NCT02600494

Study ID: ITI-007-401

Title: A Phase 3, Randomized, Double-Blind, Placebo-Controlled, Multi-Center Study with an Open-Label Extension to Assess the Efficacy and Safety of ITI-007 Monotherapy in the Treatment of Patients with Major Depressive Episodes Associated with Bipolar I or Bipolar II Disorder (Bipolar Depression)

SAP Version 4.0 Date: 12 June 2019

STATISTICAL ANALYSIS PLAN

ITI-007-401

A PHASE 3, RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED, MULTI-CENTER STUDY WITH AN OPEN-LABEL EXTENSION TO ASSESS THE EFFICACY AND SAFETY OF ITI-007 MONOTHERAPY IN THE TREATMENT OF PATIENTS WITH MAJOR DEPRESSIVE EPISODES ASSOCIATED WITH BIPOLAR I OR BIPOLAR II DISORDER (BIPOLAR DEPRESSION)

AUTHOR: [redacted]

VERSION NUMBER AND DATE: V4.0, 12JUN2019

[redacted]

Signature

MODIFICATION HISTORY

Unique Identifier for this Version	Date of the Document Version	Author	Significant Changes from Previous Authorized Version
1.0	16Mar2016	[redacted]	Not Applicable – First Version
1.1	19Dec2017	[redacted]	Text added to describe the planned interim analysis as well as the tipping point sensitivity analysis. In addition, updates required following review of V1.0 were implemented.
1.2	22Feb2018	[redacted]	Update to address sponsor comments
1.3	30Apr2018	[redacted]	Update for protocol amendment
2.0	09Jul2018	[redacted]	Update for correction in conditional power formula in Appendix 4
3.0	12Jul2018	[redacted]	Update for correction to effect size calculation
4.0	12Jun2019	[redacted]	Update for additional analyses, sponsor comments on dry runs, and minor corrections

TABLE OF CONTENTS

[redacted]

[redacted]

1. LIST OF ABBREVIATIONS

AE	Adverse Event
AIMS	Abnormal Involuntary Movement Scale
ANCOVA	Analysis of Covariance
AR(1)	Autoregressive(1)
ARH(1)	Heterogeneous autoregressive(1)
BARS	Barnes Akathisia Rating Scale
BLQ	Below the limit of quantification
BMI	Body mass index
BPI	Bipolarity Index
C-SSRS	Columbia – Suicide Severity Rating Scale
CGI-BP-S	Clinical Global Impression Scale, Bipolar Version – Severity
CI	Confidence interval
CPK	Creatine phosphokinase
C-VISA™	Clinical Validation Inventory for Study Admission (Bracket)
DMC	Data Monitoring Committee
DSM-V	Diagnostic and Statistical Manual, Fifth Edition
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
ENR	All Subjects Enrolled
EPS	Extrapyramidal symptoms
FA0(q)	No diagonal factor analytic
FWER	Familywise error rate
GGT	Gamma-glutamyl transferase
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
LOCF	Last observation carried forward
LOE	Lack of efficacy
LS	Least squares
MADRS	Montgomery-Åsberg Depression Rating Scale
MAR	Missing at random
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple imputation
MMRM	Mixed-Effect Model Repeated Measure
MNAR	Missing not at random
NEO-FFI	Neuroticism, Extraversion and Openness to Experience - Five Factor Inventory
OLE	Open label extension
[redacted]	[redacted]
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
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
SCID-CT	Structured Clinical Interview for DSM Disorders - Clinical Trial Version
SD	Standard deviation
SDS	Sheehan Disability Scale
SMQs	Standard MedDRA Queries
SOC	System Organ Class
[redacted]	[redacted]
TEAE	Treatment-emergent adverse event
TOEPH	Heterogeneous Toeplitz structure
TOEP	Toeplitz structure
WHO	World Health Organization
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, [redacted] data for Protocol ITI-007-401. 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 1.4, dated 09Mar2018. This SAP details analyses for the double-blind portion of the study (Part A) and only references the open label extension (OLE) where it pertains to the Part A analyses.

3. STUDY OBJECTIVES FOR PART A (DOUBLE-BLIND)

3.1. PRIMARY OBJECTIVE

The primary objective of Part A of this study is to compare the efficacy of 2 doses of ITI-007 administered orally once daily to that of placebo in mean change from baseline in the total score on the rater-administered Montgomery-Åsberg Depression Rating Scale (MADRS) in subjects with bipolar depression at Week 6.

3.2. SECONDARY OBJECTIVES

3.2.1. KEY SECONDARY EFFICACY

The key secondary objective of Part A of this study is to compare the efficacy of 2 doses of ITI-007 administered orally once daily to that of placebo in time to first response, defined as the number of days from first dose of study drug to the earliest date the patient experiences a sustained ≥50% reduction from baseline in the rater-administered MADRS total score.

[redacted]

3.2.3. SECONDARY OBJECTIVES - SAFETY

The safety objectives of Part A of this study are to determine the safety and tolerability of 2 doses of ITI-007 via change in the following safety endpoints:

- Incidence of adverse events (AEs);
- Young Mania Rating Scale (YMRS);
- Columbia Suicide Severity Rating Scale (C-SSRS);
- Abnormal Involuntary Movement Scale (AIMS);
- Barnes Akathisia Rating Scale (BARS);
- Simpson Angus Scale (SAS);
- Clinical laboratory evaluations;
- Electrocardiograms (ECGs);
- Vital sign measurements;
- Physical and neurological examination findings.

[redacted]

4. STUDY DESIGN

4.1. GENERAL DESCRIPTION

The study will be conducted in two parts, Part A and Part B. Part A is a randomized, double-blind, placebo-controlled, two stage adaptive design study comparing the efficacy and safety of ITI-007 versus placebo administered orally once daily in subjects with bipolar depression. Subjects who safely complete treatment in Part A may enroll in Part B, an open-label extension.

Approximately 549 subjects (183 subjects/treatment group, if no sample size adjustment is made at the interim analysis during Part A) will be randomized in a 1:1:1 ratio to receive one of three study treatments in Part A: 40-mg ITI-007, 60-mg ITI-007, or placebo. The 549 randomized subjects will result in approximately 471 evaluable subjects (157 subjects/treatment group), assuming an approximate 14% early discontinuation rate before a post-dose assessment in the primary efficacy endpoint. The randomization will be stratified by Bipolar I or Bipolar II diagnoses at screening.

Study participation will last approximately 10 weeks (up to 9 visits) and will include the following periods: a *Screening Period,* lasting up to 2 weeks; an *On-Treatment Period* of 6 weeks of daily administration of study medication, and a *Safety Follow-up Period* of approximately 2 weeks after last dose of study medication for subjects not rolling over into Part B upon completing treatment in Part A. Subjects who will roll over directly into Part B will not have a safety follow-up visit after completing Part A. The timing and assessments during each study period are described in the Schedule of Events (Table 6.1 of the protocol and Appendix 1).

An adaptive design with a single interim analysis and a final analysis will be utilized in Part A of this study. An interim analysis will be performed after 375 subjects (125 subjects/treatment group) completed the 6-week On-Treatment Period or confirmed to have discontinued treatment or study after first post baseline MADRS assessment in Part A. The interim data will be used to evaluate the assumptions on variability and effect size, and to adjust the sample size if needed.

[redacted]

[redacted]

5.2. FINAL ANALYSIS

The analyses detailed in this SAP will be performed by [redacted] following [redacted] this SAP, Database Lock, Sponsor authorization of Analysis Sets and Unblinding of Treatment.

6. ANALYSIS SETS

For Part A, 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 interim and final analyses (and approved by ITCI) based on blinded data review.

Note: Analysis sets defined in sections below are defined for Part A of the study.

6.1. ALL SUBJECTS ENROLLED [ENR] SET

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

6.2. ALL SUBJECTS RANDOMIZED [RND] SET

The All Subjects Randomized (RND) Set will contain all subjects who signed informed consent and were randomized to study medication.

For analyses and displays based on RND, subjects will be classified according to randomized treatment.

6.3. INTENT-TO-TREAT [ITT] SET

The intent-to-treat (ITT) Set will contain all randomly assigned subjects who receive at least one dose of study medication and who have a valid (pre-dose) baseline measurement and at least one valid post-baseline measurement of rater-administered MADRS total score. All analyses using the ITT Set will classify subjects according to randomized treatment, regardless of the treatment received during the course of the study. Note that the clinical study protocol refers to this population as the Efficacy analysis set (EAS).

[redacted]

6.5. SAFETY ANALYSIS [SAF] SET

The safety analysis (SAF) Set will contain all subjects who received at least one dose of study medication. All analyses using the SAF Set will classify subjects 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 double-blind study medication in Part A, and will be used to show the start and/or stop days of treatment, 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 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 event 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 for Part A. 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 (if before study Part B start). [redacted]

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

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

Study Phase	Study Period	Range of Relative Study Days	Scheduled Target Day	Scheduled StudyWeek
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
Study Treatment		33 to 39 days relative to Day 1 date	Day 36	5
Study Treatment		≥40 days relative to Day 1 date and afterthe last dose of treatment and before the start of the Safety Follow-up Period (one day after actual Day 43 visit) or Part B study.	Day 43	6
Post-treatment**	Safety Follow-up	> 43 days relative to Day 1 date for treatment completers (42 day on-treatment) and after Day 43 post-treatment assessments.	Day 57	8

** Only for subjects not enrolling over into Part B study phase upon completing treatment in Part A.

First treatment with study medication and baseline assessments are scheduled for Visit 2 on Relative Study Day 1. For analysis purposes, baseline is defined as the last non-missing pre-treatment measurement and endpoint defined as the last Part A On-treatment measurement.

Assessments will be considered baseline if the assessment date is before the date of the first treatment of the 6-week On Treatment Period or if the assessment 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. Bracket assessments originally scheduled for Day -1 that are rescheduled for Day 1 and are confirmed as taken prior to first dose will be considered baseline.

[redacted]

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

Change from baseline 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 6 ITT subjects will be pooled as follows: The largest site with fewer than 6 ITT subjects will be pooled with the smallest site with fewer than 6 ITT subjects. If this results in a pooled site still having fewer than 6 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 12 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 the site effect will be reconsidered and may be dropped from the model. 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 listings.

The default significance level for statistical tests will be 5%; confidence intervals 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 rater-

administered MADRS, will be derived from 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 Part A of this study is to evaluate the effect of ITI-007 compared to placebo if the treatment is administered for the planned study duration. In order to evaluate this estimand in the presence of subjects that may discontinue treatment prematurely, the primary efficacy analyses will be performed based on an 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 treatment group for subjects who have similar history, i.e., the distribution of the unobserved future outcomes for subjects who had discontinued treatment is the same as the distribution of the observed outcomes for those who continued treatment, conditional on the available data prior to discontinuation.

The Mixed-Effects Model for repeated Measures (MMRM) will be used for the analysis of the primary efficacy endpoint in Part A. 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]

[redacted]

7.3. EXAMINATION OF SUBGROUPS

The following subgroups will be assessed as part of the exploratory analyses for Part A:

- Bipolar I or Bipolar II;
- Severity of illness determined by MADRS score at Screening (Mild or Absent (score of 0 to 19), Moderate or Severe (score of 20 to 60));
- Age category (≤ 40, > 40 years);
- Gender;
- Race (White or Non-White);
- Ethnicity (Hispanic or Latino or Not Hispanic or Latino);
- Study Site;
- Age at first diagnosis (<22, >=22 years)

- Number of lifetime bipolar depressive episodes (1-9, 10-20, >20 episodes).

Other subgroup analyses may be performed as deemed appropriate.

8. OUTPUT PRESENTATIONS

The tables, listings and figures (TLFs) shells provided with this SAP describe the presentations for the double-blind portion (Part A) of 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 patients 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 patient data listings.

9. DISPOSITION AND WITHDRAWALS

Subject disposition and withdrawals, will be presented for the ENR Set.

Subject disposition and withdrawals will be presented by treatment group, when applicable, and overall. The number and percentage of subjects who were screened, screen failed, randomized, completed or discontinued the 6-week On-Treatment Period, reasons for early treatment discontinuation, completed or withdrew from Part A of the study, and reasons for study withdrawal will be presented. For subjects who completed the Part A treatment period, the number who enter safety follow-up, the number who enter Part B, and the number who directly roll over to Part B will be summarized. 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 to ≤ Day 15, > Day 15 to ≤ Day 22, > Day 22 to ≤ Day 29, > Day 29 to ≤ Day 36, > Day 36 to ≤ Day 43). The reasons for study withdrawal are listed in Table B. The same reasons apply to treatment discontinuation except for lost to follow-up

and screen failure.

A subject is defined to have completed the Part A treatment period if the subject has completed the 6-week On-Treatment Period and procedures.

A subject is defined to have completed Part A of the study if the subject has completed the 6-week On-Treatment Period and the End-of-Study assessments on Study Day 57 (±2) or directly rolled over to Part B.

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

eCRF Terminology
Adverse event Adverse event associated with worsening of bipolar depression Adverse event not associated with worsening of bipolar depression
Death
Lack of efficacy
Lost to follow-up
Protocol violation
Physician decision
Screen Failure
Study terminated by sponsor
Subject withdrew consent: Personal or family reasons Refused to provide a reason and refused or End of study procedures Self-reported adverse event Self-reported lack of efficacy
Other

Adverse event preferred terms associated with worsening of bipolar depression will be

identified by the 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 will be defined in days for those subjects who are randomized and receive at least one dose of study medication but discontinue study medication prior to Day 43 as the date of discontinuation minus date of first dose of study medication plus 1. Subjects who complete the On-Treatment Period 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 discontinuation between each treatment group and the placebo group. The same analysis will be repeated for time to treatment discontinuation for any reason, where only subjects who complete the On-Treatment Period will be censored.

The number of subjects randomized by site will be summarized by treatment group and overall for all randomized subjects.

[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 Set, [redacted] and SAF Set.

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

The following demographic and other baseline characteristics will be reported for this study:

Demographics

- Age (years) calculated as (Date of Informed Consent – Date of Birth)/365.25;
- Gender;
- Race;
- Ethnicity;

Other Baseline Characteristics

- Waist circumference (cm);
- Weight (kg);
- Height (cm);
- Body Mass Index (BMI) (kg/m^2), where BMI (kg/ m^2) = weight (kg) / height $(m)^2$

Consumption Habits

- Alcohol consumption;
- Caffeine consumption;
- Tobacco use;

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;

Bipolar Disorder Baseline Efficacy

- Baseline efficacy parameters, including rater-administered MADRS total score, CGI-BP-S score, SDS total score, Q-LES-Q-SF, WHO-5 Well-being Index, and NEO-FFI;

Bipolar Disorder Baseline Safety

- Baseline safety parameters, including BARS total score, BARS global assessment of akathisia, SAS total score, AIMS total score, YMRS total score, and C-SSRS assessment of suicidal behavior and ideation;

12. MEDICAL HISTORY

Medical History information will be presented for the SAF Set.

Medical History will be coded using Medical Dictionary for Regulatory Activities (MedDRA) central coding dictionary currently in effect at the time of interim analysis or final Part A database lock.

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. Within 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 Who Drug Dictionary.

See Appendix 2 for handling of partial dates for medications, in the case where it is not possible to define a medication as prior, prior concomitant, concomitant, or post-treatment, the medication will be classified by the worst case, i.e. concomitant.

For Part A, 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 all study procedures and assessment related to the last dose of study medication.

- Part A 'Prior' medications are medications which started and stopped prior to the date of first dose of study medication.
- Part A 'Prior concomitant' medications are medications that started prior to and stopped after the first dose of study medication.
- Part A 'Concomitant' medications are medications that:
    - started on or 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 after the last dose of study

medication (if planned assessments are not performed).

- Part A 'Post-treatment' medications are medications that started after completion of Day 43 (or last treatment day for subjects who discontinue treatment) assessments that are planned to be conducted after the last dose of study medication and before the start of Part B. Post-treatment medications will only be summarized for subjects not enrolling over into Part B study phase upon completing treatment in Part A.

Prior, prior concomitant, concomitant, and post-treatment medication use will be summarized separately 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 if taken no more than 3 times per week and only during the Screening Period and first 2 weeks of the On-Treatment Period. The number and percent of subjects in the ITT Set receiving zolpidem and the total number of days on zolpidem will be summarized by treatment group for the Screening Period, for each week on study treatment (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 in Part A will be presented for the SAF Set and ITT Set.

The date of first and last study medication administration will be taken from the eCRF 'Dosing' form. If multiple entries occurred on this form, the earliest 'start date of study medication' and the latest 'stop date of study medication' will be used.

Duration of exposure (days) will be calculated as (date of last dose of study medication) – (date of first dose of study medication) + 1. 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.

For each subject, dosing compliance (%) will be calculated as 100 x (number of compliant days) / (number of days in the On-Treatment Period). A day within the On-Treatment Period will be considered a compliant day if the subject had 2 capsules. The "Number of Days in the On-Treatment Period" is equivalent to the 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 ENDPOINT AND DERIVATION

For the double-blind portion (Part A), the primary efficacy variable is change from baseline in rater-administered MADRS total score to Week 6 (Day 43). 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.

In addition to the rater-administered MADRS assessment, a computer-administered MADRS assessment will be conducted. At screening, the computer-administered MADRS assessment will be conducted prior to the rater-administered one. At all other visits, the rater-administered MADRS assessment will be conducted first. Computer-administered MADRS scores will be compared to the ones from the rater-administered MADRS assessment on an ongoing basis as part of a remote quality assurance program.

15.1.2. MISSING DATA METHODS FOR PRIMARY EFFICACY ENDPOINT

The primary analysis method for evaluating the primary efficacy endpoint is the likelihood-based analysis of repeated measures, MMRM, described in details in Section 15.1.3. The use of MMRM inherently implies that the treatment effect on the change from baseline in the rater-administered MADRS total score will be similar for the subjects who withdraw and for those who complete the study in their respective treatment groups, conditional on the outcomes observed prior to withdrawal (MAR assumption). To challenge the robustness of the MAR assumption, sensitivity analyses which utilize multiple imputations and a different assumption about unobserved outcomes will be performed, as detailed in Section 15.1.4.

15.1.3. PRIMARY ANALYSIS OF PRIMARY EFFICACY ENDPOINT

The treatment effect on the Part A primary efficacy endpoint will be evaluated using an MMRM model. The model will include the change from baseline at each pre-specified timepoint as the response variable and study visit, the bipolar disorder stratification variable, baseline MADRS total score, site (or pooled site), baseline MADRS total score-by-study visit interaction, treatment (ITI-007 40 mg, ITI-007 60 mg, placebo), and treatment-by-study visit interaction. An unstructured covariance matrix will be used to model the correlation among repeated measurements within subject. [redacted]

[redacted] The Kenward-Rogers method will be used to estimate the denominator degrees of freedom. The treatment and treatment-by-time interaction terms allow for comparisons of the treatment groups at each of the following time points: Weeks 1, 2, 3, 4, 5, and 6 (Days 8, 15, 22, 29, 36, and 43). Treatment differences will be evaluated via contrasts for the time-by-treatment factor. This methodology will be used to compare the treatment groups versus placebo for change from baseline at each time point.

The change from baseline to Week 6 (Day 43) will be used for the primary efficacy analysis, and will be performed for the ITT Set [redacted].

[redacted]

Estimates of model parameters will be presented, as well as least-squares mean (LSM) estimates for change from baseline in MADRS score, standard errors and 95% confidence intervals (CI) for LSMs will be presented by treatment group and time point. Contrast estimates (LSMs) for between-group comparisons (ITI-007 60 mg vs. placebo and ITI-007 40 mg vs. placebo), the corresponding standard errors, 95% CIs, effect sizes, and p-values will be presented for each visit. Effect size will be calculated for each ITI-007 dose group (vs. placebo) as $\frac{\text{LS Mean Difference}}{\text{Pooled estimate of within patient error standard deviation}}$. [redacted]

[redacted]

15.2. KEY SECONDARY EFFICACY

For Part A, the key secondary efficacy endpoint is time to first sustained response, defined as the number of days from first dose of study drug to the earliest date the patient experiences a sustained ≥50% reduction from baseline in the rater-administered MADRS total score (i.e. achieve responder status at ≥ 2 consecutive visits and continue to last assessment). Subjects who do not experience at least 50% reduction from baseline in the rater-administered MADRS total score, or do not maintain it for ≥ 2 consecutive visits or did not continue to last

assessment will be censored. Subjects who meet the ≥50% reduction from baseline on Day 43 for the first time will be considered a sustained responder. The corresponding censored time will be calculated using last visit where the MADRS is assessed. Time to the start of sustained response will be analyzed using Cox's proportional hazards model with terms for baseline MADRS score, site (or pooled site), treatment and the bipolar disorder stratification variable. A Kaplan-Meier graph will be presented, showing a separate line for each treatment group. An estimate and 95% CI for the hazard ratio between each dose of ITI-007 and placebo will be presented. The proportion of sustained responders and descriptive statistics for the time to start of sustained response will also be summarized.

The key secondary efficacy analysis will be performed for the ITT Set [redacted].

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

16. SAFETY OUTCOMES

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

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.

16.1. ADVERSE EVENTS

Adverse Events (AEs) will be coded using MedDRA central coding dictionary currently in effect at the time of interim analysis or final Part A database lock.

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

All AEs with an onset date after the last dose of study medication plus one day, but before the start of Part B 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, the AE will be classified as the worst case, i.e., treatment emergent.

An overall summary of the 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 adverse events, TEAEs and Non-TEAEs.

16.1.1. AEs AND TEAEs

The incidence of TEAEs will be presented by SOC and 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 having events for each PT and SOC will be summarized. A summary of TEAEs will be provided with only PTs and a separate summary for only SOCs. An additional summary of TEAEs will be provided for PTs occurring in at least 5% of subjects in any treatment group (ITI-007 60 mg, ITI-007 40 mg, or placebo). This summary will be repeated for TEAEs possibly related (possibly, probably, or definitely) to study drug. Within each subject, multiple reports of events that map to a common MedDRA PT and/or SOC will be condensed into a single AE for incidence counts. Summaries will be presented by treatment group and in decreasing frequency by decreasing dose group (ITI-007 60 mg, ITI-007 40 mg) and placebo group. AEs which occur during the follow-up period will be summarized by PT and SOC and treatment group.

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.

Intensity is classed as "not specified", "mild", "moderate", "severe", or "life-threatening" (increasing severity). 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, as indicated by the Investigator, is classed as "not specified", "unrelated", "unlikely to be related", "possibly related", "probably related", or "definitely related" (increasing severity of relationship). A "related" TEAE is defined as a TEAE with a relationship to study medication as "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 early withdrawal or discontinuation of study medication will be identified by using the Adverse Events page of the eCRF, where item 'Action taken with study treatment' indicates permanent discontinuation of study medication, i.e., "Drug withdrawal". AEs leading to early study withdrawal will be identified by using the Adverse Events page of the (e)CRF, where the answer to item 'Did the AE cause the subject to discontinue from the study?' indicates "Yes".

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 (SAEs) are those events recorded as "Serious" on the Adverse Events page of the eCRF. A summary of SAEs by SOC and PT will be prepared.

AEs leading to death are those events which are recorded as "Fatal" on the Adverse Events page of the eCRF. 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.

16.1.3. EXTRAPYRAMIDAL TEAES

Treatment-emergent extrapyramidal adverse events will be defined by the Standard MedDRA Query (SMQ) labeled as Extrapyramidal Syndrome. The number and percentage of subjects with at least one AE PT mapped to an extrapyramidal term contained in the SMQ will be presented. The extrapyramidal terms will be summarized as well. 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 along with a summary of the PTs below.

Table D: Abuse-Related PTs

Category	Applicable PTs
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, and urinalysis. A list of laboratory assessments to be included in the outputs is included in the protocol, Section 6.5. Summary statistics will be presented in SI and US conventional units. Shift tables and abnormalities based on normal range criteria will be presented using SI units only.

Local laboratory tests will be utilized if necessary for this study. Local laboratory results will be converted to the same units as central laboratory tests and summarized together with the central results. In cases where both central and local laboratory results are available for the same time point, the central result will be used for the by-visit summaries. When central results are unavailable, local results will be summarized.

Quantitative laboratory measurements reported as "< X", i.e. below the 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.

All laboratory samples are to be collected after an overnight fast. Results for glucose, insulin, and triglycerides collected non-fasting will be excluded from descriptive statistics but will appear in the data listings.

The following summaries will be provided for laboratory data:

- Actual values and change from baseline by visit (for quantitative measurements)
- Incidence of abnormal values according to normal range criteria
- Shifts from baseline according to normal range criteria (for quantitative and categorical measurements)
- Shifts from baseline according to the markedly abnormal criteria defined in Section 16.2.1 (for quantitative and categorical measurements)
- Listing of laboratory values meeting markedly abnormal criteria
- Incidence of liver function-related values meeting pre-defined criteria as defined in Section 16.2.2 during the On-Treatment Period and the entire Part A 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 the table below.

Table E: Markedly Abnormal Values for Laboratory Evaluations

Haematology Parameter	Markedly Abnormal Range
Hemoglobin	Male: ≤ 11.5 g/dL
	Female: ≤ 9.5 g/dL
Hematocrit	Male: ≤ 37%

	Female: ≤ 32%
WBC	≤ 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
GGT	≥ 3 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

HbA_{1c}	≥ 7.5%
Prolactin	≥ 5 x ULN
Total Cholesterol	≥ 300 mg/dL
LDL Cholesterol	≥ 200 mg/dL
Triglycerides	≥ 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 listed below 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 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

Results from the central Electrocardiogram (ECG) Reading will be included in the reporting of this study.

The following ECG parameters will be reported for this study:

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

ECGs which are collected in triplicate will beanalyzed as an average of the non-missing measurements at each time point. For the overall assessment, there are five possible results for each triplicate from the central cardiologist: '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 summarized according to the worst non-missing assessment and categorized in the following order, 'Abnormal Significant', 'Abnormal, Insignificant' and 'Normal'. For single ECGs, the measurements and assessments will be analyzed as collected according to the categories described previously.

The single or average of the three ECG recordings corresponding to the triplicate collected pre-dose at Visit 2 on Day 1 will be considered baseline.

The following summaries will be provided for ECG data:

- Actual and change from baseline by visit (for quantitative measurements)
- Incidence of abnormal and normal ECGs by visit
- Shifts from baseline to each visit and time point according to markedly abnormal criteria (for quantitative measurements and categorical measurements)
- Incidence of markedly abnormal results by categories defined in Section 16.3.1
- Listing of ECG values 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 interval, QTcB interval, and QTcF will be classified as:

- >30 msec increase from baseline
- >60 msec increase from baseline

16.4. VITAL SIGNS

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

- Systolic blood pressure (mmHg) after 10 minutes in supine position, 1 minute sitting, immediately upon standing and 3 minutes after standing
- Diastolic blood pressure (mmHg) after 10 minutes in supine position, 1 minute sitting, immediately upon standing and 3 minutes after standing
- Pulse rate (bpm) after 10 minutes in supine position, 1 minute sitting, immediately upon standing and 3 minutes after standing
- Respiratory rate (breaths/min)

- Oral temperature (°C)
- Weight (kg)
- BMI (kg/m^2) [derived – see Section 11]
- Waist circumference (cm)

Vital signs, including systolic blood pressure, diastolic blood pressure, pulse rate, and respiratory rate, are collected on pre-specified study days. The assessment at Visit 2 on Day 1 will be considered baseline.

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 signs data:

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

16.4.1. VITAL SIGNS MARKEDLY ABNORMAL CRITERIA

Markedly abnormal quantitative vital sign 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. PHYSICAL EXAMINATION

The following summaries will be provided for physical examination data:

- Incidence of abnormalities at screening
- Incidence of abnormalities post baseline at Weeks 3, 6, and 8 (Days 22, 43, and 57) (Follow up/early discontinuation)

16.6. OTHER SAFETY ASSESSMENTS

16.6.1. 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 score will not be calculated.

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

16.6.2. 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 will be summarized by treatment and visit.

16.6.3. SIMPSON-ANGUS SCALE (SAS)

The SAS is a measure of extrapyramidal side effects consisting of 10 items: gait, arm dropping, shoulder shaking, elbow rigidity, wrist rigidity, leg pendulousness, head rotation, glabella tap, tremor, and salivation. 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 will be summarized by treatment and visit.

16.6.4. 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.

16.6.5. 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 trial. 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 the investigator or an expert site-based rater.

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]

[redacted]

18. REFERENCES

Brannath, W., Posch, M., Bauer, P. Recursive combination tests. Journal of the American Statistical Association. 2002; 97: 236-244.

Chang, M. Adaptive Design Theory and Implementation Using SAS and R. Chapman & Hall/CRC, Boca Raton, Florida. 2008.

Cui, L., Hung, H.M., Wang, S.J. Modification of sample size in group-sequential clinical trials. Biometrics. 1999; 55: 853-857.

EMA. (n.d.). Guideline on Missing Data in Confirmatory Clinical Trials. Retrieved December 15, 2011, from European Medicines Agency: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500096793.pdf

Hochberg, Y. A sharper Bonferroni procedure for multiple tests of significance. Biometrika. 1988; 75, 800-2.

Hommel, G. A stagewise rejective multiple test procedure based on a modified Bonferroni test. Biometrika. 1988; 75, 383-386.

Jennison, C., Turnbull, B.W. Group Sequential Methods with Applications to Clinical Trials. Boca Raton: Chapman and Hall, 2000.

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

Kordzakhia, G., Dmitrienko, A., Ishida, E. Mixture-based gatekeeping procedures in adaptive clinical trials. Journal of Biopharmaceutical Statistics. 2018; 28:129-145.

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 Intention-to-treat analysis for longitudinal studies with drop-outs. Biometrics 1996; 52: 1324-1333.

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

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

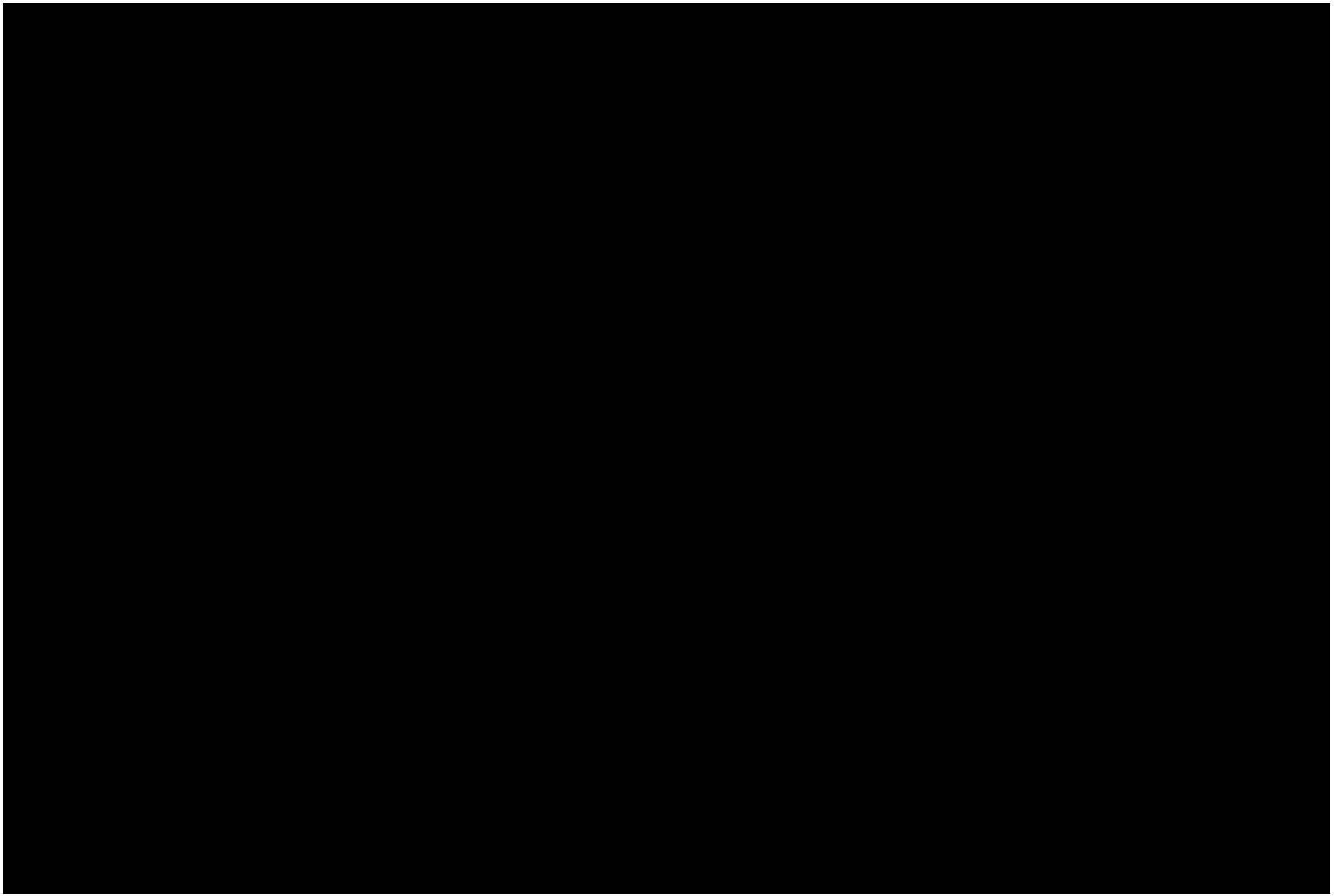
Permutt T (2016a). A taxonomy of estimands for regulatory clinical trials with discontinuations. Statist. Med. 35:2865–2875

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

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(6):337-47Simes, R.J. An improved Bonferroni procedure for multiple tests of significance. Biometrika. 1986; 73, 751-754.

Rubin, D. B. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons, 1987.

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.



APPENDIX 2. PARTIAL DATE CONVENTIONS

Imputed dates will NOT be presented in the listings.

ALGORITHM FOR TREATMENT EMERGENCE OF ADVERSE EVENTS

START DATE	STOP DATE	ACTION
Known	Known	If start date < study medication start date or start date>study medication end date + 1 day , then not TEAE If study medication start date ≤start date ≤ study medication end date + 1 day, then TEAE
	Partial	Impute stop date as latest possible date (i.e. last day of month if day unknown or 31st December if day and month are unknown), then: If start date < study med start date or start date>study med end date + 1 day, then not TEAE If study med start date ≤start date ≤ study med end date + 1 day, then TEAE
	Missing	If start date < study med start date or start date>study med end date + 1 day, then not TEAE If study med start date ≤start date ≤ study med end date + 1 day, then TEAE
Partial, but known components show that it cannot be on or after study medication start date or cannot be before study medication end date	Known	Not TEAE
	Partial	Not TEAE

START DATE	STOP DATE	ACTION
	Missing	Not TEAE
Partial, could be on or after study medication start date but before study med end date	Known	If stop date < study med start date, then not TEAE If stop date ≥ study med start date, then TEAE
	Partial	Impute stop date as latest possible date (i.e. last day of month if day unknown or 31st December if day and month are unknown), then: If stop date < study med start date, then not TEAE If stop date ≥ study med start date, then TEAE
	Missing	Assumed TEAE
Missing	Known	If stop date < study med start date, then not TEAE If stop date ≥ study med start date, then TEAE
	Partial	Impute stop date as latest possible date (i.e. last day of month if day unknown or 31st December if day and month are unknown), then: If stop date < study med start date, then not TEAE If stop date ≥ study med start date, then TEAE
	Missing	Assumed TEAE

ALGORITHM FOR PRIOR / CONCOMITANT MEDICATIONS

START DATE	STOP DATE	ACTION
Known	Known	If stop date < study med start date, assign as prior If stop date ≥ study med start date and start date < study med start date, assign as prior concomitant If stop date ≥ study med start date and start date ≤ end of treatment, assign as concomitant If stop date ≥ study med start date and start date > end of treatment, assign as post study
	Partial	Impute stop date as latest possible date (i.e. last day of month if day unknown or 31st December if day and month are unknown), then: If stop date < study med start date, assign as prior If stop date ≥ study med start date and start date < study med start date, assign as prior concomitant If stop date ≥ study med start date and start date ≤ end of treatment, assign as concomitant If stop date ≥ study med start date and start date > end of treatment, assign as post study
	Missing	If stop date is missing could never be assumed a prior medication If start date ≤ end of treatment, assign as concomitant If start date > end of treatment, assign as post treatment
Partial	Known	Impute start date as earliest possible date (i.e. first day of month if day unknown or 1st January if day and month are unknown), then: If stop date < study med start date, assign as prior If stop date ≥ study med start date and start date < study med start date, assign as prior concomitant If stop date ≥ study med start date and start date ≤ end of treatment, assign as concomitant If stop date ≥ study med start date and start date > end of treatment, assign as post study

START DATE	STOP DATE	ACTION
	Partial	Impute start date as earliest possible date (i.e. first day of month if day unknown or 1st January if day and month are unknown) and impute stop date as latest possible date (i.e. last day of month if day unknown or 31st December if day and month are unknown), then: If stop date < study med start date, assign as prior If stop date ≥ study med start date and start date < study med start date, assign as prior concomitant If stop date ≥ study med start date and start date ≤ end of treatment, assign as concomitant If stop date ≥ study med start date and start date > end of treatment, assign as post study
	Missing	Impute start date as earliest possible date (i.e. first day of month if day unknown or 1st January if day and month are unknown), then: If stop date is missing could never be assumed a prior medication If start date ≤ end of treatment, assign as concomitant If start date > end of treatment, assign as post treatment
Missing	Known	If stop date < study med start date, assign as prior If stop date ≥ study med start date, assign as concomitant Cannot be assigned as 'post treatment'
	Partial	Impute stop date as latest possible date (i.e. last day of month if day unknown or 31st December if day and month are unknown), then: If stop date < study med start date, assign as prior If stop date ≥ study med start date, assign as concomitant Cannot be assigned as 'post treatment'
	Missing	Assign as concomitant

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

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]