc:

- No (≤500 msec)
- Yes (>500 msec)

Change from baseline:

- No concern: QTc ≤30 msec
- Concern: QTc >30 – ≤ 60 msec
- Clear concern: QTc >60 msec

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

Listings of treatment-emergent abnormal ECG values, QTc intervals >450 msec for male and QTc intervals >470 for female, and QTc interval changes >30 msec will also be provided.

6.5. Other Safety Parameters

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

Emergence of suicidal ideation will be assessed using the C-SSRS. The C-SSRS is a low-burden measure of the spectrum of suicidal ideation and behavior. It is a semi-structured clinician-administered questionnaire designed to solicit the occurrence, severity, and frequency of suicide-related ideation and behaviors during the assessment period.

Using the C-SSRS, potentially suicide-related events will be categorized using the following scores:

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 (non-fatal)
10: Completed Suicide (only applicable for post baseline)

If no events qualify for a score of 1 to 10, a score of 0 will be assigned (0="no event that can be assessed on the basis of C-SSRS"). Higher scores indicate greater severity.

A frequency distribution at each time point by treatment group will be provided. Shifts from baseline to the maximum postbaseline score during the double-blind phase will be summarized by treatment group.

The maximum postbaseline score during the double-blind phase assigned for each subject will be summarized into one of three broad categories: No suicidal ideation or behavior (0), Suicidal ideation (1-5), Suicidal behavior (6-10). Shifts from baseline to the maximum category during the double-blind phase will be summarized by treatment group.

In addition, a frequency distribution at each time point for whether the "subject has engaged in non-suicidal self-injurious behavior" will be presented by treatment group.

A listing of C-SSRS items throughout the study for subjects with Suicidal Ideation or Behavior at any time point will be provided.

6.5.2. Arizona Sexual Experiences Scale (ASEX)

Effect on sexual functioning will be assessed using the ASEX at Baseline, Week 12, and Week 24. The ASEX is a five-item rating scale that quantifies sex drive, arousal, vaginal lubrication/penile erection, ability to reach orgasm, and satisfaction from orgasm. Each of the 5 items is rated on a 6-point Likert scale, ranging from 1 to 6. The 5 items are summed to create a total score, ranging from 5 to 30, with the higher scores indicating 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 treatment group and gender at each time point. In addition, for each of the 5 items, a frequency distribution will be provided by treatment group, combining the responses for the genders. For this analysis, "vaginal lubrication/penile erection" will be summarized as 1 question. The ASEX total score at each time point and the change from baseline will be summarized with descriptive statistics by treatment group.

The number and percentage of subjects 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 treatment group.

6.5.3. Extrapyramidal Symptom Rating Scale-Abbreviated (ESRS-A)

The ESRS-A is an abbreviated manualized version of the ESRS, a semi-structured interview that rates parkinsonian symptoms, dystonia, dyskinesias, and akathisia over the previous 7 days. The ratings include a motor examination for rigidity, tremor, reduced facial expression or speech, impaired gait/posture, postural instability, and bradykinesia/hypokinesia. Twenty-four individual items are rated on a 6-point scale: 0=Absent, 1=Minimal, 2=Mild, 3=Moderate, 4=Severe, or 5=Extreme. Frequency is included as an index of severity. Symptoms are divided into the 4 corresponding subscales and each subscale is summarized in a Clinical Global Impression of Movement Severity (CGI-MS) score.

For each of the 24 individual items, as well as for the 4 subscales, a frequency distribution will be provided by treatment group at each time point.

6.5.4. Physicians Withdrawal Checklist (PWC)

The Physician Withdrawal Checklist (20 items; PWC-20) is a simple and accurate method used to assess potential withdrawal symptoms following cessation of treatment and will be measured at the end of study or follow-up visit. Each of the 20 items is rated on a 4-point Likert scale, with 0=not present, 1=mild, 2=moderate, 3=severe.

For each of the 20 items, a frequency distribution will be provided by treatment group for subjects with assessments at the time of discontinuation, and for subjects with assessments at the follow-up visit.

A total score (0 to 24) will be calculated by adding the scores of the following 8 items: Nausea-Vomiting, Diarrhea, Poor Coordination, Diaphoresis, Tremor-Tremulousness, Dizziness-Lightheadedness, Increased Acuity Sound Smell Touch, Paresthesias. If 1 or more items are missing, the total score will be left missing. Higher scores represent a more severe condition. The total score will be summarized with descriptive statistics by treatment group.

7. SUMMARY OF PLASMA CONCENTRATIONS FOR JNJ-42847922, M12, AND M16

Plasma concentrations for JNJ-42847922, and metabolites (M12, and M16) will be summarized by dose, day and time point, using descriptive statistics.

8. BIOMARKERS

Details of the biomarker analysis are provided in a separate document.

9. HEALTH ECONOMICS

9.1. Healthcare Resource Use Questionnaire (HRUQ)

Medical resource utilization data, associated with medical encounters, will be collected using the HRUQ during the double-blind and follow-up phases of the study.

The number and percentage of subjects who had at least one outpatient visit to a healthcare professional, related to their depression, in the period since the most recent assessment preceding each time point, will be summarized for each type of healthcare professional and for any healthcare professional by time point.

The number and percentage of subjects who had at least one emergency room visit without hospitalization, related to their depression, in the period since the most recent assessment preceding each time point, will be summarized by time point.

The number and percentage of subjects who had at least one day hospital or partial hospital visit (no overnight stay), related to their depression, in the period since the most recent assessment preceding each time point, will be summarized by time point.

The number and percentage of subjects who had at least one hospitalization (minimum one overnight stay), related to their depression, in the period since the most recent assessment preceding each time point, will be summarized by time point.

The total number of outpatient visits to each type of healthcare professional, related to depression, and the total number of outpatient visits to any healthcare professional, related to depression, in the period since the most recent assessment preceding each time point, will be summarized with descriptive statistics at each time point.

Likewise, the total number of emergency room visits without hospitalization, related to depression, the total number of day hospital or partial hospital visits (no overnight stay), related to depression, and the total number of hospitalizations (minimum one overnight stay), related to depression, will be summarized with descriptive statistics at each time point.

All summaries will be presented by treatment group.

REFERENCES

1. Buysse DJ, Ancoli-Israel S, Edinger JD, Lichstein KL, Morin CM. Recommendations for a standard research assessment of insomnia. Sleep; 2006; 29(9):1155-1173.
2. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. Psychological Methods; 2010; Vol. 15, No. 4, pp. 309–334.
3. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia; 1985; 28:412–419.
4. 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

ATTACHMENTS

ATTACHMENT 1 SCORING OF QLDS ITEMS

I just want time to pass	True	1
	Not true	0
I feel hopeful about the future	True	0
	Not true	1
I find it hard to hold a conversation	True	1
	Not true	0
I like to know what is going on in the world	True	0
	Not true	1
I feel as if my life is wasting away	True	1
	Not true	0
I feel as if I am not in control of my life	True	1
	Not true	0
I am reluctant to leave the house	True	1
	Not true	0
On the whole, I enjoy the things I do	True	0
	Not true	1
I have lost all pleasure in life	True	1
	Not true	0
I feel as if I have nothing to offer anyone	True	1
	Not true	0
I turn away from people I care about	True	1
	Not true	0
I take good care of myself	True	0
	Not true	1
I am able to think about the future	True	0
	Not true	1
I just want to hide away	True	1
	Not true	0
I look forward to things	True	0
	Not true	1
I've forgotten what it's like to enjoy myself	True	1
	Not true	0

I can't be bothered with my friends	True	1
	Not true	0
I can cope easily with everyday tasks	True	0
	Not true	1
I cut myself off from other people	True	1
	Not true	0
It is difficult for me to make even simple decisions	True	1
	Not true	0
I feel as if I am a burden to people	True	1
	Not true	0
Most of the time I just sit and stare into space	True	1
	Not true	0
I can't face anyone	True	1
	Not true	0
I shut everything out	True	1
	Not true	0
I'm neglecting my appearance	True	1
	Not true	0
I can see the funny side of things	True	0
	Not true	1
I don't take in what people say to me	True	1
	Not true	0
I feel as if I'm letting everyone down	True	1
	Not true	0
I dread each coming day	True	1
	Not true	0
I enjoy my food	True	0
	Not true	1
I avoid people if I can	True	1
	Not true	0
I am reluctant to answer the door or the telephone	True	1
	Not true	0
My life has no meaning	True	1
	Not true	0
I am able to cope with everyday problems	True	0
	Not true	1

ATTACHMENT 2 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

ATTACHMENT 3 CONVERSION OF RAW SCORE TO T-SCORE FOR PROMIS-SRI

Sleep-Related Impairment 8a *Short Form Conversion Table*		
Raw Score	**T-Score**	**SE***
8	30.0	5.4
9	35.2	4.6
10	38.7	4.2
11	41.4	3.8
12	43.6	3.6
13	45.5	3.4
14	47.3	3.1
15	48.9	2.9
16	50.3	2.7
17	51.6	2.6
18	52.9	2.6
19	54.0	2.5
20	55.1	2.5
21	56.1	2.5
22	57.2	2.5
23	58.2	2.4
24	59.3	2.4
25	60.3	2.4
26	61.3	2.4
27	62.3	2.3
28	63.3	2.3
29	64.3	2.3
30	65.3	2.3
31	66.3	2.3
32	67.3	2.3
33	68.4	2.3
34	69.5	2.3
35	70.7	2.4
36	71.9	2.5
37	73.4	2.6
38	75.0	2.8
39	76.9	3.1
40	80.1	3.9

*SE= Standard Error on T-score

Adult version

ATTACHMENT 4 CRITERIA FOR TREATMENT-EMERGENT MARKEDLY ABNORMAL LABORATORY VALUES

	Markedly Abnormal Limits	
Laboratory Parameter (unit)	**Low**	**High**
Clinical Chemistry		
Albumin (g/dL) but SI unit = g/L	2.4→24	6.0 →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) but SI unit=mmol/L	15.1→15.1	34.9→34.9
Bilirubin (direct) (mg/dL) but SI unit =µmol/L	N/A	3.0 mg/dL→51.3 umol/L
Bilirubin (total) (mg/dL) but SI unit =µmol/L	N/A	3.0 mg/dL→51.3 umol/L
Blood urea nitrogen (mg/dL) but SI unit=mmol/L	N/A	50 mg/dL→17.9 mmol/L
Calcium (mg/dL) but SI unit=mmol/L	6→1.497 mmol/L	12→2.994 mmol/L
Chloride (mEq/L or mmol/L)	94	112
Cholesterol (mg/dL) but SI unit=mmol/L	N/A	300→7.758 mmol/L
Creatine kinase (U/L)	N/A	990
Creatinine (mg/dL) SI unit=µmol/L	N/A	3→ 265.2 µmol/L
Gamma glutamyl transferase (U/L)	N/A	300 U/L
Glucose Plasma (mg/dL) but SI unit=mmol/L	40→ 2.204 mmol/L	300→16.653 mmol/L
High-density lipoprotein cholesterol (HDL) (mg/dL) but SI unit=mmol/L	35→0.905	N/A
Lactic acid dehydrogenase (LDH) (U/L)	N/A	500
Low-density lipoprotein cholesterol (LDL) (mg/dL) but SI unit=mmol/L	89→ 2.3015	160→4.1376 mmol/L
Phosphate (mg/dL) but SI unit=mmol/L	2.2→0.71038 mmol/L	8.1→ 2.61549 mmol/L
Potassium (mmol/L)	3.0	5.8
Sodium (mEq/L) but SI unit = mmol/L	125→125	155→155
Total protein (g/L)	50	N/A
Triglycerides (mg/dL) but SI unit=mmol/L	N/A	500→5.645 mmol/L
Uric acid (mg/dL) but SI unit=µmol/L	1.5→89.22	10→594.8 µmol/L
Hematology		
Hematocrit (%) - female	0.28	0.50
- male	0.24	0.55
Hemoglobin (g/dL) but SI unit=g/L	8→80	19→190
Hemoglobin A1c (%)	4	8
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 (10^{12}/L; tera/L) - female	3.0	5.5
- male	3.0	6.4
White blood cell (WBC) count (10^9/L; giga/L)	2.5	15.0
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.