Janssen Research & Development *

Statistical Analysis Plan

A Randomized, Double-blind, Multicenter, Active-Controlled Study of Intranasal Esketamine Plus an Oral Antidepressant for Relapse Prevention in Treatment-resistant Depression

Sustenance of Esketamine Treatment Response with Repeated Doses at Intervals Determined by Symptom Severity (SUSTAIN-1)

Protocol ESKETINTRD3003; Phase 3

JNJ-54135419 (esketamine)

Status:	Approved
Date:	16 February 2018
Prepared by:	Janssen Research & Development, LLC
Document No.:	EDMS-ERI-126947689, 8.0

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	8 June 2017
Amendment 1	20 June 2017
Amendment 2	21 June 2017
Amendment 3	16 Feb. 2018

Amendments below are listed beginning with the most recent amendment.

Amendment 3 (16 Feb. 2018)

The overall reason for the amendment: Modified the language pertaining to the primary estimand, corrected typos in the log-rank test statistics calculation, added details for sensitivity analysis, and updated AE terms of special interest and added analyses of TEAEs associated with discharge readiness.

Applicable Section(s)	Description of Change (s)
Section 2.4.2	Removed the sentences that transferred-entry subjects who continue to receive an oral antidepressant plus intranasal placebo will not be included for efficacy analyses during the optimization phase or maintenance phase. Clarified that these subjects will not be included in the full analysis sets.
Section 2.5	Added consented protocol (pre/post Protocol amendment 4) and entry source (direct-entry, transferred-entry) to subgroups.
Section 5.1.1	Clarified that a positive interim efficacy analysis result Z_{IA} favors intranasal esketamine plus oral antidepressant in delaying relapse compared to oral antidepressant plus intranasal placebo.
Section 5.2.1	Modified the language on the variable and intercurrent event for the primary estimand.
Section 5.2.2	Clarified that if the study is not terminated at the time of the interim analysis the final analysis will be performed on the Full (stable remitters) analysis set including all the relapse events accumulated prior to final database lock. Corrected typos in the formulas for log-rank test statistics, and further clarified the notations. Added that the estimate of the hazard ratio and confidence interval will calculated using software R and ADDPLAN.
Section 5.2.3	Added details of the multiple imputations procedures. Added unweighted log-rank tests on the Full (stable remitters) analysis set with all the relapse events accumulated prior to database lock and 59 events.
Section 6.1	Updated terms for two of the adverse events of special interest categories: drug abuse, dependence and withdrawal, and impaired cognition. Added analyses of TEAEs associated with discharge readiness.
References	Added a reference paper on multiple imputation approach

Amendment 2 (21 June 2017)

The overall reason for the amendment: Added amendment history section and included the updated version of Cogstate SAP

Applicable Section(s)	Description of Change (s)
Amendment History	Added Amendment History Section

Attachment 1	Included the updated version of the Cogstate SAP

Amendment 1 (20 June 2017)

The overall reason for the amendment: Modified the language pertaining to Interim analysis to be consistent with IA SAP Amendment 1.

Applicable Section(s)	Description of Change (s)
Section 1.4	Modified the timing of Interim Analysis to be consistent with IA SAP Amendment 1.
Section 3	Clarified that the data cut-off date for the interim analysis will only be known by Cytel, the independent statistical support group conducting the interim analysis. The changed language is consistent with IA SAP Amendment 1.

ABBREVIATIONS

AD	Antidepressant
AE	adverse event
ANCOVA	analysis of covariance
ASA	American Society of Anesthesiologists
BMI	body mass index
BP	Blood Pressure
BPIC-SS	Bladder Pain / Interstitial Cystitis Symptom Score
BPRS+	Four-item positive symptom subscale of the Brief Psychiatric Rating Scale
CADSS	Clinician Administered Dissociative States Scale
CGADR	Clinical Global Assessment of Discharge Readiness
CGI-S	Clinical Global Impression – Severity
CI	confidence interval
CRF	case report form
CSR	Clinical Study Report
C-SSRS	Columbia-Suicide Severity Rating Scale
DB	Double-blind
D/C	Discontinued
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders (4th edition)
ECG	Electrocardiogram
eCRF	electronic case report form
ECT	electroconvulsive therapy
EQ-5D-5L	EuroQol Group; 5 dimension; 5 level
EQ-VAS	EuroQol Group: visual analogue scale
EU	European Union
FDA	Food and Drug Administration
FU	Follow-Up
GAD-7	Generalized Anxiety Disorder 7-item scale
HVLT-R	Hopkins Verbal Learning Test-Revised
IDMC	Independent Data Monitoring Committee
ICH	International Conference on Harmonization
IDS-C_{30}	Inventory of Depressive Symptoms-Clinician rated, 30 item
IWRS	Interactive web response system
LOCF	last observation carried forward
MADRS	Montgomery-Asberg Depression Rating Scale
MDD	major depressive disorder
MedDRA	Medical Dictionary for Regulatory Activities
MINI	Mini International Neuropsychiatric Interview
MGH-ATRQ	Massachusetts General Hospital Antidepressant Treatment Response Questionnaire
MMRM	Mixed-effects model using repeated measures
MOAA/S	Modified Observer's Assessment of Alertness/Sedation
OL	Open-label
PD	Pharmacodynamics
PHQ-9	Patient Health Questionnaire – 9 item
PK	pharmacokinetic(s)
PWC-20	Physician Withdrawal Checklist; 20 item
SAE	serious adverse event
SAP	Statistical Analysis Plan
SD	standard deviation
SDS	Sheehan Disability Scale
SE	Standard error
SOC	System Organ Class
SNRI	Serotonin and Norepinephrine Reuptake Inhibitors
SSRI	Selective Serotonin Reuptake Inhibitors
TEAEs	Treatment-emergent adverse events
TEMA	treatment-emergent markedly abnormal
TRD	Treatment Resistant Depression

UPSIT	University of Pennsylvania Smell Identification Test
XR	extended release

1. INTRODUCTION

This statistical analysis plan (SAP) contains definitions of analysis sets, derived variables, and statistical methods for all planned analyses for study JNJ54135419-ESKETINTRD3003.

1.1. Trial Objectives

Primary Objective

The primary objective of this study is to assess the efficacy of intranasal esketamine plus an oral antidepressant compared with an oral antidepressant (active comparator) plus intranasal placebo in delaying relapse of depressive symptoms in subjects with treatment resistant depression (TRD) who are in stable remission (see Definitions of Terms below) after an induction and optimization course of intranasal esketamine plus an oral antidepressant.

Secondary Objectives

- To assess the efficacy of intranasal esketamine plus an oral antidepressant compared with an oral antidepressant (active comparator) plus intranasal placebo in delaying relapse of depressive symptoms in subjects with TRD with stable response (but who are not in stable remission) (see Definitions of Terms below) after an induction and optimization course of intranasal esketamine plus an oral antidepressant
- To assess the effect of intranasal esketamine plus an oral antidepressant compared with an oral antidepressant (active comparator) plus intranasal placebo on:
  - Depressive symptoms
  - Overall severity of depressive illness
  - Functional impairment and associated disability
  - Anxiety symptoms
  - Health-related quality of life and health status
- To investigate the safety and tolerability of intranasal esketamine plus an oral antidepressant compared with an oral antidepressant (active comparator) plus intranasal placebo in subjects with TRD, with special attention to the following:
  - Treatment-emergent adverse events (TEAEs), including AEs of special interest
  - Local nasal tolerability
  - Effects on heart rate, blood pressure, respiratory rate, and blood oxygen saturation
  - Effects on alertness and sedation
  - Potential psychosis-like symptoms
  - Dissociative symptoms
  - Potential effects on cognitive function
  - Potential effects on suicidal ideation/behavior
  - Potential treatment-emergent symptoms of cystitis and/or lower urinary tract symptoms

- Potential withdrawal and/or rebound symptoms following cessation of intranasal esketamine treatment
- Potential effects on sense of smell

Exploratory Objectives

- To assess the potential relationship of biomarkers with response/non-response to intranasal esketamine plus an oral antidepressant or to an oral antidepressant plus intranasal placebo in adult subjects with TRD
- To assess medical resource utilization

1.2. Trial Design

This is a randomized, double-blind, parallel-group, active-controlled, multicenter study to evaluate the efficacy, safety, and tolerability of intranasal esketamine plus an oral antidepressant compared with an oral antidepressant (active comparator) plus intranasal placebo in delaying relapse of depressive symptoms in adult men and women with TRD who are in stable remission after an induction and optimization course with intranasal esketamine plus an oral antidepressant.

Approximately 211 subjects in stable remission (see Definition of Terms below) at the end of the optimization phase will be randomized in a 1:1 ratio to either continue intranasal esketamine (same dose) or be switched to intranasal placebo; all subjects will continue the same oral antidepressant at the same dose.

An Independent Data Monitoring Committee (IDMC) will be commissioned for this study.

Subjects will enter the study either directly (referred to as direct-entry subjects) or after completing the double-blind induction phase of a short-term study (ESKETINTRD3001 or ESKETINTRD3002) (referred to as transferred-entry subjects).

The study has up to 5 phases:

- A 4-week screening/prospective observational phase, with an optional taper of up to 3 weeks for oral antidepressant medication(s) (direct-entry subjects only)
- A 4-week open-label induction phase (direct-entry subjects only)
- A 12-week optimization phase (open-label for direct-entry subjects and double-blind for transferred-entry subjects)
- A maintenance phase (double blind and variable duration)
- A 2-week follow-up phase

The maximum duration of a subject's participation will be variable, depending on whether he or she enters the study directly or is transferred from one of the double-blind short-term studies, and whether he or she meets phase-specific criteria (e.g., meets criteria for response at the end of the induction phase, is in stable remission/response at the end of the optimization phase, and when and if he or she relapses in the maintenance phase). Direct-entry subjects may participate in up to 5 phases and transferred-entry subjects may participate in up to 3 phases.

The study will be stopped once 84 relapses (in the subjects with stable remission) occur during the maintenance phase, or earlier based on the results of the interim analysis for efficacy. At the time the study is stopped, subjects in the induction phase will be able to complete the induction phase. Those subjects who are responders after completing an Early Withdrawal Visit may proceed to the 54135419TRD3008 study without completing the follow up phase. Those who are not responders will have an Early Withdrawal Visit and proceed directly to the follow-up phase. Subjects who are in the optimization or maintenance phase at the time the study is terminated will have an Early Withdrawal Visit/End of Maintenance Visit conducted and proceed directly to the follow-up phase.

Screening/prospective observational phase (4-week duration)

Direct-entry subjects will participate in this phase, which prospectively assesses treatment response to the subject's current oral antidepressant treatment regimen.

At the start of the screening/prospective observational phase, the subject must have had documented non-response to at least 1 antidepressant treatment (based on MGH-ATRQ) in the current episode of depression, and the subject is taking a different oral antidepressant treatment (listed on the MGH-ATRQ) for at least the previous 2 weeks at or above the minimum therapeutic dose. This antidepressant treatment, as well as any other ongoing medications being taken for depression at screening (including adjunctive/augmentation therapies), will continue from the start of Week 1 through the end of Week 4 of the screening/prospective observational phase. Dose adjustment is permitted per clinical judgment, but the oral antidepressant treatment is to remain at or above the minimum therapeutic dose (per the MGH-ATRQ) through the end of Week 4.

After 4 weeks, subjects who are non-responders to their current oral antidepressant treatment may be eligible to proceed to the open-label induction phase. Non-response at the end of the screening/prospective observational phase is defined as ≤25% improvement in the MADRS total score from Week 1 to Week 4 and a MADRS total score of ≥28 on Week 2 and Week 4.

Eligible subjects who are entering the open-label induction phase will discontinue all of their current antidepressant treatment(s), including adjunctive/augmentation therapies. Of note, subjects taking benzodiazepines (at dosages equal to or less than the equivalent of 6 mg/day of lorazepam) and/or permitted non-benzodiazepine sleep medications (eg, zolpidem, zaleplon) during the screening/prospective observational phase can continue these medications. No dose increases beyond the equivalent of 6 mg/day lorazepam, or new benzodiazepine are permitted during the screening/prospective observational phase. If clinically indicated, a subject's current antidepressant treatment(s) may be tapered and discontinued over an additional, optional period of up to 3 weeks per the local prescribing information or clinical judgment.

Open-Label induction phase (4-week duration)

Eligible direct-entry subjects will receive intranasal esketamine (flexible dose: 56 mg or 84 mg) treatment sessions twice weekly for 4 weeks. In addition, all subjects will initiate a new, open-label oral antidepressant on Day 1 that will be taken daily for the duration of the induction phase. The assigned oral antidepressant will be 1 of 4 oral antidepressant medications (duloxetine, escitalopram, sertraline, or venlafaxine extended release [XR]), that the subject has not previously had a nonresponse to in the current depressive episode, has not been previously intolerant to (lifetime), and is available in the participating country.

At the end of the induction phase, subjects who are responders (defined as ≥50% reduction in the MADRS total score from baseline [Day 1 prior to the first intranasal dose] to the end of the 4-week open-label induction phase) may be eligible to proceed to the optimization phase. All subjects who do not proceed to the optimization phase will have an Early Withdrawal visit conducted and proceed to the follow-up phase.

At the time the study is stopped, subjects in the induction phase will be able to complete the phase. Those who are responders, after completing an Early Withdrawal Visit, may proceed to the 54135419TRD3008 study without completing the follow up phase. Those who are not responders will have an Early Withdrawal Visit conducted and proceed directly to the follow-up phase.

Optimization Phase

Eligible direct-entry subjects from the open-label induction phase and transferred-entry subjects from the 2 double-blind short-term studies (ESKETINTRD3001 and ESKETINTRD3002) will participate in this 12-week phase.

The intranasal treatment session frequency will be reduced from that in the induction phase (twice weekly) to weekly for the first 4 weeks of this phase. After the first 4 weeks, the frequency of intranasal treatment sessions will be individualized to either once weekly or once every other week based on the severity of depressive symptoms, as assessed by the MADRS total score. The dose of intranasal esketamine will remain unchanged from the dose at the end of the induction phase. All subjects will continue taking the same oral antidepressant treatment (at the same dosage) that was initiated during the induction phase.

At the end of the optimization phase, subjects in stable remission and those with stable response (but who are not in stable remission) may be eligible to continue into the maintenance phase; all other subjects will have an Early Withdrawal visit conducted and proceed to the follow-up phase.

For subjects in stable remission and those with stable response at the end of this phase, the last visit of the optimization phase (Visit 3.13; Week 16) also serves as the baseline visit (Visit 4.1; Week 16) of the maintenance phase. Subjects eligible for the maintenance phase will be randomized and receive their first double-blind intranasal treatment session of the maintenance phase at this visit.

At the time the study is stopped, subjects in the optimization phase will have an Early Withdrawal visit conducted and proceed directly to the follow-up phase.

Definitions of terms:

- Stable remission: MADRS total score ≤12 for at least 3 of the last 4 weeks of the optimization phase, but one excursion of a MADRS total score >12 or one missing MADRS assessment is permitted at Optimization week 13 or 14 only. The MADRS total score at weeks 15 and 16 must be ≤12.
- Stable response: ≥50% reduction in the MADRS total score from baseline (Day 1 of induction phase; pre-randomization/prior to the first intranasal dose) in each of the last 2 weeks of the optimization phase, but does not meet criteria for stable remission. Note: For transferred-entry subjects, Day 1 of the induction phase will take place in ESKETINTRD3001 or ESKETINTRD3002.

Maintenance Phase

On Day 1 of this phase:

- Approximately 211 subjects in stable remission at the end of the optimization phase (after treatment with intranasal esketamine plus an oral antidepressant) will be randomized in a 1:1 ratio to either continue with intranasal esketamine (same dose) and the same oral antidepressant or to continue with the same oral antidepressant (active comparator) but switch to intranasal placebo. The primary efficacy analysis will be performed for these subjects only.
- Additionally, subjects with stable response (but who are not in stable remission) at the end of the optimization phase (after treatment with intranasal esketamine plus an oral antidepressant) will be randomized in a 1:1 ratio (using a separate randomization list) to either continue with intranasal esketamine (same dose) and the same oral antidepressant or to continue with the same oral antidepressant (active comparator) but switch to intranasal placebo (for a secondary efficacy analysis only).
- Transferred-entry subjects who achieve stable remission or stable response at the end of the optimization phase after treatment with an oral antidepressant plus intranasal placebo will continue to receive the same treatment in order to maintain the blinding for the ongoing short-term studies. These subjects will not be included in the efficacy analyses, but will be included in safety analyses.

The frequency of intranasal treatment sessions will be further individualized during the maintenance phase to once weekly or once every other week based on the severity of depressive symptoms, as assessed by the MADRS total score. Subjects will only be permitted to switch from weekly to every other week dosing a total of 3 times during the maintenance phase. After this time, if a given subject is unable to sustain improvement on every other week dosing they will remain on a weekly dosing regimen for the duration of this phase.

This phase will have a variable duration, continuing until 84 relapses occur in the subjects with stable remission, or earlier based on interim analysis results.

Subjects who meet the relapse criteria and subjects who remain relapse-free at study termination will have an End of Maintenance Phase visit conducted and may proceed to the follow-up phase. If clinically indicated, subjects who have met relapse criteria after completing the end of maintenance visit may proceed to the open-label safety extension study, 54135419TRD3008, without completing the follow up phase. Subjects who are participating in the maintenance phase

at the time the study is stopped will have an Early Withdrawal Visit conducted and proceed directly to the follow-up phase.

Relapse is defined as any of the following:

- MADRS total score ≥22 for 2 consecutive assessments separated by 5 to 15 days. The date of the second MADRS assessment will be used for the date of relapse
- Hospitalization for worsening depression or any other clinically relevant event determined per clinical judgment to be suggestive of a relapse of depressive illness such as suicide attempt, completed suicide, or hospitalization for suicide prevention. If hospitalized for any of these events, the start date of hospitalization will be used for the date of relapse. Otherwise the date of the event will be used if the subject is not hospitalized.

In case both relapse criteria are met, the earlier date will be defined as the date of relapse for this subject

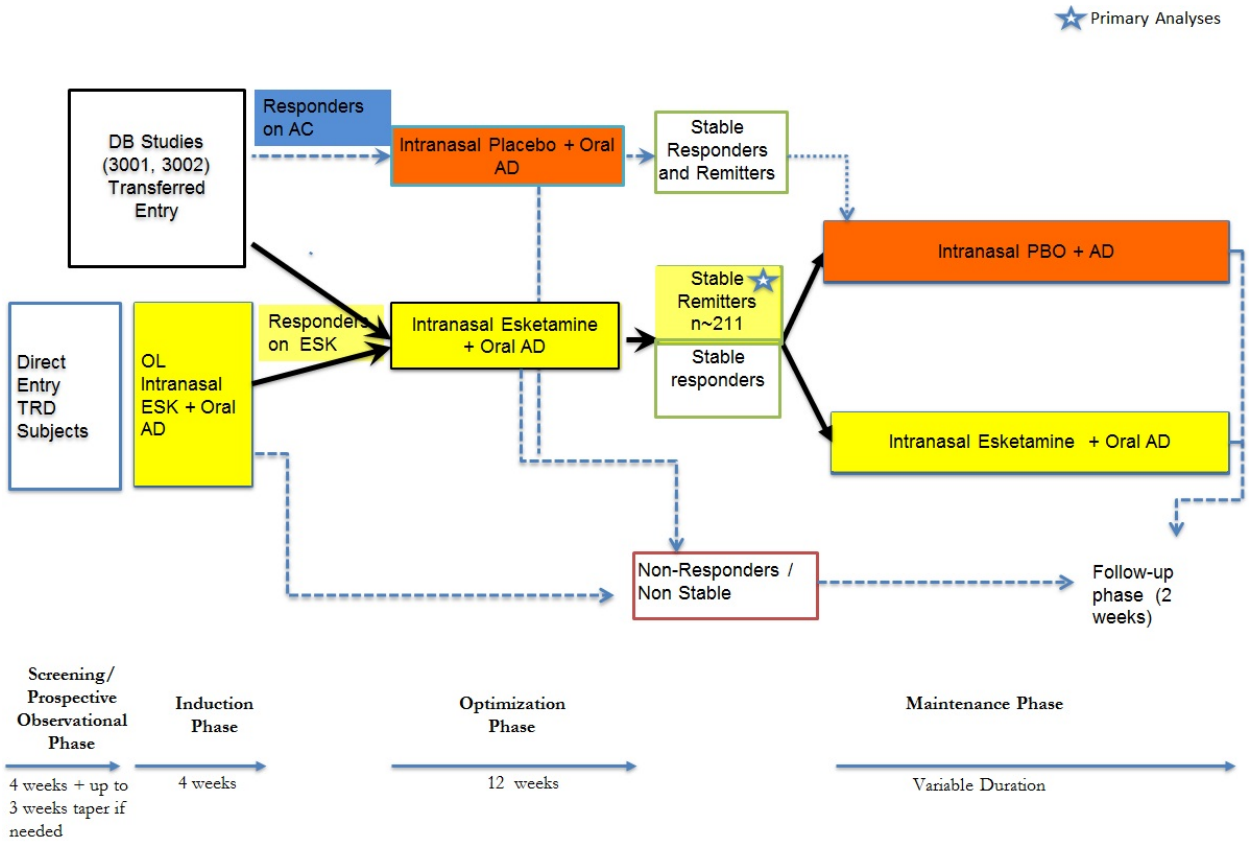
Follow-up Phase

This phase will include all subjects who have received at least 1 dose of intranasal study medication in this study. Follow-up visits will be performed at 1 and 2 weeks after the last clinic visit.

At the start of the follow-up phase, further clinical/standard of care for the treatment of depression will be arranged by the study investigator and/or the subject's treating physician. There will be no intranasal treatment administered during this phase. Subjects will be provided with an additional 2-week supply of the oral antidepressant medication to ensure that there is no interruption of oral antidepressant therapy during the transition to further clinical/standard of care. The decision to continue the oral antidepressant in this phase will be at the discretion of the investigator; however, in order to better assess potential withdrawal symptoms from intranasal study drug, it is recommended that the oral antidepressant be continued during the 2-week follow-up phase unless determined as not clinically appropriate.

A diagram of the study design is provided in Figure 1

Figure 1: Schematic Overview of the Study



AC = active comparator; AD = antidepressant; DB = double-blind; ESK = esketamine; OL = open-label; PBO = placebo. TRD = treatment-resistant depression.
The study will end when 84 relapses are reached. An interim analysis will be performed at 30 relapses.

1.3. Statistical Hypotheses for Trial Objectives

The hypothesis for this study is that intranasal esketamine plus an oral antidepressant is more effective than treatment with an oral antidepressant plus intranasal placebo in delaying relapse of depressive symptoms in subjects with TRD in stable remission.

1.4. Sample Size Justification

The maximum number of relapses (in the subjects with stable remission) required by this study is 84, which provides 90% power to detect a hazard ratio of 0.493 at the 1-sided significance level of 0.025 for a fixed-sample design to detect superiority of esketamine plus oral antidepressant over antidepressant alone in delaying relapse of depressive symptoms in subjects with TRD who are in stable remission. The calculation of sample size assumed that the time to the first relapse follows an exponential distribution, with a median time of 6 months for oral antidepressant alone and 12.17 months for intranasal esketamine plus oral antidepressant (hazard ratio = 0.493). The corresponding 6-month relapse rates are 50% for oral antidepressant alone and 28.95% for oral antidepressant plus intranasal esketamine.

Assumptions were made for accrual period and rate, maximum study duration, and dropout rate. Based on such assumptions, a total of approximately 211 subjects in stable remission need to be randomized (in a 1:1 ratio) in order to obtain 84 relapses.

Interim Analysis

To evaluate the assumptions used in the sample size calculation, relapse rates will be monitored sequentially during the maintenance phase. In particular, a 2-stage group-sequential design will be adopted, with 1 interim analysis (IA) to be performed when at least 33 relapse events have occurred in stable remitters with at least 30 relapses from subjects treated with intranasal esketamine plus an oral antidepressant in the optimization phase. If 33 relapses are reported in stable remitters and the notification is not triggered by the IWRS system, further determination of the timing of interim analysis will be made at every third relapse reported in stable remitters. Making this assessment at every third relapse will help in maintaining the blind. Early termination of the maintenance phase for efficacy will be based on interim analysis results. At the interim analysis, if the study is not stopped for efficacy then a sample size re-estimation will be performed. The study will continue until a maximum of 84 relapses are observed.

If the study is not terminated at the interim analysis, the sample size will be re-estimated to ensure a conditional power at Stage 2 of at least 90% with a minimum number of relapses after interim to be 29 and a maximum number of relapses after interim to be 54 (with 30 relapses having occurred before the IA).

A rigorous interim statistical analysis plan (SAP) and charter will be developed detailing the algorithm for a sample size re-estimation based on the interim data and how the analysis will be executed. An IDMC will perform the interim analysis and will make recommendations for any sample size adjustment based on the rules defined in the interim SAP. Any changes to sample size will be communicated by the IDMC (or the statistician from the Statistical Support Group) to the IWRS vendor to ensure that the appropriate number of subjects is enrolled in the study. None of the esketamine team members or staff members at the investigational sites conducting the clinical study will be informed of the results of the interim analysis and any adjustments that will be made to the sample size; however, the clinical supplies group will be informed of the decision made at the interim analysis so that only the required amount of study medication will be packaged.

Procedures will be in place to ensure that the results of the interim analysis do not influence the conduct of the study, investigators, or subjects.

1.5. Randomization and Blinding

Blinded intranasal treatment will be used in the optimization phase (for transferred-entry subjects only) to avoid unblinding the short-term studies at the individual level, and in the maintenance phase (for all subjects) to reduce potential bias during data collection and evaluation of clinical endpoints. An intranasal placebo control will be used in the maintenance phase to establish the frequency and magnitude of changes in clinical endpoints that may occur in the absence of intranasal active treatment.

Central randomization will be implemented in the maintenance phase of this study. At the start of the maintenance phase, subjects in stable remission after treatment with intranasal esketamine plus an oral antidepressant will be randomly assigned to 1 of 2 intranasal medication treatment

groups (intranasal esketamine or intranasal placebo) in a 1:1 ratio based on a computer-generated randomization schedule prepared before the study by or under the supervision of the sponsor. The randomization will be balanced by using randomly permuted blocks and will be stratified by country. All subjects will continue their current oral antidepressant.

Although not part of the primary analysis, subjects in stable response (but who are not in stable remission) after treatment with intranasal Esketamine plus an oral antidepressant will also be randomly assigned to 1 of 2 intranasal medication treatment groups in a 1:1 ratio based on a separate computer-generated randomization schedule (ie, different schedule for the subjects in stable remission) prepared before the study by or under the supervision of the sponsor. The randomization will be balanced by using randomly permuted blocks and will be stratified by country. All subjects will continue their current oral antidepressant.

Transferred-entry subjects who achieve stable remission or stable response after treatment with intranasal placebo plus an oral antidepressant will continue to receive the same treatment in order to maintain the blinding for the ongoing short-term studies. These subjects will not be randomized during the maintenance phase.

The investigator will not be provided with randomization codes. The codes will be maintained within the IWRS, which has the functionality to allow the investigator to break the blind for an individual subject.

Data that may potentially unblind the treatment assignment (e.g., study drug plasma concentrations, study drug accountability data) will be handled with special care to ensure that the integrity of the blind is maintained and the potential for bias is minimized.

2. GENERAL ANALYSIS DEFINITIONS

2.1. Analysis Phases

There are 5 analysis phases defined in this study:

- A 4-week screening/prospective observational phase, with an optional taper of up to 3 weeks for oral antidepressant medication(s) (direct-entry subjects only)
- Open-label induction phase (direct-entry subjects only)
- Optimization phase (open-label for direct-entry subjects and double-blind for transferred-entry subjects)
- Maintenance phase (double blind and of variable duration)
- Follow-up (post treatment) phase

Each analysis phase has its own analysis phase start and end dates.

2.1.1. Study Reference Start and End Dates

For direct-entry subjects, the reference start date for the study is defined as the earlier of date of the first dose of intranasal study medication or oral antidepressant (the date is missing for

screened subjects who did not take any intranasal study drug or oral antidepressant). For transferred-entry subjects, the reference start date of the study is the date of first dose of oral antidepressant in the optimization phase, if available; in instances where the start date of oral antidepressant medication is not available, the reference start date is the informed consent date. The reference end date for the study is the end of trial date including the last follow-up visit.

2.1.2. Analysis Phase Start and End Dates

Screening/Prospective Observational Phase

The start date of the screening/prospective observational phase is the informed consent date. For subjects who continue to the open-label induction phase, the screening/prospective observational phase end date is the open-label induction phase start date.

The start and end dates for the screening/prospective observational phase are defined for direct entry subjects only.

Open-Label Induction Phase

The open-label induction phase will begin on the day (referred to as, 'IND start date') of the first dose of intranasal medication or oral antidepressant taken in the open-label induction phase (only for direct-entry subjects). For subjects who continue to the optimization phase, the open-label induction phase end date (referred to as, 'IND end date') (only for direct-entry subjects) is the date of first dose of oral antidepressant medication taken in the optimization phase or the date of the last visit in the open-label induction phase if the date of the first oral medication is missing. For subjects who discontinue in the open-label induction phase, the IND end date is the maximum of the date of last visit in the open-label induction phase and the date of early termination.

The start date/time of the open-label induction phase (referred to as, 'IND start date/time') is the IND start date and the time of the first dose of intranasal study medication in this phase. If no intranasal study medication is administered or it is administered after the start of oral antidepressant, then the time will be left blank.

The start and end dates for the open-label induction phase are defined for direct entry subjects only.

Optimization Phase

The optimization phase will begin on the day of first dose of oral antidepressant medication in this phase (referred to as, 'OP start date'). If the date of first dose of oral antidepressant is missing, the OP start date for direct-entry subjects is the date of last visit in the open label induction phase and for the transfer-entry subjects it is the date of their informed consent. For subjects who continue to the maintenance phase, the optimization phase end date (referred to as, 'OP end date') is the date when the intranasal medication was started in the maintenance phase. For subjects who discontinue in the optimization phase, the OP end date is the maximum of the date of last visit in the optimization phase and the date of early termination.

Maintenance Phase

The maintenance phase will begin on the day of the first dose of intranasal medication or oral antidepressant taken in the maintenance phase (referred to as, 'MA start date'). For subjects who complete/discontinue from the maintenance phase, the maintenance phase end date (referred to as, 'MA end date') is the maximum of the date of last visit in the maintenance phase, date of completion of the maintenance phase due to relapse or study termination, and date of early withdrawal in the maintenance phase.

Follow-Up Phase

The start date of the Follow-up (post-treatment) phase (referred to as 'F/U start date') is the day after the end date for the last treatment phase the subject participated in. The follow-up phase end date (referred to as 'F/U end date') is the maximum of the last follow-up visit date or the end of trial date.

2.1.3. Study Day and Relative Day

Study day is calculated relative to the reference start date for the study. Relative day is calculated relative to the analysis phase start date of the analysis phase in which the data are captured. A minus (-) sign indicates days prior to the start of study or prior to the start of the analysis phase.

Study day for an event on or after the start of the study is calculated as:

event date- reference start date+1.

Study day for an event prior to the start of the study is calculated as:

event date-reference start date

Relative day for an event on or after the analysis phase start date is calculated as:

event date – analysis phase start date + 1.

Relative day for an event prior to the analysis phase start date is calculated as:

event date – analysis phase start date.

There is no study day 0 or relative day 0.

2.2. Baseline and End Point

Baseline is defined for each parameter/assessment.

- Open-label induction phase:
  - Direct-entry subjects: The last observation prior to or on the start date of open-label induction phase is denoted as, 'Baseline (IND)'. This pertains to direct-entry subjects.
  - Transferred-entry subjects: The last observation prior to or on the start date of the double-blind induction phase will be used as the baseline. This is also denoted as

'Baseline (IND)'. (Baseline (IND) is computed in the databases for studies 3001 and 3002 and will be copied into study 3003).

- The 'Average Predose' value for the ECG measurements is defined as the average of all non-missing assessments on or before the first dose of study medication. For transferred-entry subjects the 'Average Predose' is computed in the databases for studies 3001 and 3002 and will be copied into study 3003.

- Optimization phase: The last observation prior to or on the start date of optimization phase is denoted as, 'Baseline (OP)'.
- Maintenance phase: The last observation prior to or on the start date of maintenance phase is denoted as, 'Baseline (MA)'.

For each variable measured over time, the 'End Point (IND)' value is defined as the last postbaseline assessment value during the open-label induction Phase. This value will be the same as the Baseline (OP) value for subjects who continue into the optimization phase.

The 'End Point (OP)' value is defined as the last postbaseline assessment value during the optimization phase. This value will be the same as the Baseline (MA) value for subjects who continue into the maintenance phase for scales without pre/post dose measurement.

The 'End Point (MA)' value is defined as the last postbaseline assessment value during the maintenance phase.

2.3. Visit Windows

As subjects do not always adhere to the protocol visit schedule, the following rules are applied to assign actual visits to protocol visits. Listed below are the visit windows for analysis and the target days for each visit. The reference day is Study Day 1 (which is the first day that study drug was taken in the open-label induction phase for direct-entry subjects and the first day that study drug was taken in the double-blind induction phase for transferred-entry subjects).

If a subject has 2 or more scheduled or unscheduled visits in one visit window, the visit closest to the target day will be used as the protocol visit for that visit window. If 2 actual visits are equidistant from the target day within a visit window, the later visit is used. If a visit window has no scheduled visits but does have unscheduled visits, then the unscheduled visit closest to the scheduled visit will be used.

All assignments will be made in chronological order. Once a visit is assigned to a visit window, it will no longer be used for a later time point except for the end point.

Listed below are the visit windows and the target days (if applicable) for each visit defined in the protocol for all phases (Table 1).

Table 1: Analysis Visits

Parameter	Analysis Phase	Scheduled Day	Time Interval (label on output)	Time Interval (Day) [a]	Target Time Point from start of each phase (Day)
MADRS	Screening		Week 1 (SC)		
			Week 2 (SC)		
			Week 4 (SC)		
	IND	1	Baseline (IND)	<= 1	1
		8	Day 8 (IND)	2-11	8
		15	Day 15 (IND)	12-18	15
		22	Day 22 (IND)	19-24	22
		28	Day 28 (IND)	25 to end of IND	28
		IND final visit	End Point (IND)	2 to end of IND	
	OP	28	Baseline (OP)	<= 1	1
		32	Week 1 (OP)	2-8	5
		39	Week 2(OP)	9-15	12
		46	Week 3 (OP)	16-22	19
		53	Week 4 (OP)	23-29	26
		60	Week 5 (OP)	30-36	33
		67	Week 6 (OP)	37-43	40
		74	Week 7 (OP)	44-50	47
		81	Week 8 (OP)	51-57	54
		88	Week 9 (OP)	58-64	61
		95	Week 10 (OP)	65-71	68
		102	Week 11 (OP)	72-78	75
		109	Week 12 (OP)	79 to end of OP	82
		OP final visit	End Point (OP)	2 to end of OP	
	MA Weekly x=1,2, 3…	109	Baseline (MA)	<= 1	1
		116	Week 1 (MA)	2-11	8
		116 + x*7	Week (x+1) (MA)	5 + x*7 to 11 + x*7	8 + x*7
		MA final visit	End Point (MA)	2 to end of MA	
	F/U	F/U Day7 (Last dose +7 days)	F/U Week 1	1 - 10	7
		F/U Day 14 (Last dose +14 days)	F/U Week 2	11 to end of F/U	14
		F/U final visit	End Point (F/U)	1 to end of F/U	
CGI-S	Screening		Week 1 (SC)		
	IND	1	Baseline (IND)	<=1	1
		4	Day 4 (IND)	2-6	4
		8	Day 8 (IND)	7-9	8
		11	Day 11 (IND)	10-13	11
		15	Day 15 (IND)	14-18	15
		22	Day 22 (IND)	19-24	22
		28	Day 28 (IND)	25 to end of IND	28
		IND final visit	End Point (IND)	2 to end of IND	

Table 1: Analysis Visits

Parameter	Analysis Phase	Scheduled Day	Time Interval (label on output)	Time Interval (Day) [a]	Target Time Point from start of each phase (Day)
	OP	28	Baseline (OP)	<= 1	1
		32	Week 1 (OP)	2-8	5
		39	Week 2(OP)	9-15	12
		46	Week 3 (OP)	16-22	19
		53	Week 4 (OP)	23-29	26
		60	Week 5 (OP)	30-36	33
		67	Week 6 (OP)	37-43	40
		74	Week 7 (OP)	44-50	47
		81	Week 8 (OP)	51-57	54
		88	Week 9 (OP)	58-64	61
		95	Week 10 (OP)	65-71	68
		102	Week 11 (OP)	72-78	75
		109	Week 12 (OP)	79 to end of OP	82
		OP final visit	End Point (OP)	2 to end of OP	
	MA x = 1,2,3…	109	Baseline (MA)	<= 1	1
		123	Week 2 (MA)	2 – 22	15
		137	Week 4 (MA)	23 – 32	29
		144	Week 5 (MA)	33 – 39	36
		144 + x*7	Week (x + 5) (MA)	33+ x*7 to 39 + x*7	36 + x*7
		MA final visit	End Point (MA)	2 to end of MA	
	F/U	F/U Day 14 (Last dose +14 days)	F/U Week 2	1 to end of F/U	14
PHQ-9 and EQ-5D-5L	Screening		Week 1 (SC)		
	IND	1	Baseline (IND)	<= 1	1
		15	Day 15 (IND)	2-21	15
		28	Day 28 (IND)	22 to end of IND	28
		IND final visit	End Point (IND)	2 to end of IND	
	OP	28	Baseline (OP)	<= 1	1
		39	Week 2 (OP)	2-19	12
		53	Week 4 (OP)	20-33	26
		67	Week 6 (OP)	34-47	40
		81	Week 8 (OP)	48-61	54
		95	Week 10 (OP)	62-75	68
		109	Week 12 (OP)	76 to end of OP	82
		OP final visit	End Point (OP)	2 to end of OP	
	MA Every 2 weeks x= 1,2,3…	109	Baseline (MA)	<= 1	1
		123	Week 2 (MA)	2-22	15
		137	Week 4 (MA)	23-36	29
		137+14*x	Week (4 + x*2) (MA)	23 + x*14 to 36 + x*14	29 + x*14
		MA final visit	End Point (MA)	2 to end of MA	
	F/U	F/U Day 14 (Last dose +14 days)	F/U Week 2	1 to end of F/U	14

Table 1: Analysis Visits

Parameter	Analysis Phase	Scheduled Day	Time Interval (label on output)	Time Interval (Day) [a]	Target Time Point from start of each phase (Day)
GAD-7	Screening		Week 1 (SC)		
	IND	1	Baseline (IND)	<= 1	1
		28	Day 28 (IND)	2 to end of IND	28
	OP	28	Baseline (OP)	<= 1	1
		53	Week 4 (OP)	2-33	26
		67	Week 6 (OP)	34-47	40
		81	Week 8 (OP)	48-68	54
		109	Week 12 (OP)	69 to end of OP	82
		OP final visit	End Point (OP)	2 to end of OP	
	MA Every 4 weeks x=1,2,3...	109	Baseline (MA)	<= 1	1
		137	Week 4 (MA)	2-43	29
		137+x*28	Week (4+x*4) (MA)	16+ x*28 to 43+x*28	29 + x*28
		MA final visit	End Point (MA)	2 to end of MA	
	F/U	F/U Day 14 (Last dose +14 days)	F/U Week 2	1 to end of F/U	14
SDS	Screening		Week 1 (SC)		
	IND	1	Baseline (IND)	<=1	1
		15	Day 15 (IND)	2-21	15
		28	Day 28 (IND)	22 to end of IND	28
		IND final visit	End Point (IND)	2 to end of IND	
	OP	28	Baseline (OP)	<= 1	1
		53	Week 4 (OP)	2-40	26
		81	Week 8 (OP)	41-68	54
		109	Week 12 (OP)	69 to end of OP	82
		OP final visit	End Point (OP)	2 to end of OP	
	MA Every 4 weeks x=1,2,3…	109	Baseline (MA)	<= 1	1
		137	Week 4 (MA)	2-43	29
		137+x*28	Week (4+x*4) (MA)	16+ x*28 to 43 + x*28	29 + x*28
		MA final visit	End Point (MA)	2 to end of MA	
	F/U	F/U Day 14 (Last dose +14 days)	F/U Week 2	1 to end of F/U	14
C-SSRS (BL/SC version and since last visit version)	Screening		Week 2 (SC)		
			Week 4 (SC)		
	IND	1	Baseline (IND)	<=1	1
		4	Day 4 (IND)	2-6	4
		8	Day 8 (IND)	7-9	8
		11	Day 11 (IND)	10-13	11
		15	Day 15 (IND)	14-16	15
		18	Day 18 (IND)	17-20	18
		22	Day 22 (IND)	21- 23	22
		25	Day 25 (IND)	24-26	25
		28	Day 28 (IND)	27 to end of IND	28
		IND final visit	End Point (IND)	2 to end of IND	

Table 1: Analysis Visits

Parameter	Analysis Phase	Scheduled Day	Time Interval (label on output)	Time Interval (Day) [a]	Target Time Point from start of each phase (Day)
	OP	28	Baseline (OP)	<= 1	1
		32	Week 1 (OP)	2-8	5
		39	Week 2(OP)	9-15	12
		46	Week 3 (OP)	16-22	19
		53	Week 4 (OP)	23-29	26
		60	Week 5 (OP)	30-36	33
		67	Week 6 (OP)	37-43	40
		74	Week 7 (OP)	44-50	47
		81	Week 8 (OP)	51-57	54
		88	Week 9 (OP)	58-64	61
		95	Week 10 (OP)	65-71	68
		102	Week 11 (OP)	72 to end of OP	75
		OP final visit	End Point (OP)	2 to end of OP	
	MA Weekly x=1,2, 3…	102	Baseline (MA)	<= 1	1
		116	Week 1 (MA)	2-11	8
		116 + x*7	Week (x+1) (MA)	5 + x*7 to 11 + x*7	8 + x*7
		MA final visit	End Point (MA)	2 to end of MA	
	F/U	F/U Day 14 (Last dose +14 days)	F/U Week 2	1 to end of F/U	14
MOAA/S[c], Pulse Oximetry[d] (predose and every 15 minutes to 1.5H) BPRS+ and CADSS (predose, 40M, 1.5H) CGADR[e] (1H, 1.5H)	IND	1	Day 1 (IND)	<=1	1
		4	Day 4 (IND)	2-6	4
		8	Day 8 (IND)	7-9	8
		11	Day 11 (IND)	10-13	11
		15	Day 15 (IND)	14-16	15
		18	Day 18 (IND)	17-20	18
		22	Day 22 (IND)	21- 23	22
		25	Day 25 (IND)	24 to end of IND	25
	OP	32	Week 1 (OP)	1-8	5
		39	Week 2(OP)	9-15	12
		46	Week 3 (OP)	16-22	19
		53	Week 4 (OP)	23-29	26
		60	Week 5 (OP)	30-36	33
		67	Week 6 (OP)	37-43	40
		74	Week 7 (OP)	44-50	47
		81	Week 8 (OP)	51-57	54
		88	Week 9 (OP)	58-64	61
		95	Week 10 (OP)	65-71	68
		102	Week 11 (OP)	72 to end of OP	75
	MA Weekly x=1,2, 3…	109	Day 1 (MA)	1	1
		116	Week 1 (MA)	2-11	8
		116 + x*7	Week (x+1) (MA)	5 + x*7 to 11 + x*7	8 + x*7

Table 1: Analysis Visits

Parameter	Analysis Phase	Scheduled Day	Time Interval (label on output)	Time Interval (Day) [a]	Target Time Point from start of each phase (Day)
BPIC-SS	IND	1	Baseline (IND)	<= 1	1
		15	Day 15 (IND)	2-21	15
		28	Day 28 (IND)	22 to end of IND	28
		IND final visit	End Point (IND)	2 to end of IND	
	OP	28	Baseline (OP)	<= 1	1
		32	Week 1 (OP)	2-15	5
		53	Week 4 (OP)	16-40	26
		81	Week 8 (OP)	41-68	54
		109	Week 12 (OP)	69 to end of OP	82
		OP final visit	End Point (OP)	2 to end of OP	
	MA Every 4 weeks x=1,2, 3…	109	Baseline (MA)	<= 1	1
		137	Week 4 (MA)	2-43	29
		137+x*28	Week (4 + x*4) (MA)	16 + x*28 to 43 + x*28	x*28 + 29
		MA final visit	End Point (MA)	2 to end of MA	
	F/U	F/U Day 14 (Last dose +14 days)	F/U Week 2	1 to end of F/U	14
PWC-20	IND	IND final visit	End Point (IND)	2 to end of IND	
	OP (Only for subjects not entering MA)	OP final visit	End Point (OP)	1 to end of OP	
	MA	MA final visit	End Point (MA)	1 to end of MA	
	F/U	F/U Day 7 (Last dose +7 days)	F/U Week 1	1 to 10	7
		F/U Day 14 (Last dose +14 days)	F/U Week 2	11 to end of F/U	14
		F/U final	End Point (F/U)	1 to end of F/U	
Nasal Symptom Questionnaire	IND	1	Day 1 (IND)	<= 1	1
		4	Day 4 (IND)	2-7	4
		11	Day 11 (IND)	8-14	11
		18	Day 18 (IND)	15-21	18
		25	Day 25 (IND)	22 to end of IND	25
		IND final visit	End Point (IND)	1/postdose to end of IND	
	OP	32	Week 1 (OP)	1-8	5
		39	Week 2(OP)	9-15	12
		46	Week 3 (OP)	16-22	19
		53	Week 4 (OP)	23-29	26
		60	Week 5 (OP)	30-36	33
		67	Week 6 (OP)	37-43	40
		74	Week 7 (OP)	44-50	47
		81	Week 8 (OP)	51-57	54
		88	Week 9 (OP)	58-64	61
		95	Week 10 (OP)	65-71	68
		102	Week 11 (OP)	72 to end of OP	75
		OP final visit	End Point (OP)	1/postdose to end of OP	
	MA	109	Day 1 (MA)	1	1

Table 1: Analysis Visits

Parameter	Analysis Phase	Scheduled Day	Time Interval (label on output)	Time Interval (Day) [a]	Target Time Point from start of each phase (Day)
	Every 4 weeks x=1,2, 3…	123	Week 2 (MA)	2-22	15
		137	Week 4 (MA)	23-43	29
		165	Week 8 (MA)	44 - 71	57
		165+ x*28	Week (8+ x*4) (MA)	44 + x*28 to 71 + x*28	57 + x*28
		MA final visit	End Point (MA)	1/postdose to end of MA	
Smell Threshold Test	IND	1	Baseline (IND)	<= 1	1
		28	Day 28 (IND)	2 to end of IND	28
		IND final visit	End Point (IND)	2 to end of IND	
	OP	28	Baseline (OP)	<= 1	1
		109	Week 12 (OP)	2 to end of OP	82
		OP final visit	End Point (OP)	2 to end of OP	
	MA Every 12 weeks x=1,2, 3…	109	Baseline (MA)	<= 1	1
		193	Week 12 (MA)	2-127	85
		277	Week 24 (MA)	128-211	169
		277 + x*84	Week (24 + x*12)	128 + x*84 to 211 + x*84	169 + x*84
		MA final visit	End Point (MA)	2 to end of MA	
UPSIT	IND	1	Baseline (IND)	<= 1	1
		15	Day 15 (IND)	2 -21	15
		28	Day 28 (IND)	22 to end of IND	28
		IND final visit	End Point (IND)	2 to end of IND	
	OP	28	Baseline (OP)	<= 1	1
		81	Week 8 (OP)	2 to end of OP	54
		OP final visit	End Point (OP)	2 to end of OP	
	MA Every 8 weeks x=1,2, 3…	81	Baseline (MA)	<=1	1
		137	Week 4 (MA)	2-57	29
		193	Week 12 (MA)	58-113	85
		193 + 56*x	Week (12 + x*8)	58 + x*56 to 113 + x*56	85 + x*56
		MA final visit	End Point (MA)	2 to end of MA	
HRUQ	OP	39	Week 2 (OP)	<= 19	12
		53	Week 4 (OP)	20-33	26
		67	Week 6 (OP)	34-47	40
		81	Week 8 (OP)	48-61	54
		95	Week 10 (OP)	62-75	68
		109	Week 12 (OP)	76 to end of OP	82
		OP final visit	End Point (OP)	2 to end of OP	
	MA Every 2 weeks x=1,2, 3…	109	Baseline (MA)	<= 1	1
		123	Week 2 (MA)	2-22	15
		137	Week 4 (MA)	23-36	29
		137+14*x	Week (4+x*2) (MA)	x*14+23 to x*14+36	x*14 + 29
		MA final visit	End Point (MA)	2 to end of MA	
	F/U	F/U Day 14 (Last dose +14 days)	F/U Week 2	1 to end of F/U	14

Table 1: Analysis Visits

Parameter	Analysis Phase	Scheduled Day	Time Interval (label on output)	Time Interval (Day) [a]	Target Time Point from start of each phase (Day)
Hematology, chemistry from LAB	IND	1	Baseline (IND)	<=1	1
		28	Day 28 (IND)	2 to end of IND	28
		IND final visit	End Point (IND)	2 to end of IND	
	OP	28	Baseline (OP)	<= 1	1
		53	Week 4 (OP)	2-40	26
		81	Week 8 (OP)	41-68	54
		109	Week 12 (OP)	69 to end of OP	82
		OP final visit	End Point (OP)	2 to end of OP	
	MA Every 4 weeks x=1,2, 3…	109	Baseline (MA)	<= 1	1
		137	Week 4 (MA)	2-43	29
		137+x*28	Week (4+x*4) (MA)	16+ x*28 to 43 + x*28	29 + x*28
		MA final visit	End Point (MA)	2 to end of MA	
	F/U	F/U Day 14 (Last dose +14 days)	F/U Week 2	1 to end of F/U	14
Vital Signs (TEMP [predose at each visit], BP[b], Pulse, RR, (at each visit, predose, 40M, 1H, 1.5H)	IND	1	Baseline (IND)[f] Day 1 (IND): 40M Day 1 (IND): 1H Day 1 (IND): 1.5H	<=1/predose	1
		4	Day 4 (IND): Predose Day 4 (IND): 40M Day 4 (IND): 1H Day 4 (IND): 1.5H	2-6	4
		8	Day 8 (IND): Predose Day 8 (IND): 40M Day 8 (IND): 1H Day 8 (IND): 1.5H	7-9	8
		11	Day 11 (IND): Predose Day 11 (IND): 40M Day 11 (IND): 1H Day 11 (IND): 1.5H	10-13	11
		15	Day 15 (IND): Predose Day 15 (IND): 40M Day 15 (IND): 1H Day 15 (IND): 1.5H	14-16	15
		18	Day 18 (IND): Predose Day 18: 40M Day 18: 1H Day 18: 1.5H	17-20	18
			Day 22 (IND): Predose		

Table 1: Analysis Visits

Parameter	Analysis Phase	Scheduled Day	Time Interval (label on output)	Time Interval (Day) [a]	Target Time Point from start of each phase (Day)
		22	Day 22 (IND): 40M Day 22 (IND): 1H Day 22 (IND): 1.5H	21-23	22
		25	Day 25 (IND): Predose Day 25 (IND): 40M Day 25 (IND): 1H Day 25 (IND): 1.5H	24 to end of IND	25
		IND final visit	End Point (IND)	Day 1: 40M to end of IND	
	OP	25	Baseline (OP)	<= 1	1
		32	Week 1 (OP): Predose Week 1 (OP): 40M Week 1 (OP): 1H Week 1 (OP): 1.5H	1-8	5
		39	Week 2 (OP): Predose Week 2 (OP): 40M Week 2 (OP): 1H Week 2 (OP): 1.5H	9-15	12
		46	Week 3 (OP): Predose Week 3 (OP): 40M Week 3 (OP): 1H Week 3 (OP): 1.5H	16-22	19
		53	Week 4 (OP): Predose Week 4 (OP): 40M Week 4 (OP): 1H Week 4 (OP): 1.5H	23-29	26
		60	Week 5 (OP): Predose Week 5 (OP): 40M Week 5 (OP): 1H Week 5 (OP): 1.5H	30-36	33
		67	Week 6 (OP): Predose Week 6 (OP): 40M Week 6 (OP): 1H Week 6 (OP): 1.5H	37-43	40
		74	Week 7 (OP): Predose Week 7 (OP): 40M Week 7 (OP): 1H Week 7 (OP): 1.5H	44-50	47
			Week 8 (OP):		

Table 1: Analysis Visits

Parameter	Analysis Phase	Scheduled Day	Time Interval (label on output)	Time Interval (Day) [a]	Target Time Point from start of each phase (Day)
		81	Predose Week 8 (OP): 40M Week 8 (OP): 1H Week 8 (OP): 1.5H	51-57	54
		88	Week 9 (OP): Predose Week 9 (OP): 40M Week 9 (OP): 1H Week 9 (OP): 1.5H	58-64	61
		95	Week 10 (OP): Predose Week 10 (OP): 40M Week 10 (OP): 1H Week 10 (OP): 1.5H	65-71	68
		102	Week 11 (OP): Predose Week 11 (OP): 40M Week 11 (OP): 1H Week 11 (OP): 1.5H	72-79	75
		109 (only for Predose)	Week 12 (OP) Predose	80 to end of OP	82
		OP final visit	End Point (OP)	Day 1: 40M to end of OP/Day 1(MA):Predose	
	MA Weekly x=1,2, 3…	109	Baseline (MA) [g] Day 1 (MA): 40M Day 1 (MA): 1H Day 1 (MA): 1.5H	<=1/predose	1
		116	Week 1 (MA) Predose Week 1 (MA): 40M Week 1 (MA): 1H Week 1 (MA): 1.5H	2-11	8
		116 + x*7	Week (x+1) (MA) Predose Week (x+1) (MA): 40M Week (x+1) (MA): 1H Week (x+1) (MA): 1.5H	x*7 + 5 to x*7 + 11	x*7 + 8
		MA final visit	End Point (MA)	Day 1: 40M to end of MA	
	F/U	F/U Day 14 (Last dose +14 days)	F/U Week 2	1 to end of F/U	14

Table 1: Analysis Visits

Parameter	Analysis Phase	Scheduled Day	Time Interval (label on output)	Time Interval (Day) [a]	Target Time Point from start of each phase (Day)
Urinalysis	IND	1	Baseline (IND)	<= 1	1
		15	Day 15 (IND)	2-21	15
		28	Day 28 (IND)	22 to end of IND	28
		IND final visit	End Point (IND)	2 to end of IND	
	OP	28	Baseline (OP)	<= 1	1
		53	Week 4 (OP)	2-40	26
		81	Week 8 (OP)	41-68	54
		109	Week 12 (OP)	69 to end of OP	82
		OP final visit	End Point (OP)	2 to end of OP	
	MA Every 4 weeks x=1,2, 3…	109	Baseline (MA)	<= 1	1
		137	Week 4 (MA)	2-43	29
		137+x*28	Week (4+x*4) (MA)	16+ x*28 to 43+x*28	x*28 + 29
		MA final visit	End Point (MA)	2 to end of MA	
	F/U (only for Urinalysis)	F/U Day 14 (Last dose +14 days)	F/U Week 2	1 to end of F/U	14
12-lead ECG	IND	Screening, 1	Average Predose	<=1/Predose	1
		1	Day 1 (IND):1H	1	1
		8	Day 8 (IND): 1H	2-11	8
		15	Day 15 (IND): 1H	12-20	15
		25	Day 25 (IND): 1H	21 to end of IND	25
		IND final visit	End Point (IND)	2 to end of IND	
	OP	25	Baseline (OP)	<= 1, Predose	1
		32	Week 1 (OP): 1H	1-15	5
		53	Week 4 (OP): 1H	16-29	26
		60	Week 5 (OP): 1H	30-40	33
		74	Week 7 (OP): 1H	41-50	47
		81	Week 8 (OP): 1H	51-57	54
		88	Week 9 (OP): 1H	58-68	61
		102	Week 11 (OP): 1H	69 to end of OP	75
		OP final visit	End Point (OP)	2 to end of OP	
	MA Every 4 weeks x=1,2, 3…	109	Baseline (MA)	<= 1, Predose	1
		109	Day 1(MA): 1H	1	1
		123	Week 2 (MA): 1H	2 - 29	15
		151	Week 6 (MA): 1H	30 - 57	43
		151 + x*28	Week (6 + x*4) (MA): 1H	30 + x*28 to 57+x*28	x*28 + 43
		MA final visit	End Point (MA)	2 to end of MA	
	F/U	F/U Day 14 (Last dose +14 days)	F/U Week 2	1 to end of F/U	14

Table 1: Analysis Visits

Parameter	Analysis Phase	Scheduled Day	Time Interval (label on output)	Time Interval (Day) [a]	Target Time Point from start of each phase (Day)
Nasal Exam, Weight and BMI	IND	1	Baseline (IND)	<=1	1
		28	Day 28 (IND)	2 to end of IND	28
		IND final visit	End Point (IND)	2 to end of IND	
	OP	28	Baseline (OP)	<= 1	1
		53	Week 4 (OP)	2-43	26
		81	Week 8 (OP)	44-68	54
		109	Week 12 (OP)	69 to end of OP	82
		OP final visit	End Point (OP)	2 to end of OP	
	MA Every 4 weeks x=1,2, 3…	109	Baseline (MA)	<= 1	1
		137	Week 4 (MA)	2-43	29
		137+x*28	Week (4+x*4) (MA)	16+ x*28 to 43+x*28	x*28 + 29
		MA final visit	End Point (MA)	2 to end of MA	
	F/U (not for Weight and BMI)	F/U Day 14 (Last dose +14 days)	F/U Week 2	1 to end of F/U	14
Cogstate/ HVLT-R	IND	1	Baseline (IND)	<=1	1
		28	Day 28 (IND)	2 to end of IND	28
	OP	28	Baseline (OP)	<= 1	1
		109	Week 12 (OP)	2 to end of OP	82
	MA Every 12 weeks x=1,2, 3…	109	Baseline (MA)	<= 1	1
		193	Week 12 (MA)	2-127	85
		277	Week 24 (MA)	128-211	169
		277 + x*84	Week (24 + x*12)	128 + x*84 to 211 + x*84	169 + x*84
	F/U	F/U Day 14 (Last dose +14 days)	F/U Week 2	1 to end of F/U	14

[a] For each phase, the time interval is relative to the first day of that phase.

[b] During each phase, at 1.5 hours post dose if the SBP is ≥160 and/or DBP≥100, assessments should continue every 30 minutes until the blood pressure is <160 and/or <100 or in the investigator's clinical judgment, the subject is clinically stable and can be discharged from the clinical site, or the subject is referred for appropriate medical care, if clinically indicated.

[c] If the MOAA/S score is ≤ 3 at any time during the 1.5 hour postdose interval, the MOAA/S will be performed every 5 minutes until a score of 4 is reached (at which point a frequency of every 15 minutes can be resumed until t=+1.5 hours post dose).) If a subject does not have a score of at least 5 at t=+1.5 hours postdose, they should continue to be monitored. For subjects with a score of 4, the assessment should be repeated every 15 minutes. And for subjects with a score of ≤3, the assessment should be repeated every 5 minutes until the score returns to 5 or the subject is referred for appropriate medical care, if clinically indicated.

[d] If pulse oximetry is <93% at any time during the 1.5 hour postdose interval, pulse oximetry will be performed every 5 minutes until oxygen saturation returns to ≥93% or until the subject is referred for appropriate medical care, if clinically indicated.

[e] If the response is not "Yes" at 1.5 hour postdose, the assessment will be repeated every 15 minutes until a "Yes" response is achieved or until the subject is referred for appropriate medical care if clinically indicated.

[f] This would be labeled as Day 1(IND): Predose for changes from Predose summaries.

[g] This would be labeled as Day 1(MA): Predose for changes from Predose summaries.

2.4. Analysis Sets

Subjects will be classified into the following analysis sets: all enrolled, full analysis sets, interim full analysis set, safety analysis sets and follow-up analysis set. Due to Good Clinical Practice (GCP) issues, [redacted] subjects will not be included in any of the analysis sets. However, data for this site will be presented in listings.

2.4.1. All Enrolled Analysis Set

This analysis set will include all transferred-entry and direct-entry subjects who are not screen failures.

2.4.2. Efficacy Analysis Sets

2.4.2.1. Full Analysis Sets

Full (IND): All subjects who receive at least 1 dose of intranasal study drug and 1 dose of oral antidepressant in the open-label induction phase (for direct-entry subjects only).

Full (OP): All subjects who receive at least 1 dose of intranasal esketamine study drug and 1 dose of oral antidepressant in the optimization phase.

Maintenance phase: There are 2 full analysis sets defined for this phase.

One set will be used to perform primary and secondary efficacy evaluations on randomized subjects who are in stable remission at the end of the optimization phase and who receive at least 1 dose of intranasal study drug and 1 dose of oral antidepressant during the maintenance phase. This analysis set will be referred to as 'Full (stable remitters)'.

Another set will be used to perform secondary efficacy evaluations on randomized subjects who are stable responders (who are not stable remitters) at the end of the optimization phase and who receive at least 1 dose of intranasal study drug and 1 dose of oral antidepressant during the maintenance phase. This analysis set will be referred to as 'Full (stable responders)'.

Transfer entry oral AD plus intranasal placebo subjects who are stable remitters or stable responders will not be included in the above analysis sets.

2.4.3. Interim Full Analysis Set

Interim Full (stable remitters): All subjects who are in stable remission at the end of the optimization phase and who receive at least 1 dose of intranasal study drug and 1 dose of oral antidepressant during the maintenance phase at the time of the interim analysis data cutoff (see IA SAP for details).

2.4.4. Safety Analysis Sets

The following safety analysis sets are defined for each phase. Analyses of change from baseline will include only those subjects who have baseline and at least 1 post-baseline observation in that phase.

Safety (IND) analysis set: All subjects who receive at least 1 dose of intranasal study medication or 1 dose of oral antidepressant in the open-label induction phase (for direct-entry subjects only).

Safety (OP) analysis set: All subjects who receive at least 1 dose of intranasal esketamine study medication or 1 dose of oral antidepressant in the optimization phase (who are not in Safety (OP TEP) analysis set).

Safety (OP_TEP) analysis set: Transferred-entry subjects who continue to receive an oral antidepressant plus intranasal placebo will be summarized separately in the optimization phase.

Safety (MA) analysis set: All randomized subjects who receive at least 1 dose of intranasal study medication or 1 dose of oral antidepressant during the maintenance phase. This analysis set includes both stable remitters and stable responders (who are not stable remitters).

Safety (stable remitters) analysis set: All randomized subjects who receive at least 1 dose intranasal study medication or 1 dose of oral antidepressant during the maintenance phase. This analysis set includes only stable remitters and only AE summaries will be provided for this set.

Safety (MA_TEP) analysis set: Transferred-entry subjects who continue to receive an oral antidepressant plus intranasal placebo will be summarized separately in the maintenance phase. Safety (MA_TEP) analysis set and Safety (MA) analysis set are mutually exclusive.

2.4.5. Follow-up Analysis Set

The Follow-up analysis set includes all subjects who enter the follow-up phase. This analysis set will be used to summarize all efficacy and safety evaluations during the follow-up phase.

2.5. Definition of Subgroups

Analyses will be provided for the primary endpoint, time to relapse, for the Full (stable remitters) analysis set by the following subgroups.

- Sex
- Race (White, Black, Other)
- Age Group in years (18-44 years, 45-64 years)
- Region: North America (US and Canada), Europe (Belgium, Czech Republic, Estonia, France, Germany, Hungary, Italy, Poland, Slovakia, Spain, Sweden, Turkey), Other (Mexico and Brazil)
- Country (Brazil, Belgium, Canada, Czech Republic, Estonia, France, Germany, Hungary, Italy, Mexico, Poland, Slovakia, Spain, Sweden, Turkey, US)
- Number of Previous antidepressant Treatment Failures in Current Episode (based on ATRQ data)
- Functional Impairment based on Baseline (IND) SDS Total Score: not impaired (0-3), mild (4-11), moderate (12-19), marked (20-26) or extreme (27-30)
- Class of antidepressant study treatment (SNRI or SSRI)

- Consented Protocol (pre/post protocol amendment 4)
- Entry source (direct-entry, transferred-entry)

2.6. Imputation Rules for Missing AE Dates

Treatment-emergent adverse events (AEs)

- For direct-entry subjects, treatment-emergent AEs are those events with an onset date/time on or after the start of IND phase study medication, and occurred on or before the end of the maintenance phase. A conservative approach will be used to handle the missing dates for AEs
- For transferred-entry subjects, treatment-emergent AEs are those events with an onset date/time on or after the start of optimization phase study medication, and occurred on or before the end of the maintenance phase. A conservative approach will be used to handle the missing dates for AEs

Onset Date

AEs for each phase are those events with an onset date/time on or after the start of that particular phase, and occurred before the end of that phase. The rules for estimating incomplete AE onset dates will be as follows:

Direct-entry subjects:
If the onset date of an adverse event is missing the day only, it will be set to:

i) First day of the month that the AE occurred, if month/year of the onset of AE is different than the month/year of IND start date

ii) The day of IND start date, if the month/year of the onset of AE is the same as month/year of the IND start date and month/year of the AE resolution date is different

iii) The day of IND start date or day of AE resolution date, whichever is earliest, if month/year of the onset of AE and month/year of the IND start date and month/year of the AE resolution date are the same.

If the onset date of an adverse event is missing both day and month, it will be set to the earliest of:

i) January 1 of the year of onset, as long as this date is after the IND start date

ii) One day after the IND start date, if this date is the same year that the AE occurred.

A completely missing onset date of an adverse event will be set to the IND start date.

Transferred-entry subjects:
If the onset date of an adverse event is missing day only, it will be set to:

i) First day of the month that the AE occurred, if month/year of the onset of AE is different than the month/year of OP start date

ii) The day of OP start date, if the month/year of the onset of AE is the same as month/year of the OP start date and month/year of the AE resolution date is different

iii) The day of OP start date or day of AE resolution date, whichever is earliest, if month/year of the onset of AE and month/year of the OP start date and month/year of the AE resolution date are the same.

If the onset date of an adverse event is missing both day and month, it will be set to the earliest of:

i) January 1 of the year of onset, as long as this date is after the OP start date

ii) One day after the OP start date, if this date is the same year that the AE occurred.

A completely missing onset date of an adverse event will be set to the OP start date.

Similar rules will be applied for MA missing onset date

Resolution Date

The missing day of resolution of an adverse event will be set to the last day of the month of resolution.

If the resolution date of an adverse event is missing both day and month, it will be set to December 31 of the year.

A completely missing resolution date of an adverse event that is not recorded as ongoing will be set to the date of withdrawal or study completion.

Direct-entry subjects:

If the time of onset is missing, it will be imputed as follows:

(i) 00:00 if the date of onset is after IND start date

(ii) 00:00 if the date is the same as IND start date, but the intranasal study medication in the (open-label) induction phase was started after the oral antidepressant medication in this phase

(iii) The time of intranasal medication start in the (open-label) induction phase if the intranasal medication was started on or before the oral antidepressant medication in this phase

Transferred-entry responder subjects:

If the time of onset is missing, it will be imputed as follows:

(i) 00:00 if the date of onset is after OP start date

(ii) 00:00 if the date is the same as OP start date, but the intranasal study medication in the optimization/maintenance phase was started after the oral antidepressant medication in this phase

(iii) The time of intranasal medication start in the optimization/maintenance phase if the intranasal medication was started on or before the oral antidepressant medication in this phase

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

2.6.1. Imputation Rules for Missing Concomitant Medication Dates

If a partial date is reported, it is assumed the medication (or therapy) was taken in all phases that overlap with the partial date. If both start and end dates are missing but this concomitant medication was taken both prior to the study entry and still ongoing at study end, it is assumed medication was taken in all phases.

The rules for estimating an incomplete concomitant medication start date are as follows:

Direct-entry subjects: If the month of the concomitant medication start date is equal to the month of the start of the induction phase, then the estimated start date is the IND start date;

If the month of the concomitant medication start date is greater than the month of the start of the induction phase and earlier than the study end date, then the estimated start date of the concomitant medication is the first day of the month;

If the month of the concomitant medication start date is greater than the month of the study end date, then no imputation will be done;

If the month and year of the concomitant medication start date are known and the IND start date is after the month of the concomitant medication start date, then no imputation will be done;

If either the month or year of the concomitant medication start date is missing, no imputation is to be performed.

Transferred-entry subjects:

If the month of the concomitant medication start date is equal to the month of the start of the optimization phase, then the estimated start date is the OP start date;

If the month of the concomitant medication start date is greater than the month of the start of the optimization phase and earlier than the study end date, then the estimated start date of the concomitant medication is the first day of the month;

If the month of the concomitant medication start date is greater than the month of the study end date, then no imputation will be done;

If the month and year of the concomitant medication start date are known and the OP start date is after the month of the concomitant medication start date, then no imputation will be done;

If either the month or year of the concomitant medication start date is missing, no imputation is to be performed.

3. INTERIM ANALYSIS AND DATA MONITORING COMMITTEE REVIEW

An IDMC will be established to monitor data on an ongoing basis to ensure the continuing safety of the subjects enrolled in this study. In addition, the committee will review the interim analysis data. The committee will meet every 6 months to review safety data and will meet once to review

efficacy data after the interim analysis has been completed. After the reviews, the IDMC will make recommendations regarding the continuation of the study, or in the case of the interim analysis for efficacy, to either stop the study due to efficacy or to adjust the sample size (i.e., number of relapses) to achieve the desired power while maintaining control of the overall Type I error.

The data cut-off date for the interim analysis is defined as the date when the 30th relapse event is observed in the maintenance phase for stable remitters treated with intranasal esketamine plus an oral antidepressant in the optimization phase (IDMC SAP). This date will only be known by Cytel, the independent statistical support group conducting the interim analysis. If the result of the interim analysis using the log-rank test is significant for the Interim Full (stable remitters) analysis set at a level of 0.0086 (two-sided), the study will be stopped based on the IDMC recommendation. Note that the trial will be ongoing during the period of conducting the interim analysis (data cleaning, analysis and holding the interim analysis committee meeting), which leads to overrunning data, including possible succeeding events. In this case, a final analysis using all data will be analyzed using the log-rank test to support the findings from the interim analysis. The interim analysis of time to relapse is considered as primary and the final analysis is considered as supportive if the study is stopped due to the efficacy at the time of the interim analysis. If the study is not stopped based upon the interim analysis results, a sample size re-estimation will be performed. The study will continue until the total number of relapses is at least 59, but no more than 84. The final analysis using the weighted log-rank test (Section 5.2.2) for the Full (stable remitters) analysis set will be conducted. In this case, the final analysis to compare intranasal ESK+AD with PBO+AD in terms of delaying time to relapse will be considered as primary.

4. SUBJECT INFORMATION

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 for the All Enrolled, Full (Stable Remitters and Stable Responders), Safety (MA and Stable Remitters) analysis sets (described in Section 2.4). Continuous variables will be summarized using descriptive statistics (N, mean, standard deviation [SD], median, minimum, and maximum). Categorical variables will be summarized using a frequency distribution with the number and percentage of subjects in each category. Baseline (IND) values will be used for these summaries.

Table 2: Demographic Variables and Baseline Characteristics

Continuous Variables:

- Age (years) (informed consent date (for transfer entry subjects use 3001/3002 informed consent date) – date of birth + 1) / 365.25
- Baseline weight (kg)
- Baseline height (cm)
- Baseline BMI (kg/m^2) calculated as Weight (kg)/[Height (m)]^2

Categorical Variables:

- Age in years (18-44, 45-64)
- Sex (male, female)
- Race[a] (White, Black or African American, Asian, American Indian or Alaskan native, Native Hawaiian or other Pacific islander, other)
- Ethnicity (Hispanic or Latino, not Hispanic or Latino)
- Baseline BMI (underweight <18.5kg/m^2; normal: 18.5 to <25 kg/m^2, overweight: 25 kg/m^2 to <30 kg/m^2, obese:30 to <40 kg/m^2; morbidly obese: ≥40 kg/m^2)
- Employment Status
- Hypertension Status
- Class of antidepressant (SSRI/SNRI)
- Oral antidepressant
- Country
- Region

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

Table 3: Psychiatric History at Baseline Variables

Continuous Variables:

- Baseline MADRS total score
- Baseline IDS-C_{30} total score
- Baseline CGI-S score
- Baseline PHQ-9 total score
- Age (years) when diagnosed with MDD

Categorical Variables:

- Baseline CGI-S score
- Baseline C-SSRS category (no event, suicidal ideation, suicidal behavior)
- Antidepressant treatment history (number of medications with non-response taken for at least 6 weeks during the current episode as obtained in the MGH-ATRQ)
- Family history of
  - Depression
  - Anxiety Disorder
  - Bipolar Disorder
  - Schizophrenia
  - Alcohol Abuse
  - Substance Abuse

4.2. Disposition Information

The number of subjects who enrolled from study ESKETINTRD3001, study ESKETINTRD3002 and direct-entry into ESKETINTRD3003 will be provided.

The following disposition summaries by treatment groups will be provided for each phase separately. These summaries will be provided for each of the Full and Safety analysis sets described in Section 2.4

- The number of subjects who entered a specific treatment phase
- The number of subjects who completed a specific treatment phase and their reasons for completion (only applicable for the maintenance phase). The reasons for completion for the maintenance phase include relapse during the maintenance phase or completion of the maintenance phase without a relapse.
- The number of subjects who discontinued a specific treatment phase prematurely and their reasons for discontinuation
- The number of subjects who are ongoing in each phase at the time the sponsor terminated the study

The number of subjects who terminated the trial and the reasons for ending study participation will also be summarized.

4.3. Extent of Exposure

Extent of exposure in terms of total duration of exposure and number of dosing sessions of intranasal study medication will be summarized by phase and treatment group for Full (Stable Remitters and Stable Responders) and Safety (IND, OP, and MA) analysis sets, and across phases for the All Enrolled described in Section 2.4.

The total duration of exposure for the intranasal study drug and for each type of oral antidepressant (AD) during each phase is defined as the time between the first and the last dose of each type of study medication in that specific phase (last day of study medication-first day of study medication +1). If a subject only receives a partial dose it is considered as a day of dosing.

Modal dose for a subject is defined as the most frequently taken dose by a subject during that phase. Mean dose of a subject is calculated as the sum of doses during the phase divided by the total number of days exposed. The final dose is the last non-zero dose received during that phase. The calculation of mean, modal and final dose will exclude days off study drug.

Descriptive statistics (N, mean, SD, median, minimum and maximum) of modal dose, mean dose and final dose will be presented for the induction phase for intranasal study drug.

Doses of oral AD will be summarized using descriptive statistics of the mean dose (days on drug), final dose and mode dose (days on drug), by each type of oral AD during the induction phase.

At the end of the induction phase and the optimization phase, the number and percentage of subjects at each dose (56 mg and 84 mg) will be provided.

At Weeks 8 and 12 of the optimization phase, the number and percentage of subjects at each dosing frequency (weekly or every other week) will be provided.

During the maintenance phase, the number and percentage of subjects at each dosing frequency (weekly or every other week) and the number of subjects who changed their frequency (weekly to every other week and every other week to weekly) will be summarized every 4 weeks starting at Week 16. A frequency distribution of the dosing frequency used the majority of the time during the maintenance phase will be provided.

4.4. Protocol Deviations

Deviations that occurred during the study will be tabulated for the All Enrolled analysis set. Major deviations will be tabulated as they are grouped prior to unblinding in the following categories: subject not withdrawn as per protocol, selection criteria not met, excluded concomitant treatment, treatment deviation, non-compliance, regulatory requirement. More categories may be included depending on the nature of the protocol deviation.

4.5. Prior and Concomitant Medications

The number and percent of subjects receiving prior antidepressant medications will be summarized by treatment group for the Safety (IND) analysis set.

The number and percent of subjects who receive concomitant therapies will be summarized by phase and treatment group using the generic term of the medication for the Safety (IND, OP, and MA) analysis sets and Follow-up analysis set described in Section 2.4.

5. EFFICACY

5.1. Analysis Specifications

5.1.1. Level of Significance

A 2-stage group sequential design with 1 interim analysis will be implemented to allow for early stopping if there is significant evidence of efficacy based upon the interim analysis after 30 relapse events have occurred for randomized stable remitters who were treated with esketamine in the optimization phase. In either case of stopping at the interim analysis or continuing with sample size re-estimation, control of overall Type I error will be maintained at 5% level (2-sided).

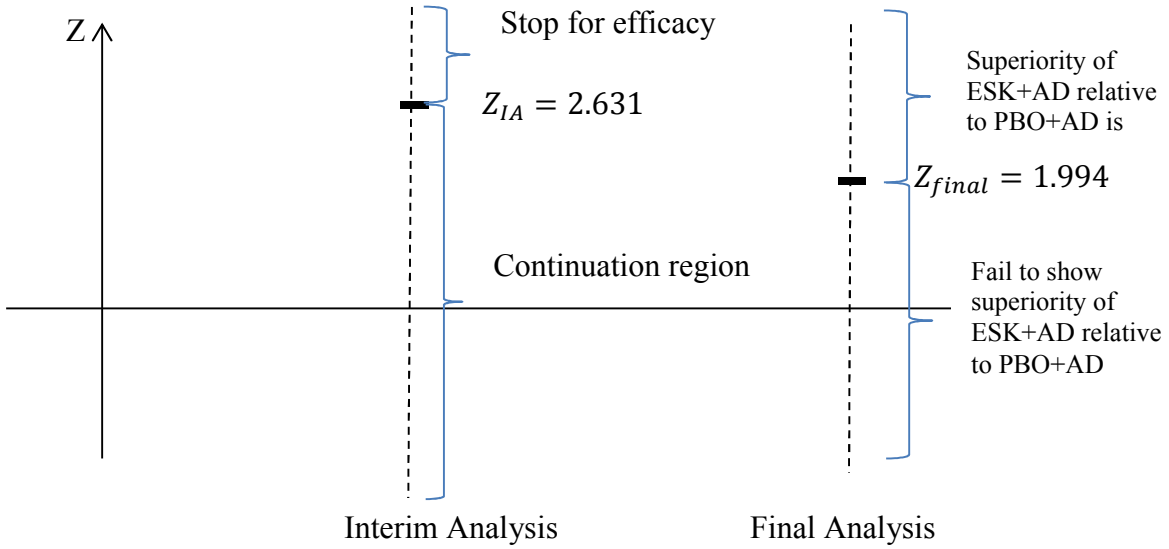
At the time of the interim analysis, time to relapse will be evaluated and compared between intranasal esketamine plus oral antidepressant and oral antidepressant plus intranasal placebo. The Wang-Tsiatis boundary[11] with shape parameter Δ=0.1 will be used for detection of early efficacy.

In accordance with the design given above, the interim efficacy analysis will be performed at a significance level of 0.0086 (two-sided). If the result of the interim efficacy analysis is significant (i.e., $Z_{IA} \geq 2.631$ where a positive Z_{IA} favors intranasal esketamine plus oral antidepressant in delaying relapse compared to oral antidepressant plus intranasal placebo) the study will be terminated and intranasal esketamine plus oral antidepressant will be declared superior to oral antidepressant plus intranasal placebo in delaying relapse. Otherwise, this study

is in the continuation region (Figure 2) and the sample size will be re-estimated using the observed value of Z_{IA}.

If the study is not stopped at the interim analysis, then the study will be stopped once the number of relapses determined by the sample size re-estimation (in the subjects with stable remission) occur during the maintenance phase. The final efficacy analysis will be performed at a significance level of 0.046 (two-sided). If the result of the final efficacy analysis is significant (i.e.,$Z_f \geq 1.994$), intranasal esketamine plus oral antidepressant will be declared superior to oral antidepressant plus intranasal placebo in delaying relapse.

Figure 2: Two-stage Adaptive Group Sequential Design for Randomized Stable Remitters Who Received Esketamine During the Optimization Phase (List 1)



5.1.2. Data Handling Rules

For the efficacy scales MADRS, CGI-S, PHQ-9, GAD-7 and SDS, both observed case and last observation carried forward (LOCF) values will be determined for the induction, optimization and maintenance phases. These imputed time points will be labeled as 'DAY X LOCF' or 'WEEK X LOCF'. Because it is possible for more than 1 visit to occur during the same interval for a protocol-specified visit, rules for choosing the visit to use for the analysis are those given in Section 2.4. Imputed time points Day 28 (IND) LOCF, Week 16 (OP) LOCF and Week XX (MA) LOCF (where 'XX" corresponds to the last window in the Maintenance Phase, excluding the end point visit) are not needed as they are essentially equivalent to End Point (IND), End Point (OP), and End Point (MA), respectively. If there are multiple visits in a time interval with non-missing values, the visit closest to the protocol-specified time is used as both observed case and LOCF. If there is no visit in a time interval with a non-missing value, then the observed case is missing and the last non-missing, post-baseline value prior to the interval is used for LOCF. For example, if a subject has a visit on Day 7 and then a final visit on Day 10. The data on Day

7 will be slotted for the 'Day 8 (IND)', because Day 7 is closer to the target Day 8 than Day 10 is. The data on Day 10 will be used for finding the End Point (IND) and all LOCF windows from 'Day 15 (IND) LOCF' through 'Day 22 (IND) LOCF'.

5.1.3. Imputation Methods for Missing Items

Imputation of the MADRS total score is described in Section 5.3.2. 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 considered missing.

5.2. Primary Efficacy Endpoint

The primary efficacy endpoint is the time from randomization to the first relapse during the maintenance phase in esketamine-treated subjects who achieved stable remission at the end of optimization phase.

Relapse is defined as any of the following:

- MADRS total score ≥22 for 2 consecutive assessments separated by 5 to 15 days. The date of the second MADRS assessment will be used for the date of relapse
- Hospitalization for worsening depression or any other clinically relevant event determined per clinical judgment to be suggestive of a relapse of depressive illness such as suicide attempt, completed suicide, or hospitalization for suicide prevention. If hospitalized for any of these events, the start date of hospitalization will be used for the date of relapse. Otherwise the date of the event will be used if the subject is not hospitalized.
- In case both relapse criteria are met, the earlier date will be defined as the date of relapse for this subject.

The date of relapse will be the date of the first assessment of relapse. The time to relapse and censoring are defined as follows:

Subject status during maintenance phase	Time to relapse/Censoring	Censoring indicator
Randomized subjects who relapse during maintenance phase	(Date of relapse – maintenance phase start date) + 1	No
Randomized subjects who remained relapse-free at the end of the maintenance phase	(End of maintenance phase date – maintenance phase start date)+1	Yes
Early withdrawal/discontinued during the maintenance phase without relapse	(Date of early withdrawal – maintenance phase start date) + 1	Yes

5.2.1. Primary Estimand

The primary estimand, the main clinical quantity of interest to be estimated in the study, is defined by the following variable and summary measure in the population, under the specified intervention effect:

Population: subjects with treatment-resistant depression who are in stable remission on esketamine at the end of the optimization phase

Variable: time to relapse during the maintenance phase, while on their initially randomized treatment

Intercurrent Event: the intercurrent event of treatment discontinuation is captured through the variable definition

Summary Measure: Kaplan–Meier estimate of the survival function.

The primary analysis will be based on the Full (stable remitters) analysis set, as described in Section 2.4.2.1 and the relapse defined above collected during the maintenance phase.

5.2.2. Analysis Methods

At the interim analysis, the cumulative distribution function of the time to relapse will be estimated by the Kaplan-Meier method and the treatment groups will be compared using a 2-sided log-rank test for the Interim Full (stable remitters) analysis set. For further details refer to IA SAP.

If the result of the interim analysis using the log-rank test is significant (i.e. two-sided significance level of 0.0086) and based on the recommendation of the IDMC, the study is stopped, then the interim analysis of time to relapse is considered as the primary analyses.

Note that the trial will be ongoing during the period of conducting the interim analysis (data cleaning, analysis and holding the interim committee meeting), which leads to overrunning data, including possible succeeding events. In this case, a final analysis using all data will be analyzed using the log-rank test and will be considered as supportive analysis to the interim primary analysis.

If the study is not terminated at the time of the interim analysis, the sample size will be re-estimated to ensure a conditional power at Stage 2 of at least 90% based on the interim analysis data, using the approach proposed by Wassmer[12]. The interim analysis SAP provides details on the sample size re-estimation. The final analysis, if a sample size re-estimation occurred, will be based on a 2-stage group sequential survival design with the decision based on the log rank test on accumulated information from both stages. The Full (stable remitters) analysis set will be used and the analysis will include all the relapse events accumulated prior to final database lock.

In this 2-stage group sequential design the decision is based on the following test on accumulated data on both stages. Under the null hypothesis, the following test statistics are approximately standard normal. This test statistic is performed on the Full (stable remitters) analysis set mentioned in Section 2.4.2.1 with the final number of events determined by the sample size re-estimation (between 59 and 84), including any additional events that occurred after the notification that the required number of events have been met and completion of the study. Final test statistic Z_f, is weighted combination of the one-sided log-rank test statistics LR_1 and LR_2. LR_2 is performed on the full analysis set and LR_1 is calculated on the interim full analysis set

$$Z_f = \sqrt{\frac{30}{59}}\,LR_1 + \sqrt{\frac{29}{59}}\left(\left(\sqrt{\frac{d_2}{d_2 - d_{IA}}}\right)LR_2 - \left(\sqrt{\frac{d_{IA}}{d_2 - d_{IA}}}\right)LR_1\right)$$

where, $\mathbf{LR_1} = \frac{\sum_{i=1}^{d_{IA}}\left(I_{2i1} - \frac{n_{2i1}}{n_{1i1}+n_{2i1}}\right)}{\sqrt{\sum_{i=1}^{d_{IA}}\frac{n_{1i1}n_{2i1}}{(n_{1i1}+n_{2i1})^2}}}$ and $\mathbf{LR_2} = \frac{\sum_{i=1}^{d_2}\left(I_{2i2} - \frac{n_{2i2}}{n_{1i2}+n_{2i2}}\right)}{\sqrt{\sum_{i=1}^{d_2}\frac{n_{1i2}n_{2i2}}{(n_{1i2}+n_{2i2})^2}}}$

Let n_{1i1} and n_{2i1} be the number of randomized stable remitters at risk in treatment group ESK+AD and PBO+AD, respectively, when the i^{th} event occurred during stage 1, and $I_{2i1} = 1$ if the event occurred in the treatment group PBO+AD and 0 otherwise, d_{IA} = number of events observed in stage 1. Let n_{1i2} and n_{2i2} be the number of randomized stable remitters at risk in treatment group ESK+AD and PBO+AD, respectively, when the i^{th} event occurred during stage 2, and $I_{2i2} = 1$ if the event occurred in the treatment group PBO+AD and 0 otherwise, d_2=accumulated events for stage 1 and stage 2.

The treatment groups will be compared using the test statistic Z_f and the corresponding two-sided p-value will be provided. Time to relapse will be summarized (number of events, number of censored subjects and quartiles of time to relapse). The cumulative distribution function of the time to relapse will be estimated by the Kaplan-Meier method. The estimate of the hazards ratio and its 95% confidence interval will be based on Wassmer[12] and will be calculated using ADDPLAN Adaptive Designs - Plans and Analyses® software[1] and software R.

5.2.3. Sensitivity Analyses

The primary analysis relies on the assumption of ignorable censoring. Therefore, sensitivity analyses will be performed to stress-test the robustness of results to deviations from ignorable censoring. Specifically, it is assumed that subjects on experimental treatment who discontinue prematurely from the maintenance phase have a higher relapse hazard starting from the discontinuation time, compared with similar subjects who remain in this phase. The higher relapse hazard is determined by the single sensitivity parameter Delta, representing the ratio of subject-specific hazard at any given time point t following discontinuation compared to that

same subject's hazard at the same time t if he or she had continued the study. A multiple imputation non-parametric approach (based on the Kaplan-Meier method) will be used for the imputation of relapse events, as described in Taylor, Murray and Hsu (2002)[13] and Lipkovich, Ratitch and O'Kelly (2016)[14]. A sequence of Delta values will be used for all subjects with non-administrative censoring from the esketamine group (i.e. subjects censored due to other reasons than the study cut-off date, which occurs when the pre-planned number of events is accumulated), starting with 1 (ignorable censoring) and increasing by 1 until the tipping point, where the results are no longer significant according to the level of significance specified in Section 5.1.1. For the control group, the sensitivity parameter Delta will be set to one, i.e. maintaining the ignorable censoring assumption. For the multiple imputation procedure under each delta adjustment, the seed will be set to 234 and 1000 multiple imputations will be used.

Notes:

- If the study is stopped at the interim analysis, the approximately normal log-rank test LR_1 will be used for the primary and sensitivity analyses.

- If the study is not stopped at the interim analysis, the approximately normal log-rank type test statistic z_f will be used for the final primary and sensitivity analyses. For sensitivity analyses, the multiple imputation procedure under each Delta adjustment will be performed separately for each stage. If the imputed times for the interim stage exceed the cut-off date for the interim analysis, these events will be censored at the interim cut-off date. Similarly, if the imputed times for the final stage exceed the final cut-off date, these events will be censored at the final cut-off date. The imputations performed at the interim stage will not be carried over to the final stage. A subject who discontinues prematurely from the maintenance phase, prior to the interim analysis cut-off, will be imputed differently in the interim and final stages, according to the information at each stage.

In addition, unweighted log-rank tests will be performed on the Full (stable remitters) analysis set with the accumulated events d_2 and 59 events. The estimate of the hazards ratio and its 95% confidence interval will be based on the Cox proportional hazards model with treatment as a factor.

5.2.4. Model Diagnostics

To assess the appropriateness of the proportional hazards assumption, a log-log survival plot of the Kaplan-Meier estimates will be generated. If the proportional hazards assumption is correct, this plot should present approximately parallel lines corresponding to the two treatment groups. Cumulative sums of Schoenfeld residuals over time may also be used to assess the proportional hazards assumption.

5.2.5. Subgroup Analyses

Subgroup analyses will be performed using the Cox proportional hazards model. The model will include treatment and one subgroup (defined in Section 2.5) and treatment-by-subgroup at a time. A forest plot with the corresponding hazard ratios and 95% CI will be presented for the subgroups. If there are insufficient numbers in the subgroups to draw meaningful conclusions, the subgroup analysis will not be performed.

5.3. Secondary Endpoints

No multiplicity adjustments will be made for secondary endpoints. All p-values are considered nominal.

5.3.1. Time to Relapse in Stable Responders (but not remitters)

The cumulative distribution function of the time to relapse will be estimated by the Kaplan-Meier method and the treatment groups will be compared using a 2-sided log-rank test for the Full (stable responders) analysis set. Time to relapse will be summarized (number of relapses, number of censored subjects, median, 25^{th} and 75^{th} percentile, if estimable). Confidence intervals of 25^{th}, 50^{th} and 75^{th} percentile of time to relapse will also be provided. Standard error estimates will be based on Greenwood's formula. The estimate of the hazards ratio and its 95% confidence interval will be based on the Cox proportional hazards model with treatment as a factor.

5.3.2. MADRS

5.3.2.1. Definition

The Montgomery-Asberg Depression Rating Scale (MADRS)[6] is a clinician-rated scale designed to measure depression severity and to detect changes due to antidepressant treatment. The scale consists of 10 items, each of which is scored from 0 (item is not present or is normal) to 6 (severe or continuous presence of the symptoms), for a total possible score of 60. Higher scores represent a more severe condition. The MADRS evaluates apparent sadness, reported sadness, inner tension, sleep, appetite, concentration, lassitude, interest level, pessimistic thoughts, and suicidal thoughts. The test exhibits high inter-rater reliability.

If 2 or more items are missing, no imputation will be performed and the total score will be left missing. Otherwise, the total score will be calculated as sum of the non-missing items multiplied by the ratio of the maximum number of items (i.e., 10) to the number of non-missing items

5.3.2.2. Analysis Methods

Descriptive statistics of the total score and change from baseline (of the respective phase) will be provided for each visit during the induction phase, optimization phase, maintenance phase for the Full analysis sets and follow-up phase for the Follow-up analysis set defined in Section 2.4. Summaries of both observed and LOCF data will be presented.

The change from baseline (MA) at each visit, including observed case and LOCF data, during the double-blind maintenance phase and at end point (MA) will be analyzed using an ANCOVA

model with factors for treatment and country and baseline (MA) score as a covariate. Least-squares estimates of the treatment differences and 95% confidence intervals will be presented.

5.3.3. PHQ-9

5.3.3.1. Definition

The PHQ-9 is a 9-item, self-report scale assessing depressive symptoms[8]. 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), with a total score range of 0-27. A higher score indicates greater severity of depression. The recall period is 2 weeks. The scale scores each of the nine symptom domains of the Diagnostic and Statistical Manual of Mental Disorders (DSM) Major Depressive Disorder criteria and it has been used both as a screening tool and a measure of response to treatment for depression. 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).

5.3.3.2. Analysis Methods

Descriptive statistics of the total score and change from baseline (of the respective phase) will be provided for each visit during the induction phase, optimization phase and maintenance phase for the Full analysis sets defined in Section 2.4. In addition, a frequency distribution by severity will be provided for each assessment visit for all phases, including the follow-up phase. Summaries of both observed and LOCF data will be presented.

The change from baseline (MA) at each visit, including observed case and LOCF data, during the double-blind maintenance phase and at end point (MA) will be analyzed using an ANCOVA model with factors for treatment and country and baseline (MA) score as a covariate. Least-squares estimates of the treatment differences and 95% confidence intervals will be presented.

5.3.4. CGI-S

5.3.4.1. Definition

The Clinical Global Impression of Severity (CGI-S)[5] 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 0 to 7. Considering total clinical experience, a subject is assessed on severity of mental illness at the time of rating according to: 0=not assessed; 1=normal (not at all ill); 2=borderline mentally ill; 3=mildly ill; 4=moderately ill; 5=markedly ill; 6=severely ill; 7=among the most extremely ill patients. The CGI-S permits a global evaluation of the subject's condition at a given time.

5.3.4.2. Analysis Methods

Descriptive statistics of the total score and change from baseline (of the respective phase) will be provided for each visit during the induction phase, optimization phase, maintenance phase for the Full analysis sets and follow-up phase for the Follow-up analysis set defined in Section 2.4. In

addition, a frequency distribution by severity will be provided for each assessment visit for all phases, including the follow-up phase. Summaries of both observed and LOCF data will be presented.

The ranks of the change from baseline (MA) for the CGI-S score at each visit, including observed case and LOCF data, during the double-blind maintenance phase and at end point (MA) will be analyzed using an ANCOVA model with factors for treatment and country and baseline (MA) score as a covariate.

5.3.5. GAD-7

5.3.5.1. Definition

The GAD-7 (Generalized Anxiety Disorder - 7 Items)[9] is a brief and validated 7-item self-report assessment of overall anxiety. Subjects respond to each item using a 4 point scale with response categories of 0=not at all, 1=several days, 2=more than half the days, and 3=nearly every day. Item responses are summed to yield a total score with a range of 0 to 21, where higher scores indicate more anxiety. The recall period is 2 weeks. The severity of the GAD-7 is categorized as follows: None (0-4), Mild (5-9), Moderate (10-14) and Severe (15 -21).

5.3.5.2. Analysis Methods

Descriptive statistics of the actual values and change from baseline (of the respective phase) will be provided for each visit during the induction phase, optimization phase and maintenance phase for the Full analysis sets defined in Section 2.4. In addition, a frequency distribution of GAD-7 severity categories will be provided for each assessment visit for all phases, including the follow-up phase.

The change from baseline (MA) for the GAD-7 total score at each visit, during the double-blind maintenance phase and at end point (MA) will be analyzed using an ANCOVA model with factors for treatment and country and baseline (MA) score as a covariate. Least-squares estimates of the treatment differences and 95% confidence intervals will be presented.

5.3.6. Sheehan Disability Scale (SDS)

5.3.6.1. Definition

The SDS is a subject-reported outcome measure and is a 5-item questionnaire which has been widely used and accepted for assessment of functional impairment and associated disability. The first three items assess disruption of (1) work/school, (2) social life, and (3) family life/home responsibilities using a 0-10 rating scale. The score for the first three items are summed to create a total score of 0-30 where a higher score indicates greater impairment. It also has one item on days lost from school or work and one item on days when underproductive. The recall period for this study is 7 days. Scores $\leq$4 for each item and $\leq$12 for the total score are considered response. Scores $\leq$2 for each item and $\leq$6 for the total score are considered remission. If any of the first three items are missing, the total score will be set to missing as well as response and remission status.

5.3.6.2. Analysis Methods

Descriptive statistics of the actual values and change from baseline (of the respective phase) will be provided for each visit during the induction phase, optimization phase, maintenance phase for the Full analysis sets and follow-up phase for the Follow-up analysis set defined in Section 2.4. The total score as well as the individual item scores will be summarized separately. Summaries of both observed and LOCF data will be presented.

The change from baseline (MA) in SDS total score at each visit, including observed case and LOCF data, during the double-blind maintenance phase and at end point (MA) will be analyzed using an ANCOVA model with factors for treatment and country and baseline (MA) score as a covariate. Least-squares estimates of the treatment differences and 95% confidence intervals will be presented.

5.3.7. EuroQol Group; 5 Dimension; 5 Level (EQ-5D-5L)

5.3.7.1. Definition

The EQ-5D-5L (EuroQol Group - 5 Dimension - 5 Level)[3,4] is a standardized 2-part instrument for use as a measure of health outcome, primarily designed for self-completion by respondents. It essentially 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).

The time taken to complete the questionnaire varies with age, health status, and setting but is likely to be around 1 minute.

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 or health state: 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 for various values of 5L profile scores is published for Canada in the following website: https://www.ncbi.nlm.nih.gov/pubmed/26492214

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).

5.3.7.2. Analysis Methods

Descriptive statistics of actual values and changes from baseline for the weighted EQ-5D health status index, the EQ-VAS, and the sum score by treatment group will be summarized for each visit for all phases including the follow-up phase. In addition, individual dimension responses using a frequency distribution by treatment group will also be summarized for each visit for the all phases, including the follow-up phase.

6. SAFETY

Safety summaries for each phase will be based on the Safety analysis sets described in Section 2.4.4.

6.1. Adverse Events

The Medical Dictionary for Regulatory Activities (MedDRA) (version 18.1 or above) will be used to classify AEs by system organ class and preferred term. Treatment-emergent adverse events (TEAEs) that occurred in each study phase will be summarized by system organ class and preferred term.

The number (%) of subjects with TEAEs, serious TEAEs (SAEs), and TEAEs that led to study drug discontinuation will be summarized by system organ class and preferred term. Data listings will be generated for deaths, other SAEs, and discontinuations due to AEs.

A TEAE is an event that is new in onset or increased in severity following treatment initiation. An event that starts prior to, and ends after the initiation of study medication will be considered treatment-emergent only if the severity increases after the start of medication.

TEAEs are defined as follows for each study phase:

- TEAEs in the induction phase:
  a. If AE onset time is not missing:
     i. If subjects continue to OP phase: IND phase start date/time <= AE onset date and time < IND phase end date

     ii. If subjects discontinue in the IND phase: IND phase start date/time <= AE onset date and time <= IND phase end date

  b. If AE onset time is missing:

     i. If subjects continue to OP phase: IND phase start date <= AE onset date < IND phase end date

     ii. If subjects discontinue in the IND phase: IND phase start date <= AE onset date <= IND phase end date

- TEAEs in the optimization phase:
  a. If AE onset time is not missing:

     i. If subjects continue to MA phase: OP phase start date <= AE onset date and time < OP phase end date

     ii. If subjects discontinue in the OP phase: OP phase start date <= AE onset date and time <= OP phase end date

  b. If AE onset time is missing:

     i. If subjects continue to MA phase: OP phase start date <= AE onset date < OP phase end date

     ii. If subjects discontinue in the OP phase: OP phase start date <= AE onset date <= OP phase end date

- TEAEs in the maintenance phase:
  a. If AE onset time is not missing:

     MA start date/time <= AE onset date and time <= MA end date

  b. If AE onset time is missing: MA start date <= AE onset date <= MA end date
- AEs in follow-up phase: F/U start date <= AE onset date <= F/U end date
- For the AEs that have both day and month missing, treatment emergent flag is assigned based on the rules presented in Section 2.6

In addition, TEAEs will be summarized by severity and relationship to study medication using the preferred term. For the summaries of AEs by severity/relationship to study medication, the observation with the most severe occurrence/closest relationship to study medication will be chosen if there is more than one incident of an adverse event reported during the analysis phase by the subject. AE duration for transient dizziness/vertigo and anxiety will also be summarized.

Adverse Events of Special Interest

Clinically relevant TEAEs of special interest will be examined separately grouped in the following categories:

- drug abuse, dependence and withdrawal (Aggression, Confusional state, Decreased activity, Dependence, Disorientation, Dissociation, Dissociative disorder, Dizziness, Drug abuse, Drug abuser, Drug dependence, Drug use disorder, Drug detoxification, Drug diversion, Drug rehabilitation, Drug tolerance, Drug tolerance increased, Drug withdrawal convulsions, Drug withdrawal headache, Drug withdrawal syndrome, Euphoric mood, Feeling abnormal, Feeling drunk, Feeling of relaxation, Hallucination, Hallucination, auditory, Hallucination, gustatory, Hallucination, olfactory, Hallucination, synaesthetic, Hallucination, tactile, Hallucination, visual, Hallucinations, mixed, Inappropriate affect, Mental impairment, Product tampering, Psychomotor hyperactivity, Psychotic disorder, Rebound effect, Somatic hallucination, Somnolence, Substance abuser, Substance dependence, Substance use, Substance use disorder, Substance-induced mood disorder, Substance-induced psychotic disorder, Thinking abnormal, Withdrawal arrhythmia, Withdrawal syndrome);

- increased blood pressure (Blood pressure increased, Blood pressure diastolic increased, Blood pressure systolic increased, Hypertensive crisis, Hypertensive emergency, Hypertension)
- increased heart rate (Heart rate increased, Tachycardia)
- transient dizziness/vertigo (Dizziness, Dizziness exertional, Dizziness postural, Dizziness procedural, Vertigo, Vertigo labyrinthine, Vertigo positional, Vertigo CNS origin);
- impaired cognition (Cognitive disorder);
- cystitis (Allergic cystitis, Chemical cystitis, Cystitis, Cystitis erosive, Cystitis haemorrhagic, Cystitis interstitial, Cystitis noninfective, Cystitis ulcerative, Cystitis-like symptom);
- anxiety (Anticipatory anxiety, Anxiety, Anxiety disorder).

The number and percentage of subjects taking concomitant medication for dissociation events (preferred term of Dissociation) at any time during each treatment phase will be provided.

Summary statistics for the duration of all episodes of TEAEs associated with discharge readiness (Dissociation, Dizziness, Feeling abnormal, Feeling drunk, Nausea, Somnolence, Vertigo, and Vomiting) with an onset on the day of intranasal study drug administration is summarized for each treatment phase. These summaries are presented for each dosing session during each treatment phases. In addition, the number of occurrences of TEAEs associated with discharge readiness and the proportion of subjects with TEAEs associated with discharge readiness will be presented by dosing session for each treatment phase.

6.2. Clinical Laboratory

Descriptive statistics (N, mean, median, minimum, maximum and range) for values and changes from baseline will be provided for clinical laboratory tests (hematology, chemistry and urinalysis) at each scheduled time point for all the phases, including the follow-up phase. Baseline (IND) will be used for the change summaries and to determine abnormal values during the treatment phases (IND, OP, MA) and during the follow-up phase for those subjects who discontinue during the induction and optimization phases. Baseline (MA) values will be used for summaries during the maintenance phase and follow-up phase for those subjects who discontinue during the maintenance phase. Baseline values are defined in Section 2.2.

Clinical laboratory tests that meet the criteria for markedly abnormal will be listed by subject for each treatment phase. The incidence of treatment-emergent markedly abnormal laboratory values that occurred at any time during each treatment phase will be presented. Clinical laboratory test values will be considered "TEMA using the criteria defined by the Sponsor (Janssen Research & Development, LLC)" listed in Attachment 2. The identification of TEMA laboratory values is based on the postbaseline value being out of range while the baseline value (defined above) is either missing or within the range given in Attachment 2. If post-baseline laboratory results are above the upper limit and the baseline value is below the lower limit, then the post-baseline 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.

The incidence of subjects with ALT values > 3*upper normal limit (ULN) will be presented for each study phase. Additionally, incidence of hepatic toxicity (Hy's Law)[10] defined as ALT values > 3*ULN and total bilirubin values >2*ULN will be presented for each study phase. Similar to the markedly abnormal analysis, only subjects with baseline (IND) ALT values < =3*ULN (and baseline (IND) total bilirubin values < =2*ULN for hepatic toxicity or if the baseline (IND) value is missing) will be eligible for these analyses.

6.3. Vital Signs, Weight, and BMI

Descriptive statistics for values and changes from baseline at each scheduled time-point during each treatment phase (IND, OP and MA) only will be presented for temperature, systolic blood pressure, diastolic blood pressure, pulse rate, respiratory rate, oxygen saturation, weight, and BMI. Baseline (IND) will be used for change summaries and to determine abnormal values during all treatment phases (IND, OP, MA) and during the follow-up phase for those subjects who discontinued study agent during the induction and optimization phases. Baseline (MA) will be used for summaries during the maintenance phase and during the follow-up phase for those subjects who discontinued during the maintenance phase. Baseline values are defined in Section 2.2.

The proportion of subjects who have a treatment-emergent abnormality, as defined in Table 4 below, during each treatment phase will be presented. A listing of subjects meeting any of the following criteria during each treatment phase will also be provided.

Table 4: Treatment-Emergent Abnormality Categories for Vital Signs

	Post-baseline value outside of normal limit if:	
Vital Parameter	**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

BP = blood pressure

The proportion of subjects who experienced treatment-emergent acute hypertension (systolic BP≥180 or diastolic BP≥110) at any time during the study will be summarized by study phase, treatment group and hypertension status.

A listing of subjects with oxygen saturation less than 93% will be provided.

6.4. Electrocardiogram

ECG variables that will be analyzed include heart rate, RR, PR interval, QRS interval, QT interval and QTc intervals. The corrected QT (QTc) intervals will include QTcB (Bazett) and QTcF (Fridericia).

Baseline ECG is defined as the average of all ECG results collected up to and including the day of first dose of study medication (either intranasal or oral AD). Baseline (IND) will be used as the baseline for change summaries and to determine abnormal values during all treatment phases (IND, OP, MA) and during the follow-up phase for those subjects who discontinued during the induction and optimization phases. Baseline (MA) will be used for change summaries and to determine abnormal values during the maintenance phase and during the follow-up phase for those subjects who discontinued during the maintenance phase. The baseline values are defined in Section 2.2.

The maximum post-baseline value during each treatment phase will be computed for each ECG parameter using data from both scheduled and unscheduled visits.

Summary tables for observed values and changes from baseline will be presented by treatment at each scheduled time point during each treatment phase and the follow-up phase.

The frequency of treatment-emergent abnormalities will be tabulated and presented for all treatment phases. The identification of treatment-emergent abnormal ECG values is based on the post-baseline value (a value occurring after the start of the phase) being out of range while the baseline value is either missing or within the limits given in Table 5. If post-baseline ECG results are above the upper limits (abnormally high) and the baseline 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 baseline value being above the upper limits (abnormally high). Abnormal ranges for the HR, PR, QRS and QT intervals are given in Table 5.

Table 5: Limits for HR, PR, QRS and QT Interval Abnormality

ECG parameter	Abnormally Low	Abnormally High
HR (bpm)	≤ 50	≥ 100
PR interval (msec)	--	≥ 210
QRS interval (msec)	≤ 50	≥ 120
QTc interval (msec)	≤ 200	≥ 500

Based on the maximum QTc value for each subject during a given phase (separate for each QTc correction, QTcB and QTcF) the incidence of abnormal QTc values and changes from baseline will be summarized by treatment group. Criteria for abnormal corrected QT intervals and changes from baseline are given in Table 6 and are derived from the ICH E14 Guidance (the same criteria apply to all QT corrections).

Table 6: Criteria for Abnormal QTc Values and Changes From Baseline

Parameter	Classification	Criteria
Clinically Significant QTc Value	No	≤500
	Yes	>500
QTc change from baseline[a]	No concern	≤30
	Concern	>30 – 60
	Clear concern	> 60
QTc value	Normal	≤450
	> 450 – 480	>450 - ≤480
	> 480 – 500	>480 – ≤500
	> 500	> 500

These criteria are based on ICH E14 Guideline
[a] Baseline is defined as the average pre-dose.

The proportion of subjects with treatment-emergent abnormalities will be presented for all treatment phases. A listing of subjects with abnormalities will also be provided.

6.5. Nasal Examination

Targeted nasal examinations (including the upper respiratory tract/throat) will be conducted by a qualified healthcare practitioner. The objective of the examination at Screening is to rule out any subjects with anatomical or medical conditions that may impede drug delivery or absorption.

Subsequent examinations will consist of a visual inspection of the nostrils, nasal mucosa, and throat for nasal erythema, rhinorrhea, rhinitis, capillary/blood vessel disruption and epistaxis and graded as follows: absent, mild, moderate, or severe.

Changes in findings from baseline (of the respective phase) for each examination (including the upper respiratory tract/throat) will be listed by treatment group for all phases.

6.5.1. Nasal Symptom Questionnaire

Subjects will complete a nasal symptom questionnaire on every dosing day at predose and again at 1 hour postdose. The questionnaire was developed to assess nasal tolerability following intranasal administration of study drug. Subjects will rate nasal symptoms as none, mild, moderate, or severe for the following items: stuffy nose, blocked nose, runny nose, itching nose, crusting discharge in or on nose, dryness of nose, burning sensation in the nose, discomfort of nose, bleeding from the nose, postnasal drip, cough, sore throat, taste disturbance and sneezing.

Frequency distributions will be provided by treatment group for each of the items at each timepoint during the each of the treatment phases. Shift from predose to postdose questionnaires during each time point throughout the study will be provided by treatment group to see if there is any change after repeated administrations of study drug. Frequency of subjects who report moderate or severe symptoms at any postdose timepoint during each treatment phase will be presented by treatment group. In addition, a listing of severe symptoms will also be presented.

6.6. Other Safety Parameters

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

The Columbia Suicide Severity Rating Scale (C-SSRS)[7] is a low-burden measure of the spectrum of suicidal ideation and behavior that was developed in the National Institute of Mental Health Treatment of Adolescent Suicide Attempters Study to assess severity and track suicidal events through any treatment. 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.

The summaries of the C-SSRS outcomes will be based on the Safety analysis set for a given treatment phase for subjects who have at least 1 post-baseline C-SSRS measurement and a pre-treatment C-SSRS assessment (assessment at Baseline (IND) visit).

A frequency distribution at each scheduled time point by treatment will be provided. Shifts from the Baseline (IND) value to the most severe/maximum score during each phase will be summarized by treatment.

The maximum score assigned for each subject will also be summarized into one of three broad categories: No suicidal ideation or behavior (0), Suicidal ideation (1-5), Suicidal behavior (6-10).

Shifts from the Baseline (IND) value to the maximum category during each phase, including the follow-up phase, will be summarized by treatment.

6.6.2. Clinician Administered Dissociative States Scale (CADSS)

The Clinician Administered Dissociative States Scale (CADSS)[2] is an instrument for the measurement of present-state dissociative symptoms, and is administered to assess treatment-emergent dissociative symptoms. The CADSS comprises 23 subjective items and participant's responses are coded on a 5-point scale (0 = "Not at all", 1 = "Mild", 2 = "Moderate", 3 = 'Severe" and 4 = "Extreme"). The CADSS is divided into 3 components using the following scoring method:

Component	Questions	Range
Depersonalization	Sum of 3, 4, 5, 6, 7, 20, 23	0-28
Derealization	Sum of 1, 2, 8, 9, 10, 11, 12, 13, 16, 17, 18, 19, 21	0-52
Amnesia	Sum of 14, 15, 22	0-12
Total Score	Sum of 1 through 23	0-92

For the total score and each component, a higher score represents a more severe condition. If any response is missing the total score is set to missing. The CADSS is measured prior to each dose, at 40 minutes, and at 1.5 hours postdose.

Descriptive statistics (N, median, minimum, and maximum) of the total scores and component scores at each time point and visit, changes from predose and proportion of subjects with an increase in CADSS total score from the predose value at any time during the phase will be summarized by treatment phase. Mean (SD) CADSS values will be presented graphically for each intranasal dosing day.

6.6.3. Brief Psychiatric Rating Scale (BPRS)

The Brief Psychiatric Rating Scale (BPRS) is an 18 item rating scale which is used to assess potential treatment-emergent psychotic symptoms. The BPRS assesses a range of psychotic and affective symptoms rated from both observation of the subject and the subject's own report. Only the four-item positive symptom subscale ($BPRS^{+}$) will be used in the study to assess treatment-emergent psychotic symptoms. The $BPRS^{+}$ consists of: suspiciousness, hallucinations, unusual thought content and conceptual disorganization. Each symptom is rated on a scale of 0 to 6 as follows: 0: not present, not evident or absent; 1: very mild; 2: mild; 3: moderate; 4: moderate severe; 5: severe; or 6: extreme. A total score will be derived by summing the individual items, with a range of 0 to 24 with a higher score representing a more severe condition.

The BPRS^{+} is measured prior to each dose, at 40 minutes, and at 1.5 hours post dose during each of the treatment phases.

Descriptive statistics (N, median, minimum, and maximum) of the total scores at each time point and the change from the pre-dose time point within each visit, and the proportion of subjects with an increase in BPRS^{+} from the predose value at any time will be provided for each treatment phase. Mean (SD) BPRS^{+} values will be presented graphically for each intranasal dosing day.

6.6.4. Modified Observer's Assessment of Alertness/Sedation (MOAA/S)

The Modified Observer's Assessment of Alertness/Sedation (MOAA/S) will be used to measure treatment-emergent sedation with correlation to levels of sedation defined by the American Society of Anesthesiologists (ASA) continuum. The MOAA/S scores range from 0 [No response to painful stimulus; corresponds to ASA continuum for general anesthesia] to 5 [Readily responds to name spoken in normal tone (awake); corresponds to ASA continuum for minimal sedation].

The MOAA/S is measured on each dosing day every 15 minutes from predose to 1.5 hours postdose or longer, if necessary, until the subject has a score of 5.

- If the score is ≤3 at any time during the 1.5 hour postdose interval, the MOAA/S will be performed every 5 minutes until a score of 4 is reached (at which point a frequency of every 15 minutes can be resumed until t=+1.5 hours post dose).
- If a subject does not have a score of 5 at t=+1.5 hours postdose, they should continue to be monitored. For subjects with a score of 4, the assessment should be repeated every 15 minutes. And for subjects with a score of ≤3, the assessment should be repeated every 5 minutes until the score returns to 5 or the subject is referred for appropriate medical care, if clinically indicated.

Descriptive statistics of the MOAA/S score, changes from predose, and the proportion of subjects experiencing sedation (score less than or equal to 3) will be summarized at each scheduled time point.

Mean MOAA/S scores will be presented graphically for each intranasal dosing day.

6.6.5. Clinical Global Assessment of Discharge Readiness (CGADR)

The Clinical Global Assessment of Discharge Readiness (CGADR) will be used to measure a subject's current clinical status and is the clinician's assessment of the readiness to be discharged from the study site.

The clinician will answer "Yes" or "No" to the question "Is the subject considered ready to be discharged based on their overall clinical status (e.g., sedation, blood pressure, and other adverse events)?"

On each intranasal dosing day, the CGADR will be performed at 1 hour and 1.5 hours postdose, repeated every 15 minutes if necessary until the response is 'Yes'. A subject should not be discharged prior to the 1.5-hour time point.

The proportion of subjects with a response of 'No' at each time point will be presented by treatment during each of the treatment phases.

6.6.6. Physician Withdrawal Checklist (PWC-20)

The PWC-20 is a 20-item simple and accurate method to assess potential withdrawal symptoms following cessation of intranasal esketamine treatment. The PWC-20 will be performed for all subjects on Day 25 to establish a baseline prior to discontinuation of esketamine treatment – although only relevant for those subjects not continuing to the optimization phase. For those subjects who proceed to the optimization and maintenance phases, the PWC-20 is conducted at the End of Study Visit. If subjects withdraw early from the study during any phase, the PWC-20 will be conducted at the Early Withdrawal Visit.

The proportion of subjects with withdrawal symptoms at the end of each treatment phase (IND, OP and MA) and during the follow-up phase will be presented by treatment. In addition, symptoms at follow-up will be compared to the last assessment in the relevant treatment phase (IND, OP and MA) for subjects who discontinue during that treatment phase and will be summarized using the following categories: new or worsened symptoms, symptoms present and unchanged, no symptoms, and improved.

6.6.7. Computerized Cognitive Battery and Hopkins Verbal Learning Test-Revised (HVLT-R)

The effect of intranasal esketamine on cognition will be assessed using the computerized cognitive battery and the HVLT-R.

The computerized cognitive battery provides assessment of multiple cognitive domains, including attention, visual learning and memory, and executive function. The tests use culture-neutral stimuli, enabling use in multilingual/multicultural settings. The HVLT-R is a measure of verbal learning and memory and is a 12-item word list recall test. The total number of correct responses are captured for 4 trials as well as the number of true-positive responses and false-positive errors in Trial 4 (Delayed Recall). The Total Recall will be derived as the sum of trials 1, 2 and 3. Retention % will be derived as (the number of correct responses in Trial 4)/(higher score of Trials 2 and 3) X 100. Recognition Discrimination Index will be derived as the total number of true-positives – the total number of false-positives.

The computerized cognitive battery and HVLT-R will be assessed at Day 1 predose, Day 28/ET of the open-label induction phase, Week 16 of the optimization phase, and every 12 weeks during the maintenance phase. It will also be assessed again at 2 weeks post-treatment during the follow-up phase.

See Attachment 3 for details of this analysis.

6.6.8. University of Pennsylvania Smell Inventory Test (UPSIT)

The UPSIT is a 40-item standardized test to assess a subject's ability to identify odors. The UPSIT will be administered bilaterally (i.e., both nostrils at the same time); testing will occur during screening to establish a subject's baseline sensitivity.

The UPSIT total score is defined as the total number of correct responses; therefore, the total score ranges from 0-40, where higher scores indicate greater smell function. Descriptive statistics of the observed values for UPSIT total score (%) and percent changes from baseline will be summarized at each scheduled time point during each treatment phase. Baseline (IND) will be used for computing change. UPSIT total score (%) will be defined as [(# correct responses)/(# completed responses)]*100.

6.6.9. Smell Threshold Test

The Smell Threshold Test will assess the smell threshold using a forced-choice single staircase threshold procedure. This test will be administered bilaterally (i.e., both nostrils at the same time); testing will occur during screening to establish a subject's baseline sensitivity and also throughout each treatment phase.

The smell threshold will be derived as the average of the non-missing results for the fourth through seventh reversals, as recorded on the CRF. If there are any missing results, the average will be set to missing. A small deficit is defined as thresholds above -2.40. Descriptive statistics for observed values and changes from baseline (IND) will be summarized at each scheduled time point during each treatment phase.

7. HEALTH ECONOMICS

7.1. Healthcare Resource Use Questionnaire (HRUQ)

Medical resource utilization data, associated with medical encounters, will be collected using the Healthcare Resource Use Questionnaire (HRUQ) during the optimization, maintenance, and follow-up phases bi-weekly.

The number and percentage of subjects who visited at least one healthcare professional or had a hospital emergency room visit because of their depression in the 2 weeks preceding each time point will be summarized for each type of healthcare professional. The total number of visits to each type of healthcare professional, total number of visits to any healthcare professional, and the total number of hospital emergency room visits related to depression in the 2 weeks preceding each time point will be summarized at each time point with descriptive statistics. All summaries will be presented by treatment group.

REFERENCES

1. ADDPLAN Adaptive Designs - Plans and Analyses®, ICON, Ireland, 2016; software available at http://www.iconplc.com/innovation/addplan.
2. Bremner JD, Krystal JH, Putnam FW, Southwick SM, Marmar C, Charney DS, Mazure CM. Measurement of dissociative states with the Clinician-Administered Dissociative States Scale (CADSS). J Trauma Stress. 1998;11(1):125-136.
3. EuroQol Group. About EQ-5D. http://www.euroqol.org/about-eq-5d.html (accessed 18 June 2014).
4. EuroQol Group. EQ-5D-5L User Guide: basic information on how to use the EQ-5D-5L instrument. Version 2.0; October 2013. http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/UserGuide_EQ-5D-5L_v2.0_October_2013.pdf (accessed 6 August 2014)
5. Guy W. ECDEU Assessment Manual for Psychopharmacology. Rockville, MD: U.S. Department of Health, Education, and Welfare; 1976.
6. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry. 1979;134:382-389.
7. Posner K, Oquendo MA, Gould M, Stanley B, Davies M. Columbia Classification Algorithm of Suicide Assessment (C-CASA): classification of suicidal events in the FDA's pediatric suicidal risk analysis of antidepressants. Am J Psychiatry. 2007;164:1035-1043.
8. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. JAMA. 1999; 282:1737-1744.
9. Spitzer RL, Kroenke K, Williams JB, Lowe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med. 2006;166(10):1092-1097.
10. 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.
11. Wang, S and Tsiatis, A. Approximately optimal one-parameter boundaries for group sequential trials. Biometrics 1987; 43:193-199.
12. Wassmer G. Planning and analyzing adaptive group sequential survival trials. Biometrical Journal 48(2006):4, 714-729.
13. Taylor, J. M., Murray, S., & Hsu, C. H. (2002). Survival estimation and testing via multiple imputation. Statistics & probability letters, 58(3), 221-232.
14. Lipkovich, I., Ratitch, B., & O'Kelly, M. (2016). Sensitivity to censored-at-random assumption in the analysis of time-to-event endpoints. Pharmaceutical statistics, 15(3), 216-229.

LIST OF ATTACHMENTS

Attachment 2: Criteria of Markedly Abnormal Laboratory Values

	Markedly Abnormal Limits	
Laboratory Parameter	Low	High
Albumin [g/L]	24	60
Alkaline phosphatase [U/L]	N/A	250
Alanine transaminase (SGPT) [U/L]	N/A	200
Alanine transaminase (SGPT) [U/L]	N/A	>3X ULN
Aspartate transaminase (SGOT) [U/L]	N/A	250
Bicarbonate [mmol/L]	15.1	34.9
Blood urea nitrogen [mmol/L]	N/A	17.9
Calcium [mmol/L]	1.5	3
Chloride [mmol/L]	94	112
Creatine kinase (U/L)	N/A	990
Creatinine [µmol/L]	N/A	265.2
Gamma glutamyl transferase [U/L]	N/A	300
Glucose [mmol/L]	2.2	16.7
Phosphate [mmol/L]	0.7	2.6
Potassium [mmol/L]	3.0	5.8
Sodium [mmol/L]	125	155
Bilirubin, total [µmol/L]	N/A	51.3
Protein, total [g/L]	50	N/A
Urine pH	N/A	8.0
Hematocrit [fraction] - female	0.28	0.5
- male	0.24	0.55
Hemoglobin [g/L]	80	190
Neutrophils, segmented [%]	30	90
Monocytes [%]	N/A	20
Eosinophils [%]	N/A	10
Basophils [%]	N/A	6
Lymphocytes [%]	10	60
Platelet count [$x10^9$/L]	100	600
Erythrocytes (RBC) [$x10^{12}$/L] -- female	3.0	5.5
-- male	3.0	6.4
Leukocytes(WBC) [$x10^9$/L]	2.5	15.0
Hy's Law criteria:		
Alanine transaminase (SGPT) [U/L]		> 3X ULN
AND		
Bilirubin, total [µmol/L]		>2X ULN

Note: The same limits apply to both males and females unless gender is indicated; N/A = Not applicable.

Attachment 3: Statistical Analysis Plan for COGSTATE

STATISTICAL ANALYSIS PLAN

RANDOMIZED, DOUBLE-BLIND, MULTICENTER, ACTIVE-CONTROLLED STUDIES TO EVALUATE THE EFFICACY, SAFETY, AND TOLERABILITY OF FIXED DOSES OF INTRANASAL ESKETAMINE PLUS AN ORAL ANTIDEPRESSANT IN ADULT SUBJECTS WITH TREATMENT-RESISTANT DEPRESSION

AND

AN OPEN-LABEL, LONG-TERM, SAFETY AND EFFICACY STUDY OF INTRANASAL ESKETAMINE IN TREATMENT-RESISTANT DEPRESSION

PROTOCOL ESKETINTRD3001/2/3/4/5; PHASE 3

Prepared for: Janssen Research & Development, LLC

Prepared by: Cogstate Biostatistics Group
Level 8, 195 Church Street
New Haven, CT, USA, 06510

Version: V3
Date: 27-Jul-2016
Protocol Version: Approved Date: 17 Feb 2016

1 NOTE

Cogstate has prepared a Statistical Analysis Plan (SAP) for the Sponsor to review and sign-off for all ESKETINTRD studies. Analyses will be provided after this document has been finalized and officially signed. In this SAP, anything in italics is taken directly from the protocol. For more details, please refer to the study protocols and SAPs.

Notice that this SAP will be used for all the five ESKETINTRD studies (3001, 3002, 3003, 3004 and 3005).

COGSTATE

2 SIGNATURE PAGE FOR SAP APPROVAL

The following signatures indicate the approval of the statistical analysis plan for ESKETINTRD studies.

Approved by:

Name (print): Phillip Phiri

Position: [redacted]

Signature:

Date (ddmmmyyyy): 30 Aug 2016

Approved by:

Name (print): Jaskaran Singh

Position:

Signature: JASKARAN SINGH

Digitally signed by JASKARAN SINGH
DN: c=US, o=JNJ, ou=Subscribers, 0.9.2342.19200300.100.1.1=364610, cn=JASKARAN SINGH
Reason: I am approving this document.
Date: 2017.06.14 14:22:10 -07'00'

Date (ddmmmyyyy):

Name (print): Pilar Lim

Position:

Signature: PILAR LIM

Digitally signed by PILAR LIM
DN: c=US, o=JNJ, ou=Subscribers, cn=PILAR LIM, 0.9.2342.19200300.100.1.1=74243
Reason: I am approving this document.
Date: 2017.06.14 19:01:45 -04'00'

Date (ddmmmyyyy):

3 TABLE OF CONTENTS

4 ABBREVIATIONS

Abbreviation	Description
DET	Detection test
HVLT-R	Hopkins Verbal Learning Test-Revised
IDN	Identification test
ISLT	International Shopping List test
OCL	One Card Learning test
ONB	One Back Memory test
SAP	Statistical Analysis Plan
SAS	Statistical Analysis System
SD	Standard Deviation
SPP	Statistical Programming Plan

5 INTRODUCTION

This Statistical Analysis Plan (SAP) contains technical and detailed elaboration of the principal features of the analysis described in the protocol related to computerized cognitive battery and HVLT-R and includes detailed procedures for executing the statistical analysis of the data.

The SAP will be finalized and signed prior to database lock. If needed, revisions to the approved SAP may be made prior to database lock. Revisions will be version controlled.

This statistical analysis is coordinated by the responsible Cogstate statistician. Any changes from the analyses planned in the SAP will be justified in the Cogstate statistical report.

6 VISIT SCHEDULE

Scheduled Visits

Table 1: Scheduled visits related to Cogstate battery and HVLT-R

Period	Visit Number	Study Day	Study Week by phase	Computerized test battery & HVLT-R Assessment Included
Screening/prospective Observation Phase	**1.2**	-	**2**	**Computerized test battery Practice Session**
Double-blind Induction Phase	**2.1**	**1**	**1**	**Yes**
	2.10 or 2.9	**28**	**4**	**Yes**
	EW [b]	EW	-	**Yes**
Follow-up Phase	**3.2**	-	**2 after last intranasal dose**	**Yes**
Follow-up Phase	**3.6 to 3.x**	**Every 84 days**	**Every 12 Weeks**	**Yes for ESKETINTRD3004**

[b] If a subject withdraws before the end of the double-blind induction phase (i.e., before completing Visit 2.10/Day 28) for reasons other than withdrawal of consent, an early withdrawal visit should be conducted within 1 week of the date of discontinuation, followed by the follow-up phase. If the early withdrawal visit is conducted on the same day as a scheduled visit, duplicate assessments are not required.

7 STUDY OBJECTIVES RELEVANT TO COGSTATE ANALYSIS

The objective is to assess the effect of intranasal esketamine on cognition.

8 STUDY DESIGN

The study designs for each study are described in their respective protocols.

9 SAMPLE SIZES

The sample sizes for each study are as follows:

- *116 subjects per treatment for ESKETINTRD3001.*
- *98 subjects per treatment for ESKETINTRD3002.*
- *A total of 211subjects for ESKETINTRD3003.*
- *There is no formal sample size calculation for ESKETINTRD3004 study (Note: the total number of subjects will be based on subjects from this study and subjects from other intranasal esketamine Phase 3 studies).*
- *74 subjects per treatment for ESKETINTRD3005.*

10 ANALYSIS SETS

***Safety Analysis Set**: All randomized subjects who receive at least 1 dose of intranasal study medication or 1 dose of oral antidepressant medication in the double-blind induction phase. This analysis set will be used for ESKETINTRD3001/2/5.*

For ESKETINTRD3003, the safety analysis set for each phase is defined as all subjects who receive at least 1 dose of intranasal study drug or 1 dose of oral antidepressant during that phase.

***Full Analysis Set for open-label induction phase:** will be defined as all subjects who receive at least one dose of intranasal esketamine or 1 dose of oral antidepressant during this phase.*
***Full Analysis Set for optimization/maintenance phase:** will be defined as all subjects who receive at least one dose of intranasal esketamine or 1 dose of oral antidepressant during this phase.*

These analysis sets will be used for ESKETINTRD3004

11 ANALYSIS VARIABLES

11.1 Computerized cognitive battery and Hopkins Verbal Learning Test-Revised (HVLT-R) Endpoints

The endpoints for computerized cognitive battery (DET, IDN, ONB, OCL, and GML) and HVLT-R are the scores and change from baseline (Day 1 prior to randomization) scores at each scheduled post baseline time point.

11.2 The computerized tests from the Cogstate Battery.

11.2.1 Summary of the Cogstate Battery

Detection (DET; Psychomotor Function)

The Detection test is a measure of psychomotor function and uses a well-validated simple reaction time paradigm with playing card stimuli. In this test, the playing cards all depict the same joker. The subject is asked to press the **Yes** key as soon as the card in the center of the screen turns face up. The software

measures the speed and accuracy of each response. The test terminates when a subject has correctly responded to 35 trials. The time to respond (to a maximum) is recorded for each trial.

Duration of test 3 minutes

Identification (IDN; Attention)

The Identification test is a measure of visual attention and uses a well-validated choice reaction time paradigm with playing card stimuli. In this test, the playing cards are all either red or black jokers. The subject is asked whether the card displayed in the center of the screen is red. The subject responds by pressing the **Yes** key when the joker card is red and **No** when it is black. The software measures the speed and accuracy of each response. The time to respond (to a maximum) is recorded for each trial. Wrong responses are counted but do not have an effect on the number of correct responses required for the test to terminate.

Duration of test: 2 minutes

One Card Learning (OCL; Visual Learning)

The One Card Learning test is a measure of visual learning and uses a well-validated pattern separation paradigm with playing card stimuli. In this test, the playing cards are identical to those found in a standard deck of 52 playing cards (without the joker cards). The subject is asked whether the card displayed in the center of the screen was seen previously in this test. The subject responds by pressing the **Yes** or **No** key. The software measures the speed and accuracy of each response. Because no card has been presented yet, the first response is always "No". Eighty trials are displayed during this test.

Duration of test 3 minutes

One Back (ONB; Working Memory)

The One Back test is a measure of working memory and uses a well-validated n-back paradigm with playing card stimuli. In this test, the playing cards are identical to those found in a standard deck of 52 playing cards (without the joker cards). The subject is asked whether the card displayed in the center of the screen is the same as the card presented immediately previously. The subject responds by pressing the **Yes** or **No** key. Because no card has been presented yet on the first trial, a correct first response is always **No**. The software measures the speed and accuracy of each response. The time to respond (to a maximum) is recorded for each trial. Wrong responses are counted but do not have an effect on the number of correct responses required for the test to terminate.

Duration of test: 4 minutes

The Groton Maze Learning test (GML; Executive Function)

The Groton Maze Learning test is a measure of problem solving and reasoning and uses a well-validated maze learning paradigm. In this test, the subject is shown a [10 x 10] grid of boxes on a computer screen. A [28]-step pathway is hidden among these [100] possible locations. Each box represents move locations, and the grid refers to the box array (i.e., [10 × 10]). Subjects are required to find the hidden pathway guided by [four] search rules. These rules are: do not move diagonally, do not move more than one box (i.e., do not jump), do not move back on the pathway, and (return to the last correct location after an error). At each step only the most recently selected box is shown. Feedback is given with visual and auditory cues (green check marks and red crosses) to indicate whether the selected box is correct or incorrect. The head of path, or the last correct location, flashes with a green check when two errors are made in succession to indicate to the subject that they must return to this location. [A delayed recall condition is available for this test and requires the subject to find the hidden pathway after a 10-30 minute delay]. There are [20] well-matched alternate pathways available. The software records each move as an error or as a correct move.

Duration of test: 7 minutes

11.2.2 The outcome measures for the Cogstate battery

Although each of these cognitive tests yields multiple outcome measures, research by Cogstate has identified a set of measures that are optimal for the detection of cognitive change in clinical trials at both the group and individual level (Faletti et al., 2006; Maruff et al., 2009; Bland & Altman, 1996).

For each cognitive test, a single primary outcome measure was selected prior to data analysis from each test in the battery. Each primary outcome measure was selected because it has been shown to be optimal for the detection of change:

a) it is drawn from a data distribution that contains only a small probability of floor or ceiling effects and no restriction in the range of possible performance values (Faletti et al., 2006; Bland & Altman, 1996).

b) it is drawn from a distribution that is distributed normally or which can be corrected to normal through the use of appropriate mathematical transformation (e.g., logarithmic base 10, or arcsine) (Faletti et al., 2006; Bland & Altman, 1996).

Table 1 below summarizes the outcome measures for the Cogstate battery, with the tests from which they were derived, the operational definition, and the variable code.

Table 2: Cogstate tests Administered in ESKETINTRD studies, the Cognitive Domains they Assess, and their Primary Outcome Measures

Cogstate test	Cognitive Domain	Primary Outcome Measure	Interpretation of Primary Outcome Score
Detection test(DET)	Attention (simple reaction time)	Speed of performance (mean of the log10 transformed reaction times for correct responses)	Lower score = better performance
Groton Maze Learning test(GML)	Executive Function	Number of errors across all learning trials	Lower score = better performance
Identification test (IDN)	Attention (choice reaction time)	Speed of performance (mean of the log10 transformed reaction times for correct responses)	Lower score = better performance
One Card Learning test (OCL)	Visual Learning	Accuracy of performance (arcsine square root proportion correct)	Higher score = better performance
One Back test (ONB)	Working Memory	Speed of performance (mean of the log10 transformed reaction times for correct responses)	Lower score = better performance

11.2.3 Data Quality Assurance

Data from the Cogstate battery will be collected on computers at all sites and uploaded to the Cogstate database for processing. Cogstate data management staff will query any data discrepancies. Queries will be confirmed and resolved with the sponsor.

11.2.4 Test Completion Criteria

For each of the Cogstate tests, subjects must provide sufficient responses to allow computation of reliable performance measures. For the majority of Cogstate tests, the term "sufficient" has been defined as a test Completion criterion. The number of trials required for test Completion is unique to each test. They do not vary for different patient groups or study samples.

The completion criteria set forth a priori for each test were as follows:

- DET: The number of responses provided by the subject is ≥ 75% of the desired number of trials (responses ≥ 27).
- IDN: The number of responses provided by the subject is ≥ 75% of the desired number of trials (responses ≥ 23).
- ONB: The number of responses provided by the subject is ≥ 75% of the desired number of trials (responses ≥ 24).
- OCL: 75% of the desired numbers of trials were displayed to the subject (trials ≥ 60).

- GML: The subject provided 28 correct moves in each of the 5 learning trials.

11.3 The HVLT-R

The HVLT-R, a measure of verbal learning and memory, is a 12-item word list recall test. Administration includes 3 learning trials, a 24-word recognition list (including 12 target and 12 foil words), and a delayed recall (20-minute) trial. Administration is computer-assisted; instructions and word lists appear on-screen. The test administrator records each word correctly recalled, and scores for learning, short-term, and delayed recall are generated via the test software. The HVLT-R is a well-validated and widely used measure of verbal episodic memory. The tests will be administered in the following order: HVLT-R, computerized cognitive test battery, and HVLT-R Delayed.

12 STATISTICAL METHODOLOGY

12.1 Analysis Overview

All statistical analyses and summary information will be generated according to this Statistical Analysis Plan (SAP).

For continuous variables, descriptive statistics will include number of subjects (n), mean, standard deviation, median, minimum and maximum. Frequencies and percentages will be displayed for categorical data.

Listings will be produced and displayed.

12.1.1 Analysis of Cogstate Endpoints and HVLT-R

Descriptive statistics (n, mean, SD, min, median, max) of scores and change from baseline scores will be summarized and plotted at each scheduled time point.

For HVLT-R, Scores and change from baseline scores will include the following:
- Total Recall (sum of total correct responses for Trials 1, 2, & 3),
- Delayed Recall (Trial 4)
- Total number of true-positive errors and
- Recognition Discrimination Index (Total no. of true-positives)-(Total no. of false-positives).

12.1.2 Baseline Definition

Baseline is defined as Induction phase, Day 1.

12.1.3 Handling of Missing Data and data transformation

Missing Data

No imputations will be performed in the event of missing data due to dropouts or omitted visits. All incomplete subject profiles, consisting of time points that passed Test completion criteria, will be retained in the analysis. In view of issues of reliability, based on the recommendation of Cogstate scientists, all analyses will be conducted with values which failed completion criteria (listed in Section 11.2.3) removed.

Data Transformation

Speed outcome measure

Since Cogstate speed outcome measures are skewed to the right, $\log_{10}$ transformation will be used to normalize the data. The transformed data will be used in the analyses which are specified in this SAP:

LMN = mean (log10 transformed reaction times for correct responses) (1)

Proportion of accuracy outcome measure

Cogstate correct responses follow a binomial distribution. Since a binomial distribution is not well approximated by a normal distribution, especially when n is small, arcsine square root transformation (which is a variance stabilizing transformation) will be used on Cogstate data on the proportion of correct responses measure and this transformed data will be used in the analysis which is specified in this SAP:

ACC = arcsine (sqrt [proportion of correct responses]) (2)

12.2 Unblinding Procedure

Once all data discrepancies within the Cogstate database are resolved with the clinical research units, the database will be locked and Cogstate will receive the randomization codes from the sponsor.

Table, listing and figure (TLF) shells will be generated in a separate document (Statistical Programming Plan<SPP>).

13 REFERENCES

1. Falleti M.G., Maruff P., Collie A., Darby D.G., & McStephen M. (2003). Qualitative similarities in cognitive impairment associated with 24h of sustained wakefulness and a blood alcohol concentration of 0.05%. Journal of Sleep Research, 12, 265-274.

2. Falleti M.G., Maruff P., Collie A., & Darby D.G. (2006). Practice effects associated with the repeated session of cognitive function using the CogState battery at 10- minute, one week and one-month test retest intervals. Journal of Clinical and Experimental Neuropsychology, 28, 1095-1012.

3. Maruff P., Thomas E., Cysique L., Brew B., Collie A., Snyder P., Pietrzak RH (2009). Validity of the CogState brief battery: Relationship to standardized tests and sensitivity to cognitive impairment in mild traumatic brain injury, schizophrenia, and AIDS dementia complex. Archives of Clinical Neuropsychology, 24, 165-178.