

Janssen Research & Development \*

Statistical Analysis Plan

---

**A Randomized, Double-blind, Multicenter, Active-controlled Study to Evaluate the Efficacy, Safety, and Tolerability of Intranasal Esketamine Plus an Oral Antidepressant in Elderly Subjects with Treatment-resistant Depression**

Trial of Rapid-acting Intranasal Esketamine for Treatment-resistant Major Depressive Disorder (TRANSFORM-3)

---

**Protocol ESKETINTRD3005; Phase 3**

**JNJ-54135419 (esketamine)**

**Status:** Approved  
**Date:** 10 January 2017  
**Prepared by:** Janssen Research & Development, LLC  
**Document No.:** EDMS-ERI-122279246, 2.0

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

---

**Confidentiality Statement**

The information in this document contains trade secrets and commercial information that are privileged or confidential and may not be disclosed unless such disclosure is required by applicable law or regulations. In any event, persons to whom the information is disclosed must be informed that the information is privileged or confidential and may not be further disclosed by them. These restrictions on disclosure will apply equally to all future information supplied to you that is indicated as privileged or confidential.

**TABLE OF CONTENTS**

<b>TABLE OF CONTENTS .....</b>	<b>2</b>
<b>LIST OF IN-TEXT TABLES .....</b>	<b>4</b>
<b>LIST OF IN-TEXT FIGURES .....</b>	<b>4</b>
<b>ABBREVIATIONS .....</b>	<b>5</b>
<b>1. INTRODUCTION.....</b>	<b>7</b>
1.1. Trial Objectives .....	7
1.2. Trial Design .....	8
1.3. Statistical Hypotheses for Trial Objectives.....	11
1.4. Sample Size Justification .....	11
1.5. Interim Analysis for Sample Size Re-Estimation or Stopping for Futility .....	12
1.6. Randomization and Blinding .....	12
<b>2. GENERAL ANALYSIS DEFINITIONS .....</b>	<b>13</b>
2.1. Pooling Algorithm .....	13
2.2. Analysis Phases.....	13
2.2.1. Study Reference Start and End Dates .....	13
2.2.2. Analysis Phase Start and End Dates.....	13
2.2.3. Study Day and Relative Day.....	14
2.3. Baseline and End Point.....	14
2.4. Visit Windows .....	14
2.5. Analysis Sets.....	19
2.5.1. All Randomized Analysis Set.....	19
2.5.2. Full Analysis Set .....	19
2.5.3. Safety Analysis Set.....	19
2.5.4. Follow-up Analysis Set .....	19
2.6. Definition of Subgroups.....	19
2.7. Imputation Rules for Missing AE Dates .....	20
2.8. Imputation Rules for Missing AE time of Onset/Resolution .....	21
2.9. Imputation Rules for Missing Concomitant Medication Dates .....	21
<b>3. INTERIM ANALYSIS AND DATA MONITORING COMMITTEE REVIEW.....</b>	<b>21</b>
<b>4. SUBJECT INFORMATION.....</b>	<b>22</b>
4.1. Demographics and Baseline Characteristics .....	22
4.2. Disposition Information.....	23
4.3. Extent of Exposure .....	23
4.4. Protocol Deviations .....	24
4.5. Prior and Concomitant Medications .....	24
<b>5. EFFICACY .....</b>	<b>24</b>
5.1. Analysis Specifications.....	25
5.1.1. Level of Significance.....	25
5.1.2. Data Handling Rules.....	25
5.1.3. Imputation Methods for Missing Items.....	25
5.2. Primary Efficacy Endpoint.....	25
5.2.1. Definition.....	25
5.2.2. Primary Estimand .....	26
5.2.3. Analysis Methods.....	26
5.3. Other Efficacy Variables.....	31
5.3.1. Responders .....	31
5.3.1.1. Definition.....	31

5.3.1.2.	Analysis Methods .....	31
5.3.2.	Remitters .....	32
5.3.2.1.	Definition .....	32
5.3.2.2.	Analysis Methods .....	32
5.3.3.	CGI-S .....	32
5.3.3.1.	Definition .....	32
5.3.3.2.	Analysis Methods .....	32
5.3.4.	EuroQol Group; 5 Dimension; 5 Level (EQ-5D-5L) .....	33
5.3.4.1.	Definition .....	33
5.3.4.2.	Analysis Methods .....	33
5.3.5.	Sheehan Disability Scale (SDS) .....	34
5.3.5.1.	Definition .....	34
5.3.5.2.	Analysis Methods .....	34
5.3.6.	PHQ-9 .....	34
5.3.6.1.	Definition .....	34
5.3.6.2.	Analysis Methods .....	34
<b>6.</b>	<b>SAFETY .....</b>	<b>35</b>
6.1.	Adverse Events .....	35
6.2.	Clinical Laboratory Tests .....	36
6.3.	Vital Signs, Weight, and BMI .....	37
6.4.	Electrocardiogram .....	38
6.5.	Nasal Examination .....	39
6.5.1.	Nasal Symptom Questionnaire .....	39
6.6.	Other Safety Parameters .....	40
6.6.1.	Columbia Suicide Severity Rating Scale (C-SSRS) .....	40
6.6.2.	Clinician Administered Dissociative States Scale (CADSS) .....	41
6.6.3.	Brief Psychiatric Rating Scale (BPRS) .....	41
6.6.4.	Modified Observer's Assessment of Alertness/Sedation (MOAA/S) .....	42
6.6.5.	Clinical Global Assessment of Discharge Readiness (CGADR) .....	42
6.6.6.	Physician Withdrawal Checklist (PWC-20) .....	43
6.6.7.	Computerized Cognitive Battery and Hopkins Verbal Learning Test-Revised (HVLT-R) .....	43
6.6.8.	University of Pennsylvania Smell Inventory Test (UPSIT) .....	43
<b>7.</b>	<b>PHARMACOKINETICS/PHARMACODYNAMICS .....</b>	<b>44</b>
	<b>REFERENCES .....</b>	<b>45</b>
	<b>ATTACHMENTS .....</b>	<b>46</b>
Attachment 1:	Criteria of Markedly Abnormal Laboratory Values .....	46
Attachment 2:	Statistical Analysis Plan for Cogstate .....	47

**LIST OF IN-TEXT TABLES**

Table 1:	Analysis Visits .....	16
Table 2:	Demographic Variables and Baseline Characteristics.....	22
Table 3:	Psychiatric History at Baseline Variables .....	23
Table 4:	Efficacy Variables .....	25
Table 5:	Analysis Types and Assumptions .....	28
Table 6:	Treatment-Emergent Abnormality Categories for Vital Signs.....	37
Table 7:	Limits for HR, PR, QRS and QT Interval Abnormality .....	38
Table 8:	Criteria for Abnormal QTc Values and Changes From Baseline.....	39

**LIST OF IN-TEXT FIGURES**

Figure 1:	Study Design for ESKETINTRD3005 .....	11
-----------	---------------------------------------	----

**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
CMH	Cochran-Mantel-Haenszel
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
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
HVLT-R	Hopkins Verbal Learning Test-Revised
IDMC	Independent Data Monitoring Committee
ICH	International Conference on Harmonization
IDS-C <sub>30</sub>	Inventory of Depressive Symptoms-Clinician rated, 30 item
IWRS	Interactive web response system
LOCF	last observation carried forward
MADRS	Montgomery-Asberg Depression Rating Scale
MCMC	Markov Chain Monte Carlo
MDD	major depressive disorder
MedDRA	Medical Dictionary for Regulatory Activities
MGH-ATRQ	Massachusetts General Hospital Antidepressant Treatment Response Questionnaire
MI	multiple imputation
MINI	Mini International Neuropsychiatric Interview
MMRM	Mixed-effects model using repeated measures
MOAA/S	Modified Observer’s Assessment of Alertness/Sedation
OL	open-label
PBO	placebo
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
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
ULN	upper limit of normal

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-ESKETINTRD3005.

### 1.1. Trial Objectives

#### Primary Objective

The primary objective of this study is to evaluate the efficacy of switching elderly subjects with treatment-resistant depression (TRD) from a prior antidepressant treatment (to which they have not responded) to flexibly dosed intranasal esketamine (28 mg, 56 mg or 84 mg) plus a newly initiated oral antidepressant compared with switching to a newly initiated oral antidepressant (active comparator) plus intranasal placebo, in improving depressive symptoms, as assessed by the change from baseline in the Montgomery-Asberg Depression Rating Scale (MADRS) total score from Day 1 (pre-randomization) to the end of the 4-week double-blind induction phase.

#### Secondary Objectives

- To assess the effect of intranasal esketamine plus a newly initiated oral antidepressant compared with a newly initiated oral antidepressant (active comparator) plus intranasal placebo on the following parameters in elderly subjects with TRD:
  - Depression response rates
  - Depression remission rates
  - Overall severity of depressive illness
  - Health-related quality of life and health status
- To investigate the safety and tolerability of intranasal esketamine plus a newly initiated oral antidepressant compared with a newly initiated oral antidepressant (active comparator) plus intranasal placebo in elderly subjects with TRD, including 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 effects
  - 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

- To assess the pharmacokinetics (PK) of intranasal esketamine in elderly subjects with TRD receiving intranasal esketamine plus a newly-initiated oral antidepressant

### **Exploratory Objectives**

- To assess the PK/pharmacodynamic (PK/PD) relationship of intranasal esketamine and MADRS total score in elderly subjects with TRD
- To assess the potential relationship of biomarkers with response/non-response to intranasal esketamine or oral antidepressants in elderly subjects with TRD

### **1.2. Trial Design**

This is a randomized, double-blind, active-controlled, multicenter study in male and female elderly subjects with TRD to assess the efficacy, safety, and tolerability of flexibly dosed intranasal esketamine (28 mg, 56 mg or 84 mg) plus a newly initiated oral antidepressant compared with a newly initiated oral antidepressant (active comparator) plus intranasal placebo.

The study has 3 phases:

#### **Screening/prospective observational phase (4-week duration + optional 3-week taper period)**

This phase will prospectively assess 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 nonresponse to at least one antidepressant treatment (based on MGH-ATRQ criteria) in the current episode of depression, and the subject is taking a different oral antidepressant treatment(s) (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 and 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 double-blind induction phase. Non-response at the end of the screening/prospective observational phase is defined as  $\leq 25\%$  improvement in the MADRS total score from Week 1 to Week 4 and a MADRS total score of  $\geq 24$  on Week 2 and Week 4.

Eligible subjects who are entering the double-blind induction phase will discontinue all of their current medication(s) being used for depression treatment, 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, but no dose increases beyond the equivalent of 6 mg/day of lorazepam or new

benzodiazepine medications 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. This 3 week period may also be used to optimize medical management if needed to facilitate subject participation (e.g., treatment of blood pressure or diabetes, wean from other medications, etc.). In such cases, if there is no planned taper, the oral regimen should be continued in the interim and then discontinued by Day 1 of the double-blind induction phase.

As a new oral antidepressant will be initiated on Day 1 of the double-blind induction phase, eligible subjects who do not require a tapered discontinuation of their antidepressant treatment(s) can proceed immediately into the double-blind induction phase.

### **Double-blind induction phase (4-week duration)**

Approximately 148 subjects will be randomly assigned at a 1:1 ratio (n=74 subjects per treatment arm) to receive double-blind treatment with either intranasal esketamine or intranasal placebo. The intranasal treatment sessions (esketamine or placebo) will occur twice weekly. In addition, all subjects will initiate a new open-label oral antidepressant on Day 1 that will be taken daily for the duration of this 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.

Subjects who were 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 during the induction phase. No dose increases beyond the equivalent of 6 mg/day of lorazepam or new benzodiazepine medications are permitted during the induction phase, with the exception of the use of permitted benzodiazepine rescue medication.

At the end of the induction phase, regardless of response status, subjects may be eligible to participate in the subsequent study ESKETINTRD3004 if they meet all other study entry criteria (ESKETINTRD3004 is a longer-term open-label safety study involving repeated treatment sessions of intranasal esketamine).

If a subject withdraws from the study before the end of the double-blind induction phase 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.

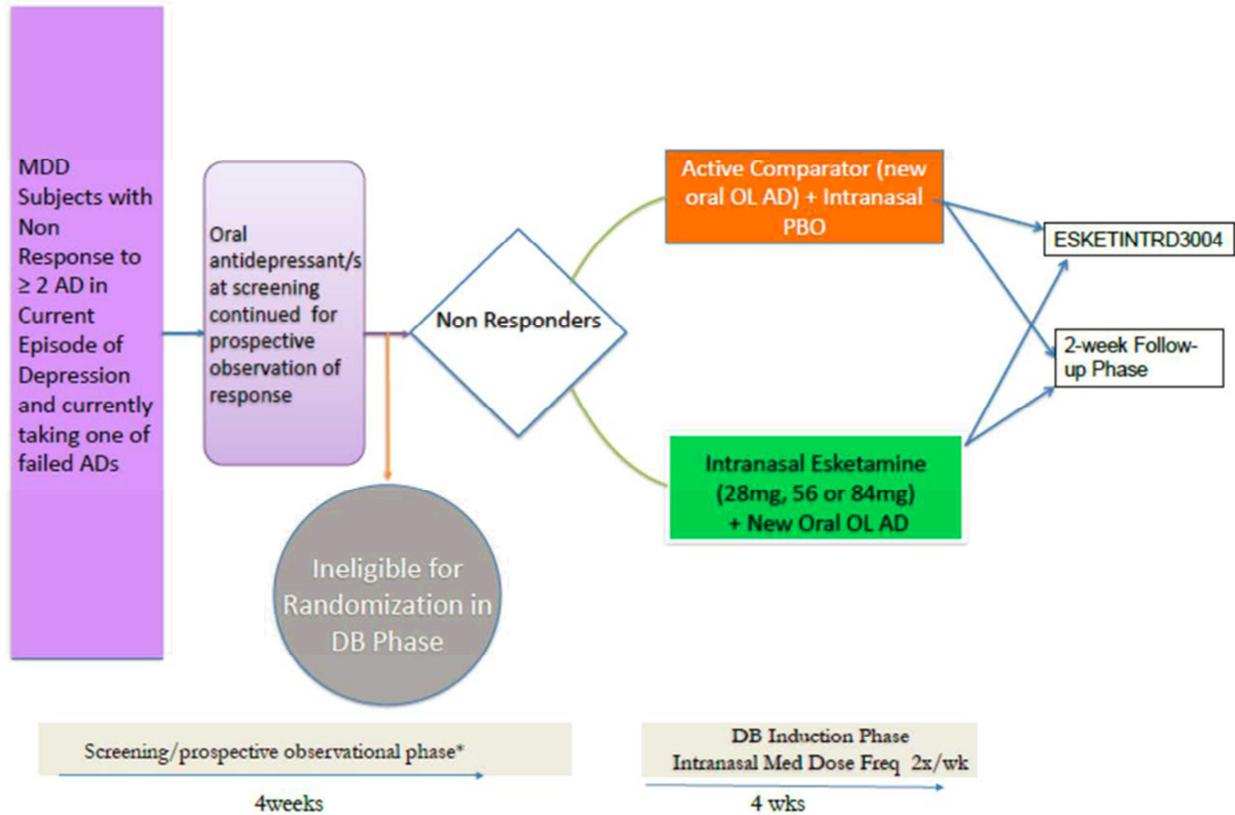
**Follow-up phase (2-week duration)**

This phase will include all subjects who are not eligible or who choose to not participate in the subsequent ESKETINTRD3004 safety study, and have received at least 1 dose of intranasal study medication in the double-blind induction phase. There will be no intranasal treatment administered during this phase.

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. All 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 medication, the oral antidepressant medication should be continued for the 2 weeks of the follow-up phase unless determined as not clinically appropriate.

Taking into consideration the up to 3-week optional taper period, the duration of a subject's study participation will be 8 to 11 weeks (for subjects continuing into ESKETINTRD3004) or 13 weeks (for subjects completing the follow-up phase).

A diagram of the study design is provided in [Figure 1](#).

**Figure 1: Study Design for ESKETINTRD3005**

\* followed by optional, up to 3 week Taper if clinically indicated.

AD=antidepressant. DB=double-blind. D/C=discontinued. Esk=Esketamine. MDD=major depressive disorder. OL=open-label. PBO=placebo.

Note: Subjects who withdraw early from the double-blind induction phase and receive at least 1 dose of intranasal study medication in the double-blind induction phase will have an Early Withdrawal visit performed and then proceed into the follow-up phase.

### 1.3. Statistical Hypotheses for Trial Objectives

The hypothesis for this study is that, in elderly subjects with TRD, switching from a failed antidepressant treatment to intranasal esketamine plus a newly initiated oral antidepressant is superior to switching to a newly initiated oral antidepressant treatment (active comparator) plus intranasal placebo in improving depressive symptoms.

### 1.4. Sample Size Justification

The maximum sample size planned for this study was calculated assuming a treatment difference for the double-blind induction phase of 6.5 points in MADRS total score between esketamine and the active comparator, a standard deviation of 12, a 1-sided significance level of 0.025, and a drop-out rate of 25%. A maximum of about 74 subjects will need to be randomized to each treatment group to achieve 80% power using a fixed design with no interim analysis. The treatment difference and standard deviation used in this calculation were based on results of Panel A of the ESKETINTRD2003 study and on clinical judgment.

As detailed below, an interim analysis is planned to re-estimate sample size or to stop the study due to futility.

### **1.5. Interim Analysis for Sample Size Re-Estimation or Stopping for Futility**

One unblinded interim analysis will be performed 4 weeks after randomizing 50 subjects in the study (approximately 25 per treatment group). It is projected that at that time approximately 36 subjects in the full analysis set would have completed the double-blind induction phase of the study (approximately 18 subjects per treatment group). The dropout rate will be monitored to ensure a sufficient number of subjects are included in the interim analysis. As the assumptions of the expected treatment difference and variability may or may not be upheld, the purpose of the interim analysis is to either re-estimate sample size or to stop the study due to futility. The sample size may be adjusted to achieve the desired power while maintaining control of the overall Type I error. The maximum sample size planned for this study is 74 subjects per treatment group.

A separate rigorous interim analysis 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 Independent Data Monitoring Committee (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.

### **1.6. Randomization and Blinding**

Randomization will be used to minimize bias in the assignment of subjects to treatment groups, to increase the likelihood that known and unknown subject attributes (eg, demographic and baseline characteristics) are evenly balanced across treatment groups, and to enhance the validity of statistical comparisons across treatment groups.

Central randomization will be implemented in this study. Subjects will be randomly assigned to 1 of 2 treatment groups (placebo or flexible dose esketamine) in a 1:1 ratio based on a computer-generated randomization schedule prepared before the study under the supervision of the sponsor. The randomization will be balanced by using randomly permuted blocks and will be stratified by country and class of oral antidepressant (SNRI or SSRI) to be initiated in the double-blind induction phase. The interactive web response system (IWRS) will assign a unique treatment code, which will dictate the treatment assignment and matching study drug kits for the subject.

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.

An intranasal placebo control will be used in the double-blind induction phase to establish the frequency and magnitude of changes in clinical endpoints that may occur in the absence of intranasal active treatment. Blinded intranasal treatment will be used to reduce potential bias during data collection and evaluation of clinical endpoints.

## **2. GENERAL ANALYSIS DEFINITIONS**

### **2.1. Pooling Algorithm**

A total of 15 countries (Brazil, Belgium, France, Poland, Spain, US, Italy, South Africa, United Kingdom, Sweden, Australia, Bulgaria, Finland, Lithuania, Republic of Korea) are to enroll subjects. Due to low enrollment in some countries, countries would be pooled by region: North America, Europe and Other. Region will be used as a factor in the statistical models to analyze efficacy to account for region variability.

### **2.2. Analysis Phases**

There are 2 analysis phases defined in this study: Double-Blind Induction and Follow-up (post treatment). Each analysis phase has its own analysis phase start and end dates.

#### **2.2.1. Study Reference Start and End Dates**

The reference start date for the study is defined as the earlier date of the first dose of intranasal study drug, or the oral antidepressant study drug (the date is missing for screened subjects who did not take any intranasal study drug or oral antidepressant study drug). The reference end date for the study is the end of trial date including the last follow-up visit.

#### **2.2.2. Analysis Phase Start and End Dates**

##### **Double-Blind Induction Phase**

The start date of the double-blind induction phase (referred to as ‘DB start date’) is the earlier of date of the first dose of intranasal study medication, or the oral antidepressant study medication. The double-blind induction phase end date (referred to as ‘DB end date’) is the maximum of the date of the last visit in the double-blind induction phase and the date of completion/withdrawal from the double-blind induction phase.

The start date/time of the double-blind induction phase (referred to as, ‘DB start date/time’) is the DB start date and the time of the first dose of intranasal study medication. If no intranasal study medication is administered or it is administered after the start of oral antidepressant, then the time will be left blank.

**Follow-up Phase**

The start date of the follow-up phase (referred to as ‘F/U start date’) is the day after the DB end date. The follow-up phase end date (referred to as ‘F/U end date’) is the maximum of the last follow-up visit date or end of trial date.

**2.2.3. 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 start date of the analysis phase in which the data are captured. A minus (-) sign indicates days prior to the double-blind phase.

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

$$\text{event date} - \text{reference start date} + 1.$$

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

$$\text{event date} - \text{reference start date}$$

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

$$\text{event date} - \text{analysis phase start date} + 1.$$

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

$$\text{event date} - \text{analysis phase start date}.$$

There is no study day 0 or relative day 0.

**2.3. Baseline and End Point**

Baseline is defined as the last observation before receiving the first dose of any study medication in the double-blind induction phase. Baseline is defined for each parameter/assessment.

The double-blind end point value will be the final post baseline value assessed during the double-blind induction phase.

**2.4. Visit Windows**

As subjects do not always adhere to the protocol visit schedule (including permitted visit windows), 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 any study drug was taken in the double-blind phase).

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 <sup>a</sup>	Time Interval (label on output)	Time Interval <sup>a</sup> (Day)	Target Time Point <sup>a</sup> (Day)
MADRS	Screening		Week 1 (SC)		
			Week 2 (SC)		
			Week 4 (SC)		
	DB	1	Baseline	≤1	1
		8	Day 8	2-11	8
		15	Day 15	12-18	15
		22	Day 22	19-24	22
		28	Day 28	25 to end of DB	28
		DB final visit	End Point (DB)	2 to end of DB	
	Follow-up	7	Week 1(F/U)	1 to 10	7
14		Week 2(F/U)	11 to end of F/U	14	
F/U final visit		End Point (F/U)	1 to end of F/U		
PHQ-9	Screening		Week 1 (SC)		
	DB	1	Baseline	≤1	1
		15	Day 15	2-21	15
		28	Day 28	22 to end of DB	28
		DB final visit	End Point (DB)	2 to end of DB	
Follow-up	14	Week 2 (F/U)	1 to end of F/U	14	
SDS	Screening		Week 1 (SC)		
	DB	1	Baseline	≤1	1
		15	Day 15	2-21	15
		28	Day 28	22 to end of DB	28
		DB final visit	End Point (DB)	2 to end of DB	
Follow-up	14	Week 2 (F/U)	1 to end of F/U	14	
CGI-S	Screening		Week 1 (SC)		
	DB	1	Baseline	≤1	1
		4	Day 4	2-6	4
		8	Day 8	7-9	8
		11	Day 11	10-13	11
		15	Day 15	14-16	15
		18	Day 18	17-20	18
		22	Day 22	21-23	22
		25	Day 25	24-26	25
		28	Day 28	27 to end of DB	28
DB final visit	End Point (DB)	2 to end of DB			
Follow-up	14	Week 2 (F/U)	1 to end of F/U	14	
EQ-5D-5L, EQ-VAS and Health Status Index	Screening		Week 1 (SC)		
	DB	1	Baseline	≤1	1
		15	Day 15	2-21	15
		28	Day 28	22 to end of DB	28
		DB final visit	End Point (DB)	2 to end of DB	
Follow-up	14	Week 2 (F/U)	1 to end of F/U	14	
Vital Signs (TEMP [predose at each visit], BP <sup>b</sup> , HR, RESP [at each visit, predose, 40M, 1H, 1.5H])	DB	1	Day 1: Predose Day 1: 40M Day 1: 1H Day 1: 1H30M	≤ 1 / predose	1
		4	Day 4: Predose Day 4: 40M Day 4: 1H Day 4: 1H30M	2-6	4
		8	Day 8: Predose Day 8: 40M Day 8: 1H	7-9	8

**Table 1: Analysis Visits**

Parameter	Analysis Phase	Scheduled Day <sup>a</sup>	Time Interval (label on output)	Time Interval <sup>a</sup> (Day)	Target Time Point <sup>a</sup> (Day)
			Day 8: 1H30M		
		11	Day 11: Predose Day 11: 40M Day 11: 1H Day 11: 1H30M	10-13	11
		15	Day 15: Predose Day 15: 40M Day 15: 1H Day 15: 1H30M	14-16	15
		18	Day 18: Predose Day 18: 40M Day 18: 1H Day 18: 1H30M	17-20	18
		22	Day 22: Predose Day 22: 40M Day 22: 1H Day 22: 1H30M	21-23	22
		25	Day 25: Predose Day 25: 40M Day 25: 1H Day 25: 1H30M	24 to end of DB	25
		DB final visit	End Point (DB)	Day 1: 40M to end of DB	
	Follow-up	14	Week 2 (F/U)	1 to end of F/U	14
Weight and BMI	DB	Screening, 1	Baseline	≤ 1 / predose	1
		28	Day 28	2 to end of DB	28
		DB final visit	End Point (DB)	2 to end of DB	
ECG	DB	Screening, 1	Average Predose	≤ 1 / predose	1
		1	Day 1: 1H	1	1
		4	Day 4: 1H	2-6	4
		8	Day 8: 1H	7-9	8
		11	Day 11: 1H	10-13	11
		15	Day 15: 1H	14-16	15
		18	Day 18: 1H	17-20	18
		22	Day 22: 1H	21-23	22
		25	Day 25: 1H	24 to end of DB	25
	DB final visit	End Point (DB)	1 to end of DB		
	Follow-up	14	Week 2 (F/U)	1 to end of F/U	14
Lab	DB	Screening, 1	Baseline	≤ 1 / predose	1
		28	Day 28	1 to end of DB	28
		DB final visit	End Point (DB)	1 to end of DB	
	Follow-up	14	Week 2 (F/U)	1 to end of F/U	14
C-SSRS	Screening		Week 1 (SC)		
			Week 2 (SC)		
			Week 4 (SC)		
	DB	1	Baseline	≤1	1
		4	Day 4	2-6	4
		8	Day 8	7-9	8
		11	Day 11	10-13	11
		15	Day 15	14-16	15
		18	Day 18	17-20	18
		22	Day 22	21-23	22
		25	Day 25	24-26	25
		28	Day 28	27 to end of DB	28
		DB final visit	End Point (DB)	2 to end of DB	

**Table 1: Analysis Visits**

Parameter	Analysis Phase	Scheduled Day <sup>a</sup>	Time Interval (label on output)	Time Interval <sup>a</sup> (Day)	Target Time Point <sup>a</sup> (Day)
	Follow-up	14	Week 2 (F/U)	1 to end of F/U	14
MOAA/S <sup>c</sup> Oxygen Saturation <sup>d</sup> (predose and every 15 minutes to 1.5H)	DB	1	Day 1	1	1
		4	Day 4	2-6	4
		8	Day 8	7-9	8
		11	Day 11	10-13	11
		15	Day 15	14-16	15
		18	Day 18	17-20	18
		22	Day 22	21-23	22
BPRS+ and CADSS (predose, 40M, 1.5H) CGADR <sup>e</sup> (1H, 1.5H)		25	Day 25	24 to end of DB	25
PWC-20	DB	25	Day 25	2 to end of DB	25
	Follow-up	7	Week 1 (F/U)	1 to 10	7
		14	Week 2 (F/U)	11 to end of F/U	14
Nasal exam	DB	1	Baseline	1	1
		28	Day 28	2 to end of DB	28
		DB final visit	End Point (DB)	2 to end of DB	
	Follow-up	14	Week 2 (F/U)	1 to end of F/U	14
Nasal questionnaire	DB	1	Baseline	1	1
		4	Day 4	2-7	4
		11	Day 11	8-14	11
		18	Day 18	15-21	18
		25	Day 25	22 to end of DB	25
		DB final visit	End Point (DB)	2 to end of DB	
BPIC-SS	DB	1	Baseline	1	1
		15	Day 15	2-21	15
		28	Day 28	22 to end of DB	28
		DB final visit	End Point (DB)	2 to end of DB	
	Follow-up	14	Week 2 (F/U)	1 to end of F/U	14
UPSIT	Screening		Week 2 (SC)		
	DB	28	Day 28	2 to end of DB	28
		DB final visit	End Point (DB)	2 to end of DB	
Cogstate / HVLT-R	DB	1	Baseline	1	1
		28	Day 28	2 to end of DB	28
	Follow-up	14	Week 2 (F/U)	1 to end of F/U	14

<sup>a</sup> For DB phase, time interval is relative to the first day of the double-blind induction phase.

For follow-up phase, time interval is relative to the first day of the follow-up phase.

<sup>b</sup> During the DB phase, at 1,5 hours post dose if the SBP is  $\geq 160$  and/or DBP  $\geq 100$ , assessments should continue every 30 minutes until the blood pressure is  $< 160$  and  $< 100$  or investigator's clinical judgment the subject it is clinical stable and can be discharged from the clinical site.

<sup>c</sup> If the MOAA/S score is  $\leq 3$  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.)

<sup>d</sup> 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  $\geq 93\%$  or until the subject is referred for appropriate medical care, if clinically indicated.

<sup>e</sup> 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.

## 2.5. Analysis Sets

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

### 2.5.1. All Randomized Analysis Set

This analysis set will include all subjects who were randomized (i.e., subjects who reported a randomization date, or were assigned a randomization number) regardless of whether or not treatment was received. This analysis set will be used for summarizing the overall study completion/withdrawal information.

### 2.5.2. Full Analysis Set

The efficacy analyses of data in the double-blind induction phase will be based on the full analysis set. The full analysis set is defined as all randomized subjects who received at least 1 dose of intranasal study medication and 1 dose of oral antidepressant medication during the double-blind induction phase.

### 2.5.3. Safety Analysis Set

The safety analysis set is defined for the double-blind induction phase. The safety analysis set includes all randomized subjects who received at least 1 dose of intranasal study medication or 1 dose of oral antidepressant medication during the double-blind induction phase. Analyses of change from baseline will include only subjects who have both baseline and at least 1 post-baseline observation during that phase. Screen failures and randomized subjects who received no double-blind study medication will be excluded from the safety analysis set. Subjects who received an incorrect treatment will be analyzed under the planned treatment.

### 2.5.4. Follow-up Analysis Set

The Follow-up analysis set is defined as all subjects who enter the follow-up phase. This analysis set will be used for both efficacy and safety analyses.

## 2.6. Definition of Subgroups

Analyses will be provided for the primary endpoint, change from baseline in MADRS total score, using the following subgroups.

- Sex
- Race (White, Black, Other)
- Age Group (65-74 years,  $\geq 75$  years)
- Region (North America, Europe, Other)
- Country (Brazil, Belgium, France, Poland, Spain, US, Italy, South Africa, United Kingdom, Sweden, Australia, Bulgaria, Finland, Lithuania, Republic of Korea)
- Number of Previous Treatment Failures in Current Episode (based on ATRQ)

- Baseline MADRS total score ( $\leq$ / $>$  median)
- Class of antidepressant study medication (SNRI or SSRI)

## 2.7. Imputation Rules for Missing AE Dates

Treatment-emergent adverse events (TEAEs) for the double-blind induction phase are those events with an onset date/time on or after the start of study medication, and occurred on or before the end of the double-blind phase. Adverse events (AEs) for the follow-up phase are those events with an onset date on or after the start of the follow-up phase, and occurred on or before the end of the follow-up phase. A conservative approach will be used to handle the missing dates for adverse events.

### Onset Date

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 DB start date
- ii) The day of DB start date, if the month/year of the onset of AE is the same as month/year of the DB start date and month/year of the AE resolution date is different
- iii) The day of DB start date or day of AE resolution date, whichever is earliest, if month/year of the onset of AE and month/year of the DB 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 DB start date.
- ii) One day after the DB 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 DB start 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 earlier of the date of withdrawal, study completion, or 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.

## 2.8. Imputation Rules for Missing AE time of Onset/Resolution

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

- (i) 00:00 if the date of onset is after DB start date
- (ii) 00:00 if the date is the same as DB start date, but the intranasal study medication in the double-blind induction phase was started after the oral antidepressant medication in this phase
- (iii) The time of intranasal medication start in the double-blind induction phase if the intranasal medication was started on or before the oral antidepressant medication in this phase

If the time of resolution is missing, it will be set to 23:59.

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

## 2.9. 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:

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 DB 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 DB 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 1 interim analysis for sample size re-estimation. 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 to the team regarding the continuation of the

study, or, in the case of the interim analysis for efficacy, to either stop the study due to futility or to adjust the sample size to achieve the desired power while maintaining control of the overall Type I error. Recommendation of adjustment to sample size will be relayed only to the IWRS vendor and to clinical supplies.

To protect the integrity of the clinical study, the IDMC members (medical and statistical experts) are not study team personnel or otherwise directly involved in the study conduct, data management, or statistical analysis for the study. Details about the review to be performed and the roles and responsibilities of the IDMC are presented in a separate IDMC charter.

## 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 Safety and Full analysis sets. 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.

**Table 2: Demographic Variables and Baseline Characteristics**

---

Continuous Variables:

- Age (years)
- Baseline weight (kg)
- Baseline height (cm)
- Baseline BMI ( $\text{kg}/\text{m}^2$ ) calculated as  $\text{Weight (kg)}/[\text{Height (m)}]^2$

Categorical Variables:

- Age (65-74,  $\geq 75$ )
  - Sex (male, female)
  - Racea (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:  $<18.5 \text{ kg}/\text{m}^2$ , normal:  $18.5$  to  $<25 \text{ kg}/\text{m}^2$ , overweight:  $25 \text{ kg}/\text{m}^2$  to  $<30 \text{ kg}/\text{m}^2$ , obese:  $30$  to  $<40 \text{ kg}/\text{m}^2$ , morbidly obese:  $\geq 40 \text{ kg}/\text{m}^2$ )
  - Hypertension Status
  - Class of antidepressant (SSRI/SNRI)
  - Oral antidepressant
  - Country
  - Region
- 

<sup>a</sup> 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
- Screening IDS-C30 total score
- Baseline CGI-S 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 taken for at least 2 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 distribution of the number of subjects who are randomized, receive double-blind treatment, and complete the double-blind induction phase will be presented by treatment group. In addition, the distribution of reasons for discontinuation will be presented. These summaries will be provided for the All Randomized and Safety analysis sets.

A subject will be considered to have completed the double-blind induction phase if he or she has completed the MADRS assessment at the end of the 4-week double-blind induction phase (ie Day 28 MADRS).

The distribution of the number of subjects who enter the follow-up phase and complete all follow-up visits will be presented by treatment group and overall. The reasons for discontinuation will be presented. In addition, the number of subjects who enroll in study ESKETINTRD3004 will be provided.

A subject will be considered to have completed the follow-up phase if he or she has completed the MADRS assessment at Week 2 of the follow-up phase.

**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 for the Full analysis set and the Safety analysis set.

The total duration of exposure for the intranasal study drug and for each type of oral antidepressant (AD) during the double-blind induction phase is defined as the time between the first and the last dose of each type of study medication.

Descriptive statistics (N, mean, SD, median, minimum and maximum) of total duration of exposure of intranasal study drug will be presented. The total duration of intranasal study drug exposure in the double-blind induction phase will be presented using the following categories:  $\leq 7$  days, 8-14 days, 15-21 days, 22-25 days,  $>25$  days. A frequency distribution of the total number of dosing sessions of intranasal study medication during the double-blind induction phase will be presented. A frequency distribution of dose level will be presented for each dosing session. The total duration of exposure of oral AD will be summarized similarly to the intranasal study drug, however the following categories will be used for the double-blind induction phase:  $\leq 7$  days, 8-14 days, 15-21 days, 22-28 days,  $>28$  days, and the following categories will be used for the follow-up phase:  $\leq 7$  days,  $>7$  days. Each type of oral AD will be summarized separately.

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

Descriptive statistics (N, mean, SD, median, minimum and maximum) of modal dose, mean and final dose of intranasal study drug will be presented. Doses of oral AD will be summarized using descriptive statistics of the mean dose (days on drug), mode dose (days on drug) and the final dose, by each type of oral AD for both the double-blind induction and follow-up phases. In addition, percent compliance of the oral AD calculated as days actually dosed/days expected to be dosed\*100, will be summarized.

#### **4.4. Protocol Deviations**

Deviations that occurred during the study will be tabulated for the All Randomized analysis set by treatment group. 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**

Antidepressant medications taken prior to the baseline visit will be summarized by treatment group for the Safety analysis set.

The number and percent of subjects who receive concomitant therapies will be summarized by treatment group and phase using the generic term of the medication for the Safety analysis set and for the Follow-Up analysis set.

### **5. EFFICACY**

The efficacy variables for this study are listed in [Table 4](#).

**Table 4: Efficacy Variables**

		Endpoint
MADRS	Change in MADRS from Baseline to Day 28 or End Point (DB)	Primary
	Response rates over time (at least 50% improvement from baseline)	Secondary
	Remission rates over time (MADRS $\leq$ 12)	Secondary
CGI-S	Change in CGI-S from Baseline to End Point (DB)	Secondary
EQ-5D-5L, EQ-VAS and health status index	Change from Baseline to End Point (DB)	Secondary
SDS <sup>†</sup>		N/A
PHQ-9 <sup>†</sup>		N/A

†: SDS and PHQ-9 were efficacy variables collected prior to Protocol Amendment 3 but are no longer in the current protocol (Protocol Amendment 3).

## 5.1. Analysis Specifications

### 5.1.1. Level of Significance

Statistical analysis tests will be conducted at a 1-sided 0.025 level of significance unless specified otherwise. One unblinded interim analysis will be performed with the purpose of either re-estimating the sample size or stopping the study for futility.

To adjust for the interim analysis, the primary endpoint will be analyzed using a weighted combination test as described in Section 5.2.3.

### 5.1.2. Data Handling Rules

For endpoints using analysis of covariance (ANCOVA), the last observation carried forward (LOCF) method will be applied to the MADRS total score and CGI-S for the double-blind induction phase. The last post baseline observation during the double-blind induction phase will be carried forward as the “End Point” for that phase. Besides the observed cases and the end point assessment, the LOCF values will be created for intermediate post-baseline time points as well. These imputed time points will be labeled ‘DAY X LOCF’.

### 5.1.3. Imputation Methods for Missing Items

Imputation of the MADRS total score is described in Section 5.2.1. For all other scales where multiple items are summed to create a total, if any item of the scale is missing at a visit, the total score for that scale at that visit will be left blank.

## 5.2. Primary Efficacy Endpoint

### 5.2.1. Definition

The primary efficacy endpoint is the change in MADRS total score from Day 1 to Day 28. The MADRS is a clinician-rated scale designed to measure depression severity and to detect changes due to antidepressant treatment<sup>9</sup>. 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.

### **5.2.2. Primary Estimand**

The primary estimand, the main clinical quantity of interest to be estimated in the study, is defined by the following 3 components:

**Population:** elderly subjects with treatment-resistant depression

**Endpoint:** change from baseline to Day 28 in the MADRS total score (see previous section)

**Measure of Intervention:** the effect of the initially randomized treatment together with the oral AD that would have been observed had all subjects remained on their treatment throughout the DB induction phase.

The primary analysis will be based on the full analysis set, as described in Section 2.4.2, and the MADRS total scores collected during the DB induction phase.

### **5.2.3. Analysis Methods**

#### **MMRM**

With the exception of the European Union (EU) dossier, the primary efficacy variable, change from baseline in MADRS total score at Day 28 in the double-blind induction phase, will be analyzed using a Mixed-Effect Model for Repeated Measures (MMRM) based on observed case data. The MMRM analysis will be performed for each stage (Stage 1-all data collected on subjects used for sample size re-estimation at the interim analysis and Stage 2-all data collected on the remaining subjects). The models will include baseline MADRS total score as a covariate, and treatment, region, class of antidepressant (SNRI or SSRI), day (see [Table 1](#)), and day-by-treatment interaction as fixed effects. The within-subject covariance between visits will be estimated via an unstructured variance-covariance matrix. In case of convergence problems, alternative variance-covariance structures will be tried in the following order, with the first structure that converges being used in the analysis: heterogeneous Toeplitz; standard Toeplitz; and AR(1) with separate subject random effect. Comparison of the esketamine plus oral antidepressant arm versus oral antidepressant plus intranasal placebo will be performed using the appropriate contrast.

#### **ANCOVA**

For the EU dossier, the primary efficacy analysis will be based on an analysis of covariance (ANCOVA) model using change from baseline to Day 28 based on LOCF data. The ANCOVA analysis will be performed for each stage separately (Stage 1- all data collected on subjects used

for sample size re-estimation at the interim analysis and Stage 2-all data collected on the remaining subjects). The models will include factors for treatment, region, and class of oral antidepressant (SNRI or SSRI) and baseline MADRS total score as a covariate.

### Weighted Combination Test

To account for sample size reassessment, the weighted combination test will be used for the comparison of interest. The combination test will be defined by an approach described in Lehman et al<sup>8</sup>(also Cui et al<sup>3</sup>.) which defines the test statistics as a weighted sum of the Stage 1 (before the interim analysis) and Stage 2 (after the interim analysis) test statistics. The two stages will be weighted equally in the combination test.

$$Z_C = \sqrt{0.5} \times Z_1 + \sqrt{0.5} \times Z_2$$

where  $Z_1 = \Phi^{-1}(1-p_1)$  and  $Z_2 = \Phi^{-1}(1-p_2)$  denote the z-values corresponding to the one-sided stage-wise p-values  $p_1$  and  $p_2$ , respectively, for the hypothesis of interest based on the MMRM or ANCOVA analysis of Stage 1 and Stage 2 data separately. The null hypothesis will be rejected for the large positive values of  $Z_C$ .

The weighted combination test will be provided for both the MMRM comparisons for all dossiers with the exception of the EU and for the ANCOVA comparisons for the EU dossier.

The median-unbiased estimate will be used for point estimation for the treatment difference from placebo, refer to Brannath et al<sup>1</sup>. For example, let  $\hat{\delta}_1$  and  $\hat{\delta}_2$  denote the estimated treatment effects (difference from placebo) at Stage 1 and Stage 2, and  $SE(\hat{\delta}_1)$  and  $SE(\hat{\delta}_2)$  are the standard errors of the estimates obtained from Stage 1 and Stage 2, respectively. Using the pre-specified combination test weights  $w_1 = \sqrt{0.5}$  and  $w_2 = \sqrt{0.5}$  the median-unbiased estimate of the treatment effect  $\delta$  is given by

$$\hat{\delta} = \frac{\frac{w_1}{SE(\hat{\delta}_1)} \hat{\delta}_1 + \frac{w_2}{SE(\hat{\delta}_2)} \hat{\delta}_2}{\frac{w_1}{SE(\hat{\delta}_1)} + \frac{w_2}{SE(\hat{\delta}_2)}}$$

The two-sided flexible confidence interval at level  $(1-2\alpha)$  will be symmetric about the median-unbiased point estimator. In order to have a proper coverage probability in the design with SSR, the flexible confidence interval will be based on the duality of confidence intervals and hypothesis tests<sup>1</sup>. That is, for the parameter  $\delta$  of interest, the one-sided flexible interval with a coverage probability  $1 - \alpha$  is defined by the set  $\{\delta : Z_C < z_\alpha\}$ . For analysis of the primary endpoint, the two-sided confidence interval for the treatment effect at level  $(1-2\alpha)$  will be defined by

$$\left[ \hat{\delta} - \frac{z_\alpha}{\frac{w_1}{SE(\hat{\delta}_1)} + \frac{w_2}{SE(\hat{\delta}_2)}}, \hat{\delta} + \frac{z_\alpha}{\frac{w_1}{SE(\hat{\delta}_1)} + \frac{w_2}{SE(\hat{\delta}_2)}} \right]$$

where  $z_\alpha$  denotes the  $100(1-\alpha)$ -th percentile of the standard normal distribution.

The median-unbiased estimate of the treatment effect and the corresponding confidence interval will be provided for both the MMRM comparisons for all dossiers with the exception of the EU and for the ANCOVA comparisons for the EU dossier.

For the MADRS total score, descriptive statistics of actual values and changes from baseline will be provided for observed case and LOCF data during the double-blind phase and for observed case data during the follow-up phase.

In addition, to explore the course of treatment effect over time, a standard ANCOVA (i.e. no adjustments for the sample size re-estimation) on both LOCF and observed case data will be performed for each scheduled assessment time point with factors for treatment, region, class of oral antidepressant (SNRI or SSRI) and baseline MADRS total score as a covariate. In addition, contrasts at each scheduled assessment time point based on the MMRM analysis will be provided.

Means ( $\pm$ standard error [SE]), mean changes ( $\pm$ SE) from baseline, and least square mean changes ( $\pm$ SE) from baseline will be presented graphically for the double-blind induction phase for the observed cases and separately for the LOCF evaluations.

### Model Diagnostics

The normality and equal variance assumptions underlying both the primary MMRM and ANCOVA models will be assessed graphically for the MADRS total score at end point. Residuals from the primary models will be plotted against the predicted values and a QQ plot of the residuals versus the expected quantiles of the standard normal distribution will be presented. If either the equal variance or the normality assumption appears to be grossly violated, other methods including an ANCOVA on ranks model or an appropriate transformation of the primary endpoint might be considered.

### Missing Data Sensitivity Analysis

The following table (Table 5) shows the assumptions of each considered analysis for the primary efficacy endpoint, all applied to the same full analysis set defined in Section 2.5.2:

**Table 5: Analysis Types and Assumptions**

Analysis Type	Analysis Method	Assumption
Primary Analysis (non-EU)	MMRM	Missing at Random – MAR
Primary Analysis (EU)	ANCOVA model using change from baseline to Day 28 based on LOCF	Efficacy scores at time of discontinuation (DC) from the DB induction phase remain constant up to Day 28
Sensitivity Analysis	Delta worsening adjustments applied to standard multiple imputation regression	Efficacy scores worsen after study discontinuation

To evaluate the robustness of the MMRM analysis to increasing deviations from the MAR assumption, a delta adjustment multiple imputation method will be used for sensitivity analysis. This type of method is regarded to be an informative sensitivity analysis in clinical trials (2010 National Research Council report on missing data<sup>10</sup> and T Permutt<sup>11</sup>).

This method will employ the following 3 steps:

### Step 1 – Multiple Imputation (MI)

*Note: Most missing data will be a result of subjects dropping out of the study and having observations at every scheduled visit up to the point they dropped but no observations thereafter. This is a monotone missing data pattern. Some subjects may have ‘intermediate missing’ data between non-missing observations. This is a non-monotone missing data pattern.*

If there are subjects with a non-monotone missing data pattern, datasets with only monotone missing data patterns will be created first by imputing the intermediate missing values using the MCMC (Markov Chain Monte Carlo) method. This will be done using SAS PROC MI and the MCMC statement with the following specifications:

```
PROC MI DATA=INPUT NIMPUTE=500 SEED=234 OUT=IN_MCMC;
```

```
VAR ...; (treatment group, region, class of antidepressant (SNRI or SSRI), and the preceding non-missing values in the order of clinical visits: baseline, Day 2, Day 8, Day 15, Day 22 and Day 28)
```

```
MCMC CHAIN=SINGLE NBITER=200 NITER=100 IMPUTE=MONOTONE;
```

```
RUN;
```

*Note: Graphical diagnostic tools available in the MCMC statement (TIMEPLOT(WLF) and ACFPLOT(WLF)) will be used to assess if convergence has not been achieved.*

If all subjects have a monotone missing data pattern (either directly from the study or created by the previous step), the MAR-based multiple imputation with the regression option will be used to impute missing values. This analysis will be performed using SAS PROC MI with the MONOTONE statement and the REGRESSION option with the following specifications:

```
PROC MI DATA=IN_MCMC NIMPUTE=1 (see note) SEED=234 OUT=OUTPUT;
```

```
VAR ...; (treatment group, region, class of antidepressant (SNRI or SSRI), and the preceding non-missing values in the order of clinical visits: baseline, Day 2, Day 8, Day 15, Day 22 and Day 28)
```

```
CLASS...; (treatment group, region, class of antidepressant (SNRI or SSRI))
```

```
MONOTONE REGRESSION;
```

```
RUN;
```

*Note: NIMPUTE=500 if MCMC was NOT applied at the previous step.*

## Step 2 – Delta Adjustments

The imputed values for subjects who discontinued will be adjusted by adding  $\delta_P$  to the imputed values for subjects randomized to placebo and adding  $\delta_A$  to the imputed values for subjects randomized to an active esketamine arm. Delta-adjusted fully imputed datasets will be generated for different combinations of  $\delta_P$  and  $\delta_A$  values as defined below:

- $\delta_P = 0$  and  $\delta_A = 0$  to  $\Delta_1^*$  in increments of 1 (active-only adjustment analysis)
- $\delta_P = \delta_A/2$  and  $\delta_A = 0$  to  $\Delta_2^*$  in increments of 1 (analysis with a control adjustment that is half of the active adjustment)
- $\delta_P = \delta_A = 0$  to  $\Delta_3^*$  in increments of 1 (all arms adjustment analysis).

Adding positive values results in higher (worse) scores.  $\Delta_1^*$ ,  $\Delta_2^*$  and  $\Delta_3^*$  represent the adjustments leading to the ‘tipping point’, so the smallest delta adjustments values at which conclusions change from *favorable to drug* (i.e. statistically significant: one-sided p-value  $\geq 0.025$ , with the p-value associated to the weighted combination test described above; see note below on delta adjustment and combination test) to *unfavorable* (acceptance of the null hypothesis of no treatment difference).

## Step 3 – Analysis and Pooling

For each  $(\delta_P, \delta_A)$  combination:

- Same MMRM as described for the primary efficacy analysis will be performed for each set of the adjusted fully imputed datasets;
- Multiple imputation combining rules in PROC MIANALYZE will be applied to the MMRM results from the imputed datasets to produce final inferences.

Between-group comparisons to placebo at Day 28 (e.g. p-values, point estimates for treatment difference) will be displayed graphically for each considered  $(\delta_P, \delta_A)$  combination, up to the ‘tipping point’ adjustment.

*Note - Delta Adjustment and the Combination Test: Due to the interim analysis, a weighted combination test will be used for each comparison of interest as presented in the previous section. Each  $(\delta_P, \delta_A)$  delta adjustment as described above in Steps 1-3 will be applied first to each of the 2 stages and then the combination test will be applied to produce the p-value corresponding to this delta adjustment.*

The delta adjustment MI method as described above will be applied to all subjects who discontinued from the DB induction phase and have missing efficacy scores in this phase. In addition, another version of this method will be applied. In this version, the delta adjustments from Step 2 will be applied to all subjects who discontinued, except those with a discontinuation reason that was not considered related to the study drug, including Lost to Follow-Up, Withdrawal by Subject or Other. A process is in place to obtain clarification on reasons of Withdrawal by Subject and Other to confirm they are not related to the study drug.

## Subgroup Analysis

Forest plots will be provided displaying analysis results for each subgroup listed in Section 2.6. The point estimate of the treatment difference and its 95% confidence interval for each subgroup will be based on an MMRM analysis for the primary endpoint using the appropriate contrast. The model will include baseline MADRS total score as a covariate, and treatment, region, class of antidepressant (SNRI or SSRI), day, subgroup, day-by-treatment interaction, treatment-by-subgroup and day-by-treatment-by-subgroup interaction as fixed effects. A similar forest plot will be provided using an ANCOVA analysis. The model will include factors for treatment, region, class of oral antidepressant (SNRI or SSRI), subgroup, treatment-by-subgroup and baseline MADRS total score as a covariate. The terms in the models (both ANCOVA and MMRM) will be adjusted for the subgroup of country and baseline MADRS total score ( $\leq$ / $>$ median). Region will not be included as a term when country is included in the model and baseline MADRS total score (as a continuous covariate) will not be included in the model when the dichotomized baseline MADRS total score is included in the model.

### 5.3. Other Efficacy Variables

#### 5.3.1. Responders

##### 5.3.1.1. Definition

The percentage change from baseline at Day X is calculated as  $100 * (\text{MADRS total score at Day X} - \text{Baseline MADRS total score}) / (\text{Baseline MADRS total score})$ . Negative percent changes in MADRS total score indicate improvement (e.g., percent change  $\leq -50\%$  indicates improvement  $\geq 50\%$ ).

A subject is defined a responder (yes=1 and no=0) at a given time point if the percent improvement in MADRS total score is  $\geq 50\%$ .

##### 5.3.1.2. Analysis Methods

The proportion of subjects who achieve a response will be summarized at each time point during the double-blind induction phase for both observed case and LOCF data. The proportion of subjects who achieve a response will be also be summarized by class of antidepressant study medication (SNRI or SSRI) at each time point.

The cumulative response rate, defined as the percentage of subjects experiencing at least a given value of percent reduction from baseline to Day 28 in MADRS total score will be presented graphically.

Response rates at Week 2 of the follow-up phase will also be summarized.

The cumulative distribution function of the time to sustained response will be estimated by the Kaplan-Meier method. Time to sustained response will be summarized (number of sustained responders, number of censored subjects, median, 25<sup>th</sup> and 75<sup>th</sup> percentile, if estimable) by treatment group. Sustained response is defined as the first occurrence of response that is maintained through the Day 28 assessment. Subjects who discontinue early are not considered to have sustained response.

### **5.3.2. Remitters**

#### **5.3.2.1. Definition**

Subjects who have a MADRS total score of  $\leq 12$  will be considered remitters.

#### **5.3.2.2. Analysis Methods**

The proportion of subjects who achieve remission will be summarized at each time point during the double-blind induction phase for both observed case and LOCF data. The proportion of subjects who achieve remission will also be summarized by class of antidepressant study medication (SNRI or SSRI) at each time point.

The number and percentage of subjects meeting criteria for remission will also be provided at Week 2 of the follow-up phase.

The cumulative distribution function of the time to sustained remission will be estimated by the Kaplan-Meier method. Time to sustained remission will be summarized (number of sustained responders, number of censored subjects, median, 25<sup>th</sup> and 75<sup>th</sup> percentile, if estimable) by treatment group. Sustained remission is defined as the first occurrence of remission that is maintained through the Day 28 assessment. Subjects who discontinue early are not considered to have sustained remission.

### **5.3.3. CGI-S**

#### **5.3.3.1. Definition**

The Clinical Global Impression of Severity (CGI-S) provides an overall clinician-determined summary measure of the severity of the subject's illness that takes into account all available information, including knowledge of the subject's history, psychosocial circumstances, symptoms, behavior, and the impact of the symptoms on the subject's ability to function<sup>6</sup>. 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.

#### **5.3.3.2. Analysis Methods**

Descriptive statistics of actual values and changes from baseline by treatment group will be provided for observed case and LOCF data. Frequency distributions will be provided at each assessment time point during the double-blind and follow-up phases.

The ranks of the change from baseline for the CGI-S score in the double-blind induction phase will be analyzed at each time point based on observed case and LOCF data using an ANCOVA model, with treatment, region and class of antidepressant (SNRI or SSRI) as factors, and unranked baseline score as the covariate.

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

#### 5.3.4.1. Definition

The EQ-5D-5L (EuroQol Group - 5 Dimension - 5 Level<sup>4,5</sup>) 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>
- (iii) The Canadian value set will be used to get the HSI values for all the countries participating in the study

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

#### 5.3.4.2. Analysis Methods

Descriptive statistics of actual values and changes from baseline by treatment group will be provided for the weighted EQ-5D health status index, the EQ-VAS, and the sum score at each time point for the double-blind and follow-up phases.

Individual dimension responses will also be summarized at each visit using a frequency distribution by treatment group for the double-blind and follow-up phases.

### **5.3.5. Sheehan Disability Scale (SDS)**

#### **5.3.5.1. 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. SDS was only collected for subjects enrolled prior to implementation of Amendment 3 of the protocol.

#### **5.3.5.2. Analysis Methods**

Descriptive statistics of actual values and changes from baseline by treatment group will be provided at each time point for the double-blind and follow-up phases. The total score as well as the individual item scores will be summarized separately. In addition, the proportion of subjects who achieve response and remission will be summarized at each time point during the double-blind induction phase.

### **5.3.6. PHQ-9**

#### **5.3.6.1. Definition**

The PHQ-9 is a 9-item, self-report scale assessing depressive symptoms<sup>13</sup>. 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). PHQ-9 was only collected for subjects enrolled prior to implementation of Amendment 3 of the protocol.

#### **5.3.6.2. Analysis Methods**

For the PHQ-9 total score, descriptive statistics of actual values and changes from baseline will be provided at each time point for the double-blind and follow-up phases. Frequency distributions by severity will be provided at each assessment time point during the double-blind and follow-up phases.

## 6. SAFETY

All safety summaries for the double-blind induction phase will be based on the Safety analysis set. Safety summaries for the follow-up phase will be based on the Follow-up analysis set.

### 6.1. Adverse Events

The Medical Dictionary for Regulatory Activities (MedDRA) (version 18.1 or above) will be used to classify adverse events (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.

Treatment-emergent adverse events (TEAEs) are defined as follows for each study phase:

- TEAEs in Double-Blind induction phase:
  - a. If AE onset time is not missing:
    - i. If intranasal study medication was started on the same day as the oral antidepressant: DB start date/time  $\leq$  AE onset date and time  $\leq$  DB end date
    - ii. If intranasal study medication was not taken or was started after oral antidepressant: DB start date  $\leq$  AE onset date  $\leq$  DB end date
  - b. If AE onset time is missing: DB start date  $\leq$  AE onset date  $\leq$  DB end date
- AEs in Follow-up phase: F/U start date  $\leq$  AE onset date  $\leq$  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 2.7.

In addition, TEAEs will be summarized by severity and relationship to study medication using the preferred term. For the summaries of TEAEs 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 AE reported during the analysis phase by the subject.

### 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 (Dependence, Drug abuse, Drug abuser, Drug dependence, Drug detoxification, Drug diversion, Drug rehabilitation, Drug tolerance increased, Drug withdrawal convulsions, Drug withdrawal headache, Drug withdrawal syndrome, Rebound effect, Substance abuse, Substance abuser, Substance dependence, Substance use, Substance-induced mood disorder, 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, Minor cognitive motor 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 the double-blind phase will be provided.

## 6.2. Clinical Laboratory Tests

Descriptive statistics (N, mean, median, minimum, and maximum) for values and changes from baseline will be provided for clinical laboratory tests (hematology, chemistry and urinalysis) at each scheduled time point in the double-blind induction and follow-up phases. Baseline laboratory result is defined as the last result collected prior to Day 1 predose. This will be used to calculate change for the double-blind induction phase summary. Changes from baseline for the follow-up phase will also be calculated using the double-blind baseline.

Clinical laboratory tests that meet the criteria for markedly abnormal will be listed by subject for the double-blind induction phase. The incidence of treatment-emergent markedly abnormal (TEMA) laboratory values that occurred at any time during the double-blind induction 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 1](#). The identification of TEMA laboratory values is based on the postbaseline value being out of range while the baseline value is either missing or within the range given in [Attachment 1](#). 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  $>3 \times$  upper normal limit (ULN) will be presented for each study phase. Additionally, incidence of hepatic toxicity (Hy's Law<sup>14</sup>) defined as ALT values  $>3 \times$ ULN AND total bilirubin values  $>2 \times$ ULN will be presented for the double-blind

induction phase. Similar to the markedly abnormal analysis, only subjects with baseline ALT values  $\leq 3 \times \text{ULN}$  (AND baseline total bilirubin values  $\leq 2 \times \text{ULN}$  for hepatic toxicity) (or if baseline 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 the double-blind induction phase will be presented for temperature, systolic blood pressure, diastolic blood pressure, pulse rate, respiratory rate, oxygen saturation, weight, and BMI. In addition, descriptive statistics of pulse rate and blood pressure (systolic and diastolic) values, changes and percent changes from predose will be provided for each dosing day. These summaries will also be provided by hypertension status (history of hypertension recorded in medical history or concomitant use of antihypertensive medications, Yes/No). Frequency distributions of maximum percent change increase from predose and time of maximum percent change increase will also be presented. Note that if the maximum value within a phase occurs at multiple time points, the earliest time point is selected.

The proportion of subjects who have a treatment-emergent abnormality, as defined in Table 6 below, during the double-blind induction phase will be presented. Both the double-blind baseline and the predose assessment will be used to determine abnormal values. A listing of subjects meeting any of the criteria will also be provided for the double-blind induction phase.

**Table 6: Treatment-Emergent Abnormality Categories for Vital Signs**

Vital Parameter	Post-baseline value outside of normal limit if:	
	Abnormally low	Abnormally high
Pulse (bpm)	A decrease from baseline of $\geq 15$ to a value $\leq 50$	An increase from baseline of $\geq 15$ to a value $\geq 100$
Systolic BP (mmHg)	A decrease from baseline of $\geq 20$ to a value $\leq 90$	An increase from baseline of $\geq 20$ to a value $\geq 180$
Diastolic BP (mmHg)	A decrease from baseline of $\geq 15$ to a value $\leq 50$	An increase from baseline of $\geq 15$ to a value $\geq 105$

BP = blood pressure

The proportion of subjects who experienced acute hypertension (systolic BP  $\geq 180$  or diastolic BP  $\geq 100$ ) at any time during the double-blind induction phase will be summarized by treatment group and hypertension status.

Mean (+/-SE) values for systolic BP, diastolic BP and heart rate will be presented graphically over the double-blind induction phase by treatment group and hypertension status. In addition, for subjects with hypertension who receive antihypertensive medication, the same graphs will be summarized by medication type (beta-blockers, all other agents).

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 the first dose of study medication (either intranasal or oral AD). The maximum post-baseline value during the double-blind induction 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 the double-blind induction and follow-up phases.

The frequency of treatment-emergent abnormalities will be tabulated and presented for the double-blind induction phase. The identification of treatment-emergent abnormal ECG values is based on the post-baseline value (a value occurring after the first study drug administration) being out of range while the baseline value is either missing or within the limits given in Table 7. 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). The average predose value will be used as baseline for the double-blind induction summary. Abnormal ranges for the HR, PR, QRS and QT intervals are given in Table 7.

**Table 7: Limits for HR, PR, QRS and QT Interval Abnormality**

ECG parameter	Abnormally Low	Abnormally High
HR (bpm)	$\leq 50$	$\geq 100$
PR interval (msec)	--	$\geq 210$
QRS interval (msec)	$\leq 50$	$\geq 120$
QT interval (msec)	$\leq 200$	$\geq 500$

Based on the maximum QTc value for each subject during the double-blind induction phase (separate for each QTc correction) the incidence of abnormal QTc values and changes from average predose will be summarized by treatment group. Criteria for abnormal corrected QT intervals and changes from baseline are given in Table 8 and are derived from the ICH E14 Guidance<sup>7</sup> (the same criteria apply to all QT corrections).

**Table 8: Criteria for Abnormal QTc Values and Changes From Baseline**

Parameter	Classification	Criteria
Clinically Significant QTc Value	No	≤500
	Yes	>500
QTc change from baseline <sup>a</sup>	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 Day 1 predose value for the double-blind induction phase.

The proportion of subjects with treatment emergent abnormalities will be presented for the double-blind induction phase. 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. Any treatment-emergent change or worsening from baseline examination will be recorded as an AE.

Changes in findings from the double-blind baseline for each examination (including the upper respiratory tract/throat) will be listed for the double-blind and follow-up phases by treatment group.

### 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 for each of the items at each time point during the double-blind induction phase. 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 time point during the double-blind induction 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 C-SSRS (Columbia Suicide Severity Rating Scale) 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<sup>12</sup>. 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 subjects who have at least 1 post-baseline C-SSRS measurement and a pre-treatment C-SSRS assessment (assessment at Baseline visit).

A frequency distribution at each scheduled time point by treatment will be provided. Shifts from the baseline visit to the most severe/maximum score during the double-blind induction and follow-up phases 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 visit to the maximum category during the double-blind induction and follow-up phases will be summarized by treatment.

### 6.6.2. Clinician Administered Dissociative States Scale (CADSS)

The CADSS (Clinician Administered Dissociative States Scale) is an instrument for the measurement of present-state dissociative symptoms<sup>2</sup>, 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 study will be summarized. Mean (SD) CADSS values will be presented graphically for each 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 postdose during the double-blind induction phase.

Descriptive statistics (N, median, minimum, and maximum) of the total scores at each time point, change from the predose time point within each visit, and the proportion of subjects with an increase in BPRS+ from the predose value at any time during the study will be provided. The

proportion of subjects with a total score of 3 or more at any time during the study will also be provided. Mean (SD) BPRS+ values will be presented graphically for each 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  $\leq 3$  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 postdose).
- 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. For subjects with a score of  $\leq 3$ , 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 values will be presented graphically for each 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 the double-blind induction phase.

**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 (in the event a subject does not continue into ESKETINTRD3004).

The proportion of subjects with withdrawal symptoms at the end of double-blind induction therapy or during the follow-up phase will be presented by treatment. In addition, symptoms at follow-up will be compared to the end of therapy visit 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 (HVLTR)**

The effect of intranasal esketamine on cognition over the 4-week double-blind induction phase will be assessed using the computerized cognitive battery and the HVLTR.

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 HVLTR 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 HVLTR will be assessed at Day 1 predose and Day 28/EW of the double-blind induction phase. It will also be assessed again at 2 weeks post-treatment during the follow-up phase.

See [Attachment 2](#) for details of the 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 screening Week 2 (SC) will be summarized at each scheduled time point during the double-blind induction phase. UPSIT total score (%) will be defined as  $[(\# \text{ correct responses})/(\# \text{ completed responses})]*100$ .

**7. PHARMACOKINETICS/PHARMACODYNAMICS**

Details of the pharmacokinetic and pharmacodynamic analysis are provided in a separate document.

**REFERENCES**

1. Brannath W, Konig F, Bauer P. Estimation in flexible two stage designs. *Statistics in Medicine* 2006; 25:3366-3381.
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. Cui L, Hung HMJ, Wang S-J. Modification of sample size in group sequential clinical trials. *Biometrics*. 1999 Sept; 55:853-857.
4. EuroQol Group. About EQ-5D. <http://www.euroqol.org/about-eq-5d.html> (accessed 18 June 2014).
5. 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](http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/UserGuide_EQ-5D-5L_v2.0_October_2013.pdf) (accessed 6 August 2014)
6. Guy W. ECDEU Assessment Manual for Psychopharmacology. Rockville, MD: U.S. Department of Health, Education, and Welfare; 1976.
7. International Conference on Harmonization (ICH) Tripartite Guideline E14. London : s.n., 2005. pp. The clinical evaluation of QT/QTc interval prolongation and proarrhythmic for non-antiarrhythmic drugs.
8. Lehman W, Wassmer G. Adaptive sample size calculations in group sequential trials. *Biometrics* 1999;55:1286–1290.
9. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry*. 1979;134:382-389.
10. National Research Council, Panel on Handling Missing Data in Clinical Trials [2010]. The prevention and treatment of missing data in clinical trials. Available: <http://www.nap.edu/catalog/12955.html>
11. Permutt, T. Sensitivity analysis for missing data in regulatory submissions. *Stat Med*. 2015; Nov 15 (Epub ahead of print)
12. 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.
13. 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.
14. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER). [2009] Guidance for industry drug induced liver injury: premarketing clinical evaluation. Available: <http://www.fda.gov/downloads/Drugs/.../Guidances/UCM174090.pdf>.

**ATTACHMENTS****Attachment 1: Criteria of Markedly Abnormal Laboratory Values**

Laboratory Parameter	Markedly Abnormal Limits	
	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 [ $\mu$ mol/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 [ $\mu$ mol/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 [ $\times 10^9$ /L]	100	600
Erythrocytes (RBC) [ $\times 10^{12}$ /L] -- female	3.0	5.5
-- male	3.0	6.4
Leukocytes(WBC) [ $\times 10^9$ /L]	2.5	15.0
Hy's Law criteria:		
Alanine transaminase (SGPT) [U/L]		>3X ULN
AND		
Bilirubin, total [ $\mu$ mol/L]		>2X ULN

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

**Attachment 2: 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:** V2  
**Date:** 30-Mar-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).



**2 SIGNATURE PAGE FOR SAP APPROVAL**

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

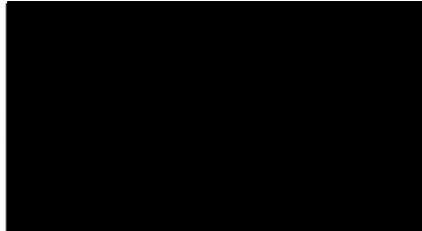
**Approved by:**

**Name (print):**

**Position:**

**Signature:**

**Date (ddmmmyyyy):**



30MAR2016

**Approved by:**

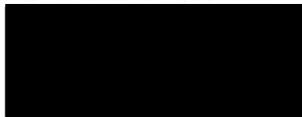
**Name (print):**

Jaskaran Singh

**Position:**

**Signature:**

**Date (ddmmmyyyy):**



Digitally signed by 3D900151-  
EFFD-4D56-9844-D29916A7A91E  
DN: cn=3D900151-  
EFFD-4D56-9844-  
D29916A7A91E, o=Johnson &  
Johnson  
Date: 2016.03.31 09:05:21 -05'00'



**3 TABLE OF CONTENTS**

1 Note ..... 2

2 Signature Page for SAP Approval ..... 3

3 Table of Contents ..... 4

4 Abbreviations ..... 5

5 Introduction ..... 6

6 Visit Schedule ..... 6

7 Study Objectives relevant to cogstate analysis ..... 6

8 Study Design ..... 6

9 Sample Sizes ..... 7

10       Analysis SetS ..... 7

11       Analysis VARIABLES ..... 7

11.1 Computerized cognitive battery and Hopkins Verbal Learning Test-Revised (HVLTR) Endpoints ..... 7

11.2 The computerized tests from the Cogstate Battery ..... 7

    11.2.1 Summary of the Cogstate Battery ..... 7

    11.2.2 The outcome measures for the Cogstate battery ..... 9

    11.2.3 Data Quality Assurance ..... 10

    11.2.4 Test Completion Criteria ..... 10

11.3 The HVLTR ..... 11

12       STATISTICAL METHODOLOGY ..... 11

12.1 Analysis Overview ..... 11

    12.1.1 Analysis of Cogstate Endpoints and HVLTR ..... 11

    12.1.2 Baseline Definition ..... 11

    12.1.3 Handling of Missing Data and data transformation ..... 11

*Missing Data* ..... 11

*Data Transformation* ..... 12

12.2 Unblinding Procedure ..... 12

13       References ..... 13



#### 4 ABBREVIATIONS

<b>Abbreviation</b>	<b>Description</b>
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 HVLTR 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 HVLTR*

Period	Visit Number	Study Day	Study Week by phase	Computerized test battery & HVLTR 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 <sup>b</sup>	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

<sup>b</sup> 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 211 subjects 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 (HVLTR) Endpoints

The endpoints for computerized cognitive battery (DET, IDN, ONB, OCL, and GML) and HVLTR 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**

Version: V2  
Date: 30-Mar-2016  
Protocol: Approved Date 17 Feb 2016

8



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



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

### 11.3 The HVLТ-R

*The HVLТ-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 HVLТ-R is a well-validated and widely used measure of verbal episodic memory. The tests will be administered in the following order: HVLТ-R, computerized cognitive test battery, and HVLТ-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 HVLТ-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 HVLТ-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:

$$\text{LMN} = \text{mean}(\log_{10} \text{ transformed reaction times for correct responses}) \quad (1)$$

#### 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:

$$\text{ACC} = \text{arcsine}(\sqrt{\text{proportion of correct responses}}) \quad (2)$$

### 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.