- **Protocol number: 331-201-00081**
- **Document title: A Multicenter, Randomized, Double-blind Trial of Brexpiprazole Versus Placebo for the Acute Treatment Manic Episodes, With or Without Mixed Features, Associated With Bipolar I Disorder**
- **Version number: 2.0**
- **Date of the document: 11Dec2018**
- **NCT number: NCT03257865**

Otsuka Pharmaceutical Development & Commercialization, Inc.

Investigational New Drug

Brexpiprazole (OPC-34712)

A Multicenter, Randomized, Double-blind Trial of Brexpiprazole versus Placebo for the Acute Treatment of Manic Episodes, With or Without Mixed Features, Associated With Bipolar I Disorder

Protocol No. 331-201-00081
IND No. 134115
EudraCT No. 2017-002222-20

Statistical Analysis Plan

Version: Final

Date: December 11, 2018

CONFIDENTIAL – PROPRIETARY INFORMATION

Table of Contents

1 Introduction

This statistical analysis plan (SAP) documents the statistical methodology and data analysis algorithms and conventions to be applied for statistical analysis and reporting of efficacy and safety data of study 331-201-00081. All amendments to the protocol are taken into consideration in developing this SAP.

2 Study Objectives

Primary: To demonstrate the efficacy of brexpiprazole for the acute treatment of manic episodes, with or without mixed features, in subjects with a diagnosis of bipolar I disorder.

Secondary: To confirm the safety and tolerability of brexpiprazole in this same population.

3 Trial Details

3.1 Study Design

This will be a 3-week, multicenter, randomized, double-blind, placebo-controlled trial of brexpiprazole in subjects diagnosed with bipolar I disorder (current manic episode with or without mixed features) according to the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5; updated terminology replaces "mixed episodes" with the descriptor of "mixed features"). The total duration of the trial is up to 8 weeks, including screening, double-blind treatment, and follow-up. See Figure 3.1-1 for a schematic of the trial design.

The trial will be organized as follows:

Screening Phase: The screening period will begin after written informed consent has been obtained and will take place between Day −14 and Day −1 prior to randomization. Hospitalization will begin with the signing of the informed consent form (ICF) for subjects who are not already hospitalized at the time of the initial screening visit. The purpose of the screening period is to assess eligibility criteria and to washout prohibited concomitant medication. Subjects will be between 18 and 65 years of age, inclusive, at the time of screening, will have a diagnosis of bipolar I disorder, and will be experiencing a manic episode with or without mixed features as defined by the DSM-5 criteria. An interactive web response system (IWRS) will be used to obtain an identification (ID) number for each subject with documented consent. Although the screening period will continue up to administration of the first dose of investigational medicinal product (IMP), screening procedures should be

initiated with a sufficient amount of time allotted in order to obtain laboratory results and electrocardiogram (ECG) results from the central reader prior to randomization. The sponsor reserves the right to utilize external quality oversight methods to ensure the validity of diagnosis, severity of illness, and other factors determining appropriateness of subject selection.

All subjects must agree to discontinue all prohibited medications during the screening period in order to meet the protocol-specified washout periods.

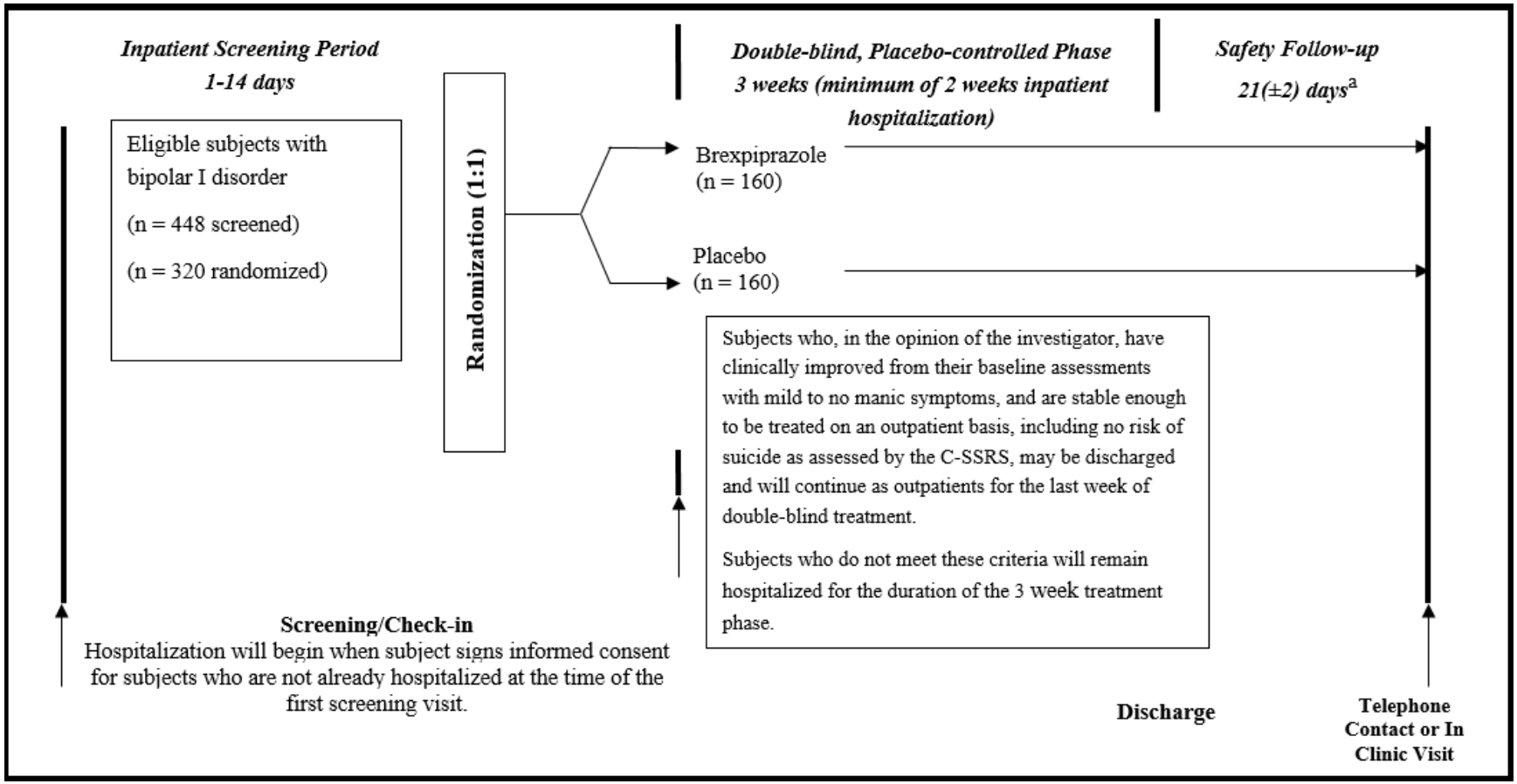
Treatment Phase: Following the screening period, subjects who meet all inclusion criteria, including a score of ≥ 24 on the Young-Mania Rating Scale (YMRS) at screening and baseline, and meet none of the exclusion criteria, will be randomized in a 1:1 ratio to receive either placebo or brexpiprazole for 3 weeks. Subjects will receive a starting dose of 2 mg/day of brexpiprazole (or corresponding placebo) from Days 1 to 3, followed by titration to 3 mg/day brexpiprazole (or placebo) on Day 4. Further dose increases up to 4 mg/day may occur no earlier than Day 7 at the investigator's discretion based on treatment response. Subjects who are unable to tolerate their current dose can be titrated down at any time to a minimum of 2 mg/day. Subjects may be re-titrated back to higher dose levels based on investigator discretion. Dose adjustments must be made in increments of 1 mg/day. Subjects who are unable to tolerate 2 mg/day brexpiprazole will be discontinued from the trial. All subjects will remain hospitalized for a minimum of the first 2 weeks of the 3-week treatment phase. However, subjects who, in the opinion of the investigator, have clinically improved from their baseline assessments with mild to no manic symptoms, and are stable enough to be treated on an outpatient basis, including no risk of suicide as assessed by the Columbia-Suicide Severity Rating Scale (C-SSRS), may be discharged and will continue as outpatients for the last week of double-blind treatment. Subjects who do not meet these criteria will remain hospitalized for the duration of the 3-week treatment phase.

A drug screen and alcohol test are required at the Day 21/Early Termination (ET) visit for subjects who are discharged from the hospital at the end of Week 2 (Day 14) to verify continued compliance with the protocol.

Subjects will be evaluated at Baseline (Day 1), Day 4, Day 7, Day 14 and Day 21 during the double-blind treatment phase.

Follow-up Phase: If any subject discontinues the trial early, every effort should be made to complete the Day 21/ET assessments as soon as possible and prior to starting any new medication or treatment. Subjects who complete all trial visits through the Day 21 visit may be offered entry into an optional open-label rollover trial. Subjects who do not enter the open-

label trial will be followed up for safety reasons via telephone contact or in clinic visit 21 (± 2) days after the last dose of IMP. This contact also applies to subjects who are withdrawn prematurely from the trial.



C-SSRS = Columbia-Suicide Severity Rating Scale.

[a]Subjects who complete all trial visits through the Day 21 visit may be offered entry into an optional open-label rollover trial.

Figure 3.1-1 Trial Design Schematic

3.2 Trial Treatments

During the double-blind treatment phase, subjects will receive IMP consisting of brexpiprazole monotherapy or placebo, depending on the subject's treatment assignment.

As shown in Table 3.2-1, subjects will receive a starting dose of 2 mg/day of brexpiprazole (or corresponding placebo) from Days 1 to 3, followed by titration to 3 mg/day brexpiprazole (or placebo) on Day 4. Subjects may be titrated (or re-titrated) to a higher dose of brexpiprazole (or placebo), up to a maximum of 4 mg/day, based on treatment response and at the investigator's discretion anytime at Day 7 or thereafter. Subjects who are unable to tolerate their current dose can be titrated down at any time to a minimum of 2 mg/day. Dose adjustments must be made in increments of 1 mg/day. Subjects who are unable to tolerate 2 mg/day brexpiprazole will be discontinued from the trial.

Table 3.2-1 Dosing Schedule

Dose	Days 1-3	Days 4[a]	Days 7-21[b]
Brexpiprazole	2 mg	3 mg	2-4 mg
Placebo	Placebo	Placebo	Placebo

[a]Down titration can occur at any time due to tolerability after Day 4. The minimum dose allowed is 2 mg/day.

[b]Option to titrate 2 to 4 mg (i.e., 2 mg, 3 mg, or 4 mg) based on clinical response and tolerability; changes must occur in 1 mg/day increments. Increases up to 4 mg/day may occur no earlier than Day 7.

All doses of IMP should be taken at the same time each day, if possible, and can be taken without regard to meals. If tolerability issues arise, the timing of administration of the IMP may be adjusted at the investigator's discretion in order to achieve optimum tolerability and compliance. Subjects will be counseled on the importance of taking the IMP.

4 Sample Size and Power Justification

It is anticipated that approximately 320 subjects will be randomized from an estimated 45 sites in the US and Europe. The primary efficacy endpoint is the change from baseline to Day 21 in the double-blind treatment phase in the YMRS Total Score. The trial will compare the placebo arm to the brexpiprazole arm, randomized at a ratio of 1:1, with an overall significance level of 0.05 (2-sided) for the primary endpoint.

CCI [redacted]

CCI [redacted] The planned sample size of 304 evaluable subjects (152 in each treatment arm) will yield at least 90% power to detect the treatment effects at a 2-tailed significance level of 0.05.

A sufficient number of subjects will be enrolled and randomized to achieve approximately 304 evaluable subjects in the double-blind treatment phase CCI [redacted] the total number of subjects to be randomized is 320 (160 in each treatment arm).

In order to ensure 304 evaluable subjects, CCI [redacted]

5 Data Sets for Analysis and Missing Data

5.1 Data Sets for Analysis

The following analysis samples are defined for this trial:

Enrolled Sample: comprises all subjects who signed an ICF for the trial and enrolled into the trial.

Randomized Sample: comprises all subjects who were randomized in the double-blind treatment phase. Subjects are considered randomized when they are assigned a treatment number by IWRS at the end of screening. A subject receiving IMP outside of the IWRS will not be considered randomized, but safety will be reported.

Safety Sample: comprises those randomized subjects in the double-blind treatment phase who received at least 1 dose of double-blind IMP as indicated on the dosing record. Subjects will be excluded from this population only if there is documented evidence (i.e., drug dispensed = drug returned or no IMP dispensed) that the subject did not take IMP. If a subject is dispensed IMP and is lost to follow up, he/she will be considered exposed.

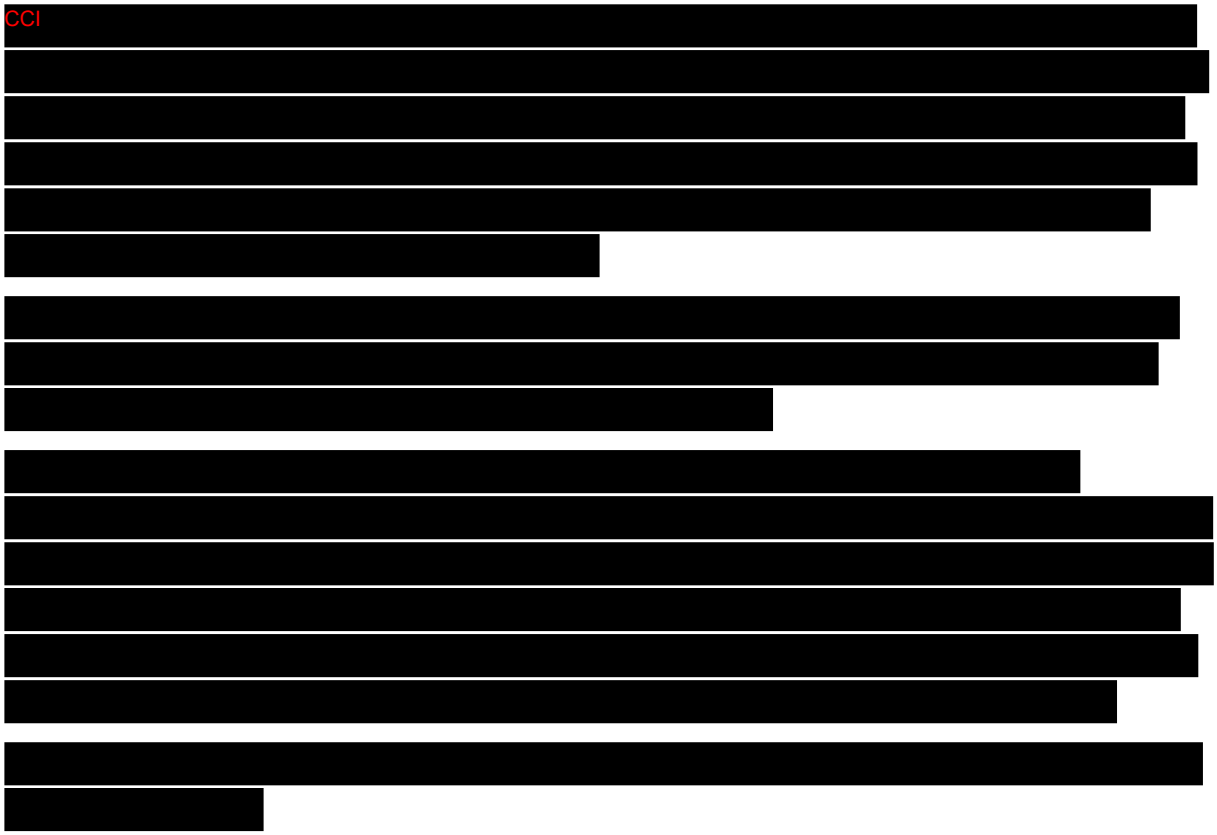
Efficacy Sample: The Full Analysis Set comprises all subjects in the Safety Sample who have a baseline value and at least 1 valid post-randomization efficacy evaluation for YMRS Total Score in the double-blind treatment phase.

5.2 Handling of Missing Data

The YMRS is utilized as the primary efficacy assessment of a subject's level of manic symptoms. The YMRS consists of 11 items: 1) elevated mood, 2) increased motor activity-energy, 3) sexual interest, 4) sleep, 5) irritability, 6) speech (rate and amount), 7) language-thought disorder, 8) content, 9) disruptive-aggressive behavior, 10) appearance, and 11) insight. Seven items are rated on a 0- to 4-scale, while four items (Items 5, 6, 8, and 9) are rated on a 0- to 8-scale with 0, 2, 4, 6, and 8 being the possible scores (twice the weight of the other items). For all items, 0 is the "best" rating and the highest score (4 or 8) is the 'worst' rating. The YMRS Total Score is the sum of ratings for all 11 items; therefore, possible total scores range from 0 to 60. The YMRS Total Score is set to be missing if less than 9 of the 11 items are recorded. If 10 of the 11 items are available and the item missing is from items 5, 6, 8 or 9, then the YMRS Total Score is the sum of scores for items 1 to 4, 7, 10 to 11 plus the mean of the recorded items from 5, 6, 8 and 9 times four. If 10 of the 11 items are available and the item missing is from items 1 to 4, 7, 10 to 11, then the YMRS Total Score is the sum of scores for items 5, 6, 8 and 9 plus the mean of the recorded items from 1 to 4, 7, 10 to 11 times seven. If 9 of the 11 items are available and both missing items are from items 5, 6, 8 and 9, then the YMRS Total Score is set to be missing. If 9 of the 11 items are available and both missing items are from items 1 to 4, 7, 10 to 11, then the YMRS Total Score is the sum of scores for items 5, 6, 8 and 9 plus the mean of the recorded items from 1 to 4, 7, 10 to 11 times seven. If 9 of the 11 items are available and one of the missing items is from items 1 to 4, 7, 10 to 11, and one of the missing items is from items 5, 6, 8 and 9, then the YMRS Total Score is the mean of the recorded items from 5, 6, 8 and 9 times four, plus the mean of the recorded items from 1 to 4, 7, 10 to 11 times seven. All imputed scores are rounded to the first decimal place.

CCI

[redacted]



6 Study Conduct

6.1 Subject Disposition, Completion Rate and Reasons for Discontinuation

Subject disposition will be summarized for the Randomized Sample by the treatment group, and by center.

Subject completion rate and reasons for discontinuation will be summarized for the Randomized Sample by treatment group.

6.2 Treatment Compliance

Based on the Investigational medicinal product (IMP) panel of the CRF, compliance in taking IMP is calculated by dividing the number of tablets/capsules taken by the total number of tablets/capsules the patients were scheduled to take during the study period. For lost-to-follow up patients, last IMP end date record will be used as the treatment end date.

6.3 Protocol Deviation

Protocol deviations will be summarized by center and type of deviation for randomized subjects by treatment group. A listing of protocol deviations will be provided.

7 Baseline Characteristics

7.1 Baseline Definition

For analyses of the double-blind treatment period data, the baseline is the Baseline measurement (expected to be at Day 1). Baseline measurement is defined as the last available measurement prior to the start of double-blind IMP.

7.2 Demographic Characteristics

Baseline demographic characteristics include age, sex, race, ethnicity, height, weight, waist circumference, and body mass index (BMI). For the Randomized Sample, demographic characteristics will be summarized by treatment group.

Mean, range and standard deviation will be used to describe continuous variables such as age. Frequency distributions will be tabulated for categorical variables such as race.

7.3 Medical and Psychiatric History

A summary of medical, psychiatric, and bipolar I disorder history will be presented for the Randomized Sample (by treatment group and overall).

7.4 Neuropsychiatric Diagnosis

A summary of the MINI International Neuropsychiatric Interview (M.I.N.I.) will be presented for the Randomized Sample (by treatment group and overall). Summarized will be the number and percentage of patients who meet each diagnosis criteria, and number and percentage of patients with each primary diagnosis.

7.5 Baseline Psychiatric Evaluation

For the Randomized Sample, baseline psychiatric scale evaluation will be summarized by treatment group and overall. The mean, median, range and standard deviation will be used to summarize the assessments of: YMRS Total Score, Clinical Global Impression – Bipolar

(CGI-BP) severity score in mania, CCI [redacted]

8 Efficacy Analysis

All efficacy analyses pertaining to the double-blind treatment period will be performed on the Efficacy Sample, and patients will be included in the treatment group as randomized.

Statistical comparisons are based on 2-sided, 0.05 significance levels.

8.1 Primary Efficacy Endpoint

The primary efficacy endpoint is the change from baseline to Day 21 of the double-blind treatment phase in YMRS Total Score. For analysis of the double-blind treatment phase data, baseline is defined as the last available measurement prior to the first dose of double-blind IMP.

8.1.1 Primary Efficacy Analysis

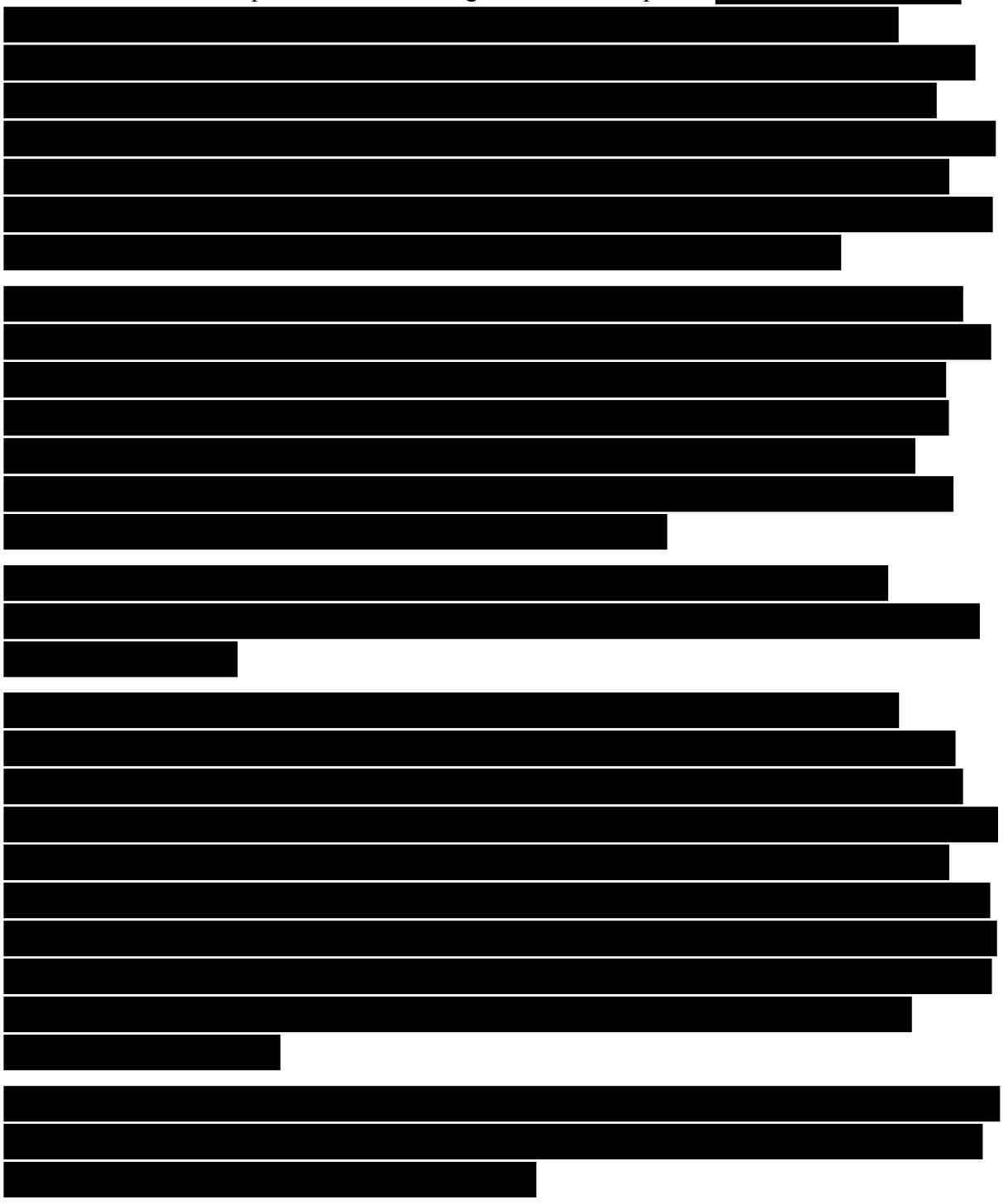
The objective of the primary analysis is to demonstrate the efficacy of brexpiprazole for the acute treatment of manic episodes, with or without mixed features, in subjects with a diagnosis of bipolar I disorder.

The estimand for the primary efficacy analysis is the difference in outcome improvement if all subjects tolerated and adhered to their treatment. This approach is Estimand #3 recommended by the 2010 National Academy of Sciences' National Research Council report on prevention and treatment of missing data and ICH E9 (R1) "Addendum to Statistical Principles for Clinical Trials on Choosing Appropriate Estimands and Defining Sensitivity Analyses in Clinical Trials".

This estimand focuses on the efficacy of brexipiprazole in change from the baseline in YMRS Total Score. The objective of this trial is consistent with the election of an efficacy rather than effectiveness estimand.

The primary analysis will be performed on the Efficacy Sample which includes all randomized subjects who took at least 1 dose of IMP in the double-blind treatment phase and who have both a baseline value and at least 1 post-randomization YMRS Total Score during the double-blind treatment phase. The primary efficacy analysis will be performed by fitting a mixed-effect model repeated measure (MMRM) analysis with an unstructured variance covariance structure in which the change from the baseline in YMRS Total Score during the

double-blind treatment phase will be the dependent variable based on the observed cases (OC) data set. The OC data set will consist of actual observations recorded at each visit during the double-blind treatment phase and no missing data will be imputed. CCI



CCI

CCI

8.1.3.2 CCI

[redacted]

8.1.4 Subgroup Analyses

Subgroup analyses of change from baseline in YMRS Total Score to every study week in the double-blind treatment period will be performed by the following factors:

- Sex (Based on the biological status)
- Race (White and All Other Races)
- Age group (Age<55 and Age≥55)
- Region (North America and Europe)

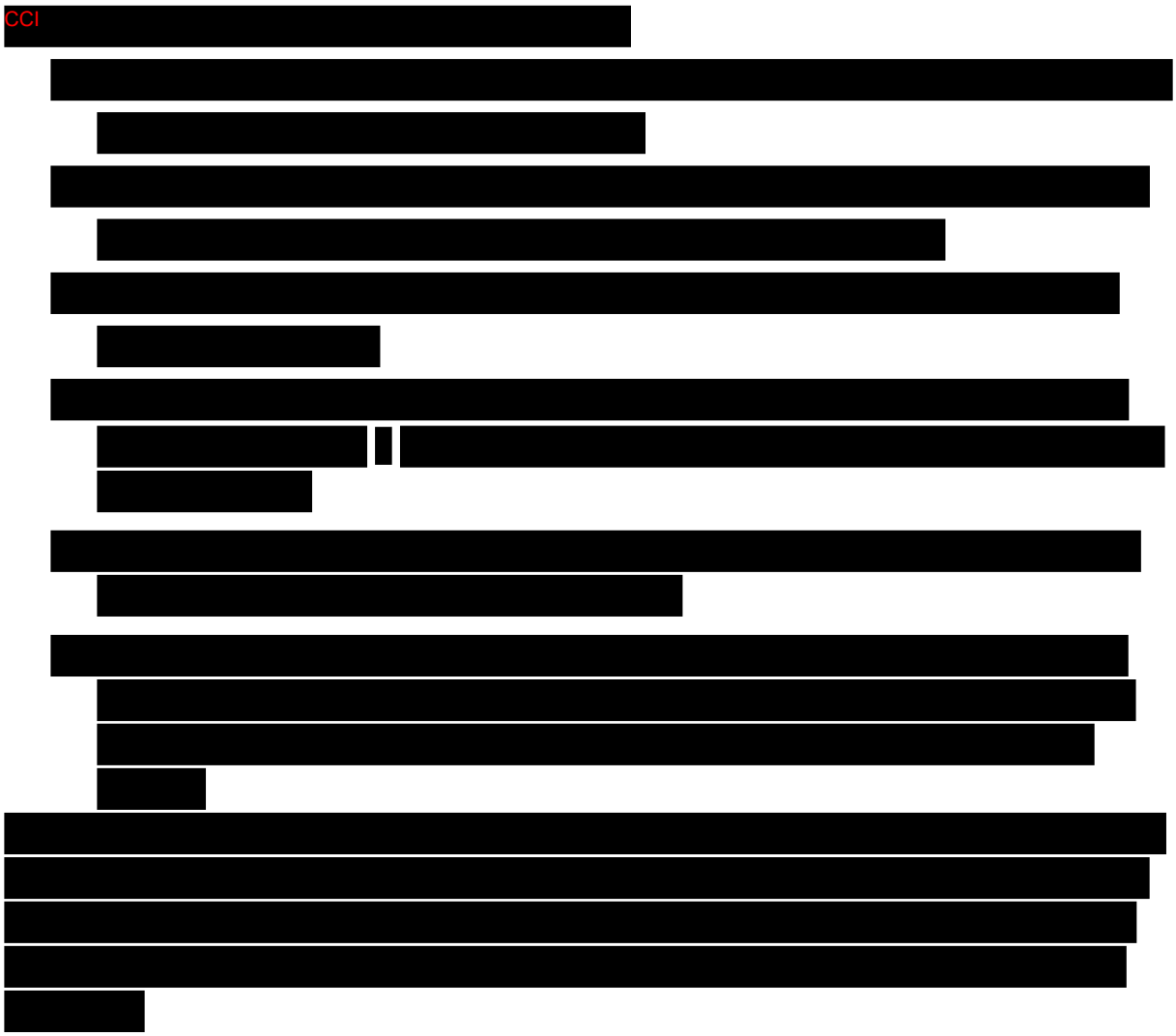
All subgroup analyses will be conducted using the same MMRM analysis as for the primary efficacy analysis except that the fixed class effect term for trial center will not be included in the model

CCI [redacted]

8.2 Secondary Endpoint Analysis

The key secondary efficacy endpoint is the change from baseline to Day 21 in the double-blind treatment phase in CGI-BP severity of illness score in mania. This endpoint will be analyzed by fitting the similar MMRM model described in the primary analysis. In order to control the overall type I error rate, a hierarchical testing procedure will be used. If the primary efficacy analysis for the YMRS Total Score yields a statistically significant result at 0.05 (2-sided) for the comparison of brexpiprazole versus placebo, then the comparison for key secondary efficacy variable (CGI-S score) will be tested at an alpha level of 0.05 (2-sided).

CCI



CCI

8.3 CCI

8.4 CCI

8.5 CCI

9 Safety Analysis

Standard safety variables to be analyzed include Adverse Events (AEs), clinical laboratory tests, vital signs, body weight, waist circumference, BMI, 12-lead electrocardiograms (ECGs), and physical examinations. CCI and Columbia-Suicide Severity Rating Scale (C-SSRS).

Safety analysis will be conducted based on the Safety Sample unless indicated otherwise. In general, baseline of a safety variable is defined as the last observation of the variable before taking the first dose of IMP, unless specified otherwise. Prospectively defined criteria will be used to identify potentially clinically relevant abnormal values for clinical laboratory tests, vital signs, ECGs, and body weight.

9.1 Adverse Events

All adverse events will be coded by System Organ Class (SOC) and Preferred Term (PT) according to the Medical Dictionary for Regulatory Activities (MedDRA). AEs that are sex-specific, e.g., ovarian cancer, will have their incidence rates evaluated for the specific sex.

Treatment-emergent AEs (TEAEs) are defined as AEs with an onset date on or after the start of double-blind treatment period. In more detail, TEAEs are all adverse events which started after start of double blind IMP; or if the event was continuous from baseline and was worsening, serious, study drug related, or resulted in death, discontinuation, interruption or reduction of study therapy. Adverse events occurring up to 30 days after the last day of IMP will be included in the summary tables.

The incidence of the following events in the double-blind treatment period will be tabulated by treatment group and overall using the Safety Sample:

a) TEAEs

b) TEAEs by severity

c) TEAEs potentially causally related to the IMP

d) TEAEs with an outcome of death

e) Serious TEAEs

f) TEAEs leading to discontinuations of the IMP

The above summaries (b), (e) and (f) will also be prepared for TEAEs potentially causally related to the IMP.

In addition, incidence of TEAE during the double-blind treatment period of at least 5% in any treatment group other than placebo group, and also greater than placebo by SOC and PT will be provided.

Incidence of TEAEs by SOC and PT will be summarized for sex, race, age and region subgroups.

CCI [redacted]

Unless otherwise specified, in general, analysis of safety data will be performed on observed case and for last visit.

9.2 Clinical Laboratory Tests

Summary statistics for mean and mean change from baseline in the routine clinical laboratory measurements, prolactin concentrations, coagulation parameters (PT, aPTT, and INR), HbA1c, and TSH will be provided by treatment and by visit.

Potentially clinically relevant laboratory measurement test results in the double-blind treatment period will be identified for the Safety Sample and will be summarized by treatment group and listed. Criteria for identifying laboratory values of potential clinical relevance are provided in Appendix 2.

9.2.1 Drug Induced Liver Injury (DILI)

Total bilirubin level should be checked for any subject with increased ALT or AST levels ≥ three times the upper normal limits (ULN) or baseline.

- Reporting all DILI as SAE to the FDA based on Hy's Law:
  - AST or ALT ≥ 3 x ULN or baseline and
  - T_Bili ≥ 2 x ULN or baseline

A separate incidence table will be provided for DILI cases, and the corresponding listing will be provided for Safety Sample during the double-blind treatment period.

9.2.2 Metabolic Change

In addition to mean change from baseline, incidence of treatment emergent significant changes in fasting lipids, fasting glucose, and metabolic syndrome will be summarized by treatment group using the following criteria.

Criteria for Treatment-Emergent Significant Change in Lipids and Glucose

LAB PARAMETER	BASELINE[1]	ANYTIME POST BASELINE
LDL Direct, Fasting (MG/DL)	Borderline 100-<160	High >=160
	Normal/Borderline <160	High >=160
	Normal <100	Borderline/High >=100
	Any Value	Increased >=30
HDL Cholesterol, Fasting (MG/DL)	Normal >=40	Low <40
	Any Value	Decreased >=20
Triglycerides, Fasting (MG/DL)	Normal <150	High 200-<500
	Borderline 150-<200	High 200-<500
	Normal/Borderline <200	High 200-<500
	Normal <150	Borderline/High/Very High >=150
	Any Value	Increased >=50
Glucose Fasting, Serum (MG/DL)	Normal <100	High >=126
	Impaired 100-<126	High >=126
	Normal/Impaired <126	High >=126
	Any Value	Increased >=10

[1] BASELINE IS CALCULATED FROM WEEK 1.

Criteria for Treatment-Emergent Metabolic Syndrome

DESCRIPTION	ANYTIME POST BASELINE[1]
Central Obesity	Waist Circumference >=102cm(MALE), >=88cm (FEMALE)
Dyslipidemia	Triglycerides >= 150mg/dl
Dyslipidemia	HDL < 40mg/dl (MALE), <50mg/dl (FEMALE)
Supine Blood Pressure	Systolic>=130mmHg and Diastolic>=85mmHg
Glucose Fasting, Serum	>=100mg/dl
Metabolic Syndrome	Met 3 Or More of the Above Criteria at a Visit

[1] BASELINE IS CALCULATED FROM WEEK 1.

9.3 Vital Signs

Summary statistics for vital signs will be provided. For the double-blind treatment period, vital signs, change from baseline will be summarized for the Safety Sample by treatment group.

Potentially clinically relevant vital signs measurements identified in the double-blind treatment period for the Safety Sample will be summarized by treatment group. Criteria for identifying vital signs of potential clinical relevance are provided in Appendix 1. All potentially clinically relevant events or changes will be listed and included in summary tables.

9.4 12-Lead ECG

Summary statistics and incidence of potentially clinically relevant changes will be provided for ECG parameters.

For the analysis of QT and QTc, data from three consecutive complexes (representing three consecutive heart beats) will be measured to determine average values. The following QT corrections will be used for reporting purposes in the clinical study report:

1) QTcB is the length of the QT interval corrected for heart rate by the Bazett formula: $QTcB=QT/(RR)^{0.5}$ and
2) QTcF is the length of the QT interval corrected for heart rate by the Fridericia formula: $QTcF=QT/(RR)^{0.33}$
3) QTcN is the length of the QT interval corrected for heart rate by the FDA Neuropharm Division formula: $QTcN=QT/(RR)^{0.37}$

Potentially clinically relevant changes in the 12-lead ECG identified in the double-blind treatment period for the Safety Sample will be listed and summarized by treatment group. Criteria for identifying ECG measurements of potential clinical relevance are provided in Appendix 3.

Categorical changes in ECG parameters during the double-blind treatment period will be summarized based on the following criteria:

Categorical Change Criteria in QT/QTc Parameters

Classification	Category	Criteria
QT	New Onset (> 450 Msec)	New onset (>450 msec) in QT means a subject who attains a value > 450 msec during treatment period but not at baseline.
QTc *	New Onset (≥ 450 Msec for men and ≥ 470 Msec	New onset (≥ 450 Msec for men and (≥ 470 Msec for women) in QTc means a subject who attains a

Categorical Change Criteria in QT/QTc Parameters		
Classification	**Category**	**Criteria**
	for women)	value ≥ 450 Msec for men or ≥ 470 Msec for women during treatment period but not at baseline.
	New Onset (≥ 450 Msec for men and ≥ 470 Msec for women) And > 10% Increase	New onset (≥ 450 Msec for men and ≥ 470 Msec for women) and > 10% increase in QTc means a subject who attains a value ≥ 450 Msec for men or ≥ 470 Msec for women, and > 10% increase during treatment period but not at baseline
	New Onset (> 500 Msec)	New onset (> 500 msec) in QTc means a subject who attains a value > 500 msec during treatment period but not at baseline.
	Increase 30 - 60 Msec	Increase from baseline value > 30 and ≤ 60 msec in QTc
	Increase > 60 Msec	Increase from baseline value > 60 msec in QTc

* QTc categorical change criteria apply to QTcB, QTcF and QTcN.

9.5 Physical Examinations

By-patient listings will be provided for physical examination.

9.5.1 Body Weight, Waist Circumference and BMI

Analyses of body weight, waist circumference and BMI will be performed for the Safety Sample. The mean change from baseline-to Day 21 (OC) and last visit in the double-blind treatment period in body weight will be tabulated and analyzed using ANCOVA. The ANCOVA models for both the OC and last visit analyses will include the baseline as a covariate and the treatment group as fixed effect.

Percentages of patients showing significant weight gain (≥ 7 % increase in weight), as well as percentages of patients showing significant weight loss (≥ 7 % decrease in weight) from baseline to Day 21 (OC and LOCF) will be analyzed using CMH General Association Test.

Body mass index is defined as weight in kilograms divided by the square of height in meters.

9.6 CCI

[redacted]

CCI [redacted]

9.7 Suicidality Data

Suicidality will be monitored during the study using the C-SSRS and will be summarized as number and percentage of subjects reporting any suicidal behavior, ideation, behavior by type (4 types), ideation by type (5 types) and treatment emergent suicidal behavior and ideation. Summary will be provided for the double-blind treatment period.

Suicidality is defined as report of at least one occurrence of any type of suicidal ideation or at least one occurrence of any type of suicidal behavior during assessment period (count each person only once).

Treatment emergent suicidal behavior and ideation is summarized by four types: Emergence of suicidal ideation, Emergence of serious suicidal ideation, Worsening of suicidal ideation, Emergence of suicidal behavior.

Emergence of suicidal behavior/ideation is defined as report of any type of suicidal behavior/ideation during treatment when there was no baseline suicidal behavior/ideation.

Emergence of serious suicidal ideation is defined as observation of suicidal ideation severity rating of 4 or 5 during treatment when there was no baseline suicidal ideation.

Worsening of suicidal ideation is defined as a suicidal ideation severity rating that is more severe than it was at baseline.

9.8 Concomitant Medications

Number and proportion of patients taking concomitant medications prior to study therapy, during the double-blind treatment period, and after study therapy are tabulated by drug classification using the WHO drug dictionary.

9.9 Extent of Exposure

The start date of double-blind study therapy - brexpiprazole or placebo - will be the first day of double-blind dosing. The number and percentage of patients who receive double-blind study medication, will be presented by week and by treatment group. Each dosing week will

be based on the actual week; i.e., Day 1-7 in Week 1, Day 8-14 in Week 2, etc. This summary will be performed on the Safety Sample.

The mean daily dosage will be summarized by week and treatment group using descriptive statistics. The mean daily dosage per patient per week will be determined for each week of the study. This will be calculated by dividing the sum of individual total doses by the number of days in the week interval. The summary will contain for each treatment group the number of patients receiving double-blind study medication`, and the mean and range of the mean daily dose for each week.

10 Conventions

10.1 Study Visit Windows

Study visit windows will be used to map visits using study day intervals. Observations at each scheduled visit and Early Termination will be assigned to Day 4, Day 7, Day 14 and Day 21 visits based on their visit windows as shown in Table 10.1A. This visit window convention applies to tables and listings for all efficacy and safety scales (YMRS, CGI-BP, CCI [redacted] This derived study window variable will be named as DAY and will be footnoted. In listings, it will be listed along with the CRF study visit.

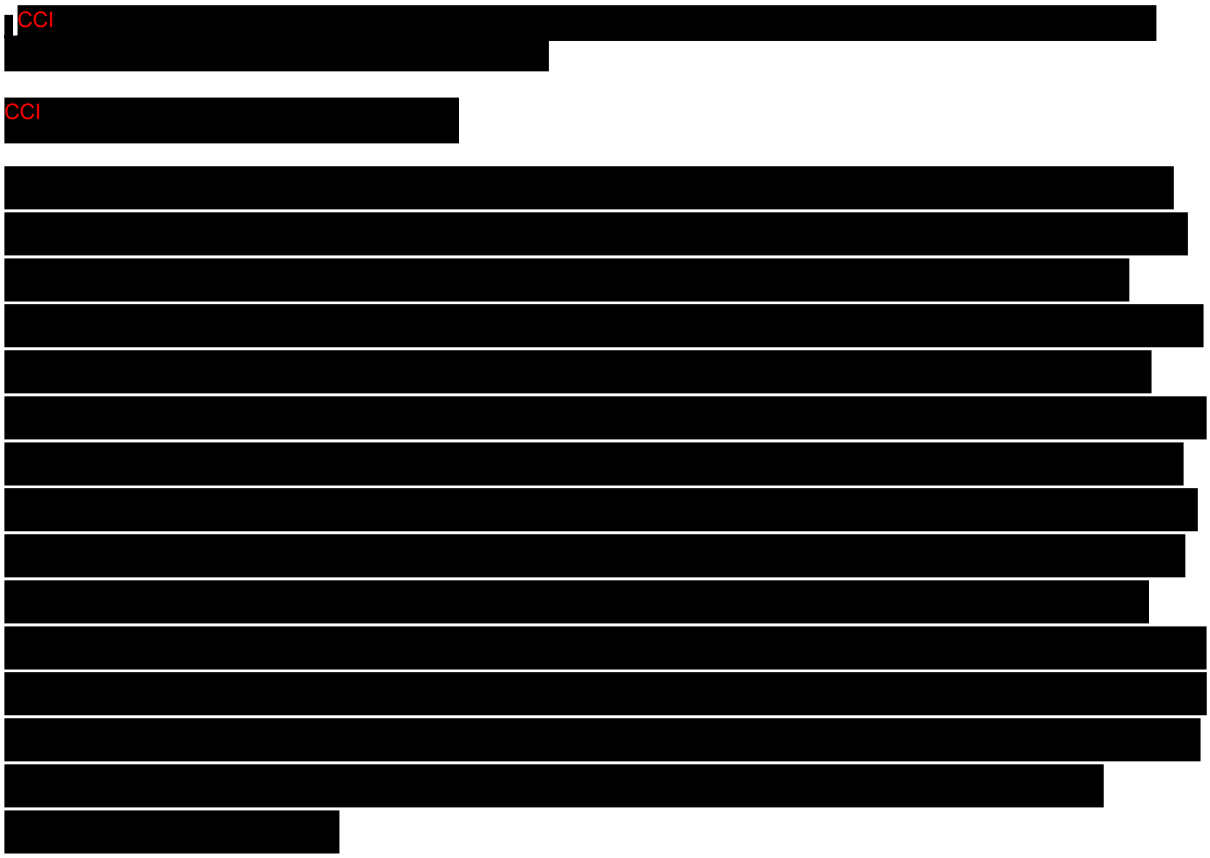
Table 10.1A shows classifications for study day intervals in the double-blind period. The variable “target day” is defined using the number of days since the start of double-blind dosing. The first day of double-blind dosing is defined as “Day 1”.

If more than one observation falls within a particular study day interval, then the last observation within that interval is used. Evaluations occurring more than 7 days after the last double-blind dosing date will not be mapped into study visit windows, and will be excluded from the analysis.

Table 10.1A: Study Day and Visit Windows

Day	Target Day [a]	Study Day Interval [a]
4	4	2-5
7	7	6-10
14	14	11-17
21	21	18-24 [b]

[a] Relative to the first day of IMP in the double-blind treatment period.



10.3 Scales: Rules for Scoring and Handling of Missing Data

10.3.1 YMRS

The YMRS is utilized as the primary efficacy assessment of a subject's level of manic symptoms. The YMRS consists of 11 items: 1) elevated mood, 2) increased motor activity-energy, 3) sexual interest, 4) sleep, 5) irritability, 6) speech (rate and amount), 7) language-thought disorder, 8) content, 9) disruptive-aggressive behavior, 10) appearance, and 11) insight. Seven items are rated on a 0- to 4-scale, while four items (Items 5, 6, 8, and 9) are rated on a 0- to 8-scale with 0, 2, 4, 6, and 8 being the possible scores (twice the weight of the other items). For all items, 0 is the "best" rating and the highest score (4 or 8) is the 'worst' rating. The YMRS Total Score is the sum of ratings for all 11 items; therefore, possible total scores range from 0 to 60. The YMRS Total Score is set to be missing if less than 9 of the 11 items are recorded. If 10 of the 11 items are available and the item missing is from items 5, 6, 8 or 9, then the YMRS Total Score is the sum of scores for items 1 to 4, 7, 10 to 11 plus the mean of the recorded items from 5, 6, 8 and 9 times four. If 10 of the 11 items are available and the item missing is from items 1 to 4, 7, 10 to 11, then the YMRS Total Score is the sum of scores for items 5, 6, 8 and 9 plus the mean of the recorded items from 1 to 4, 7, 10 to 11

times seven. If 9 of the 11 items are available and both missing items are from items 5, 6, 8 and 9, then the YMRS Total Score is set to be missing. If 9 of the 11 items are available and both missing items are from items 1 to 4, 7, 10 to 11, then the Total Score was the sum of scores for items 5, 6, 8 and 9 plus the mean of the recorded items from 1 to 4, 7, 10 to 11 times seven. If 9 of the 11 items are available and one of the missing items is from items 1 to 4, 7, 10 to 11, and one of the missing items is from items 5, 6, 8 and 9, then the YMRS Total Score is the mean of the recorded items from 5, 6, 8 and 9 times four, plus the mean of the recorded items from 1 to 4, 7, 10 to 11 times seven. All imputed scores are rounded to the first decimal place.

10.3.2 CGI-BP

The CGI-BP scale refers to the global impression of the subject with respect to bipolar disorder. The scale rates the subject's Severity of Illness (CGI-BP severity of illness: mania, depression, and overall bipolar illness) and based on a 7-point scale and rates the subject's Change from Baseline (CGI-BP Change from Baseline: mania, depression, and overall bipolar illness) based on a 7-point scale. CGI-BP severity of illness items are: 1 = normal, not at all ill, 2 = minimally ill, 3 = mildly ill, 4 = moderately ill, 5 = markedly ill, 6 = severely ill, 7 = very severely ill. CGI-BP Change from Baseline items are: 1 = very much improved, 2 = much improved, 3 = minimally improved, 4 = no change, 5 = minimally worse, 6 = much worse, 7 = very much worse. The CGI-BP Change from Baseline is not assessed at the baseline visit. At each visit other than baseline, the Change from Baseline will be judged with respect to subject's condition at baseline.

10.3.3 CCI [redacted]

[redacted]

10.3.4 CCI [redacted]

[redacted]

CCI [redacted]

10.3.5 CCI [redacted]

[redacted]

10.3.6 CCI [redacted]

[redacted]

10.3.7 CCI [redacted]

[redacted]

10.3.8 C-SSRS

Suicidality will be monitored during the trial using the C-SSRS. This trial will use the "baseline/screening" and "Since Last Visit" versions of the scale. The "baseline/screening" version, which assesses the lifetime experience of the subject with suicide events and suicidal ideation and the occurrence of suicide events and/or ideation within a specified time period prior to entry into the trial, will be completed for all subjects at screening to determine eligibility. Any subject with active suicidal ideation within the last 6 months, suicidal behaviors within the last 2 years, or who in the clinical judgment of the investigator presents a serious risk of suicide should be excluded from the trial. The "Since Last Visit" C-SSRS form will also be completed at all visits after screening.

11 References

[1] Siddiqui O, Hung JHM, O'Neill R. MMRM vs. LOCF: A comprehensive comparison based on simulation study and 25 NDA datasets. J Biopharmaceutical Stats. 2009; 19(2):227-46.

[2] Diggle P, Kenward MG. Informative drop-out in longitudinal data analysis. Applied Statistics. 1994; 43:49-93.

[3] Little RJA. Pattern-mixture models for multivariate incomplete data. J Am Stat Assoc. 1993; 88:125-34.

[4] Little RJA. Modeling the drop-out mechanism in repeated measures studies. J Am Stat Assoc. 1995; 90:1112-21.

[5] Hedeker D, Gibbons RD. Application of random effects pattern-mixture models for missing data in longitudinal studies. Psychological Methods. 1997; 2:64-78.

[6] Ali MW, Siddiqui O. Multiple imputation compared with some information dropout procedures in the estimation and comparison of rates of change in longitudinal clinical trials with dropouts. J Biopharmaceutical Stats. 2000;10(2):165-81.

[7] Wu MC, Bailey KR. Estimation and comparison of changes in the presence of informative right censoring: Conditional linear model. Biometrics. 1989; 45:939-55.

[8] van Elteren, PH. On the combination of independent two sample tests of Wilcoxon. Bull Int Stat Inst. 1960; 37:351-61.

[9] Mehrotra D, Li X, Liu J, Lu K. Analysis of Longitudinal Clinical Trials with Missing Data Using Multiple Imputation in Conjunction with Robust Regression. Biometrics 2012; 68:1250-1259.

[10] Cortina, J. M., & Nouri, H. (2000). *Effect size for ANOVA designs*. Thousand Oaks, Calif.: Sage Publications.

12 Potential Clinical Relevance Criteria from Protocol

Appendix 1 Criteria for Identifying Vital Signs of Potential Clinical Relevance

Variable	Criterion Value[a]	Change Relative to Baseline[a]
Heart Rate[b]	> 120 bpm < 50 bpm	≥ 15 bpm increase ≥ 15 bpm decrease
Systolic Blood Pressure[b]	> 180 mmHg < 90 mmHg	≥ 20 mmHg increase ≥ 20 mmHg decrease
Diastolic Blood Pressure[b]	> 105 mmHg < 50 mmHg	≥ 15 mmHg increase ≥ 15 mmHg decrease
Orthostatic Hypotension	≥ 20 mmHg decrease in systolic blood pressure and a ≥ 25 bpm increase in heart rate from supine to sitting/standing	Not Applicable (baseline status not considered)
Weight	-	≥ 7% increase ≥ 7% decrease

[a] In order to be identified as potentially clinically relevant, an on-treatment value must meet the "Criterion Value" and also represent a change from the subject's baseline value of at least the magnitude shown in the "Change Relative to Baseline" column.

[b] As defined in "Supplementary Suggestions for Preparing an Integrated Summary of Safety Information in an Original NDA Submission and for Organizing Information in Periodic Safety Updates," FDA Division of Neuropharmacological Drug Products draft (2/27/87).

Appendix 2 Criteria for Identifying Laboratory Values of Potential Clinical Relevance

Laboratory Tests	Criteria
Chemistry	
AST (SGOT)	≥ 3 x upper limit of normal (ULN)
ALT (SGPT)	≥ 3 x ULN
Alkaline phosphatase	≥ 3 x ULN
Lactate dehydrogenase (LDH)	≥ 3 x ULN
Blood urea nitrogen (BUN)	≥ 30 mg/dL
Creatinine	≥ 2.0 mg/dL
Uric Acid	
Men	≥ 10.5 mg/dL
Women	≥ 8.5 mg/dL
Bilirubin (total)	≥ 2.0 mg/dL
Creatine phosphokinase (CPK)	≥ 3 x ULN
Prolactin	> ULN
Hematology	
Hematocrit	
Men	≤ 37 % and decrease of ≥ 3 percentage points from Baseline
Women	≤ 32 % and decrease of ≥ 3 percentage points from Baseline
Hemoglobin	
Men	≤ 11.5 g/dL
Women	≤ 9.5 g/dL
White blood count	≤ 2,800/ mm^3 or ≥ 16,000/ mm^3
Eosinophils	≥ 10%
Neutrophils	≤ 15%
Absolute neutrophil count	≤ 1,000/ mm^3
Platelet count	≤ 75,000/ mm^3 or ≥ 700,000/ mm^3
Urinalysis	
Protein	Increase of ≥ 2 units
Glucose	Increase of ≥ 2 units
Casts	Increase of ≥ 2 units
Additional Criteria	
Chloride	≤ 90 mEq/L or ≥ 118 mEq/L
Potassium	≤ 2.5 mEq/L or ≥ 6.5 mEq/L
Sodium	≤ 126 mEq/L or ≥ 156 mEq/L
Calcium	≤ 8.2 mg/dL or ≥ 12 mg/dL
Glucose	
Fasting	≥ 100 mg/dL
Non-Fasting	≥ 200 mg/dL
Total Cholesterol, Fasting	≥ 240 mg/dL
LDL Cholesterol, Fasting	≥ 160 mg/dL
HDL Cholesterol, Fasting	
Men	< 40 mg/dL
Women	< 50 mg/dL
Triglycerides, Fasting	≥ 150 mg/dL

Appendix 3 Criteria for Identifying ECG Measurements of Potential Clinical Relevance

Variable	Criterion Value[a]	Change Relative to Baseline[a]
Rate		
Tachycardia	≥ 120 bpm	increase of ≥ 15 bpm
Bradycardia	≤ 50 bpm	decrease of ≥ 15 bpm
Rhythm		
Sinus tachycardia[b]	≥ 120 bpm	increase of ≥ 15 bpm
Sinus bradycardia[c]	≤ 50 bpm	decrease of ≥ 15 bpm
Supraventricular premature beat	all	not present → present
Ventricular premature beat	all	not present → present
Supraventricular tachycardia	all	not present → present
Ventricular tachycardia	all	not present → present
Atrial fibrillation	all	not present → present
Atrial flutter	all	not present → present
Conduction		
1° atrioventricular block	PR ≥ 200 msec	increase of ≥ 50 msec
2° atrioventricular block	all	not present → present
3° atrioventricular block	all	not present → present
Left bundle-branch block	all	not present → present
Right bundle-branch block	all	not present → present
Pre-excitation syndrome	all	not present → present
Other intraventricular conduction block[d]	QRS ≥ 120 msec	increase of ≥ 20 msec
Infarction		
Acute or subacute	all	not present → present
Old	all	not present → present ≥ 12 weeks post study entry
ST/T Morphological		
Myocardial Ischemia	all	not present → present
Symmetrical T-wave inversion	all	not present → present
Increase in QTc	QTcF ≥ 450 msec (men) QTcF ≥ 470 msec (women)	

[a] In order to be identified as potentially clinically relevant, an on-treatment value must meet the "Criterion Value" and also represent a change from the subject's baseline value of at least the magnitude shown in the "Change Relative to Baseline" column.

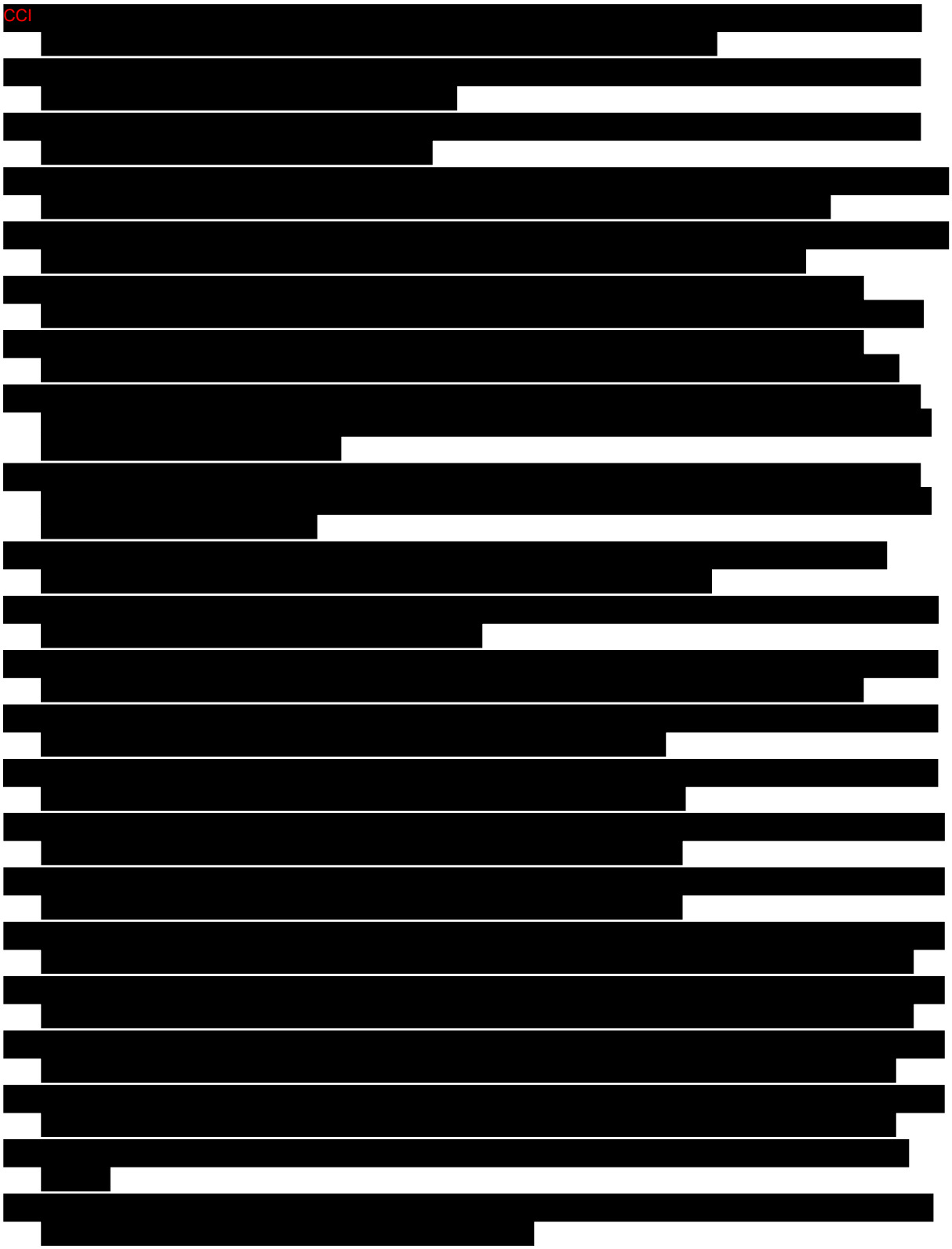
[b] No current diagnosis of supraventricular tachycardia, ventricular tachycardia, atrial fibrillation, atrial flutter, or other rhythm abnormality.

[c] No current diagnosis of atrial fibrillation, atrial flutter, or other rhythm abnormality.

[d] No current diagnosis of left bundle branch block or right bundle branch block.

13 CCI [redacted]

[redacted]


CCI

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

CCI

[redacted]