Janssen Research & Development *

Statistical Analysis Plan
Amendment 2

A Multicenter, Double-Blind, Randomized, Parallel-Group, Placebo-Controlled, Adaptive Dose-Finding Study to Evaluate the Efficacy and Safety of JNJ-42847922 as Adjunctive Therapy to Antidepressants in Adult Subjects With Major Depressive Disorder Who Have Responded Inadequately to Antidepressant Therapy

Protocol 42847922MDD2001; Phase 2b

JNJ-42847922 (Selective Orexin-2 Receptor Antagonist)

Approved, Date: 6 February 2019

Status: Approved

Date: 6 February 2019

Prepared by: Janssen Research & Development, LLC

Document No.: EDMS-ERI-133786387 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	21 May 2018
Amendment 1	10 Jan. 2019
Amendment 2	6 Feb. 2019

Amendments below are listed beginning with the most recent amendment.

Amendment 2 (6 Feb. 2019)

The overall reason for the amendment: removed analysis windows for salivary cortisol; added sensitivity analysis for MCP-Mod; added conversion rules for lab values and added analysis for plasma concentrations

Applicable Section(s)	Description of Change (s)
Section 2.5	Removed analysis windows for salivary cortisol from Table 1 (this is moved to a separate biomarker SAP)
Section 2.8	Removed association analysis between cortisol and efficacy outcome (this is moved to a separate biomarker SAP)
Section 5.2.3.1	Added sensitivity analysis using the model based on the maximum of the candidate model trend test statistics; added sensitivity analysis using bootstrap for MCP-Mod CI
Section 6.2	Added conversion rules for chemistry laboratory values
Section 7	Added that Plasma concentrations for JNJ-42847922, M12, and M16 will be summarized using descriptive statistics.
Attachment 3	Updated conversion factor for Triglycerides

Amendment 1 (10 Jan. 2019)

The overall reason for the amendment: added MADRS-6, remission/response analyses for observed case, descriptive analyses for ISI and SDS; clarified definitions for sustained response, Hy's law and identifications of treatment-emergent abnormal ECGs; and removed biomarker analysis from the SAP.

Applicable Section(s)	Description of Change (s)
Section 2.5	Table 1 minor edits to the analysis windows
Section 2.8	Added prior medication use of adjunctive antipsychotics based on concomitant medication and MGH (Yes/No) to subgroup
Section 4.1	Added WOCBP vs. WONCBP vs Men to demographic categorical variables in Table 2; added baseline ISI per IWRS and per eDC to baseline categorical variables in Table 3.
Section 4.6	Clarified that prior medications will be summarized by base preferred term.
Section 5	Added MADRS-6 Total score to the efficacy variable. Added section 5.3.15.
Section 5.1.2, 5.3.1.1, 5.3.2.1, 5.3.5.1 and 5.3.8.2	Added observed case analyses for remission/response analyses.
Section 5.2.3.2	Added a sensitivity analysis to exclude major protocol deviations.
Section 5.3.1.3	Clarified sustained response definitions
Section 5.3.3.2	Added that descriptive statistics will also be provided for change from baseline in ISI total score by baseline ISI score (≥15 versus <15)

Section 5.3.7.2	Added that the percentage of subjects who have not worked/studied at all during the past 7 days will be summarized.
Section 6.1	Removed the definition of next-day somnolence AE
Section 6.2	Clarified Hy's law with AST included to the criteria.
Section 6.4	Clarified the identification of treatment-emergent abnormal ECG values
Section 6.5.2	Added PWC-20 total score
Section 8	Biomarker analysis are moved out to a separate document.

ABBREVIATIONS

AE	adverse event
AIC	Akaike Information Criterion
ANCOVA	analysis of covariance
ASEX	Arizona Sexual Experiences Scale
CGI-S	Clinical Global Impression-Severity
CI	confidence interval
CRF	case report form
CSR	Clinical Study Report
C-SSRS	Columbia Suicide Severity Rating Scale
DSM-5	Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition
ECG	electrocardiogram
EQ-5D-5L	European Quality of Life (EuroQol) Group, 5-Dimension, 5-Level questionnaire
HAM-A	Hamilton Anxiety Rating Scale
HbA1c	hemoglobin A1c
HPA	hypothalamic-pituitary-adrenal
HSI	Health Status Index
IAC	Interim Analysis Committee
ISI	Insomnia Severity Index
IWRS	interactive web response system
LOCF	last observation carried forward
LS	least-squares
MADRS	Montgomery-Asberg Depression Rating Scale
MCP-Mod	Multiple Comparison Procedure-Modeling
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
PD	pharmacodynamic(s)
PGI-S	Patient Global Impression-Severity
PHQ-9	Patient Health Questionnaire 9-item
PK	pharmacokinetic(s)
PRO	patient-reported outcome(s)
PROMIS-SD	Patient Reported Outcome Measurement Information System-Sleep Disturbance
PROMIS-Fatigue	Patient Reported Outcome Measurement Information System- Fatigue
PWC	Physician Withdrawal Checklist
QTcB	QT interval corrected using Bazett's formula
QTcF	QT interval corrected using Fridericia's formula
RRS	Ruminative Response 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
SDS	Sheehan Disability Scale
SHAPS	Snaith-Hamilton Pleasure Scale
SIGH-A	structured interview guide for the Hamilton Anxiety Scale
SNRI	serotonin-norepinephrine reuptake inhibitor
SSRI	selective serotonin reuptake inhibitor
TEAE	treatment-emergent adverse event
TEMA	treatment-emergent markedly abnormal
US	United States
WHO-DD	World Health Organization Drug Dictionary
WLQ	Work Limitations Questionnaire
WOCBP	women of childbearing potential
WONCBP	women of non-childbearing potential

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 42847922MDD2001.

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

1.1. Trial Objectives

Objectives	Endpoints
Primary	
• To assess the dose-response relationship of up to 3 doses of JNJ-42847922 (20 and 40 mg, with 10 mg potentially added at the interim analysis) compared to placebo as adjunctive therapy to an antidepressant drug in improving depressive symptoms 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).	• Change from baseline to the end of Week 6 in the Montgomery-Asberg Depression Rating Scale (MADRS) total score.
• To assess the safety and tolerability of JNJ-42847922 compared to placebo as adjunctive therapy to an antidepressant in subjects with MDD.	• Safety assessments, including adverse events (AEs), laboratory values, electrocardiogram (ECG), vital signs, physical exam, the Columbia Suicide Severity Rating Scale (C-SSRS), the Arizona Sexual Experiences Scale (ASEX), and the Physician Withdrawal Checklist (PWC).

Objectives	Endpoints
Secondary	
• To assess the efficacy of JNJ-42847922 compared to placebo as adjunctive therapy to an antidepressant in improving depressive symptoms in subjects with MDD with significant insomnia symptoms (baseline Insomnia Severity Index [ISI] score ≥15) versus those without significant insomnia symptoms (baseline ISI score <15).	• Correlation between baseline ISI score as a continuous variable and change from baseline to the end of Week 6 in the MADRS total score. • Change from baseline to the end of Week 6 in the MADRS total score in subjects with baseline ISI score ≥15 versus subjects with baseline ISI score <15. • Shift in ISI score category from baseline to the end of Week 6.
• To assess the efficacy of JNJ-42847922 compared to placebo as adjunctive therapy to an antidepressant in improving: – Response and remission of depressive symptoms	– Proportion of responders on depressive symptoms scale, defined as a ≥50% improvement in MADRS total score from baseline to the end of Week 6. – Proportion of subjects with remission of depressive symptoms, defined as a MADRS total score ≤8, ≤10, or ≤12 at the end of Week 6.
– Anxiety symptoms	– Change from baseline to the end of Week 6 on the 14-item Hamilton Anxiety Rating scale (HAM-A) total score.
– Response of anxiety symptoms	– Proportion of responders on anxiety symptoms scale, defined as a ≥50% improvement in the HAM-A total score from baseline to the end of Week 6.
– Clinical severity	– Change from baseline to the end of Week 6 in the Clinical Global Impression-Severity (CGI-S) score.
– Global functioning (work/school, social and family life).	– Change from baseline to the end of Week 6 in the Sheehan Disability Scale (SDS).
• To assess the efficacy of JNJ-42847922 compared to placebo as adjunctive therapy to an antidepressant in improving depressive symptoms in the subpopulation of subjects with MDD with anxious distress.	• Change from baseline to the end of Week 6 in the MADRS total score in subjects with MDD with anxious distress versus subjects with MDD without anxious distress.
• To evaluate the effect of JNJ-42847922 exposure on the hypothalamic-pituitary-adrenal (HPA) axis in subjects with MDD.	• Change from baseline to Days 8, 22, and 42 in salivary cortisol levels, as measured upon awakening.
• To assess the exposure of JNJ-42847922 and metabolites M12 and M16 in subjects with MDD.	• Observed plasma concentrations of JNJ-42847922 and metabolites and estimated exposure parameters for JNJ-42847922 from population based PK modeling.

Objectives	Endpoints
• To capture patient-reported assessment of JNJ-42847922 compared to placebo as adjunctive therapy to an antidepressant.	• Change from baseline to the end of Week 6 in: – Depressive symptoms using the Patient Health Questionnaire 9-item (PHQ-9) – Anhedonia using the Snaith-Hamilton Pleasure Scale (SHAPS) – Sleep disturbance using the Patient Reported Outcome Measurement Information System-Sleep Disturbance (PROMIS-SD) Short Form – Fatigue using the PROMIS-Fatigue Short Form – Severity of depression using the Patient Global Impression-Severity (PGI-S) – Health related quality of life and utility using the European Quality of Life (EuroQol) Group, 5-Dimension, 5-Level (EQ-5D-5L) questionnaire – Work productivity and limitations using the Work Limitations Questionnaire (WLQ) Short Form.

In addition, the exploratory objectives are:

- To identify diagnostic biomarkers and to investigate changes in MDD-related biomarkers (HPA axis function, biomarkers of immune system activation and oxidative stress) in relation to clinical response on depression symptoms upon adjunctive treatment with JNJ-42847922
- To explore the exposure/response relationship of JNJ-42847922 in subjects with MDD
- To identify genetic factors that may influence the PK, safety, or tolerability of JNJ-42847922
- To assess the predictive capability of salivary cortisol levels assessed during the circadian nadir (predose in the evening) on the antidepressant effect of JNJ-42847922.

1.2. Trial Design

This is a multicenter, double-blind, randomized, parallel-group, placebo-controlled, 6-week adaptive dose-finding study to assess the efficacy and safety of JNJ-42847922 as adjunctive therapy in adult subjects with MDD who have had an inadequate response to current antidepressant therapy with an SSRI/SNRI. Approximately 280 subjects will be randomized.

The doses of JNJ-42847922 that will be studied prior to the interim analysis are known (i.e., 20 and 40 mg), whereas the doses of JNJ-42847922 to be studied after the interim analysis will be adaptively chosen based on the dose-response curve observed at the interim analysis. During the interim analysis review period, subjects will continue to be randomized according to the initial randomization scheme. Regardless of whether subjects are randomized before or after the interim analysis, the schedule of events and study procedures will remain the same. The final analysis will include all subjects.

Initially, the study will assess the efficacy, dose- and exposure-response relationship, safety, and tolerability of JNJ-42847922 20 and 40 mg doses compared to placebo, when administered as adjunctive therapy to an SSRI/SNRI antidepressant. At the start of the study, subjects will be randomized in a 2:1:1 ratio to receive 1 of 3 treatments: placebo: JNJ-42847922 20 mg: JNJ-42847922 40 mg.

The data cutoff for the interim analysis will be 6 weeks after randomization of 160 subjects. An Interim Analysis Committee (IAC) will be established to review the interim data to examine the dose-response relationship and determine the doses of JNJ-42847922 and the allocation ratio of the doses to be used after the interim analysis.

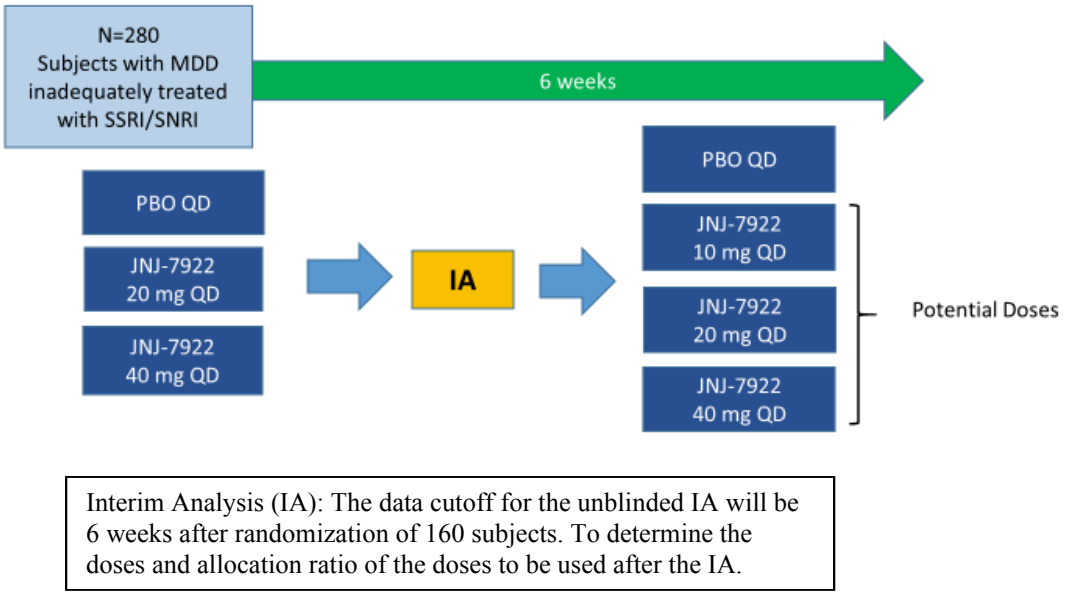
After the interim analysis, subjects will be randomly assigned to study drug (placebo or 1 of the JNJ-42847922 doses [potential doses include 10 mg, 20 mg, 40 mg]), with the allocation ratio adapted according to the dose-response curve seen at the interim analysis.

For all subjects, the study will consist of 3 phases: a screening phase (up to 4 weeks), a double-blind treatment phase (6 weeks), and a posttreatment follow-up phase (2 weeks). The planned total study duration for each subject will be approximately 12 weeks (84 days). Subjects will continue to take their baseline SSRI/SNRI antidepressant (at the same dose, without change, and at approximately the same time as prior to entering the study) throughout the screening, double-blind, and follow-up periods (the entire 12 weeks of the study duration).

In order to collect PK and salivary cortisol samples after the Day 1 dose, subjects will stay at the study site from at least 1 hour before the bedtime dosing on Day 1 until the morning of Day 2.

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

Figure 1: Schematic Overview of the Study Design



Abbreviations: IA=interim analysis, MDD=major depressive disorder, PK=pharmacokinetics, QD=once daily, SSRI= selective serotonin reuptake inhibitor, SNRI=serotonin-norepinephrine reuptake inhibitor.

1.3. Statistical Hypotheses for Trial Objectives

The primary efficacy endpoint is the change from baseline to the end of Week 6 in the MADRS total score. The null hypothesis to be tested to address the primary objective of the study is that there is no difference between any of the JNJ-42847922 doses (10, 20 and/or 40 mg) and placebo, as adjunctive treatment to an SSRI/SNRI, in the improvement of depressive symptoms of MDD, based on the primary efficacy endpoint.

1.4. Sample Size Justification

The calculation of the sample size is based on the primary efficacy endpoint (change from baseline to the end of Week 6 for the MADRS total score). The assumptions for the effect size were based on published literature and the clinically meaningful difference between groups to be detected.

The total sample size is calculated on the basis of the generalized Multiple Comparison Procedure-Modeling (MCP-Mod)[4] test applied to the placebo and the JNJ-42847922 dose groups at the final analysis. Approximately 280 randomized subjects will provide an average weighted power of approximately 85% depending on the underlying true dose-response shape, assuming a 1-sided significance level of 0.05, a treatment difference from placebo of 4.5 points in change in MADRS total score, a standard deviation of 10, and a 25% overall dropout rate.

The data cutoff for the interim analysis will be 6 weeks after randomization of 160 subjects, which corresponds to approximately 57% of the sample size of the study. This number of subjects provides approximately 93% power to detect a dose-response relationship, depending on

the underlying true dose-response shape, assuming a treatment difference of 4.5, and a standard deviation of 10 when tested by means of the MCP-Mod approach at the 30% significance level (1-sided).

Further details regarding the interim analysis will be specified in a separate charter and interim analysis SAP.

1.5. Randomization and Blinding

Central randomization will be implemented in this study. Subjects will be randomly assigned based on a computer-generated randomization schedule prepared before the study by or under the supervision of the sponsor.

At the start of study enrollment, subjects will be randomly assigned to receive 1 of 3 treatments in a 2:1:1 ratio to placebo:JNJ-42847922 20 mg:JNJ-42847922 40 mg.

After the interim analysis, subjects will be randomly assigned to study drug, with the allocation ratio adapted according to the dose-response curve observed at the time of the interim analysis. Subjects will be assigned to receive placebo or one of the JNJ-42847922 doses (potential doses include 10 mg, 20 mg, 40 mg).

The randomization will be balanced by using randomly permuted blocks and will be stratified by region (United States [US], Europe, and Japan) and by insomnia status (significant insomnia symptoms [ISI score ≥15] at baseline versus no significant insomnia symptoms [ISI score <15] at baseline). In order to randomize approximately equal proportions of subjects with and without significant insomnia symptoms at baseline, the number randomized in either stratum will be capped at 60% of the planned total sample size.

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.

In general, randomization codes will be disclosed fully only when the study is completed and the clinical database is closed. However, for the interim analysis, the randomization codes and the translation of randomization codes into treatment and control groups will be disclosed to those authorized and only for those subjects included in the interim analysis.

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 drug in the double-blind treatment phase. The screening phase end date is left missing for those subjects who did not receive study drug.

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. Study Day

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

Study day for a visit is defined as:

a. Visit date - (date of Study Day 1) +1, if visit date is >= date of Day 1

b. Visit date - Date of Day 1, if visit date < date of Day 1

There is no 'Day 0'.

2.4. Baseline and End Point (DB)

The baseline measurement is defined as the closest measurement taken prior to or at the time of the first dose of study drug, with the exception of the average predose ECG measurement, which is defined as the average of all ECG results collected up to and including the day of the first dose of study drug.

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 (DB) value if the unscheduled visit result is the last postbaseline result available within the double-blind phase.

2.5. 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 drug 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 end points. 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.

Assessments prior to Day 1 of the double-blind phase that are not considered a baseline assessment will be labeled as 'Screening'.

Table 1: Visit Windows

Parameter	Analysis Phase	Scheduled Visit Number	Time Interval (label on output)	Time Interval (Day)*	Target Time Point (Day/Time)
MADRS	Screening	1	Screening	< 1	-28 to -3
	DB	2	Baseline	≤1	1
		3	Day 8	2 to 15	8
		4	Day 22	16 to 32	22

Parameter	Analysis Phase	Scheduled Visit Number	Time Interval (label on output)	Time Interval (Day)*	Target Time Point (Day/Time)
		5	Day 42	33 to end of DB	42
		DB last visit	End point (DB)	2 to end of DB	
HAM-A, CGI-S, SDS, PGI-S, SHAPS, PROMIS-Fatigue, PROMIS-SD, WLQ, EQ-5D-5L, PHQ-9	DB	2	Baseline	≤1	1
		3	Day 8	2 to 15	8
		4	Day 22	16 to 32	22
		5	Day 42	33 to end of DB	42
		DB last visit	End point (DB)	2 to end of DB	
Weight, BMI, Waist Circumference, ISI	Screening	1	Screening	<1	-28 to -3
	DB	2	Baseline	≤1	1
		5	Day 42	2 to end of DB	42
		DB last visit	End point (DB)	2 to end of DB	
ECG	DB	1, 2	Predose	≤1	-28 to 1
		1, 2	Average Predose	≤1	-28 to 1
		5	Day 42	2 to end of DB	42
		DB last visit	End point (DB)	2 to end of DB	
Physical Exam	Screening	1	Screening	<1	-28 to -3
	DB	2	Baseline	≤1	1
		5	Day 42	2 to end of DB	42
		DB last visit	End point (DB)	2 to end of DB	
Vital signs	Screening	1	Screening	<1	-28 to -3
	DB	2	Baseline	≤1	1
		3	Day 8	2 to 15	8
		4	Day 22	16 to 32	22
		5	Day 42	33 to end of DB	42
		DB last visit	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	6	Follow-up	End of DB + 1 to end of FU	49-56
Serum Chemistry, Urinalysis	Screening	1	Screening	≤1	-28 to -3
	DB	2	Baseline	≤1	1
		3	Day 8	2 to 25	8
		5	Day 42	26 to end of DB	42
		DB last visit	End point (DB)	2 to end of DB	
Hematology	Screening	1	Screening	<1	-28 to -3
	DB	2	Baseline	≤1	1
		3	Day 8	2 to 15	8
		4	Day 22	16 to 32	22
		5	Day 42	33 to end of DB	42
		DB last visit	End point (DB)	2 to end of DB	
Lipid Panel, ASEX	DB	2	Baseline	≤1	1
		5	Day 42	2 to end of DB	42
		DB last visit	End point (DB)	2 to end of DB	
HbA1c	DB	1	Baseline	≤1	-28 to -3
		5	Day 42	2 to end of DB	42
		DB last visit	End point (DB)	2 to end of DB	
PWC	DB	5	Day 42	2 to end of DB	42
		DB last visit	End point (DB)	2 to end of DB	
	Follow-up	TC	Telephone Contact	use Day 43 visit	43
		6	Follow-up	use Day 49 to 56 visit	49-56
C-SSRS	Screening	1	Screening	<1	-28 to -3
	DB	2	Baseline	≤1	1
		3	Day 8	2 to 15	8
		4	Day 22	16 to 32	22
		5	Day 42	33 to end of DB	42

Parameter	Analysis Phase	Scheduled Visit Number	Time Interval (label on output)	Time Interval (Day)*	Target Time Point (Day/Time)
		DB last visit	End point (DB)	2 to end of DB	
	Follow-up	6	Follow-up	End of DB + 2 to end of FU	49-56

*Relative to Study Day 1; DB=double-blind; FU=follow-up, TC=telephone contact

2.6. Pooling Algorithm for Analysis Centers

Subjects will be enrolled at sites in Europe (Bulgaria, Finland, Germany, Russia, Ukraine), the US, and Japan. Actual site enrollment rates will be monitored to avoid gross imbalances across centers. To account for region variability, region (US, Japan, Europe) will be used as a factor in the statistical models to analyze efficacy.

2.7. Analysis Sets

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

2.7.1. 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 treatment was received. This analysis set will be used for summarizing the overall study completion/withdrawal information.

2.7.2. Efficacy Analysis Set

2.7.2.1. Full Analysis Set

The efficacy analyses will be based on the full analysis set, which consists of all subjects who were randomly assigned to study drug and received at least 1 dose of study drug.

2.7.3. Safety Analysis Set

Safety analyses will be based on the safety analysis set, which consists of all subjects who were randomly assigned to study drug and received at least 1 dose of study drug. The safety analysis set is the same as the full analysis set.

2.8. Definition of Subgroups

Descriptive statistics will be provided for the primary efficacy endpoint (change from baseline to the end of Week 6 in MADRS total score) by the following subgroups:

- Baseline ISI score (≥15 versus <15, using the randomization stratification factor as entered into IWRS); this will also be used in the analyses as the baseline insomnia status

- MDD subtype based on SCID-CT (with or 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])
- Baseline Ruminative Response Scale (RRS) score (≥median score versus <median score in the full analysis set)
- Sex (male, female, undifferentiated)
- Age group (18-34, 35-54, 55-64, 65-70 years)
- Women of Childbearing Potential (WOCBP) (as determined by Reproductive Tracking CRF page) vs. Men and Women of Non-Childbearing Potential (WONCBP)
- Women of Childbearing Potential vs. Women of Non-Childbearing Potential
- Adults (18-64) male, Elderly (≥65) male, Adults (18-64) female, Elderly (≥65) female
- Region (US, Europe, Japan, as defined in Section 2.6)
- Race
- Ethnicity
- MDD with anxiety based on HAM-A (No/Yes, defined as baseline HAM-A total score <18 versus ≥18)
- Number of antidepressants used in the current episode based on the Massachusetts General Hospital-Antidepressant Treatment Response Questionnaire (MGH-ATRQ; 1, 2, 3 or more)
- Number of major depressive episodes in lifetime, including current episode (<3, ≥3)
- Prior medication use of adjunctive antipsychotics based on concomitant medication (Yes/No)
- Prior medication use of adjunctive antipsychotics based on concomitant medication and MGH (Yes/No)
- Prior medication use of benzodiazepines (Yes/No)
- Baseline body mass index (BMI; underweight: <18.5 kg/m^2, normal: 18.5 kg/m^2 to <25 kg/m^2, overweight: 25 kg/m^2 to <30 kg/m^2, obese: ≥30 kg/m^2)
- Functional impairment based on baseline SDS category (not impaired [0-3], mild [4-11], moderate [12-19], marked [20-26] or extreme [27-30]).

Additional subgroup analyses of the primary efficacy endpoint are described in Section 5.2.3.5.

2.9. Imputation Rules for Missing Adverse Event (AE) Dates

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 drug start
    - The day of study drug start, if the month/year of the onset of AE is the same as month/year of the study drug start date and month/year of the AE resolution date is different
    - The day of study drug start or day of AE resolution date, whichever is earlier, if month/year of the onset of AE and month/year of the study drug start date and month/year of the AE resolution date are the same.
- If the onset date of an AE is missing both day and month, it will be set to the earliest of:
    - December 31 of the year of onset, if the year of onset is prior to the year of the study drug start date
    - January 1 of the year of onset, as long as this date is on or after the study drug start date
    - Month and day of the study drug start date, if this date has the same year as the year that the AE occurred
    - The AE resolution date.
- Completely missing onset dates will not be imputed.

Partial AE resolution dates will be imputed as follows:

- If the resolution date of an AE is missing day only, it will be set to the earlier of the last day of the month of occurrence of resolution or the day of the date of death, if the death occurred in that month.
- If the resolution date of an AE is missing both day and month, it will be set to the earlier of December 31 of the year or the day and month of the date of death, if the 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 the earlier of:
    - 00:01 as long as the onset date is different from the study drug start date
    - The time of the study drug start if this is the same day the AE occurred.
- The missing time of resolution of an AE will be set to 23:59.

If a missing time is associated with a partial 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 drug. Medications will be classified as prior if the medication start date is complete and prior to the date of first dose of study drug or the medication end date is complete and prior to the date of first dose of study drug.

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 drug 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 drug 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 drug 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 drug 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 drug. 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 drug administration, or
- The stop date month and year (day is missing) of the medication is on or after the initial study drug administration, or
- The stop date year (day and month are missing) of the medication is on or after the initial study drug 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 drug administration, or
- The stop date month and year (day is missing) of the medication is on or after the initial study drug administration, or
- The stop date year (day and month are missing) of the medication is on or after the initial study drug 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 drug administration, or
- The stop date month and year (day is missing) of the medication is on or after the initial study drug administration, or
- The stop date year (day and month are missing) of the medication is on or after the initial study drug 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 medications are those that started after the analysis reference end date of the double-blind phase.

3. INTERIM ANALYSIS AND DATA MONITORING COMMITTEE REVIEW

The data cutoff for the interim analysis will be 6 weeks after randomization of 160 subjects, which corresponds to approximately 57% of the sample size of the study. The primary purpose of the interim analysis will be to examine the dose-response relationship, to determine the doses of JNJ-42847922 and their allocation ratio to be used in subjects randomized after the interim analysis. The IAC will review the outputs from statistical analysis, exposure, exposure-response analysis, and safety for this determination.

The interim analysis will use the generalized MCP-Mod approach to establish a dose-response signal with respect to the primary efficacy endpoint. A 1-sided 0.30 significance level will be used to test for dose-response using the MCP-Mod approach on the change in MADRS total score to the end of Week 6.

Unblinded exposure-response analyses will be performed by an independent internal pharmacometrician using data available at the interim analysis, with adequate firewalls so the study team will remain blinded. At the time of interim analysis, the IAC will review the exposure and exposure-response analyses to assist in selection of doses and the allocation ratio to be used in after the interim analysis.

No adjustment to the final significance level will be performed as a result of the interim analysis.

Further details regarding the interim analysis will be specified in a separate charter and interim analysis SAP.

4. SUBJECT INFORMATION

The number of subjects in each analysis set will be provided. In addition, the number of subjects by region (as defined in Section 2.6), country, and site in the full analysis set will be provided.

4.1. Demographics and Baseline Characteristics

Demographic and baseline characteristics (Table 2) and psychiatric history at baseline (Table 3) will be summarized by treatment group and overall for the full analysis set. The continuous variables will be summarized using descriptive statistics (N, mean, standard deviation [SD], median, minimum, and maximum). The categorical variables will be summarized using a frequency distribution with the number and percentage of subjects in each category.

Table 2: Demographic Variables and Baseline Characteristics

Continuous Variables:

- Age (years), calculated based on date of informed consent date
- Baseline weight (kg)
- Baseline height (cm)
- Baseline BMI (kg/m^2) calculated as Weight (kg)/[Height (m)]^2
- Baseline waist circumference (cm)

Categorical Variables:

- Age (18-34, 35-54, 55-64, 65-70 years)

- Sex (male, female, undifferentiated)
- WOCBP vs. WONCBP vs. Men
- Race (White, Black or African American, Asian, American Indian or Alaskan native, Native Hawaiian or other Pacific islander, Not reported); If multiple race categories are indicated, then Race is recorded as "Multiple".
- Ethnicity (Hispanic or Latino, not Hispanic or Latino)
- Baseline BMI (underweight: <18.5 kg/m^2, normal: 18.5 kg/m^2 to <25 kg/m^2, overweight: 25 kg/m^2 to <30 kg/m^2, obese: ≥30 kg/m^2)

Table 3: Psychiatric History at Baseline Variables

Continuous Variables:

- Age (years) when diagnosed with MDD
- Duration (weeks) of current depressive episode
- Baseline MADRS total score
- Baseline CGI-S score
- Baseline PHQ-9 total score

Categorical Variables:

- Current antidepressant type (SSRI, SNRI)
- 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)
- 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)
- Functional impairment based on baseline SDS total score: not impaired (0-3), mild (4-11), moderate (12-19), marked (20-26), or extreme (27-30)
- Baseline ISI per IWRS (≥15 versus <15, using the randomization stratification factor as entered into IWRS); this will also be used in the analyses as the baseline insomnia status
- Baseline ISI score per eDC (≥15 versus <15, using the eDC data)
- Number of major depressive episodes in lifetime, including current episode (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 medical history and tobacco use history, will be provided.

4.2. Disposition Information

The number of screen failures will be summarized.

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 drug
- Subjects completing 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
- Subjects who discontinued from 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 drug.

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

For the safety analysis set, a Kaplan-Meier plot of the time to discontinuation of study drug will be provided.

4.3. Treatment Compliance

4.3.1. Treatment Compliance With Study Drug

Compliance for each subject will be calculated based on the percent of the scheduled number of capsules of study drug actually taken within the double-blind phase. It is defined as:

%compliance = (number of capsules taken / number of capsules supposed to have been taken)*100%.

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 drug – date of first dose of study drug + 1) multiplied by 2 (since each "dose" of study drug 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 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 = (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 drug – date of first dose of study drug) + 1. Number of doses is defined as the total number of drug 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.

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 trial. 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 for the safety 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 drug. Concomitant medications are defined as any therapy

used on or after the same day as the first dose of study drug, including those that started before and continue after the first dose of study drug.

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 base preferred term 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
MADRS	• Change from baseline to the end of Week 6 in the MADRS total score	Primary
	• Correlation between baseline ISI score as a continuous variable and change from baseline to the end of Week 6 in the MADRS total score.	Secondary
	• Change from baseline to the end of Week 6 in the MADRS total score in subjects with baseline ISI score $\geq$15 versus subjects with baseline ISI score	Secondary

Efficacy Variable		Endpoint
	<15.	
	• Proportion of responders on depressive symptoms scale (≥50% improvement in the MADRS total score from baseline to the end of Week 6)	Secondary
	• Proportion of subjects with remission on depressive symptoms scale (MADRS total score ≤8, ≤10, ≤12 at the end of Week 6)	Secondary
	• Change from baseline to the end of Week 6 in the MADRS total score in subjects with MDD with anxious distress versus subjects with MDD without anxious distress.	Secondary
	• Change from baseline to the end of Week 6 in the MADRS-6 total score	Secondary
ISI	• Shift in ISI score category from baseline to the end of Week 6.	Secondary
HAM-A	• Change from baseline to the end of Week 6 in the HAM-A total score	Secondary
	• Proportion of responders on anxiety symptoms scale (≥50% improvement in the HAM-A total score from baseline to the end of Week 6)	Secondary
CGI-S	• Change from baseline to the end of Week 6	Secondary
SDS	• Change from baseline to the end of Week 6	Secondary
PHQ-9	• Change from baseline to the end of Week 6	Secondary
SHAPS	• Change from baseline to the end of Week 6	Secondary
PROMIS-SD	• Change from baseline to the end of Week 6	Secondary
PROMIS-Fatigue	• Change from baseline to the end of Week 6	Secondary
PGI-S	• Change from baseline to the end of Week 6	Secondary
EQ-5D-5L, VAS and health status index	• Change from baseline to the end of Week 6	Secondary
WLQ	• Change from baseline to the end of Week 6	Secondary

5.1. Analysis Specifications

5.1.1. Level of Significance

The primary efficacy endpoint will be evaluated at a 1-sided significance level of 0.05 using the MCP-Mod approach to test for dose-response. For all other analyses of the primary efficacy

endpoint and for all other efficacy endpoints, no multiplicity adjustment will be done will be presented.

For the interim analysis, a 1-sided 0.30 significance level will be used to test for dose-response using the MCP-Mod approach on the change in MADRS total score to the end of Week 6. No adjustment to the final significance level will be performed due to the interim analysis.

5.1.2. Data Handling Rules

As sensitivity analyses, the change in MADRS total score, MADRS-6 total score, HAM-A total score, and PHQ-9 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 5.2.3.3, 5.3.4.2, and 5.3.8.2.

For the analyses of response of depressive symptoms based on MADRS total score (Section 5.3.1), remission of depressive symptoms based on MADRS total score (Section 5.3.2), response of anxiety symptoms based on HAM-A total score (Section 5.3.5), and response and remission of depressive symptoms based on PHQ-9 total score (Section 5.3.8.2), both observed case and imputed case where subjects with missing values will be imputed as non-responders/non-remitters will be presented.

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 the end of Week 6 for MADRS total score (Section 5.3.1), HAM-A total score (Section 5.3.5), and PHQ-9 total score (Section 5.3.8.2), both observed and LOCF data will be presented. Since ISI is collected at only 1 postbaseline time point, the cumulative response curve for ISI (Section 5.3.3.2) will be presented using observed data.

5.1.3. Imputation Methods for Missing Items

For MADRS, imputation of the total score when there are missing items is described in Section 5.2.1. Likewise, for PROMIS-SD, PROMIS-Fatigue, WLQ and MADRS-6, imputation of the total score when there are missing items is described in Sections 5.3.10.1, 5.3.11.1, 5.3.14.1 and 5.3.15.1, respectively. 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 is the change in MADRS total score from baseline to the end of Week 6. 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 6 is calculated as (MADRS total score at Week 6 – Baseline MADRS total score). Negative changes in MADRS total score indicate improvement.

5.2.2. Primary 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: change in MADRS total score from baseline to the end of Week 6 (see Section 5.2.1);

Intervention Effect: the effect of the initially randomized treatment together with the oral SSRI/SNRI antidepressant that would have been observed had all subjects remained on their treatment throughout the double-blind phase;

Summary Measure: the difference in variable means.

The primary analysis will be based on the full analysis set and the MADRS total scores collected during the double-blind phase.

5.2.3. Analysis Methods

Descriptive statistics of the actual values on Baseline, Day 8, Day 22, and Day 42 and the change from baseline to Day 8, 22, and 42 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.

5.2.3.1. MCP-Mod Analysis

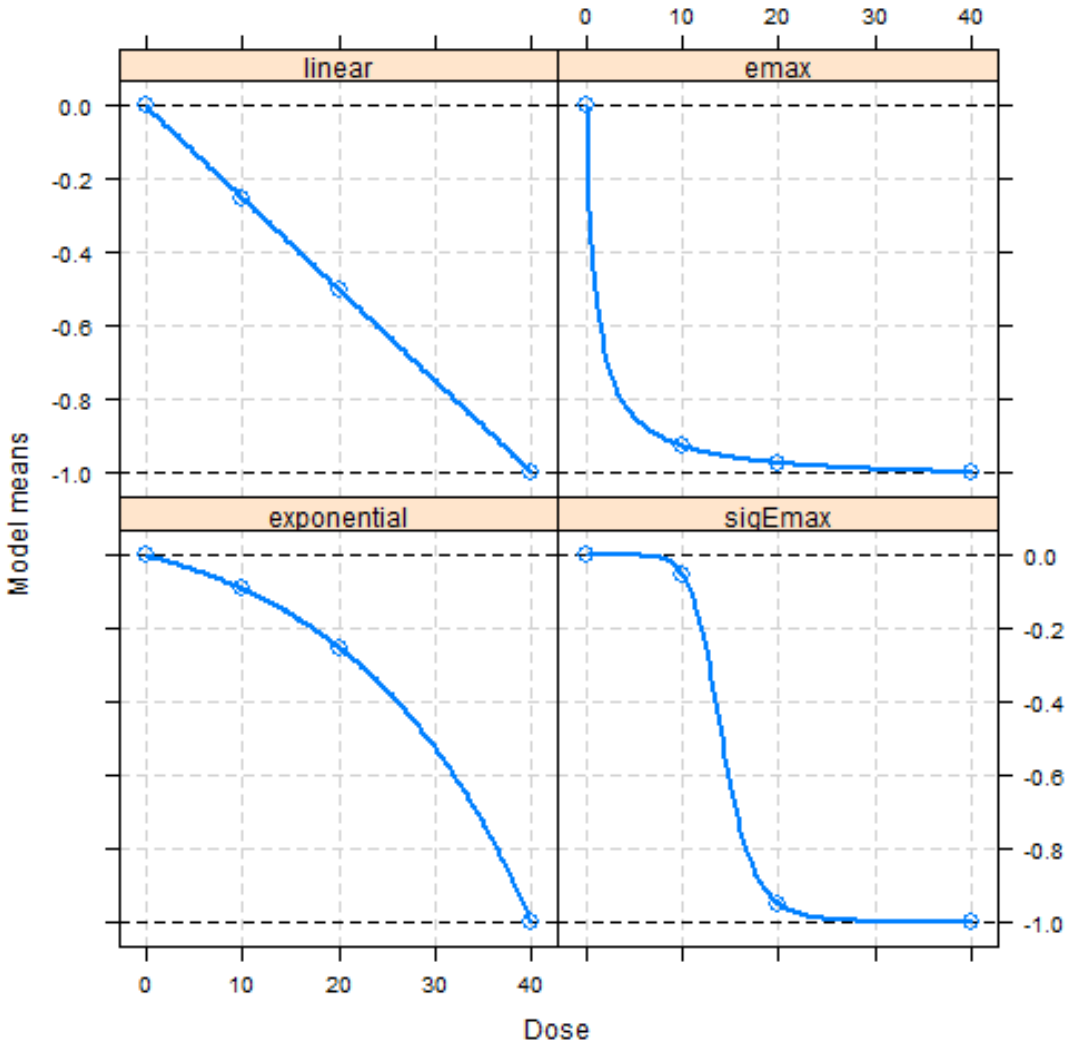
The primary efficacy endpoint will be evaluated using the generalized MCP-Mod approach. The MCP-Mod approach provides a method for dose-finding using model-based estimation rather than hypothesis testing via pairwise comparisons. The analysis will be done in two steps.

First, a mixed model for repeated measures (MMRM) analysis of the observed data from the placebo and JNJ-42847922 groups will be used as the method of addressing missing data and to

adjust for important covariates. The MMRM will include region (as defined in Section 2.6), time (scheduled Week), treatment (placebo and JNJ-42847922 dose groups), baseline insomnia status (present/absent), and treatment-by-time interaction as factors and baseline MADRS total score as a covariate. 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 the end of Week 6, 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 least-squares (LS) mean estimates for each JNJ-42847922 dose group and placebo at Week 6 and the corresponding variances will be obtained.

Subsequently, the generalized MCP-Mod approach will be applied towards the estimates obtained from the MMRM to analyze the dose-response relationship. This approach requires pre-specification of a candidate model set. The actual candidate model set will depend on an interim analysis decision with two possible scenarios. The first scenario is when the 10 mg dose is not added after the interim analysis (i.e., subjects have been randomized to placebo, JNJ-42847922 20 mg and 40 mg dose groups only). For that scenario, using notation as in Bornkamp et al (2009)[1], the candidate set consists of the following 3 standardized model profiles (with the corresponding parameters): "linear" (no parameters), "emax" (ED_{50}=1.053), and "exponential" (δ=18.20). The second scenario is realized when the 10 mg dose is added after the interim analysis. For the second scenario, the candidate set is extended to the total of 4 standardized model profiles: "linear" (no parameters), "emax" (ED_{50}=1.053), "exponential" (δ=18.20), and "sigEmax" (ED_{50}=14.14, h=8.496). Figure 2 below shows the extended set (of note, the maximum response is scaled to 1). Also, for the second scenario, the MCP-Mod model selection criteria will be based on the Akaike Information Criterion (AIC) with the k parameter equal to 0.5. As a sensitivity analysis, the maximum of the candidate model trend test statistics will be used as MCP-MOD model selection criterion.

Figure 2: MCP-Mod Candidate Model Set



The significance of the dose-response signal associated with each candidate model will be determined using trend tests with model-specific optimal contrast coefficients. The maximum of the candidate model trend test statistics will be used to evaluate the presence of a dose-response signal, properly accounting for multiplicity at an overall level of 5% (1-sided) using MCP-Mod methodology. If the maximum test statistic is not significant, no dose-response relationship will be further explored. Otherwise, the model family corresponding to the candidate model with maximum trend test statistic will be used to fit to the observed data to represent the dose-response relationship. The corresponding confidence interval (CI) for the response at each dose and placebo will be computed. A bootstrap method may be used for obtaining CI as sensitivity analysis.

As a secondary analysis, age and gender will be added to the MMRM model for the MCP-MOD analyses.

5.2.3.2. MMRM Pairwise Comparisons

In conjunction with the MCP-Mod analysis, the pairwise comparison between JNJ-42847922 doses and placebo will also be performed using the appropriate contrasts directly from the MMRM analysis described in Section 5.2.3.1, using estimates at the end of Week 6. A 90% CI for the difference in LS Means and p-value will be calculated based on the contrast test statistic for each JNJ-42847922 dose level.

As a secondary analysis, age and gender will be added to the MMRM model described in Section 5.2.3.1 for the pairwise comparison between JNJ-42847922 doses and placebo.

As a sensitivity analysis, the change in MADRS total score from baseline to the end of Week 6 will be analyzed in an MMRM analysis with the exclusion of subjects with major protocol deviations that may affect efficacy (e.g., developed withdrawal criteria but not withdrawn, entered but did not satisfy criteria, received a disallowed concomitant treatment, received wrong treatment or incorrect dose and others that may affect efficacy).

5.2.3.3. Sensitivity Analysis – ANCOVA with LOCF

As a sensitivity analysis, the change in MADRS total score from baseline to the end of Week 6 will be analyzed using an ANCOVA model of LOCF data. The ANCOVA model will include factors for treatment, region, baseline insomnia status (present/absent), and baseline MADRS total score as a continuous covariate. A 90% CI for the difference in LS Means and p-value will be calculated based on the contrast test statistic for each JNJ-42847922 dose level.

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

For example, if a subject has a visit on Day 7 for the "Day 8 visit" and then a final visit on Day 10, the visit on Day 7 will be slotted to "Day 8" because Day 7 is closer to the target Day 8 than Day 10. The Day 10 visit will be used as "Day 22 LOCF", "Day 42 LOCF" and "End Point (DB)".

5.2.3.4. Mediation Analysis

In the mediation analysis framework, natural direct effect can be conceived of as the independent treatment effect on the outcome (i.e., change in MADRS total score) that is above and beyond its effect on the mediator (i.e., change in ISI score); controlled direct effect can be conceived of as the independent treatment effect on the outcome controlling the mediator at a fixed level; and natural indirect effect can be conceived of as a treatment effect on the outcome that is accounted for by its effect on the mediator.

A mediation analysis will be performed to examine the mediating role of improvement in sleep assessed by change from baseline in ISI score on change from baseline in MADRS total score provided that both endpoints demonstrate a statistically difference between at least one dose of JNJ-42847922 and placebo. This analysis will assess the extent to which change in MADRS

total score may be mediated by or independent of change in ISI score based on observed data (i.e., data collected at each time point without carrying forward previous values) at the end of Week 6. The analysis will consider both change in MADRS total score and change in ISI score as continuous variables. The calculation of the ISI score is described in Section 5.2.3.5.

- Simulation-based Counterfactual Approach: this analysis introduced by Imai et.al [2] obtains the natural direct effect and natural indirect effect using numerical simulations. The approach uses parametric bootstrapping to construct the point estimate and its uncertainty estimates for direct effect and indirect effect from the bootstrap sampling distribution. For each bootstrapped sample, change from baseline in MADRS total score and change from baseline in ISI score will be analyzed. The ANCOVA analysis of change in MADRS total score from baseline will include factors for treatment (placebo and JNJ-42847922 dose groups), region, and baseline insomnia status (present/absent), change from baseline (Day 1, predose) in ISI score, treatment-by-change from baseline in ISI score interaction if the interaction is significant, and baseline MADRS total score as a covariate. The ANCOVA analysis of change in ISI score will include factors for treatment (placebo and JNJ-42847922 dose groups), region, and baseline insomnia status (present/absent), and baseline MADRS total score as a covariate. Estimate of the controlled direct effect will be obtained by plugging in the estimated coefficient values and the mean level of change in ISI score into the analytic expressions.

Since the relationship can be bi-directional, a mediation analysis similar to the above analysis will be performed, to assess the extent to which change in ISI score may be mediated by or independent of change in MADRS total score, based on observed data (i.e., data collected at each time point without carrying forward previous values) at the end of Week 6.

5.2.3.5. Subgroup Analyses

In addition to the analyses described below, descriptive statistics will be provided for the primary efficacy endpoint (change from baseline to the end of Week 6 in MADRS total score) by the subgroups identified in Section 2.8.

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 calculated as the sum of the 7 items.

The correlation between baseline ISI score as a continuous variable and change from baseline to the end of Week 6 in the MADRS total score will be evaluated by means of a scatter plot and if appropriate, regression analysis.

In addition, 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 the end of Week 6 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 the end of Week 6 in the MADRS total score will be presented by treatment group for these subgroups.

MDD Subtype

Subjects will be dichotomized based on whether they have MDD with anxious distress or MDD without anxious distress (as determined by the SCID-CT) to evaluate whether there are differences in these subgroups (i.e., evaluating change from baseline to the end of Week 6 in the MADRS total score in subjects with and without anxious distress). Descriptive statistics of the change from baseline to the end of Week 6 in the MADRS total score will be presented by treatment group for these subgroups.

Baseline RRS Score

The RRS consists of 22 items which are rated on 4-point Likert scales ranging from "almost never" to "almost always". The total score is calculated as the sum of the 22 items.

Subjects will be dichotomized based on their baseline RRS score to evaluate whether there are differences in these subgroups (i.e., evaluating change from baseline to the end of Week 6 in the MADRS total score in subjects with RSS score ≥the median versus subjects with RRS score <median). Descriptive statistics of the change from baseline to the end of Week 6 in the MADRS total score will be presented by treatment group for these subgroups.

In addition, the correlation between baseline RRS score as a continuous variable and change from baseline to the end of Week 6 in the MADRS total score will be evaluated by means of a scatter plot and if appropriate, regression analysis.

5.3. Secondary Efficacy Endpoints

5.3.1. Response of Depressive Symptoms

5.3.1.1. Definition

A secondary efficacy endpoint is the proportion of responders on depressive symptoms scale at the end of Week 6, defined as a ≥50% improvement in MADRS total score from baseline to the end of Week 6. The calculation of the MADRS total score is described in Section 5.2.1. The percentage change from baseline for MADRS total score is calculated as 100*(MADRS total score at Day X – Baseline MADRS total score)/(Baseline MADRS total score). Negative percent changes in MADRS total score indicate improvement (eg, percent change < -50% indicates improvement >50%).

A subject is defined a responder (yes=1) at a given time point if the percent reduction in MADRS total score is ≥50%. Subjects who do not meet such criterion will be considered as non-responders and will be assigned a value of 0 (i.e., no). Observed case analysis, i.e., no imputation for subjects with missing values, will be explored.

5.3.1.2. Utility Estimand

The secondary endpoint of proportion of responders on depressive symptoms scale at the end of Week 6 can also be considered a "utility estimand", as it considers efficacy and tolerability data together. It 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: is binary (responder or non-responder), as defined in Section 5.3.1.1; subjects who do not have a MADRS total score at the end of Week 6 (eg, subjects who have early discontinuation of study drug) will be considered non-responders;

Intervention Effect: is not applicable as the intercurrent event (discontinuation of study drug) is captured in the variable definition;

Summary Measure: the difference in response proportions.

The analysis will be based on the full analysis set and the MADRS total scores collected during the double-blind phase.

5.3.1.3. Analysis Methods

The number and percentage of subjects who achieve a response will be summarized at each time point during the double-blind phase by treatment group. The point estimate and 2-sided 90% confidence interval will be provided for the relative response using a Mantel-Haenszel test controlling for region and 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 the end of Week 6 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 on Day 8 or Day 22 that is maintained through the end of Week 6 assessment. Subjects are allowed one excursion (non-response) on a subsequent visit prior to Day 42, however the score must show at least 25% improvement. Subjects who discontinue early are not considered to have sustained response. A stratified log rank test (stratified for region and baseline insomnia status) will be used to test the hypothesis that there is no difference between the JNJ-42847922 dose groups and placebo in the probability of achieving sustained response.

A variation of definition for sustained response will be explored, for which no excursion is allowed. The analyses mentioned above will be repeated for this variation.

5.3.2. Remission of Depressive Symptoms

5.3.2.1. Definition

A secondary efficacy endpoint is the proportion of subjects with remission of depressive symptoms at the end of Week 6, defined as a MADRS total score ≤8, ≤10, or ≤12 at the end of Week 6. Subjects who do not meet such criterion will be considered as non-remitters. Observed case analysis, i.e., no imputation for subjects with missing values, will be explored. The calculation of the MADRS total score is described in Section 5.2.1.

5.3.2.2. Analysis Methods

The number and percentage of subjects who achieve remission (based on each of the 3 definitions) will be summarized at each time point during the double-blind phase by treatment group.

The point estimate and 2-sided 90% confidence interval will be provided for the relative risk of remission using a Mantel-Haenszel test controlling for region and baseline insomnia status (present/absent).

5.3.3. Shift in Insomnia Severity Index (ISI) Score Category

5.3.3.1. Definition

A secondary efficacy endpoint is the shift in ISI score category from baseline to the end of Week 6. The calculation of the ISI score is described in Section 5.2.3.5. The scores are categorized as follows: score 0-7=no clinically significant insomnia; score 8-14=sub-threshold insomnia; score 15-21=moderate insomnia, and score 22-28=clinically severe insomnia.

5.3.3.2. Analysis Methods

The shift in ISI score category from baseline to the end of Week 6 will be presented. A frequency distribution of the ISI individual item scores at each time point will be provided by treatment group. In addition, descriptive statistics of the ISI score at Baseline and Week 6, as well as the change from baseline to Week 6, will be presented by treatment group. Descriptive statistics will also be provided for change from baseline to the end of Week 6 in ISI total score by baseline ISI score (≥15 versus <15, using the randomization stratification factor as entered into IWRS)

The change in ISI score from baseline to the end of Week 6 will be analyzed using an ANCOVA model. The ANCOVA model will include factors for treatment and region, and baseline ISI score as a continuous covariate. A 90% CI for the difference in LSMeans and p-value will be calculated based on the contrast test statistic for each JNJ-42847922 dose level.

The cumulative response rate, defined as the percentage of subjects experiencing at least a given value of percent reduction from baseline to the end of Week 6 in ISI score, will be presented graphically, using observed data.

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 the end of Week 6 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 6 is calculated as (HAM-A total score at Week 6 – Baseline HAM-A total score). 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, Day 8, Day 22, and Day 42 and the change from baseline to Day 8, 22, and 42 will be presented for HAM-A total score, and HAM-A somatic and psychic factor scores by treatment group.

The change from baseline to Day 8, 22, and 42 in HAM-A total score will be analyzed using the same MMRM as described in Section 5.2.3.1 for the primary efficacy endpoint, 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 the end of Week 6 will be analyzed using an ANCOVA model, using LOCF data. The derivation of the LOCF values is described in Section 5.2.3.3. The ANCOVA model will include factors for treatment, region, baseline insomnia status (present/absent), and baseline HAM-A total score as a continuous covariate. A 90% CI for the difference in LSMeans and p-value will be calculated based on the contrast test statistic for each JNJ-42847922 dose level.

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 category 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. 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.

5.3.5. Response of Anxiety Symptoms

5.3.5.1. Definition

A secondary efficacy endpoint is the proportion of responders on anxiety symptoms scale at the end of Week 6, defined as a ≥50% improvement in the HAM-A total score from baseline to the end of Week 6. The calculation of the HAM-A total score is described in Section 5.3.4.1. The percentage change from baseline for HAM-A total score is calculated as 100*(HAM-A total score at Day X – Baseline HAM-A total score)/(Baseline HAM-A total score). Negative percent changes in HAM-A total score indicate improvement (eg, percent change < -50% indicates improvement >50%).

A subject is defined a responder (yes=1) at a given time point if the percent reduction in HAM-A is ≥50%. Subjects who do not meet such criterion will be considered as non-responders and will be assigned a value of 0 (i.e., no). Observed case analysis, i.e., no imputation for subjects with missing values, will be explored.

5.3.5.2. Analysis Methods

The number and percentage of subjects who achieve a response will be summarized at each time point during the double-blind phase by treatment group.

The cumulative response rate, defined as the percentage of subjects experiencing at least a given value of percent reduction from baseline to the end of Week 6 in HAM-A total score, will be presented graphically, for both observed and LOCF data.

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

5.3.6.1. Definition

A secondary efficacy endpoint is the change from baseline to the end of Week 6 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 6 is calculated as (CGI-S score at Week 6 – Baseline CGI-S score). Negative changes in CGI-S score indicate improvement.

5.3.6.2. Analysis Methods

A frequency distribution over time of the CGI-S scores at Baseline, Day 8, Day 22, and Day 42 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 Day 8, 22, and 42 will be performed using an ANCOVA model on the ranks of the change in score with treatment, region, 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 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.

5.3.7. Sheehan Disability Scale (SDS)

5.3.7.1. Definition

A secondary efficacy endpoint is the change from baseline to the end of Week 6 in the SDS. The SDS is a subject-reported, 5-item questionnaire which has been widely used and accepted for assessment of functional impairment and associated disability. The first 3 items assess disruption of (1) work/school, (2) social life, and (3) family life/home responsibilities using a 0-10 rating scale. The scores for the first 3 items are summed to create a total score of 0-30 where a higher score indicates greater impairment. It also has 1 item on days lost from school or work and 1 item on days when underproductive. The recall period for this study is 7 days.

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

5.3.7.2. Analysis Methods

Descriptive statistics of the actual values on Baseline, Day 8, Day 22, and Day 42 and the change from baseline to Day 8, 22, and 42 will be presented for SDS total score by treatment group. In addition, descriptive statistics of the actual values and the change from baseline, as well as a frequency distribution, will be presented by treatment group for each of the 5 SDS individual item scores at each time point.

The change from baseline to Day 8, 22, and 42 in SDS total score will be analyzed using the same MMRM as described in Section 5.2.3.1 for the primary efficacy endpoint, with the covariate “baseline MADRS total score” changed to “baseline SDS total score”.

In addition, the percentage of subjects who have not worked/studied at all during the past 7 days will be summarized at each time point during the double-blind phase by treatment group.

5.3.8. Patient Health Questionnaire 9-item (PHQ-9)

5.3.8.1. Definition

A secondary efficacy endpoint is the change from baseline to the end of Week 6 in PHQ-9. The PHQ-9 is a 9-item, patient-reported outcome (PRO) measure to assess depressive symptoms. The scale scores each of the 9 symptom domains of the DSM-5 MDD criteria and it has been used both as a screening tool and a measure of response to treatment for depression. Each item is rated on a 4 point scale (0=not at all, 1=several days, 2=more than half the days, and 3=nearly every day). The subject's item responses are summed to provide a total score (range of 0 to 27), with higher scores indicating greater severity of depressive symptoms. The recall period is 2 weeks. The severity of the PHQ-9 is categorized as follows: None-minimal (0-4), Mild (5-9), Moderate (10-14), Moderately Severe (15-19) and Severe (20-27).

The PHQ-9 change from baseline at Week 6 is calculated as (PHQ-9 total score at Week 6 – Baseline PHQ-9 total score). Negative changes in PHQ-9 total score indicate improvement.

5.3.8.2. Analysis Methods

Descriptive statistics of the actual values on Baseline, Day 8, Day 22, and Day 42 and the change from baseline to Day 8, 22, and 42 will be presented for PHQ-9 total score by treatment group.

The change from baseline to Day 8, 22, and 42 in PHQ-9 total score will be analyzed using the same MMRM as described in Section 5.2.3.1 for the primary efficacy endpoint, with the covariate "baseline total MADRS score" changed to "baseline PHQ-9 total score".

As a sensitivity analysis, the change in PHQ-9 total score from baseline to the end of Week 6 will be analyzed using an ANCOVA model, using LOCF data. The derivation of the LOCF values is described in Section 5.2.3.3. The ANCOVA model will include factors for treatment, region, baseline insomnia status (present/absent), and baseline PHQ-9 total score as a continuous covariate. A 90% CI for the difference in LS Means and p-value will be calculated based on the contrast test statistic for each JNJ-42847922 dose level.

Frequency distributions of the PHQ-9 severity categories, of the 9 individual items, and of the responses to the question "How difficult have these problems made it for you to do your work, take care of things at home, or get along with other people" will be provided at each assessment time point.

In addition, the number and percentage of subjects who achieve remission of depressive symptoms (defined as a PHQ-9 total score <5) will be summarized at each time point during the double-blind phase by treatment group. Subjects who do not meet such criterion will be considered as non-remitters.

The number of subjects who are responders on depressive symptoms based on the PHQ-9 will also be evaluated. The first definition of a responder on PHQ-9 will be based on an improvement in PHQ-9 total score of ≥6 points. Based on this definition, a subject is defined a responder (yes=1) at a given time point if the improvement in PHQ-9 total score is ≥6 points.

Subjects who do not meet such criterion will be considered as non-responders and will be assigned a value of 0 (i.e., no).

The second definition of a responder on PHQ-9 will be based on a ≥50% improvement in PHQ-9 total score from baseline. The percentage change from baseline for PHQ-9 total score is calculated as 100*(PHQ-9 total score at Day X – Baseline PHQ-9 total score)/(Baseline PHQ-9 total score). Negative percent changes in PHQ-9 total score indicate improvement (eg, percent change < -50% indicates improvement >50%). Based on this definition, a subject is defined a responder (yes=1) at a given time point if the percent reduction in PHQ-9 total score is ≥50%. Subjects who do not meet such criterion will be considered as non-responders and will be assigned a value of 0 (i.e., no).

For both definitions, the number and percentage of subjects who achieve a response will be summarized at each time point during the double-blind phase by treatment group. Observed case analysis, i.e., no imputation for subjects with missing values, will be explored.

The cumulative response rate, defined as the percentage of subjects experiencing at least a given value of percent reduction from baseline to the end of Week 6 in PHQ-9 total score, will be presented graphically, for both observed and LOCF data.

5.3.9. Snaith-Hamilton Pleasure Scale (SHAPS)

5.3.9.1. Definition

A secondary efficacy endpoint is the change from baseline to the end of Week 6 in SHAPS. The SHAPS is a reliable, valid, and unidimensional instrument to assess hedonic capacity in adults with MDD. It is a 14-item, self-report tool. Each of the items has a set of 4 response categories-Definitely Agree/Strongly Agree, Agree, Disagree, and Strongly Disagree, with either of the Disagree responses receiving a score of 1 and either of the Agree responses receiving a score of 0. The subject's item responses are summed to provide a total range (range of 0 to 14). A higher total SHAPS score indicates higher levels of present state of anhedonia.

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

5.3.9.2. Analysis Methods

Descriptive statistics of the actual values on Baseline, Day 8, Day 22, and Day 42 and the change from baseline to Day 8, 22, and 42 will be presented for SHAPS total score by treatment group.

The change from baseline to Day 8, 22, and 42 in SHAPS total score will be analyzed using the same MMRM as described in Section 5.2.3.1 for the primary efficacy endpoint, with the covariate "baseline total MADRS score" changed to "baseline SHAPS total score".

Frequency distributions of the SHAPS individual items will be provided at each assessment time point.

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

5.3.10.1. Definition

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 1. 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.

5.3.10.2. Analysis Methods

Descriptive statistics of the actual values on Baseline, Day 8, Day 22, and Day 42 and the change from baseline to Weeks 2, 4, and 6 will be presented for PROMIS-SD raw score and the T-score by treatment group.

The change from baseline to Day 8, 22, and 42 in the PROMIS-SD raw score and the T-score will be analyzed using the same MMRM as described in Section 5.2.3.1 for the primary efficacy endpoint, 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.

In addition, the correlations between baseline ISI score as a continuous variable and change from baseline to the end of Week 6 in the PROMIS-SD raw score and T-score will be evaluated by means of scatter plots and if appropriate, regression analysis.

5.3.11. Patient Reported Outcome Measurement Information System- Fatigue (PROMIS-Fatigue)

5.3.11.1. Definition

The PROMIS-Fatigue Short Form subscale consists of a static 8-item questionnaire. Using a recall period of the past 7 days, it assesses a range of symptoms from mild subjective feelings of tiredness to an overwhelming, debilitating, and sustained sense of exhaustion that likely decreases one's ability to execute daily activities and function normally in family or social roles. Fatigue is divided into the experience of fatigue (frequency, duration, and intensity) and the impact of fatigue on physical, mental, and social activities. The fatigue short form is generic rather than disease-specific.

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 fatigue.

"I feel fatigued" ranges from 1=not at all to 5=very much
"I have trouble starting things because I am tired" ranges from 1=not at all to 5=very much
"How run-down did you feel on average" ranges from 1=not at all to 5=very much
"How fatigued were you on average" ranges from 1=not at all to 5=very much
"How much were you bothered by your fatigue on average" ranges from 1=not at all to 5=very much

"To what degree did your fatigue interfere with your physical functioning" ranges from 1=not at all to 5=very much
"How often did you have to push yourself to get things done because of your fatigue" ranges from 1=never to 5=always
"How often did you have trouble finishing things because of your fatigue" ranges from 1=never to 5=always

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

5.3.11.2. Analysis Methods

Descriptive statistics of the actual values on Baseline, Day 8, Day 22, and Day 42 and the change from baseline to Day 8, 22, and 42 will be presented for PROMIS-Fatigue raw score and the T-score by treatment group.

The change from baseline to Day 8, 22, and 42 in the PROMIS-Fatigue raw score and the T-score will be analyzed using the same MMRM as described in Section 5.2.3.1 for the primary efficacy endpoint, with the covariate "baseline MADRS total score" changed to "baseline PROMIS-Fatigue raw score" or "baseline PROMIS-Fatigue T-score", respectively.

Frequency distributions of the PROMIS-Fatigue individual items will be provided at each assessment time point.

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

5.3.12.1. Definition

The PGI-S is a self-report scale to measure severity of illness (1=none, 2=mild, 3=moderate, 4=severe). Considering all aspects of depression, subjects will rate their severity on the PGI-S.

5.3.12.2. Analysis Methods

A frequency distribution over time of the PGI-S scores at Baseline, Day 8, Day 22, and Day 42 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 Day 8, 22, and 42 will be performed using an ANCOVA model on the ranks of the change in score with treatment, region, and baseline insomnia status as factors, and unranked baseline PGI-S score as a covariate.

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

5.3.13.1. Definition

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

The subject selects an answer for each of the 5 dimensions considering the response that best matches his or her health "today." The descriptive system can be represented as a health state. The EQ-VAS self-rating records the respondent's own assessment of his or her overall health status at the time of completion, on a scale of 0 (the worst health you can imagine) to 100 (the best health you can imagine.

Individual scores from the 5 dimensions will be used to obtain a weighted health status index as shown below:

(i) Scores from each dimension will be combined to obtain a 5L profile score: eg, a score of 1 for each dimension will give a 5L profile score of 11111. Dimension scores will be combined in the following order: Mobility, Self-Care, Usual Activities, Pain/Discomfort, Anxiety/Depression

(ii) The value set of the Health Status Index (HSI) for various values of 5L profile scores is published for Canada in the following website: https://www.ncbi.nlm.nih.gov/pubmed/26492214

(iii)The Canadian value set will be used to get the HSI values for all the countries participating in the study.

In addition, a sum score will be derived as follows: The scores of the five dimensions (values 1-5) will be added (sums between 5 and 25). From this score, subtract 5 (range 0-20) and multiply by 5 (range 0-100).

5.3.13.2. Analysis Methods

Descriptive statistics of actual values on Baseline, Day 8, Day 22, and Day 42 and changes from baseline to Day 8, 22, and 42 will be provided by treatment group for the weighted EQ-5D health status index, the EQ-VAS, and the sum score.

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

5.3.14. Work Limitations Questionnaire (WLQ)

5.3.14.1. Definition

The objective of the WLQ is to assess the on-the-job impact of chronic health problems and/or treatment ("work limitations") in adults. It is a PRO tool, consisting of 8 items in the short form version. The subject's impressions are captured using 5-point Likert scales. The recall period is 2 weeks.

Step 1: Assign Numerical Values to Each Response

There are 4 scales within the WLQ-Short Form, each consisting of 2 items, and each of the 8 items is given a score as follows:

Scale/Items	**Coding**
WLQ Time Management	
"Get going easily at the beginning of the workday" "Start on your job as soon as you arrived at work"	5=difficult all of the time 4=difficult most of the time 3=difficult some of the time 2=difficult a slight bit of the time 1=difficult none of the time [set to missing value]=does not apply to my job
WLQ Physical Tasks	
"Sit, stand, or stay in one position for longer than 15 minutes while working" "Repeat the same motions over and over again	1= able all of the time 2= able most of the time 3= able some of the time 4=able a slight bit of the time 5=able none of the time

while working"	[set to missing value]=does not apply to my job
WLQ Mental-Interpersonal Tasks	
"Concentrate on your work" "Speak with people in-person, in meetings or one the phone"	5=difficult all of the time 4=difficult most of the time 3=difficult some of the time 2=difficult a slight bit of the time 1=difficult none of the time [set to missing value]=does not apply to my job
WLQ Output Tasks	
"Handle the workload" "Finish work on time"	5=difficult all of the time 4=difficult most of the time 3=difficult some of the time 2=difficult a slight bit of the time 1=difficult none of the time [set to missing value]=does not apply to my job

Step 2: Score the Four WLQ Scales

After each of the 8 items is scored, the scores for the 4 WLQ scales are determined. For each the of 4 WLQ scales, the following 2 steps will be performed:

(i) Compute the average item score for a scale, including addressing missing data.

(ii) Convert the average item score to a 0-100 range.

Average Item Score

The average item score within a scale is calculated by summing the non-missing item scores within a scale and then dividing by the number of items with non-missing answers within the scale (i.e., not including items coded as missing). For example, if a respondent answers items in the Time Management scale as follows:

"Get going easily at the beginning of the workday" = 2

"Start on your job as soon as you arrived at work" = 4

First, sum the item scores: 2+4. Then divide by the number of items with valid answers, i.e., divide by 2. Your answer is therefore 6/2=3.0.

A scale score can still be calculated even if one of the responses within a scale is missing (either left blank or a "does not apply to my job" response). The scale score is calculated based on the

one non-missing item's score. In this case, the average item score is the value of the non-missing item score.

If both of the items within a scale are missing, the scale cannot be scored and the scale score will be treated as missing data.

Converting the Average Item Score to the Scale Score

To obtain the final WLQ Scale Score, convert average item score to a 0-100 range using the following formula:

WLQ Scale Score = 25*(average item score-1)

For example, using the average item score from above, the WLQ Time Management Scale Score is: 25*(3-1) = 50.

Step 3: Checking and Fixing Physical Scale Score Errors

Although the response categories in the Physical Scale are different than the other 3 scales, some people may not notice the change in wording. They report on the amount of time they had difficulty while they should be reporting on the amount of time they did not have difficulty.

If the answers to the Physical Scale items contradict the answers in the other 3 scales, it can be reasonably inferred that the respondent didn't notice the direction change.

For these cases, implement the Physical Scale score correction as follows.

If the Physical Scale score computed above is greater than or equal to 75 and the other three scale scores are less than or equal to 30, the Physical Scale needs to be reversed by subtracting the original physical scale score from 100.

For example, if the Physical Scale score is 80 and the other three scale scores are 30, 25, and 25 respectively, the corrected Physical Scale score is 100-80=20.

Step 4: Compute the Summary Score of Productivity Loss

After the 4 WLQ scale scores are determined, then the WLQ Productivity Loss Score is computed. The WLQ Productivity Loss Score indicates the percentage decrement in work output due to health problems. This is important when comparing WLQ results to other published scales, which sometimes measure productivity using the metric of time (productivity hours lost) or effort (percent effectiveness). The WLQ Productivity Loss Score tells you the estimated percent difference in an employee's at-work productivity compared to employees who do not have health-related work limitations (a healthy benchmark group).

To obtain the WLQ Productivity Loss Score, the following 2 steps are performed:

(i) Calculate the WLQ Index

(ii) Convert the WLQ Index into the WLQ Productivity Loss Score

Calculate the WLQ Index

The WLQ Productivity Loss score is based on a weighted sum of the scores from the 4 WLQ scales (Time Management, Physical, Mental-Interpersonal, and Output). The resulting score (known as the WLQ Index) is in the form of the natural log of work productivity. If a scale score is missing (in other words, it could not be computed using the procedures described previously), then Step 4 cannot be completed.

If there are scores for each of the WLQ's four scales, the following formula is used:

WLQ Index = (β1*WLQ Time Scale + β2*WLQ Physical Scale + β3*WLQ Mental-Interpersonal Scale + β4*WLQ Output Scale),
where β1 = 0.00048, β2 = 0.00036, β3 = 0.00096, and β4 = 0.00106

Convert the WLQ Index into the WLQ Summary Productivity Loss Score

The final step needed to generate the WLQ Productivity Loss Score is to convert the WLQ Index score to a percentage. The formula is as follows:

WLQ Productivity Loss Score = (1- exp (-WLQ Index))

('exp' stands for exponent. Therefore, to obtain the Productivity Loss Score, take the exponent of the negative WLQ Index score and subtract the result from 1.)

This result is then multiplied by 100 to express the score as a % of at-work productivity loss.

For example, if the four WLQ scale scores for the Time Management, Physical, Mental-Interpersonal, and Output scales were 25, 35, 25, and 40, respectively, the WLQ Index score would be calculated as:

(0.00048*25) + (0.00036*35) + (0.00096*25) + (0.00106*40) = 0.091.

Then the Productivity Loss score would be calculated as: 1-exp(-0.091)=0.08698. This is then multiplied by 100 (i.e., 0.08698*100 = 8.698%) to get the % of at-work productivity loss.

The resulting WLQ Productivity Loss score is a relative measure. It is interpreted as the percentage of productivity loss in the past two weeks due to presenteeism relative to a healthy benchmark sample. The benchmark sample consists of employees who had WLQ scale scores of zero (not limited by health). It is not necessary to further adjust scores by age, gender, or other demographic characteristics. The maximum attainable for WLQ index (with all scales at 100) is 28.6% and the maximum attainable productivity loss is 24.9%.

5.3.14.2. Analysis Methods

Descriptive statistics of actual values on Baseline, Day 8, Day 22, and Day 42 and changes from baseline to Day 8, 22, and 42 will be provided by treatment group for each of the 8 items, the 4 WLQ scales, and the WLQ Productivity Loss score.

Each of the 8 items will also be summarized at each visit with frequency counts and percentage of subjects by treatment group.

5.3.15. MADRS-6

5.3.15.1. 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 6 is calculated as (MADRS-6 total score at Week 6 – Baseline MADRS-6 total score). Negative changes in MADRS-6 total score indicate improvement.

5.3.15.2. Analysis Methods

Descriptive statistics of the actual values on Baseline, Day 8, Day 22, and Day 42 and the change from baseline to Day 8, 22, and 42 will be presented for MADRS-6 total score by treatment group.

The change from baseline to Day 8, 22, and 42 in MADRS-6 total score will be analyzed using the same MMRM as described in Section 5.2.3.1 for the primary efficacy endpoint, 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 the end of Week 6 will be analyzed using an ANCOVA model, using LOCF data. The derivation of the LOCF values is described in Section 5.2.3.3. The ANCOVA model will include factors for treatment, region, baseline insomnia status (present/absent), and baseline MADRS-6 total score as a continuous covariate. A 90% CI for the difference in LS Means and p-value will be calculated based on the contrast test statistic for each JNJ-42847922 dose level.

6. SAFETY

All safety analyses will be based on the safety analysis set.

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 drug 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 drug, and either the AE time or time of administration are missing, then the AE will be assumed to be treatment-emergent. If the AE date is recorded as partial or completely missing, then the AE will be considered to be treatment-emergent unless it is known to be prior to the first administration of study drug 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.

Summary tables will be provided for:

- TEAEs
- TEAEs occurring in ≥5% of subjects in any treatment group
- Treatment-emergent serious AEs (SAEs)
- TEAEs leading to discontinuation of study drug
- TEAEs by severity
- TEAEs by relationship to study drug

For the summaries of TEAEs by severity/relationship to study drug, the observation with the most severe occurrence/closest relationship to study drug 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 drug
- Had somnolence-related TEAEs
- Had TEAEs of special interest

6.2. Clinical Laboratory Tests

Descriptive statistics will be presented for 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, and urinalysis (pH and specific gravity) 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 3. 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 3. 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) or AST > 3* ULN will be presented for the double-blind phase. Additionally, incidence of treatment-emergent hepatic toxicity (Hy's Law[5]) defined as (ALT values >3*ULN or AST > 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 and at 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 5, 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 5: 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 6). The number and percentage of subjects who experience treatment-emergent orthostatic hypotension outside of pre-defined 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 6: 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 BP (mmHg)	< -20 mmHg
(2b) Diastolic BP (mmHg)	< -10 mmHg
BP = blood pressure 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, and Week 6. The ECGs will be read by a central reader. The ECG variables 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:
$QTcB\ (msec) = QT\ (msec) * (HR(bpm)/60)^{0.5}$;

Fridericia's formula:
$QTcF\ (msec) = QT\ (msec) * (HR(bpm)/60)^{0.33}$

Descriptive statistics for observed values and changes from average predose will be presented by treatment group for the above ECG variables 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 **Error! Reference source not found.**. 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 for the postbaseline value by treatment group.

Table 7: 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 and ≤470 msec for female)
- QTc (>450 to ≤480 msec for male and >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 >470 msec 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

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. 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 on Week 6, during the telephone contact, and at the 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 at each time point.

In addition, symptoms at follow-up (telephone contact and the follow-up visit) will be compared to the Week 6/early withdrawal visit and will be summarized as follows: new or worsened symptoms, symptoms present and unchanged, improved symptoms, and no symptoms. This analysis will be repeated including only subjects whose Week 6/early withdrawal visit is the day after their last dose of study drug.

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.

A listing of subjects with withdrawal symptoms following abrupt cessation of treatment (i.e., those subjects with new or worsened symptoms compared to the Week 6/early withdrawal visit) will be presented.

6.5.3. Arizona Sexual Experiences Scale (ASEX)

Effect on sexual functioning will be assessed using the ASEX at Baseline and Week 6. 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.

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 analyses are provided in a separate document.

REFERENCES

1. Bornkamp B, Pinheiro JC, and Bretz, F. MCPMod: An R Package for the Design and Analysis of Dose-Finding Studies. Journal of Statistical Software; 2009; 29(7):1-23.
2. Imai K, Keele L, Tingley D (2010). A general approach to causal mediation analysis. *Psychological Methods*, 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* 28:412–419, 1985.
4. Pinheiro J, Bornkamp B, Glimm E, Bretz, F. Model-based dose finding under model uncertainty using general parametric models. Stat Med. 2014;33(10):1646-1661.
5. 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: 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 2: Conversion of Raw Score to T-Score for PROMIS-Fatigue

Fatigue 8a *Short Form Conversion Table*		
Raw Score	**T-score**	**SE***
8	33.1	4.8
9	38.5	2.7
10	41.0	2.2
11	42.8	2.0
12	44.3	1.9
13	45.6	1.8
14	46.9	1.8
15	48.1	1.8
16	49.2	1.8
17	50.4	1.8
18	51.5	1.7
19	52.5	1.7
20	53.6	1.7
21	54.6	1.7
22	55.6	1.7
23	56.6	1.7
24	57.5	1.7
25	58.5	1.7
26	59.4	1.7
27	60.4	1.7
28	61.3	1.7
29	62.3	1.7
30	63.3	1.7
31	64.3	1.7
32	65.3	1.7
33	66.4	1.7
34	67.5	1.7
35	68.6	1.7
36	69.8	1.8
37	71.0	1.8
38	72.4	2.0
39	74.2	2.4
40	77.8	3.7

*SE = Standard Error

Attachment 3: 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.