NCT03249376

Study ID: ITI-007-404

Title: A Phase 3, Randomized, Double-blind, Placebo-Controlled, Multi-Center Study to Assess the Efficacy and Safety of Lumateperone Monotherapy in the Treatment of Patients With Major Depressive Episodes Associated With Bipolar I or II Disorder (Bipolar Depression) Conducted Globally

Statistical Analysis Plan Date: 12Jun2019

STATISTICAL ANALYSIS PLAN

ITI-007-404

A PHASE 3, RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED, MULTI-CENTER STUDY TO ASSESS THE EFFICACY AND SAFETY OF LUMATEPERONE MONOTHERAPY IN THE TREATMENT OF SUBJECTS WITH MAJOR DEPRESSIVE EPISODES ASSOCIATED WITH BIPOLAR I OR BIPOLAR II DISORDER (BIPOLAR DEPRESSION) CONDUCTED GLOBALLY

AUTHOR: [redacted]

VERSION NUMBER AND DATE: FINAL 1.0, 12JUN2019

STATISTICAL ANALYSIS PLAN SIGNATURE PAGE

Statistical Analysis Plan V1.0 (Dated 12JUN2019) for Protocol ITI-007-404.

[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		

Upon review of this document, the undersigned approves this version of the Statistical Analysis Plan, authorizing that the content is acceptable for the reporting of this study.

[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		
[redacted]	[redacted]		

MODIFICATION HISTORY

Unique Identifier for this Version	Date of the Document Version	Author	Significant Changes from Previous Authorized Version
Draft 1.0	27NOV2018	[redacted]	N/A
Final 1.0	12JUN2019	[redacted]	Updates following sponsor review and updates for consistency with protocol and following dry run comments

TABLE OF CONTENTS

[redacted]

[redacted]

1. LIST OF ABBREVIATIONS

AE	Adverse Event
AIMS	Abnormal Involuntary Movement Scale
ALT	Alanine aminotransferase
ANCOVA	Analysis of Covariance
AR(1)	Autoregressive(1)
ARH(1)	Heterogeneous autoregressive(1)
AST	Aspartate aminotransferase
BARS	Barnes Akathisia Rating Scale
BLQ	Below limit of quantification
BMI	Body mass index
C-SSRS	Columbia – Suicide Severity Rating Scale
CGI-BP-S	Clinical Global Impression Scale – Bipolar version, Severity
CI	Confidence interval
CPK	Creatine phosphokinase
CS	Compound symmetry
CSH	Heterogeneous compound symmetry
DNA	Deoxyribonucleic acid
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
ENR	All Subjects Enrolled
FA0(q)	No Diagonal Factor Analytic
GGT	Gamma-glutamyl transferase
HbA1c	Glycated Hemaglobin A1c
HDL	High-density lipoprotein
HIV	Human Immunodeficiency Virus
HR	Heart rate
IEC	Independent Ethics Committee
IRB	Institutional Review Board
ITCI	Intra-Cellular Therapies, Inc. (Sponsor)
ITT	Intent-to-treat
LDL	Low-density lipoprotein
LOCF	Last observation carried forward
LOE	Lack of efficacy
LSM	Least-squares Mean
MADRS	Montgomery-Åsberg Depression Rating Scale
MAR	Missing at random
MCMC	Monte Carlo Markov Chain
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple imputation
MMRM	Mixed-Effects Model for Repeated Measures
MNAR	Missing not at random
PK	Pharmacokinetic(s)
PP	Per protocol
PR	PR interval of the electrocardiogram; time duration between the P and

	R waves
PT	Preferred Term
Q-LES-Q-SF	Quality of Life Enjoyment and Satisfaction Questionnaire – Short Form
QRS	QRS interval of the electrocardiogram; duration of the QRS complex
QT	QT interval of ECG, duration between the Q and T waves
QTc	QT interval of ECG corrected for heart rate
QTcB	QT interval of ECG corrected for heart rate using Bazett's formula
QTcF	QT interval of ECG corrected for heart rate using Fridericia's formula
RNA	Ribonucleic acid
RND	All Subjects Randomized
RR	Time duration between two consecutive R waves of the electrocardiogram
SAE	Serious Adverse Event
SAF	Safety Analysis
SAP	Statistical Analysis Plan
SAS	Simpson-Angus Scale
SAS®	Statistical Analysis Software
SENS	Sensitivity
SD	Standard deviation
SMQ	Standardized MedDRA Queries
SOC	System Organ Class
SUPPP	Supportive Per Protocol
TEAE	Treatment-emergent adverse event
TLFs	Tables, Listings and Figures
TOEP	Toeplitz structure
TOEPH	Heterogeneous Toeplitz structure
ULQ	Upper limit of quantification
WBC	White blood cells
WHO	World Health Organisation
YMRS	Young Mania Rating Scale

2. INTRODUCTION

This document describes the rules and conventions to be used in the presentation and analysis of efficacy, safety, and pharmacokinetic (PK) data for Protocol ITI-007-404. It describes the data to be summarized and analyzed, including specifics of the statistical analyses to be performed.

This statistical analysis plan (SAP) is based on Protocol Version 1.3, dated November 5, 2018.

3. STUDY OBJECTIVES

3.1. PRIMARY OBJECTIVE

The primary objective of this study is to compare the efficacy of lumateperone administered orally once daily to that of placebo as measured by mean change from baseline to Day 43 in the total score on the Montgomery-Åsberg Depression Rating Scale (MADRS) in subjects with Bipolar Depression.

3.2. SECONDARY OBJECTIVES

3.2.1. KEY SECONDARY OBJECTIVE

The key secondary objective of this study is to compare the efficacy of lumateperone administered orally once daily to that of placebo as measured by mean change from baseline to Day 43 in the total score on the Clinical Global Impression Scale, Bipolar version – Severity (CGI-BP-S) in subjects with Bipolar Depression.

[redacted]

[redacted]

3.2.3. SAFETY SECONDARY OBJECTIVES

The safety secondary objectives of this study are to determine the safety and tolerability of lumateperone, administered orally once daily for 6 weeks in subjects with Bipolar Depression, compared to placebo. Safety and tolerability will be assessed in relation to:

- Incidence of adverse events (AEs);
- Mean change from baseline in the Young Mania Rating Scale (YMRS);
- Mean change from baseline in the Columbia Suicide Severity Rating Scale (C-SSRS);
- Mean change from baseline in the Abnormal Involuntary Movement Scale (AIMS);
- Mean change from baseline in the Barnes Akathisia Rating Scale (BARS);
- Mean change from baseline in the Simpson Angus Scale (SAS);
- Changes from baseline in clinical laboratory evaluations;
- Changes from baseline in electrocardiograms (ECGs);
- Changes from baseline in vital sign measurements;
- Physical and neurological examination findings.

[redacted]

[redacted]

4. STUDY DESIGN

4.1. GENERAL DESCRIPTION

Study participation will last approximately 10 weeks (up to 9 visits) and will include the following three periods: a Screening Period, lasting up to 2 weeks, unless the Sponsor or representative, in extenuating circumstances related to the subject or the caregiver, where applicable, authorizes a longer period; a 6-week double-blind On-Treatment Period of daily administration of the study medication; and a Safety Follow-up Period of approximately 2 weeks after the last dose of study medication for subjects who completed the planned 6 weeks of treatment. The timing of assessments during each study period are described in the Schedule of Events (Appendix A of the protocol and Appendix 1).

[redacted]

5. FINAL AND INTERIM ANALYSES

5.1. BLINDED SAMPLE SIZE RE-ESTIMATION

An unplanned blinded sample size re-estimation was conducted based on a blinded data review of the efficacy data during the study. The estimated pooled standard deviation (SD) was 8.82, which was larger than what was assumed at study planning. The sample size was therefore increased to a total of 350 randomized subjects, based on an assumed SD of 9.0, to ensure 90% statistical power.

5.2. PLANNED INTERIM ANALYSIS

None planned.

5.3. FINAL ANALYSIS

The final analyses detailed in this SAP will be performed by IQVIA Biostatistics following Intra-Cellular Therapies, Inc (ITCI) authorization of the SAP, ITCI authorization of Analysis Sets, final Database Lock and Unblinding of Treatment.

6. ANALYSIS SETS

Analysis of efficacy and safety endpoints will be performed based on the analysis sets defined in this section and as specified for each endpoint throughout this SAP. [redacted], the inclusion/exclusion of subjects from each analysis set will be determined prior to the final analysis (and approved by ITCI) based on blinded data review.

6.1. ALL SUBJECTS ENROLLED [ENR] SET

The All Subjects Enrolled (ENR) Set will contain all subjects who signed the informed consent for this study.

6.2. ALL SUBJECTS RANDOMIZED [RND] SET

The All Subjects Randomized (RND) Set will contain all subjects who signed the informed consent and were randomized to study medication. Subjects will be classified according to randomized treatment for analyses and displays based on the RND Set.

6.3. INTENT-TO-TREAT [ITT] SET

The Intent-To-Treat (ITT) Set will contain all randomized subjects who received at least one dose of the study medication and who had a valid baseline (pre-dose) measurement and at least one valid post baseline measurement of MADRS total score. All analyses using the ITT Set will classify subjects according to the randomized treatment, regardless of the treatment received during the course of the study.

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

6.6. SAFETY ANALYSIS [SAF] SET

The Safety Analysis (SAF) Set will contain all randomized subjects who received at least one dose of study medication. For all analyses using the SAF set, subjects will be classified according to treatment actually received. In case of incorrect dosing, the treatment most often received will be taken as the actual treatment.

[redacted]

7. GENERAL CONSIDERATIONS

Relative Study Day will be calculated from the date of Day 1, which is the day of first treatment with study medication (Study Day 1), and will be used to show the start and/or stop days of treatments, study procedures and assessments, prior, concomitant and post-treatment medications, and adverse events.

- If the date of the treatment, procedure, or event is on or after Study Day 1 date then:

  Relative Study Day = (date of variable of interest – Day 1 date) + 1.

- If the date of the treatment, procedure, or event is prior to the Day 1 date then:

Relative Study Day = (date of variable of interest – Day 1 date).

In the situation where the date is partial or missing, Relative Study Day, and any corresponding durations will appear missing in the listings.

Analyses presented by visit or study day will be based on the scheduled visits as planned in the protocol. Visit windows for unscheduled visits or early discontinuation visits are defined in Table A:, which provides the mapping of relative day ranges to the scheduled target days and the study periods. If more than one assessment is available in the same 'Range of Relative Study Days' (window), the assessment closest to the Scheduled Target Day will be selected and assigned to the Scheduled Target Day. If two or more assessments are available in the same window and are equidistant from the Scheduled Target Day, the latest assessment will be selected. [redacted]

Listings will include scheduled, unscheduled, retest and early discontinuation data.

Table A: Mapping of Relative Day Ranges to Schedule Target Day

Study Phase	Study Period	Range of Relative Study Days	Scheduled Target Day	Schedued Study Week
Pre-Treatment	Screening	1-14* days before Day 1 date	Day -14	-2
Pre-Treatment	Baseline	Day 1 date (definition of baseline varies by assessment – see descriptions below)	Day 1	0
Study Treatment	On-Treatment	2 to 11 days relative to Day 1 date	Day 8	1
Study Treatment		12 to 18 days relative to Day 1 date	Day 15	2
Study Treatment		19 to 25 days relative to Day 1 date	Day 22	3
Study Treatment		26 to 32 days relative to Day 1 date	Day 29	4
		33 to 39 days relative to Day 1 date	Day 36	5
		≥40 days relative to Day 1 date and after the last dose of treatment and before the start of the Safety Follow-up Period (one day after actual Day 43 visit)	Day 43	6
Post-treatment	Safety Follow-up	> 43 days relative to Study Day 1 date for treatment completers (42 days on-treatment) and after Day 43 post-treatment assessments.	Day 57	8

* Not to exceed maximum allowed days

Baseline assessments are scheduled for Visit 2 on Relative Study Day 1, prior to first treatment of study medication. Unless otherwise specified, baseline is defined as the last non-missing measurement before the first dose of study medication.

Assessments will be considered baseline if the date of the last pre-treatment measurement is before the date of the first treatment or if the measurement was done on Study Day 1 and, according to the Study Schedule of Events, was supposed to be performed on Day 1, prior to treatment (including unscheduled measurements).

For parameters planned to be collected multiple times at the same time point (3-positional

blood pressure), the average of the last non-missing pre-treatment measurements will be considered as baseline.

[redacted]

Unless otherwise specified, the following calculations will be used for change from baseline and percent change from baseline:

Change from baseline visit will be calculated as:

- Value at Visit X – Baseline Value

Percent change from baseline will be calculated as:

- $\frac{\text{Value at Visit X} - \text{Baseline Value}}{\text{Baseline Value}} \times 100$

All investigative sites with fewer than 4 ITT subjects will be pooled as follows: The largest site with fewer than 4 ITT subjects will be pooled with the smallest site with fewer than 4 ITT subjects . If this results in a pooled site still having fewer than 4 ITT subjects, this site will be pooled together with the next smallest investigative site, if one exists; otherwise, no further pooling is needed. Sites with the same number of ITT subjects will be ordered in ascending order of their numerical site identification number. This will serve as a tie-breaker rule in case multiple sites have the same number of ITT subjects. If the primary efficacy analysis model, described in section 15.1.3 presents convergence issues due to the too small number of subjects per site, the same site pooling algorithm will be applied again, but this time pooling sites with fewer than 8 ITT subjects. Should the primary efficacy analysis model still present convergence issues, after testing the sequence of correlation structures listed in section 15.1.3, then including the site effect in the model will be reconsidered and may be dropped. These pooled investigative sites, as determined based on the primary efficacy response variable, will be used for any analysis that has investigative site as a fixed effect in the model. The actual investigative site numbers will be included in the by-subject listings.

The default significance level for statistical tests will be 5%; confidence intervals (CIs) will be 95% and all tests will be two-sided, unless otherwise specified in the description of the analyses. All analyses will be conducted using SAS® version 9.4 or higher.

7.1. MISSING DATA

The total and subscale scores of any assessment with more than one item, such as MADRS, CGI-BP-S, YMARS, CSSR-S, Q-LES-Q-SF, AIMS, BARS and SAS, will be derived from the individual items. Any individual missing item in any scale will not be imputed. If one or more items are missing at a visit, then the associated total score or subscale score will be set to missing.

The main objective of the analyses in this study is to evaluate treatment effect of 60 mg ITI-007 compared to placebo if the treatment is administered for the planned treatment duration. In order to evaluate this estimand in the presence of subjects that may discontinue treatment prematurely, the primary efficacy analysis will be performed based on the assumption of data being missing at random (MAR). This implies that subjects discontinuing from treatment early are considered to have unobserved values similar to the observed ones

in their respective treatment group for subjects who have similar history, i.e., the distribution of the unobserved future values for subjects who had discontinued treatment is the same as the distribution of the observed values for those subjects who completed treatment, conditional on the available data prior to discontinuation.

The Mixed-Effects Model for Repeated Measures (MMRM) method will be used for the analysis of the primary and key secondary efficacy endpoints. The MMRM approach does not impute missing data but is based on all subjects included in the analysis set using all available longitudinal data (either complete or partial). This approach is based on the MAR assumption as described above.

[redacted]

8. OUTPUT PRESENTATIONS

The Tables, Listings and Figures (TLFs) shells provided with this SAP describe the presentations for this study and therefore the format and content of the summary TLFs to be provided by IQVIA Biostatistics.

Continuous variables will be summarized using descriptive statistics (number of subjects [n], mean, standard deviation [SD], median, minimum, and maximum, unless otherwise stated). The minimum and maximum will be reported to the same number of decimal places as the raw data recorded in the database. The mean and median will be reported to one more decimal place than the raw data recorded in the database. The SD will be reported to two more decimal places than the raw data recorded in the database.

Categorical variables will be summarized using frequency counts and percentages. Unless otherwise stated, the calculation of percentage will be based on the number of subjects in the analysis set of interest.

P-values will be presented to three decimal places. P-values less than 0.001 will be presented as "<0.001".

Confidence intervals (CIs) will be presented to two more decimal places than the raw data.

Source data for summary tables and statistical analyses will be presented as subject data listings.

9. DISPOSITION AND WITHDRAWALS

Subject disposition and withdrawals will be presented by treatment group, when applicable, and overall for the ENR Set.

The number and percentage of subjects who were screened, screen failed, randomized, completed or discontinued the 6-week On-Treatment Period and completed or discontinued the study, and reasons for treatment discontinuation and study discontinuation will be presented. The reasons for study withdrawal and treatment discontinuation are listed in Table B:.

The reasons for treatment discontinuation will also be presented by treatment group and by time to treatment discontinuation, categorized based on the planned visits (≤ Day 8, > Day 8 -

≤ Day 15, > Day 15 - ≤ Day 22, > Day 22 - ≤ Day 29, > Day 29 - ≤ Day 36, > Day 36 - ≤ Day 43) for all subjects in the SAF Set. The reasons for study discontinuation and treatment discontinuation are listed in Table B.

A subject is defined to have completed treatment if the subject completed the 6-week On-Treatment period as recorded on the End of Treatment CRF page.

A subject is defined to have completed the study if the subject has completed the 6-week On-Treatment Period and the end of study follow-up assessments as recorded on the End of Study CRF page.

Table B: Reasons for Study Withdrawal and Study Medication Discontinuation Terminology

Case Report Form Terminology	Study	Study Medication
Screen failure	X	
Adverse event	X	X
Adverse event associated with worsening of bipolar depression*	X	X
Adverse event not associated with worsening of bipolar depression*	X	X
Death	X	X
Lack of efficacy	X	X
Lost to Follow-up	X	
Protocol violation	X	X
Physician decision	X	X
Study terminated by sponsor	X	X
Subject withdrew consent:	X	X
Personal or family reasons;	X	X
Refused to provide a reason or refused all end-of-study procedures;	X	X
Self-reported adverse event;	X	X
Self-reported lack of efficacy	X	X
Other	X	X

* Adverse event associated with worsening of bipolar depression will be identified by a Medical Monitor review prior to database lock.

The number and percentage of randomized subjects who discontinued due to an adverse event associated with worsening of bipolar depression will be summarized. The number and percentage of randomized subjects discontinued due to an adverse event not associated with worsening of bipolar depression will also be presented.

The time to treatment discontinuation due to all reasons, adverse events, lack of efficacy, or due to any other reason of special interest will be evaluated separately using the Kaplan-Meier method. Time to treatment discontinuation of subjects in the SAF Set will be defined as the date of discontinuation minus date of first dose of study medication plus 1. Subjects who complete treatment or who discontinue for a reason other than the one being evaluated will be censored. The Log-rank test will be used to compare the time to treatment discontinuation between the 60 mg ITI-007 treatment group and placebo group. The same analysis will be repeated for time to treatment discontinuation for any reason, where only subjects who complete the 6-week Treatment Period will be censored.

The number and percentage of subjects who withdrew from the study will be summarized by visit, treatment group and overall for all subjects in the RND Set. The number of subjects randomized will also be summarized by site for the RND Set. The number of subjects remaining on treatment over time will be presented in a figure.

The number and percentage of subjects in the SAF set, the ITT set [redacted] will be summarized by treatment group and overall for all subjects in the RND set.

[redacted]

[redacted]	[redacted]
[redacted]	[redacted]

[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

11. DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS

Demographic data and other baseline characteristics will be presented for the ITT[redacted] and SAF Sets.

No formal statistical testing will be carried out for comparing demographic or other baseline characteristics between the two treatment groups.

The following demographic and other baseline characteristics will be reported:

Demographics

- Gender
- Age (years) calculated as (Date of Informed Consent – Date of Birth)/365.25
- Age category (<=40 years and >40 years)
- Race
- Ethnicity

Other Baseline Characteristics

- Height (cm)
- Weight (kg)
- Waist circumference (cm)
- Body Mass Index (BMI) (kg/m^2) calculated as weight (kg) / height $(m)^2$

Consumption Habits

- Alcohol
- Caffeine
- Tobacco
- Cannabis

Bipolar Disorder Diagnosis

- Bipolar disorder diagnosis (Bipolar I or II);
- Age at first diagnosis of Bipolar Disorder (years);
- Duration of current major depressive episode (weeks);
- Number of lifetime depressive episodes;
- Hospitalization for emotional or psychiatric problems;
- Number of hospitalizations in lifetime;
- Hospitalization in the past year;

Baseline Efficacy

- Baseline efficacy parameters, including MADRS total score, MADRS item 4 score, CGI-BP-S subscale scores, CGI-BP-S total score and Q-LES-Q-SF score

Bipolar Disorder Baseline Safety

- Baseline safety parameters, including total scores for the following: SAS, AIMS, BARS, YMRS, and CSSR-S

12. MEDICAL HISTORY

Medical history information will be presented for the SAF Set and will be coded using a current version of the Medical Dictionary for Regulatory Activities (MedDRA).

Medical History conditions are defined as those current and past conditions at Screening.

The number and percentage of subjects with at least one pre-existing condition as well as the number of subjects having pre-existing conditions will be summarized by system organ class (SOC), preferred term (PT), treatment group and overall. For each subject, multiple reports of pre-existing conditions that map to a common PT and SOC will be condensed into a single medical history for incidence counts.

13. PRIOR AND CONCOMITANT MEDICATIONS

Medications will be presented for the SAF Set and coded to preferred names using the WHO Drug Dictionary.

Medications will be classified as prior, prior concomitant, concomitant, or post-treatment based on start and end dates. See APPENDIX 2 for handling of partial dates for medications. In the case where it is not possible to classify a medication due to missing dates the medication will be classified by the worst case, i.e. concomitant.

A medication is considered to have started prior to the first dose of study medication if indicated to have started prior to the first dose of study medication on the eCRF. If the medication started prior to the first dose of study medication, it is considered to have ended prior to the first dose of study medication if indicated to have ended prior to the first dose of study medication on the eCRF.

A medication is considered to have started after the last dose of study medication if indicated as such on the eCRF, assuming it started after completion of all study procedures and assessment related to the last dose of study medication.

- 'Prior' medications are medications which started and stopped prior to date of first dose of study medication.
- 'Prior concomitant' medications are medications which started prior to and stopped after the date of first dose of study medication.
- 'Concomitant' medications are medications which:
  - started after the date of first dose of study medication and started prior to the date of last dose of study medication,
  - AND stopped after the date of first dose of study medication or were ongoing at the completion of Day 43 post-dose assessments on Day 43 (or last treatment day for subjects who discontinue treatment), or were ongoing after the last dose of study medication (if planned assessments are not performed).
- 'Post-treatment' medications are medications which started after completion of Day 43 post-dose assessments (or after last treatment day for subjects who discontinue treatment).

Prior, prior concomitant, concomitant, and post-treatment medication use will be summarized by PT using frequencies and percentages by treatment group. Medications will be sorted alphabetically by PT in summaries. Subjects with multiple occurrences of a medication in the same PT will only be counted once within the PT for each summary.

During the study, a subject may be treated with zolpidem, benzodiazepines, zopiclone, eszopiclone or zalaplon if taken no more than 3 times per week and only during the Screening Period or first 2 weeks of the On-Treatment Period. The number and percent of subjects in the ITT Set receiving zolpidem, benzodiazepines, zopiclone, eszopiclone or zalaplon and the total number of days on zolpidem, benzodiazepines, zopiclone, eszopiclone or zalaplon will be summarized by treatment group for the Screening Period, for each week during the On-treatment period (Days 1-8, Days 9-15, Days 16-22, Days 23-29, Days 30-36 and Days 37-43), and for post treatment.

Other medications will be summarized by treatment group and study period if deemed necessary.

14. STUDY MEDICATION EXPOSURE AND TREATMENT COMPLIANCE

Exposure to study medication and treatment compliance will be presented for the ITT and SAF Sets.

The date of first and last study medication administration will be taken from the eCRF's 'Dosing' and 'End of Treatment' forms, respectively. If a discontinued subject is missing the end of treatment date from the 'End of Treatment' form, the last known dispense date will be used as the date of last study medication administration.

[redacted]

Duration of exposure will be summarized by treatment group as a continuous variable, using mean, SD, median, minimum and maximum. In addition, the number and percentage of subjects within each planned Study Week (1-8 days, 9-15 days, 16-22 days, 23-29 days, 30-36 days, 37-43 days, and > 43 days) will be presented.

Compliance with study medication will be derived from the drug return CRF page. For each subject, overall compliance with study medication (%) will be calculated as 100 x (Number of compliant days) / (Number of days in the On-Treatment Period), where a compliant day is defined as any day during the On-Treatment Period on which the subject took 1 capsule, and 'Number of days in the On-Treatment Period' is equivalent to 'Duration of Exposure' as defined above. The number and percentage of compliant and non compliant subjects will be presented. Non-compliant subjects are defined as those subjects with a compliance less than 80% or greater than 120%.

15. EFFICACY OUTCOMES

15.1. PRIMARY EFFICACY

15.1.1. PRIMARY EFFICACY VARIABLE AND DERIVATION

The primary efficacy variable is change from baseline to Day 43 in the MADRS total score. The MADRS is a 10-item checklist designed to measure the overall severity of depressive symptoms. The MADRS individual items are rated by a qualified site rater at each site on a scale of 0 to 6 in which a score of 6 represents the most severe symptoms for each item assessed. The MADRS total score is the sum of all 10 items and ranges from 0 to 60. If one or more items are missing at a visit, the total score will also be set to missing for that visit.

15.1.2. MISSING DATA METHODS FOR PRIMARY EFFICACY ENDPOINT

The primary analysis method for evaluating the primary efficacy endpoint is the likelihood-based MMRM method described in detail in Section 15.1.3. The use of MMRM inherently implies that the treatment effect on the change from baseline in the MADRS total score will be similar for the subjects who withdraw and for those who complete the treatment in their respective treatment groups, conditional on the outcomes observed prior to withdrawal (MAR assumption), as described in section 7.1. Sensitivity analysis of the primary efficacy endpoint will be performed to evaluate the robustness of the MMRM method under different assumptions on the mechanism of missng data, as detailed in Section 15.1.4.

If a patient has the score of one or more items missing at a visit, the total score for that patient will be set to missing for that visit.

15.1.3. PRIMARY ANALYSIS OF PRIMARY EFFICACY VARIABLE

The treatment effect on the primary efficacy endpoint will be estimated using an MMRM method. The model will include the change from baseline at each pre-specified time point as the response variable, study visit, Bipolar Disorder stratification variable at screening (Bipolar I

or Bipolar II), treatment (60 mg ITI-007 or placebo) and site (or pooled site) as factors, baseline MADRS total score as covariate and interaction terms for baseline MADRS total score-by-study visit and treatment-by-study visit. The subject term will be included in the model as a random effect while all other terms will be considered fixed effects. An unstructured covariance matrix will be used to model the correlation among repeated measurements within subject. [redacted]

The Kenward-Rogers method will be used to estimate the denominator degrees of freedom. The treatment and treatment-by-study visit interaction terms allow for comparisons of the treatment groups at the time point of interest (Days 8, 15, 22, 29, 36 and 43). Treatment differences will be evaluated via contrasts for the study visit-by-treatment factor.

The change from baseline to Day 43 will be used for the primary efficacy analysis, and will be performed for the ITT Set [redacted].

Summary statistics for change from baseline in MADRS total score will be presented by treatment group and study visit (Day), as well as least-squares mean (LSM) estimates, standard errorsand the corresponding 95% CI for LSMs.Contrast estimates (LSM differences) for between-group comparison, the corresponding standard errors, 95% CIs, effect sizes, and p-values will also be presented for each study visit (Day). Effect size will be calculated for the ITI-007 60 mg treatment group as:

$$\frac{\text{LS Mean Difference}}{\text{Pooled estimate of within subject error standard deviation}}.$$

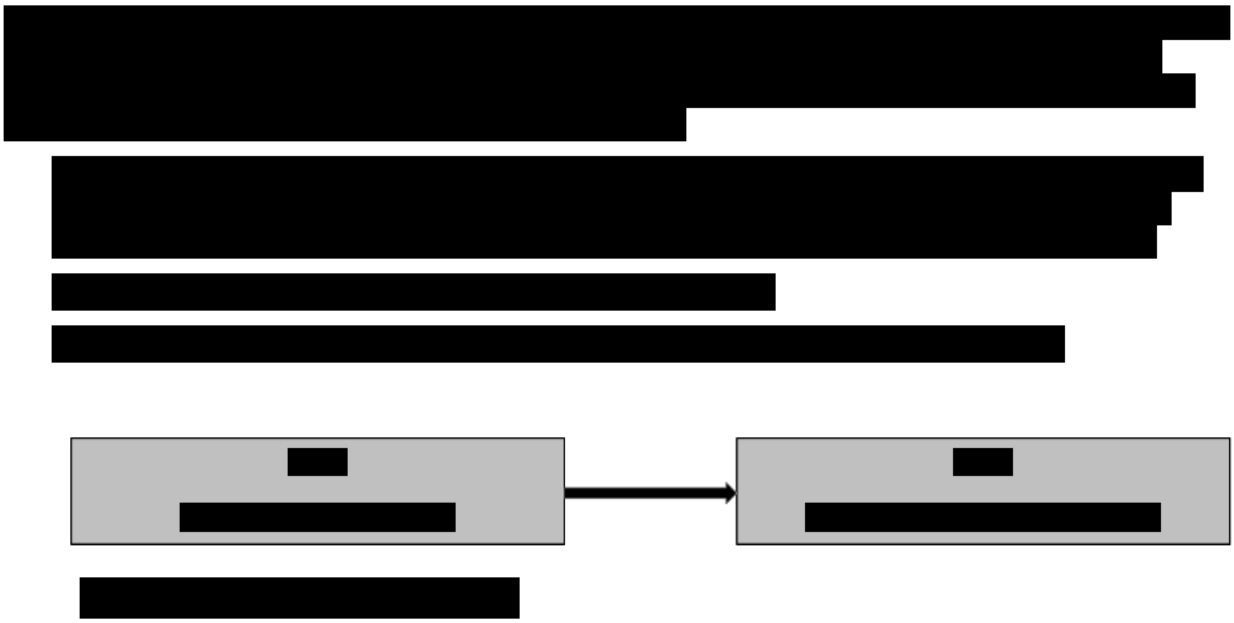
[redacted]

15.2. KEY SECONDARY EFFICACY ENDPOINT

The key secondary efficacy endpoint is the change from baseline to Day 43 in the Clinical Global Impression scale, Bipolar Version, for Severity (CGI-BP-S).

The CGI-BP-S is a single value assessment of illness severity in mania, depression and overall bipolar illness and ranges from 1='Normal, not at all ill' to 7='Very severely ill' It is used in this study at screening (as a criterion for inclusion or exclusion), at baseline, and throughout the study, at Days 8, 15, 22, 29, 36, and 43. A higher score is associated with greater illness severity. The CGI-BP-S has 3 subscales, one each for mania, depression and overall bipolar illness. These 3 subscales are included in the calculation of the CGI-BP-S total score which will be recorded on the Clinical Global Impressions – Bipolar (CGI-S-BP) CRF page.

Change from baseline in the CGI-BP-S total score will be evaluated using an MMRM method similar to the one specified for the primary analysis of the primary efficacy endpoint, where the change from baseline to each pre-specified time point in CGI-BP-S total score is the response variable, study visit, Bipolar Disorder stratification variable at screening (Bipolar I or Bipolar II), treatment (60 mg ITI-007 or placebo) and site (or pooled site) as factors, baseline CGI-BP-S total score as covariate and interaction terms for baseline CGI-BP-S total score-by-study visit and treatment-by-study visit. The subject term will be included in the model as a random effect while all other terms will be considered as fixed effects. See Section 15.1.3, the primary analysis of primary efficacy endpoint, for more details.



[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

16. SAFETY OUTCOMES

All outputs for safety outcomes will be based on the SAF Set.

16.1. FOR BY VISIT SUMMARIES OF CHANGE FROM BASELINE SUMMARIES, THE LAST VALUE ON TREATMENT (LAST ASSESSMENT ON OR BEFORE THE LAST DOSE OF STUDY MEDICATION PLUS ONE DAY) WILL BE PRESENTED.ADVERSE EVENTS

Adverse Events (AEs) will be coded using the Medical Dictionary for Regulatory Activities

(MedDRA) central coding dictionary currently in effect at the time of analysis or database lock.

Treatment emergent adverse events (TEAEs) are defined as AEs that started or worsened in severity on or after the first dose of study medication and on or before the date last dose of study medication plus one day.

All AEs with an onset date after the last dose of study medication plus one day will be listed in the AE data listing and labelled as 'Follow-up Adverse Event'.

See APPENDIX 2 for handling of partial or completely missing dates for AEs. In the case where it is not possible to define an AE as treatment emergent or not due to missing dates, the AE will be classified as the worst case, i.e. treatment emergent.

An overall summary of number of subjects within each of the categories described in the sub-sections 16.1.1 and 16.1.2 below will be provided as specified in the TLF shells.

Listings will include all AEs, TEAEs and Non-TEAEs. A seperate listing for TEAEs leading to study discontinuation will be presented.

16.1.1. AEs AND TEAEs

Incidence of TEAEs will be presented by System Organ Class (SOC) and Preferred Term (PT) and also broken down further by maximum severity and relationship to study medication.

The number and percentage of subjects with at least one TEAE and total number of subjects experiencing TEAEs will be summarized by PT, SOC and treatment group. A summary of TEAEs will be provided with only PTs and a separate summary for only SOCs. Additional summaries of TEAEs and TEAEs related to study medications ('possibly related', 'probably related', or 'definitely related') will be provided for PTs occurring in at least 5% of subjects in any treatment group (60 mg ITI-007 or placebo). Within each subject, multiple reports of events that mapped to a common MedDRA PT and SOC will be condensed into a single AE for incidence counts. Summaries will present SOCs in alphabetical order and PTs by descending (total) frequency within SOCs, by treatment group.

TEAEs will also be categorized according to the onset date of the TEAEs. The total number of subjects having events for each PT and SOC will be summarized by first onset categories, based on planned study visits (Day 1 - ≤ Day 8, > Day 8 - ≤ Day 15, > Day 15 - ≤ Day 22, > Day 22 - ≤ Day 29, > Day 29 - ≤ Day 43).

The relative risk of at least one TEAE and for each PT and SOC will also be presented along with 95% CIs and p-values obtained by the Chi-square test for association.

AE severity is classified as "mild", "moderate", "severe", "life-threatening" by the investigator. AEs and TEAEs with a missing severity will be classified as "not specified". If a subject reports a TEAE more than once within the same PT and SOC, the event with the worst case severity will be used in the corresponding severity summaries.

Relationship to study medication, as indicated by the Investigator, is classified as "unrelated", "unlikely to be related", "possibly related", "probably related", "definitely related" (increasing severity of relationship). A "related" TEAE is defined as a TEAE "possibly related", "probably related" or "definitely related" to study medication. TEAEs with a missing relationship to study medication will be regarded as "related". If a subject reports the same AE more than once within the same PT and SOC, the AE with the worst case relationship to study medication will

be used in the corresponding relationship summaries. A summary of related TEAEs by SOC and PT will be presented.

AEs leading to study discontinuation will be identified by using the Completion/Withdrawal from Study page of the eCRFs, where 'Primary Reason' indicates "AE". TEAE leading to disontinuation of study medication will be identified by using the End-of-Treatment page of the eCRFs, where 'Primary Reason' indicates "AE".

For AEs leading to early study withdrawal or discontinuation of study medication, summaries of incidence rates (frequencies and percentages) by SOC and PT will be prepared.

16.1.2. SERIOUS ADVERSE EVENTS AND DEATH

Serious adverse events are those events recorded as "Serious" on the Adverse Events page of the eCRF. SAEs will be listed and summarized by SOC and PT.

AEs leading to death are those events which are recorded as "Fatal" on the Adverse Events page of the (e)CRF. Deaths may also be recorded on the End of Treatment page and the Completion/Withdrawal from Study page of eCRF. A listing all deaths will be prepared based on these sources. A listing of all TEAEs leading to death will also be prepared.

16.1.3. EXTRAPYRAMIDAL TEAES

Extrapyramidal TEAEs will be defined by the Standard MedDRA Query (SMQ) labeled as Extrapyramidal Syndrome. The number and percentage of subjects with at least one TEAE PT mapped to an extrapyramidal term contained in the SMQ will be presented. In addition, the extrapyramidal terms will be summarized. Separate tables will be presented for the narrow and broad interpretation of the SMQ.

16.1.4. ABUSE-RELATED TEAES

TEAEs will be categorized as defined in the table below to monitor signals of potential abuse of ITI-007. The number and percentage of subjects with at least one abuse-related TEAE will be summarized by category and the applicable PTs below.

Table D: Abuse-Related PTs

Category	Applicable Preferred Terms
Euphoria	Euphoric mood; Elevated mood; Feeling abnormal; Feeling drunk; Feeling of relaxation; Dizziness; Thinking abnormal; Hallucination; Inappropriate affect
Impaired attention, cognition, and mood	Somnolence; Sedation; Mood disorders and disturbances
Dissociative/psychotic	Psychosis; Aggression; Confusion and disorientation
Other related terms	Drug tolerance; Habituation; Drug withdrawal syndrome;

	Substance-related disorders

16.2. LABORATORY EVALUATIONS

Results from the central laboratory will be included in the reporting of this study for hematology, clinical chemistry, urinalysis and screening serology. A list of laboratory assessments to be included in the outputs is included in the protocol, Section 8.5. Summary statistics will be presented in standard international (SI) and US conventional units.

Quantitative laboratory measurements reported as "< X", i.e. below the lower limit of quantitation (BLQ), or "> X", i.e. above the upper limit of quantitation (ULQ), will be converted to X for the purpose of quantitative summaries, but will be presented as recorded, i.e. as "< X" or "> X" in the listings.

Blood samples for laboratory evaluations will be collected after an overnight fast (⩾10 hours) during screening and on Days 8, 22, 43 and Day 57 safety follow-up or during the discontinuation visit. All non fasting results for glucose, insulin and triglycerides will be excluded from descriptive statistics but will appear in the data listings.

The following summaries will be provided for laboratory data:

- Actual and change from baseline by visit, and for last on-treatment assessment where applicable (for quantitative measurements)
- Incidence of abnormal values according to normal range criteria
- Shift from baseline according to normal range criteria (for quantitative measurements and categorical measurements)
- Shift from baseline according to markedly abnormal criteria defined in Section 16.2.1 (for quantitative measurements and categorical measurements)
- Listing of subjects meeting markedly abnormal criteria.
- Incidence of liver function related values meeting pre-defined criteria, as defined in Section 16.2.2, during the treatment period and the entire study

16.2.1. LABORATORY REFERENCE RANGES AND MARKEDLY ABNORMAL CRITERIA

Quantitative laboratory measurements will be compared with the relevant laboratory reference ranges in SI units and US Conventional units and categorized as:

- Low: Below the lower limit of the laboratory reference range.
- Normal: Within the laboratory reference range (upper and lower limit included).
- High: Above the upper limit of the laboratory reference range.

In addition to the high and low quantitative laboratory assignments (as identified by means of the laboratory reference ranges), markedly abnormal quantitative safety (and other) laboratory assessments will also be identified in accordance with the predefined markedly abnormal criteria as presented in Table E:

Table E: Markedly Abnormal Values for Laboratory Evaluations

Hematology Parameter	Markedly Abnormal Range
Hemoglobin	Male: ⩽ 11.5 g/dL
	Female: ⩽ 9.5 g/dL
Hematocrit	Male: ⩽ 37%
	Female: ⩽ 32%
White Blood Cell (WBC) Count	⩽ 2.8 x 10^3 cells/µL
	⩾ 16 x 10^3 cells/µL
Neutrophils (percent)	⩽ 15%
Eosinophils (percent)	⩾ 10%
Platelet Count	⩽ 75 x 10^3 cells/µL
	⩾ 700 x 10^3 cells/µL
Chemistry Parameter	**Markedly Abnormal Range**
Alkaline Phosphatase	⩾ 2 x ULN
Gamma-glutamyl transferase (GGT)	⩾ 3 x ULN
Alanine aminotransferase (ALT)	⩾ 3 x ULN
Aspartate aminotransferase (AST)	⩾ 3 x ULN
Total Bilirubin	⩾ 2 x ULN
Albumin	< 2.5 g/dL
Glucose	< 45 mg/dL
	> 160 mg/dL
Sodium	< 130 mmol/L
	> 150 mmol/L
Potassium	< 3 mmol/L
	> 5.5 mmol/L

Chloride	< 90 mmol/L
	> 115 mmol/L
Calcium	< 7 mg/dL
	> 12 mg/dL
Blood Urea Nitrogen	⩾ 30 mg/dL
Creatinine	⩾ 2.0 mg/dL
Creatine Phosphokinase (CPK)	⩾ 5 x ULN
HbA_{1c}	⩾ 7.5%
Prolactin	⩾ 5 x ULN
Total Cholesterol (fasting)	⩾ 300 mg/dL
LDL Cholesterol (fasting)	⩾ 200 mg/dL
Triglycerides (fasting)	⩾ 300 mg/dL
Urinalysis Parameter	Markedly Abnormal Range
RBC	> 10 cells/hpf
WBC	> 20 cells/hpf

16.2.2. LIVER FUNCTION RELATED CRITERIA

Liver function-related laboratory tests will be summarized in accordance to the criteria in Table F:

Table F: Liver Function Pre-defined Criteria

Liver Function Parameter	Criteria
ALT	≥ 3 x ULN
	≥ 5 x ULN
AST	≥ 3 x ULN
	≥ 5 x ULN
GGT	≥ 3 x ULN
	≥ 5 x ULN

Total Bilirubin (in combination with ALT or AST criteria)	> 1.5 x ULN > 2 x ULN
CPK	≥ 5 x ULN
Hy's Law	Criteria
ALT or AST	≥ 3 x ULN
Total Bilirubin	> 2 x ULN
Alkaline phosphatase	< 2 x ULN

16.3. ECG EVALUATIONS

ECG will be performed during screening, at baseline, Day 43 or Early Termination Visit and on the Day 57 safety follow-up for all subjects. At screening ECGs will be collected in triplicate and will be analyzed as an average of the non-missing measurements at each time point. In all cases, ECGs are conducted before other assessments scheduled in the same time window. Results from the central ECG (Electrocardiogram) Reading Centre will be included in the reporting of this study.

The following ECG parameters will be reported:

- HR (bpm)
- QRS Interval (msec)
- PR Interval (msec)
- QT Interval (msec)
- QTcF Interval (msec) [derived by central ECG]
- QTcB Interval (msec) [derived by central ECG]
- RR Interval (msec)
- Overall assessment of ECG

For the overall ECG assessment, there are five possible results for each triplicate: 'Abnormal, Significant', 'Abnormal Insignificant', 'Incomplete Analysis', 'Normal', and 'Uninterpretable'. Overall assessments with a result of 'Incomplete Analysis' or 'Uninterpretable' will be considered missing. Triplicates will be analyzed according to the worst assessment and categorized in the following order, 'Abnormal Significant', 'Abnormal, Insignificant' and 'Normal'.

The following summaries will be provided for ECG data:

- Actual and change from baseline by visit (for quantitative measurements, except unadjusted QT)
- Incidence of abnormal and normal ECGs by visit

- Shift from baseline to each visit according to markedly abnormal criteria (for quantitative measurements and categorical measurements)
- Incidence of markedly abnormal results defined in Section 16.3.1
- Listing of subjects meeting markedly abnormal criteria

16.3.1. ECG MARKEDLY ABNORMAL CRITERIA

Markedly abnormal quantitative ECG measurements will be identified in accordance with the following predefined markedly abnormal criteria:

- Absolute values for QT, QTcB and QTcF intervals will be classified as:
- > 450 msec
- > 480 msec
- > 500 msec
- Change from Baseline for QT, QTcB and QTcF intervals will be classified as:
- > 30 msec increase from baseline
- > 60 msec increase from baseline

16.4. VITAL SIGNS

The following vital signs measurements will be reported for this study:

- Systolic blood pressure (mmHg)
- Diastolic blood pressure (mmHg)
- Pulse rate (bpm)
- Respiratory rate (breaths/min)
- Oral temperature (^{0}C)
- Height (cm) at screening
- Weight (kg)
- BMI (kg/m^2) [calculated as weight (kg) / height (m)^2]

Vital signs, including systolic blood pressure, diastolic blood pressure, pulse rate, respiratory rate, oral temperature and body weight, are recorded at screening, baseline and Days 8, 15, 22, 29, 36, 43 and 57At screening, baseline and days 43 and 57 vital signs include the 3-positional blood pressure and pulse rate will be measured after 10 minutes in the supine position, 1-minute sitting, immediately after standing, and 3 minutes after standing at screening, baseline, day 43 and day 57. These 3-positional results will be averaged at each visit and used for the summary and analyses. Pulse will be measured by counting pulse over 60 seconds. Height will be measured only at screening.

BMI measurements classify a subject's weight status as underweight, normal, overweight, or obese using Table G:. Shifts from Baseline to Day 43 will be produced by treatment group for the SAF Set to show the percentage of subjects who fall into each BMI category combination.

Table G: BMI Weight Status Categories

BMI (kg/m^2)	Weight Status
< 18.5	Underweight
18.5 ≤ BMI < 25.0	Normal
25.0 ≤ BMI < 30	Overweight
30.0 ≤ BMI	Obese

The following summaries will be provided for vital sign data:

- Actual and change from baseline by visit
- Incidence of markedly abnormal values
- Listing of subjects meeting markedly abnormal criteria
- Shifts from baseline by visit and time point according to markedly abnormal criteria.
- BMI shift summaries

The analysis of change from baseline in all vital sign measurements to Day 43 will be performed using an MMRM method. The model will include visit, Bipolar Disorder stratification variable at screening (Bipolar I or Bipolar II), site (or pooled site) and treatment as factors, baseline values as covariate and interaction terms for baseline baseline value-by-study visit. The subject term will be included in the model as a random effect while all other terms will be considered fixed effects. Summary statistics for change from baseline will be presented by treatment group and study visit (Day) as well as LSM estimates and the the associated standard error.LSM differences between 60mg ITI-007 and placebo, the associated standard errors, 95% CIs and p-values for between-treatment tests of differences will be presented for each study visit (Day).

As an exploratory analysis, change from baseline in these vital signs to each assessment time point will also be evaluated using an ANCOVA with LOCF model with effects for study visit, baseline value, Bipolar Disorder stratification variable at screening (Bipolar I or Bipolar II), site (or pooled site) and treatment. LSMs and the associated standard errors for each treatment group (ITI-007 60 mg or placebo), the LSM differences for between-treatment comparisons (ITI-007 60 mg vs. placebo), the corresponding standard errors, 95% CIs and p-values for the between-treatment differences will be presented for each assessment time point.

16.4.1. VITAL SIGNS MARKEDLY ABNORMAL CRITERIA

Markedly abnormal quantitative vital signs measurements will be identified in accordance with the following predefined markedly abnormal criteria in the table below.

Table H: Markedly Abnormal Criteria for Vital Signs

Variable	Unit	Low	High
SBP	mmHg	< 90 mmHg AND decrease from baseline ≥ 20 mmHg	> 180 mmHg AND increase from baseline ≥ 20 mmHg
DBP	mmHg	< 50 mmHg AND decrease from baseline ≥ 15 mmHg	> 105 mmHg AND increase from baseline ≥ 15 mmHg
Pulse rate	bpm	< 50 bpm AND decrease from baseline ≥ 15 bpm	> 120 bpm AND increase from baseline ≥ 15 bpm
Weight	Kg	percentage change from baseline ≤ -7.0 %	percentage change from baseline ≥ 7.0 %

16.5. MODIFIED PHYSICAL AND NEUROLOGICAL EXAMINATION

The following summaries will be provided for physical examination data:

- Incidence of abnormalities at screening (baseline)
- Incidence of abnormalities post baseline

A modified physical examination, including neurological and excluding genital/rectal examinations, will be performed during screening, on Day 43 or Early Termination Visit and on Day 57 for all subjects. The examination should include evaluation of general appearance; eyes, ears, nose, and throat; skin; head and neck; lungs and chest; heart; abdomen; and extremities. Neurological examinations include an assessment of motor function, sensory function, reflexes, and gait. All physical and neurological examination findings are recorded in the eCRF.

16.6. OTHER SAFETY ASSESSMENTS

Handling of missing data for the following assessments is presented in Section 7.1.

16.6.1. YOUNG MANIA RATING SCALE (YMRS)

The YMRS is an 11-item, clinician-administered mania rating scale with established reliability, validity, and sensitivity that was designed to assess the severity of manic symptoms. Four of the YMRS items are rated on a 0 to 8 scale, with the remaining 7 items rated on a 0 to 4 scale. The total score is the sum of all 11 items and ranges from 0 to 60. The total score is appropriate both for assessing baseline severity of manic symptoms and for assessing treatment-emergent manic symptoms in subjects with Bipolar I or Bipolar II Disorder with

depression. It will be completed by the investigator or an expert site-based rater.

The observed and change from baseline in YMRS total score will be summarized by treatment and visit.

16.6.2. BARNES AKATHISIA RATING SCALE (BARS)

The BARS is a rating scale, measuring the observable, restless movements that characterize akathisia. It consists of 4 items: objective restlessness, awareness of restlessness, distress related to restlessness and global clinical assessment of akathisia. Each item is on a 4-point scale (0 to 3), except for the global clinical assessment which is on a 6-point scale (0 to 5), both using low values to represent absence of akathisia and high representing severe akathisia. The BARS total score is the sum of items 1 through 3 and ranges from 0 to 9. Higher values of the BARS total score indicate akathisia is higher in severity. If one or more items at a visit are missing the total will not be calculated.

The observed and change from baseline in BARS total scores to Day 43 will be summarized by treatment.

16.6.3. ABNORMAL INVOLUNTARY MOVEMENT SCALE (AIMS)

The AIMS is a clinician rated assessment of abnormal movements. It contains items related to: facial and oral movements; extremity movements; trunk movements; global judgments and dental status. Seven items of the AIMS range from 0= “None” to 4= “Severe”. A score of “mild” (2) in two or more categories or a score of “moderate” or “severe” in any one category results in a positive AIMS score (i.e. the scores are not averaged). The global severity score is the response to “Severity of abnormal movements” found within the global judgments section. Additionally, overall severity is scored on the basis of severity of abnormal movements and incapacitation due to abnormal movements.

The (non-global) AIMS total score is the sum of items 1 through 7. The possible range for the AIMS total score is 0 to 28. Higher values of the total AIMS score indicate increased severity in abnormal movement. If one or more of the AIMS total score items are missing at a visit, the score will be set to missing.

The observed and change from baseline in AIMS total scores to Day 43 will be summarized by treatment.

16.6.4. SIMPSON-ANGUS RATING SCALE (SAS)

The SAS is a measure of extrapyramidal side effects consisting of 10 items: gait, arm dropping, shoulder shaking, elbow rigidity, wrist rigidity, head rotation, glabella tap, tremor, salivation and akathisia. Items are rated on a scale from 0 (normal) to 4 (extreme in severity). The SAS total score is defined as the sum of all 10 items and ranges between 0 and 40. Lower values of the SAS total score indicate milder symptoms. If one or more items are missing at a visit the SAS total score will be set to missing.

The observed and change from baseline in SAS total scores to Day 43 will be summarized by treatment.

16.6.5. COLUMBIA SUICIDE SEVERITY RATING SCALE (C-SSRS)

The C-SSRS is an instrument designed to systematically assess and track suicidal behavior and suicidal ideation throughout the study. The C-SSRS includes the following four sections: Suicidal Ideation, Ideation Intensity, Suicidal Behavior and Actual Suicide Attempts. The strength of this suicide classification system is in its ability to accurately and comprehensively assess suicidality, while limiting the over-identification of suicidal behavior. The C-SSRS will be administered by a trained rater at the site.

Suicidal Ideation is assessed by 5 questions, the responses to which equate to a 6-point scale from 0="No ideation present" to 5="Active ideation with plan and intent". A score of 4 or 5 on this scale indicates serious suicidal ideation. Any score greater than 0 will be counted as having suicidal ideation.

The Ideation Intensity total score is the sum of five items from the Ideation Intensity scale: frequency, duration, controllability, deterrents, and reasons for ideation. If a subject did not endorse any suicidal ideation, a score of 0 for the intensity total score will be given. The possible range for the Intensity total score is 0 to 25.

The number and percentage of subjects with each type of suicidal ideation or any suicidal ideation during each study period will be summarized. The most severe ideation, the ideation intensity items (frequency, duration, controllability, deterrents, and reasons for ideation), and the ideation intensity total score will also be summarized descriptively.

Suicidal Behavior is collected as actual attempt, non-suicidal self-injurious behaviour, interrupted attempt, aborted attempt, preparatory acts or behavior, suicidal behavior, and suicide. The number and percentage of subjects with each type of suicidal attempt (actual attempts, interrupted attempts, aborted attempts, and preparatory acts or behavior) will be summarized for each scheduled visit and overall by study period. The number and percentage of subjects with any suicidal behavior and those completing suicide will also be summarized for each scheduled visit and overall by study period.

The number and percentage of subjects with suicidality as measured by the C-SSRS will be summarized, where suicidality is defined as having at least one occurrence of suicidal ideation or at least one occurrence of suicidal behavior. The suicidality indicator will be set to 1 if the subject exhibits suicidality for each visit, 0 if the subject does not exhibit suicidality, and missing otherwise. This data will be summarized at each scheduled visit and overall by study period (treatment and post-treatment).

Data collected on actual suicide attempts (lethality of actual attempts and potential lethality of attempts) will be presented in a data listing.

An overall summary of C-SSRS data (post-baseline data across all scheduled visits) will include the frequency and percentage of the following:

- At least one suicidal ideation post-baseline
- Emergence of suicidal ideation (no suicidal ideation at baseline, and any type of suicidal ideation post-baseline)
- Emergence of serious suicidal ideation (no suicidal ideation at baseline, and any serious suicidal ideation [ideation score of 4 or 5] post-baseline)
- Most severe type of ideation post-baseline

- Worsening of suicidal ideation (most severe suicidal ideation post-Baseline was more severe than it was at baseline)
- At least one suicidal behavior post-baseline
- Emergence of suicidal behavior (no suicidal behavior at baseline, and any type of suicidal behavior post-baseline)
- At least one actual attempt post-baseline
- At least one interrupted attempt post-baseline
- At least one aborted attempt post-baseline
- At least one preparatory acts or behaviors post-baseline
- At least one instance of suicidality [any ideation or behavior] post-baseline
- Emergence of suicidality (no suicidality at baseline, and any suicidality post-baseline)
- Any completed suicides post-baseline

[redacted]

18. REFERENCES

Koch, G. and Wienner, L.E. Commentary for the Missing Data Working Group's Perspective for Regulatory Clinical Trials, Estimands, and Sensitivity analyses. Statist. Med. 2016, 35 2887–2893.

LaVange LM and Permutt T (2015). A regulatory perspective on missing data in the aftermath of the NRC report. Statist. Med.35(17):2853-64.

Little, R., Yau, L. (1996). Intention-to-treat analysis for longitudinal studies with drop-outs. Biometrics. 52, 1324-1333.

National Research Council. The Prevention and Treatment of Missing Data in Clinical Trials. National Academies Press: Washington, D.C., 2010.

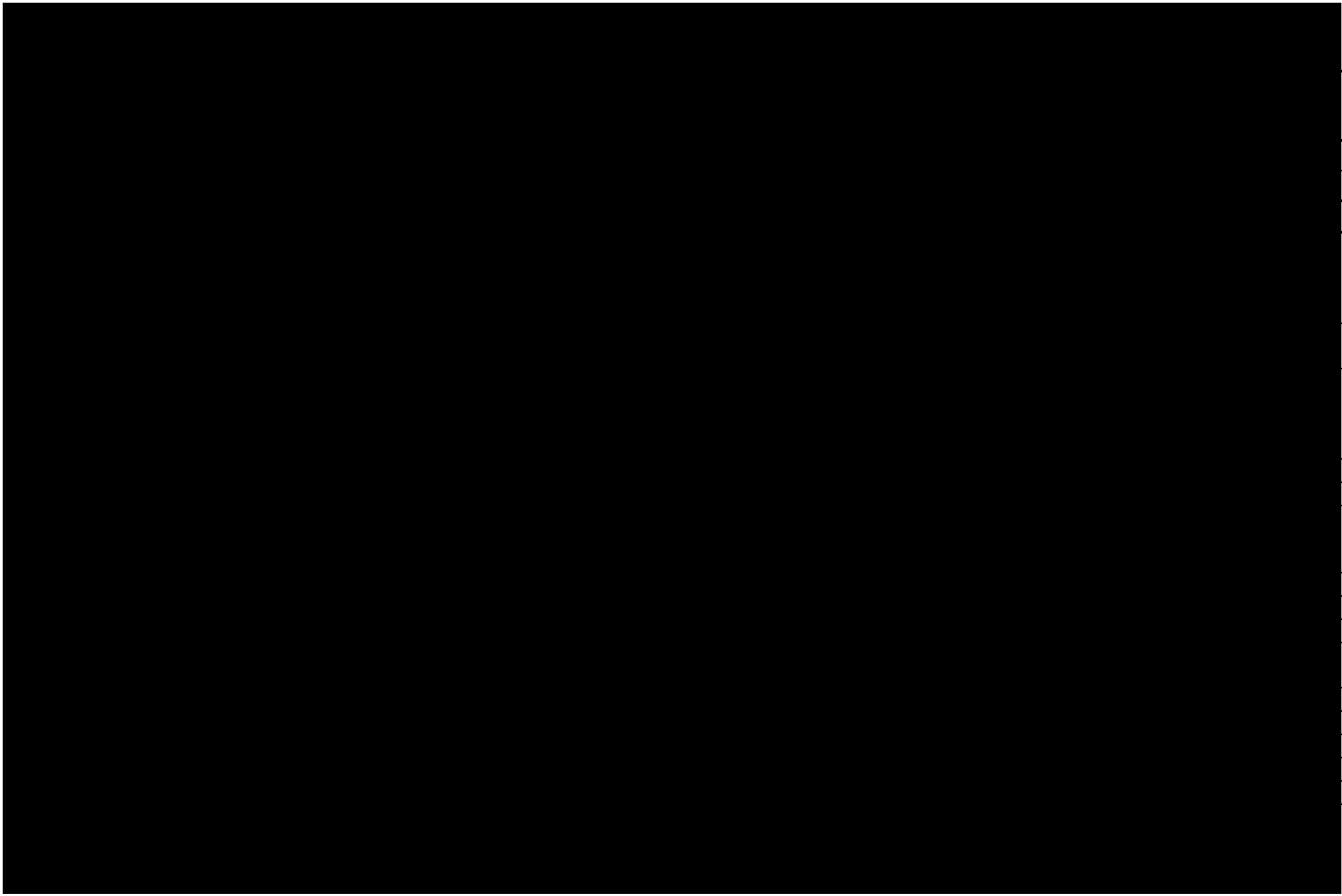
Permutt, T. (2016a) A taxonomy of estimands for regulatory clinical trials with discontinuation. Statistics in Medicine. 2016; 35(17): 2865-2875.

Permutt T for the Missing Data Working Group (2016b). Sensitivity analysis for missing data in regulatory submissions. Statist. Med. 35(17):2876–2879.

O'Kelly, M., Ratitch, B. (editors). Clinical Trials with Missing Data: A Guide for Practitioners. Wiley, 2014.

Ratitch, B., O'Kelly, M., Tosiello, R. (2013). Missing data in clinical trials: From clinical assumptions to statistical analysis using pattern mixture models. Pharmaceutical Statistics. 12, 337-347.

Shih WJ (2016). Comments on the three papers by the FDA/CDER research team on the regulatory perspective of the missing data problem. Statist. Med. 35:2880–2886.



[redacted]

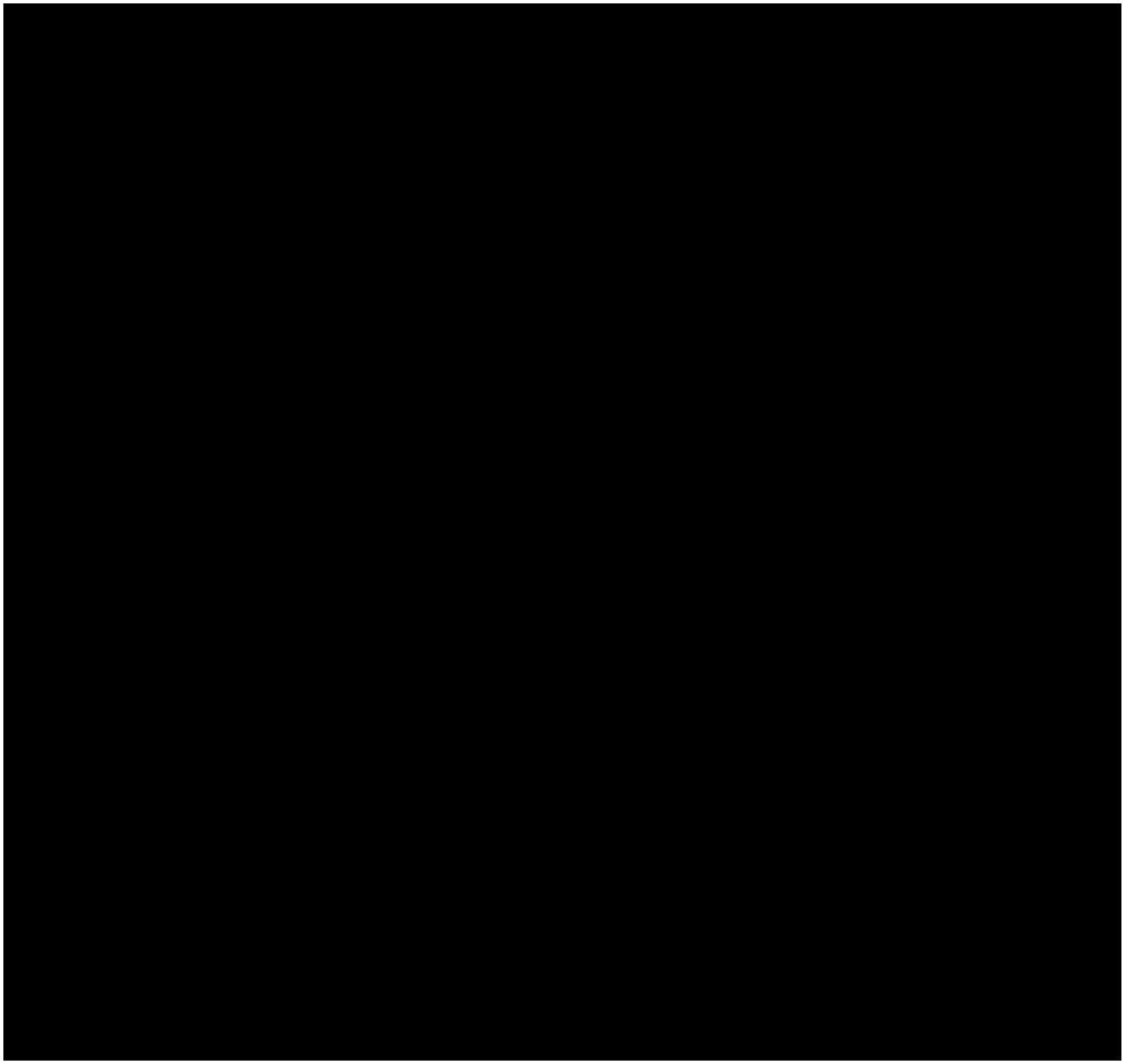
[redacted]

[redacted]

[redacted]

[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]
	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]
	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]
	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]

[redacted]



[redacted]

[redacted]

[redacted]

[redacted]