Otsuka Pharmaceutical Development & Commercialization, Inc.

Investigational New Drug
SEP-363856

Protocol No. 382-201-00001
IND No. 161135

A Phase 2/3, Multicenter, Randomized, Double blind, Placebo controlled Trial of the Safety and Efficacy of Flexible Doses of SEP-363856 as Adjunctive Therapy in the Treatment of Adults with Major Depressive Disorder

A Trial of the Safety and Efficacy of Flexible Doses of SEP-363856 as Adjunctive Therapy in the Treatment of Adults with Major Depressive Disorder

Statistical Analysis Plan

Version: 4.0
Date: 06 Nov 2025

Confidential
May not be used, divulged, published, or otherwise disclosed without the prior written consent of Otsuka Pharmaceutical Development & Commercialization, Inc.

Table of Contents

Title Page **1**
Table of Contents **2**
List of In-text Tables **6**
List of In-text Figures **7**
List of Appendices **8**
List of Abbreviations and Definitions of Terms **9**
1 Introduction **11**
2 Trial Objectives **11**
3 Trial Design **11**
3.1 Type/Design of Trial 11
3.2 Trial Treatments 16
3.3 Trial Population 16
4 Sample Size **16**
5 Statistical Analysis Datasets **17**
5.1 Baseline Definition 17
5.2 Mapped Trial Week Windows 17
5.2.1 Handling of Missing Data 19
5.3 Pooling of Small Centers 20
6 Primary and Secondary Outcome Variables **21**
6.1 Primary Outcome Variables 21
6.2 Secondary Outcome Variables 21
CCI
6.4 Safety Outcome Variables 22
7 Disposition and Demographic Analysis **22**
7.1 Subject Disposition 22
7.2 Demographic and Baseline Characteristics 22
7.3 Treatment Compliance 23
7.4 Prior and Concomitant Medications 23
7.5 Protocol Deviations 23

7.6 Medical and Psychiatric History and Antidepressant Treatment History 23
7.7 Baseline Psychiatric Evaluation 23
8 Efficacy Analysis 24
8.1 Estimands 24
8.1.1 Primary Estimand 24
8.1.2 Secondary and Other Estimands 25
8.2 Primary Efficacy Endpoint 26
8.2.1 Primary Efficacy Analysis 26
CCI
8.3 Subgroup Analyses 32
8.4 Key Secondary Efficacy Endpoint 32
8.5 Control of Experiment-wise Type 1 Error 33
CCI
8.8 Efficacy Outcome Measures in Phase A and A+ 36
9 Safety Analyses 36
9.1 Extent of Exposure 36
9.1.1 Exposure to SEP-363856 36
9.1.2 Exposure to ADT 37
9.2 Adverse Events 37
9.2.1 Adverse Events in Phase B 38
9.2.2 Adverse Events in Phase A and A+ 38
9.3 Vital Sign Data 39
9.3.1 Vital Signs in Phase B 39
9.3.2 Vital Signs in Phase A and A+ 39
9.4 Clinical Laboratory Data 40
9.4.1 Clinical Laboratory Tests in Phase B 40
9.4.2 Drug Induced Liver Injury 40

9.4.3 Clinical Laboratory Tests in Phase A and A+ 40
9.5 Physical Examination Data 41
9.5.1 Body Weight, Waist Circumference, and Body Mass Index 41
9.5.1.1 Body Weight, Waist Circumference, and Body Mass Index in Phase B 41
9.5.1.2 Body Weight, Waist Circumference, and Body Mass Index in Phase A and A Plus 41
9.6 12-Lead Electrocardiograms 42
9.6.1 12-Lead Electrocardiograms in Phase B 42
9.6.2 12-Lead Electrocardiograms in Phase A and A+ 42
9.7 Suicidality 43
9.8 Simpson-Angus Scale, Abnormal Involuntary Movement Scale, and Barnes Akathisia Rating Scale 43
9.8.1 Simpson-Angus Scale, Abnormal Involuntary Movement Scale, and Barnes Akathisia Rating Scale in Phase B 43
9.8.2 SAS, AIMS, and BARS in Phase A and A+ 44
9.9 Sexual Functioning Questionnaire 44
9.9.1 Sexual Functioning Questionnaire in Phase B 44
9.9.2 Sexual Functioning Questionnaire in Phase A and A+ 44
9.10 Pittsburgh Sleep Quality Index 45
9.10.1 Pittsburgh Sleep Quality Index in Phase B 45
9.10.2 Pittsburgh Sleep Quality Index in Phase A and A+ 45
10 Rules for Scoring 45
10.1 Montgomery-Asberg Depression Rating Scale 45
10.2 Clinical Global Impression 45
CCI [redacted]
10.6 Columbia-Suicide Severity Rating Scale 46
CCI [redacted]
10.9 Changes in Sexual Functioning Questionnaire 48
10.10 Simpson-Angus Scale 48

10.11 Abnormal Involuntary Movement Scale 48
10.12 Pittsburgh Sleep Quality Index 48
10.13 Barnes Akathisia Rating Scale 49
11 Version Change History....50
12 References 52
13 Appendices 54

List of In-text Tables

Table 5.2-1 Study Day and Visit Windows of Phase B 18
Table 5.2-2 Study Day and Visit Windows of Phase A 18
Table 5.2-3 Study Day and Visit Windows of Phase A+.....19

List of In-text Figures

Figure 3.1-1 Trial Design Schematic.....15

List of Appendices

Appendix 1 Criteria for Identifying Vital Signs and Weight of Potential Clinical Relevance 54

Appendix 2 Criteria for Identifying Laboratory Values of Potential Clinical Relevance 55

Appendix 3 Criteria for Identifying Electrocardiogram Measurements of Potential Clinical Relevance 57

List of Abbreviations and Definitions of Terms

Abbreviation	Definition
ADT	antidepressant therapy
AE	adverse event
AIM	abnormal involuntary movement
AIMS	Abnormal Involuntary Movement Scale
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
AST	aspartate aminotransferase
BARS	Barnes Akathisia Rating Scale
BMI	body mass index
C-SSRS	Columbia-Suicide Severity Rating Scale
CCI	
CGI-S	Clinical Global Impression Severity
CSFQ	Changes in Sexual Functioning Questionnaire
CMH	Cochran-Mantel-Haenszel
DILI	drug induced liver injury
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
EPS	extrapyramidal symptoms
ECG	electrocardiogram
ET	early termination
FAS	full analysis set
CCI	
FDA	United States Food and Drug Administration
GEE	generalized estimating equations
CCI	
ICF	informed consent form
ICH	International Council for Harmonisation
IMP	investigational medicinal product
ITT	intent-to-treat
LOCF	last observation carried forward
LOE	lack of efficacy
LS	least squares
MADRS	Montgomery Åsberg Depression Rating Scale
CCI	
MCMC	Monte Carlo Markov Chain
MDD	major depressive disorder
MedDRA	Medical Dictionary for Regulatory Activities
MDE	major depressive episode
MGH ATRQ	Massachusetts General Hospital Antidepressant Treatment Response Questionnaire
CCI	
OC	observed cases
CCI	
PMM	pattern mixture models
PSQI	Pittsburgh Sleep Quality Index
PT	preferred term
SAE	serious adverse event

Abbreviation	Definition
SAP	statistical analysis plan
SAS	Simpson-Angus Scale
CCI	
SOC	system organ class
TEAE	treatment-emergent adverse event
THC	tetrahydrocannabinol
TSH	thryoid-stimulating hormone
ULN	upper limit of normal
US	United States
WGEE	weighted generalized estimating equations
WHO	World Health Organization

1 Introduction

This statistical analysis plan (SAP) describes the statistical methodology and data analysis algorithms and conventions to be applied for statistical analysis and reporting of data collected under the clinical protocol 382-201-00001. The SAP is developed per Protocol Amendment 2, Protocol Amendment 3, Protocol Amendment 4 and Protocol Amendment 5.

2 Trial Objectives

Primary objective is to compare the efficacy of SEP-363856 (50 to 75 mg/day) to placebo as adjunctive therapy to an assigned open label antidepressant therapy (ADT) in subjects with major depressive disorder (MDD) who demonstrate an inadequate response to a prospective 8-week trial of the same assigned open label ADT.

Secondary objective is to evaluate the safety and tolerability of SEP-363856 (50 to 75 mg/day) as adjunctive therapy to ADT in subjects with MDD.

3 Trial Design

3.1 Type/Design of Trial

This is a phase 2/3, multicenter, randomized, double-blind, placebo-controlled, flexible dose trial designed to assess the safety and efficacy of SEP-363856 (50 to 75 mg/day) as adjunctive therapy to an assigned open-label ADT in subjects with MDD who have demonstrated an inadequate response to a prospective 8-week trial of the same assigned open-label ADT. Personnel at the site will be blinded to the randomization criteria at the end of the prospective period (ie, definition of inadequate response). It is anticipated that approximately 900 subjects will be enrolled into Phase A (single-blind prospective treatment phase) from trial sites in the United States (US) and other global trial sites in order to randomize approximately 310 subjects into Phase B (double-blind randomization phase). The randomization for Phase B is stratified by site.

The trial will be organized as follows:

Screening Period: The screening period will last up to 28 days and will begin when informed consent is signed. The purpose of the screening period is to assess eligibility criteria at 1 or more visits (as necessary to complete screening assessments) and to washout prohibited concomitant therapy, if applicable.

If applicable, eligible subjects will washout from their current ADT for a minimum of 24 hours before receiving the first dose of investigator-assigned ADT in Phase A. In instances where a subject will be assigned to the same ADT, the 24-hour washout period

is not necessary (ie, a prior course of treatment with 1 of the protocol-specified ADTs that was at an adequate dose based on the Massachusetts General Hospital Antidepressant Treatment Response Questionnaire [MGH ATRQ], but of insufficient duration [ie, < 6 weeks]).

Phase A (Single-blind Prospective Treatment Phase): Subjects meeting eligibility criteria will be enrolled into the 8-week single-blind prospective treatment phase (Phase A).

CCI [redacted]

Phase B (Double-blind Randomization Phase): Subjects with an inadequate response in Phase A who are negative for tetrahydrocannabinol (THC) and illicit drugs at the Week 6 visit and who are suitable for randomization, may be randomized at the end of prospective treatment (Week 8 visit of Phase A) in a 1:1 ratio to one of the following double-blind treatment regimens:

- Continued placebo + ADT (that is, placebo plus the final dosage of the assigned open-label marketed ADT reached during Phase A), or
- SEP-363856 + ADT (that is, adjunctive SEP-363856, 50 to 75 mg/day, plus the final dosage of the assigned open-label marketed ADT reached during Phase A).

CCI [redacted].

The dose of ADT in Phase B will be based on the final dosage of the assigned open-label ADT reached during Phase A. During Phase B, randomized subjects will attend weekly visits at Weeks 9, 10, 11, 12, 13, and 14.

The response status of each subject will be based on the following criteria for inadequate response which must ALL be met at the specified visits for the subject to be considered an inadequate responder at Week 8, and thus eligible for randomization into Phase B:

- < 50% reduction in 17-item Hamilton Depression Rating Scale ($HAM\text{-}D_{17}$) total score between baseline of Phase A and the Week 8 visit, AND
- $HAM\text{-}D_{17}$ total score ≥ 14 at the Week 8 visit, AND
- Clinical Global Impression - Change (CGI-C) Scale score of ≥ 3 at all scheduled visits at Weeks 2, 4, 6, and 8, AND
- < 50% reduction in Montgomery Åsberg Depression Rating Scale (MADRS) Total Score between baseline of Phase A and all scheduled visits at Weeks 2, 4, 6, and 8.

Continued Treatment in Phase A+ (Single-blind Phase A Responder): Subjects who are not randomized into Phase B and who were negative for THC and illicit drugs at the Week 6 visit, may continue to receive single-blind placebo + ADT for an additional 6 weeks in Phase A+, for a total of 14 weeks. CCI [redacted]

CCI [redacted]

Follow-up: If any subject discontinues the trial early, every effort should be made to complete the early termination (ET) evaluations. Following the completion of all scheduled visits, subjects will be prescribed appropriate ADT (as per investigator judgment) and subjects (completers and early withdrawals) will be followed-up to monitor for safety events via telephone or a clinic visit for 7 (± 2) days after the last dose of IMP. This contact also applies to subjects withdrawn prematurely from the trial.

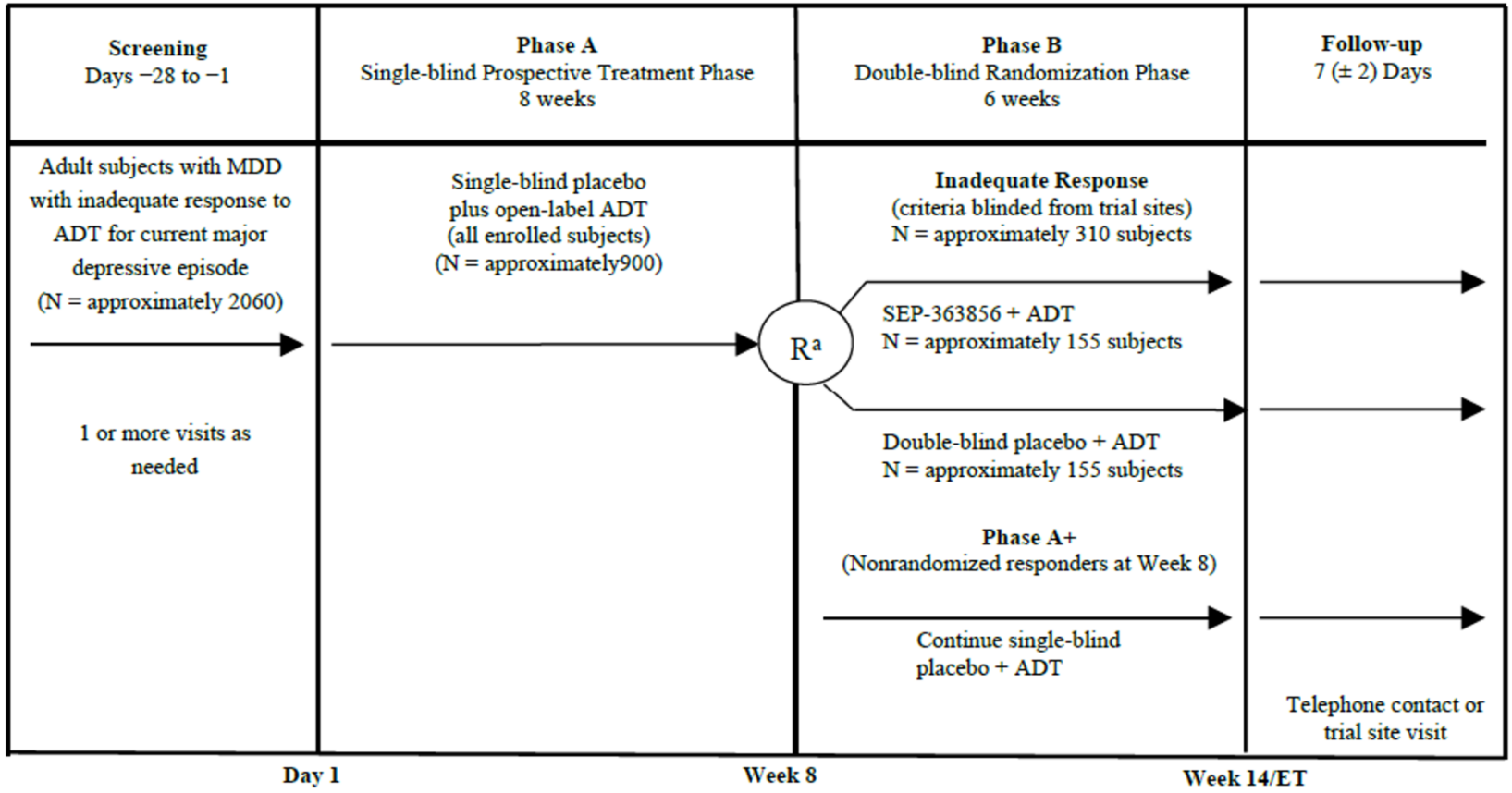


Figure 3.1-1 Trial Design Schematic

ADT = antidepressant therapy; CGI-C = Clinical Global Impressions - Change Scale; ET = early termination; HAM-D_{17} = 17-item Hamiliton Depression Rating Scale; MADRS = Montgomery Åsberg Depression Rating Scale; MDD = major depressive disorder; R= randomization.

[a]Randomization ratio of 1:1.

CCI

[redacted]

3.2 Trial Treatments

Phase A: subjects will receive single-blind placebo plus an investigator-determined, open-label ADT.

Phase B: subjects may be randomized at the end of prospective treatment (Week 8 visit of Phase A) in a 1:1 ratio to one of the following double-blind treatment regimens:

- Continued placebo + ADT (that is, placebo plus the final dosage of the assigned open-label marketed ADT reached during Phase A), or
- SEP-363856 + ADT (that is, adjunctive SEP-363856, 50 to 75 mg/day, plus the final dosage of the assigned open-label marketed ADT reached during Phase A).

Phase A+: Subjects who are not randomized into Phase B and who were negative for THC and illicit drugs at the Week 6 visit, may continue to receive single-blind placebo + ADT.

3.3 Trial Population

Approximately 900 subjects will be enrolled in Phase A and approximately 310 subjects will be randomized in Phase B. The trial population will include male and female outpatients aged 18 to 65 years with a primary Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) diagnosis of MDD and in a current major depressive episode (MDE), with reported inadequate response (< 50% improvement in symptom severity per MGH ATRQ) to at least 1 but no more than 3 ADTs in the current episode.

4 Sample Size

The primary efficacy endpoint is the change from the end of Phase A (Week 8 visit) to the end of Phase B (Week 14 visit) in MADRS total score. To compare the primary endpoint between placebo + ADT arm and the SEP-363856 + ADT arm, the trial will be randomized at a ratio of 1:1, with a 2-sided overall alpha of 0.05.

The sample size is calculated based on a CCI [redacted]

CCI [redacted]

CCI [redacted].

5 Statistical Analysis Datasets

The following analysis sets are defined for this trial:

- Enrolled Analysis Set: all subjects who signed an informed consent form (ICF) for the trial and enrolled into Phase A.
- ADT Analysis Set: all subjects who took at least 1 dose of ADT in Phase A.
- Randomized Analysis Set: all subjects who were randomized in Phase B.
- Safety Analysis Set: all randomized subjects in Phase B who received at least 1 dose of double-blind IMP. Subjects will be excluded from this population only if there is documented evidence (ie, 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.
- Full Analysis Set (FAS): all subjects in the Safety Analysis Set in Phase B, who have an end of Phase A (ie, Week 8) MADRS assessment.
- CCI [redacted]

The FAS is created based on the intent-to-treat (ITT) principle and will be used for efficacy analysis unless otherwise specified.

5.1 Baseline Definition

Baseline for Phase A/Phase A+ refers to last available measurement prior to the start of administration of ADT in Phase A.

For analyses of double-blind randomized Phase B data, the baseline is the end of Phase A measurement (expected to be at Week 8) prior to randomization. End of Phase A measurement is defined as the last available measurement prior to the start of double-blind study medication. For MADRS, if the last available measurement took place before Week 6 visit, no End of Phase A measurement is available.

5.2 Mapped Trial Week Windows

Study visit windows will be used to map visits using study day intervals. This visit window convention applies to tables and listings for all efficacy and safety scales (MADRS, Clinical Global Impression Severity [CGI-S], CCI [redacted]

CCI [redacted]

Simpson-Angus Scale [SAS], Barnes Akathisia Rating Scale [BARS], Abnormal Involuntary Movement Scale [AIMS], Pittsburgh Sleep Quality Index [PSQI], and Changes in Sexual Functioning Questionnaire [CSFQ]). This derived study window variable will be named as WEEK. In listings it will be listed along with the eSource Form study visit.

Table 5.2-1 shows classifications for study day intervals in Phase B. The variable "target day" is defined using the number of days since the start of double-blind dosing in Phase B. The first day of double-blind dosing is defined as "Day 1" in Table 5.2-1. An end of Phase A observation is the last available observation in Phase A on or before "Day 1". If that measurement took place prior to the lower bound of the Week 6 study visit window (ie, Day 39; see Table 5.2-2), it does not qualify as an End of Phase A measurement. 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 5.2-1 Study Day and Visit Windows of Phase B

Week	Target Day[a]	Study Day Interval[a]
9	7	2 - 10
10	14	11 - 17
11	21	18 - 24
12	28	25 - 31
13	35	32 - 38
14	42	39 - 49[b]

IMP = investigational medicinal product.

[a]Relative to the first day of double-blind IMP in Phase B.

[b]Evaluations occurring more than 7 days after the last double-blind dosing date will be excluded from the Phase B analyses.

CCI [redacted]

CCI



5.2.1 Handling of Missing Data

The MADRS is utilized as the primary efficacy assessment of a subject's level of depression. The MADRS consists of 10 items, all rated on a 0 to 6 scale with zero being the "best" rating and 6 being the "worst" rating. The MADRS Total Score is the sum of ratings for all 10 items; therefore, possible total scores range from 0 to 60. The MADRS Total Score will be un-evaluable if less than 8 of the 10 items are recorded. If 8 or 9 of the 10 items are recorded, the MADRS Total Score will be the mean of the recorded items multiplied by 10 and then rounded to the first decimal place.

Mixed model repeated measurements (MMRM) assumes data are missing at random (MAR), which is a reasonable assumption in longitudinal clinical trials in MDD.[1] However, the possibility of "missing not at random" (MNAR) data can never be ruled

out. For the primary analysis of the primary endpoint, unobserved future data after participant discontinuation due to lack of efficacy and AE will be imputed using control-based multiple imputation under the MNAR assumption. As sensitivity analyses, selection model[2], pattern-mixture model[3,4,5,6] and/or shared parameter model[7] will be used to explore data missing mechanisms of MNAR and investigate the response profile of dropout reason. Details are provided in Section 8.2.3.

Mixed Model Repeated Measurements, Wu-Bailey, and pattern-mixture model will be performed on the OC dataset.

The last observation carried forward (LOCF) data set will include data recorded at a scheduled Phase B visit or, if no observation is recorded at that visit, data carried forward from the previous scheduled Phase B visit. Baseline data (eg, the last visit of Phase A) will not be carried forward to impute missing values for the LOCF data set.

For CGI-C and categorical response/remission variables, OC analyses will be performed in addition to LOCF analyses. Study center will not be included in the models for OC analyses.

5.3 Pooling of Small Centers

Primary efficacy analysis will be performed on the FAS which comprises those subjects in the safety analysis set who have an End of Phase A MADRS value. Small centers will be defined as centers that do not have at least 2 FAS subject in each treatment arm in Phase B. All small centers will be pooled to form “pseudo centers” for the purpose of analysis according to the following algorithm. Small centers will be ordered from the largest to the smallest based on the number of subjects in FAS. The process will start by pooling the largest of the small centers with the smallest of the small centers until a non-small center is formed. This process will be repeated using the centers left out of the previous pass. In case of ties in center size, the center with the smallest center code will be selected. If any centers are left out at the end of this process, they will be pooled with the smallest pseudo centers, or if no pseudo centers exist, they will be pooled with the smallest non-small center. The pooling of the small centers will be performed within each region.

6 Primary and Secondary Outcome Variables

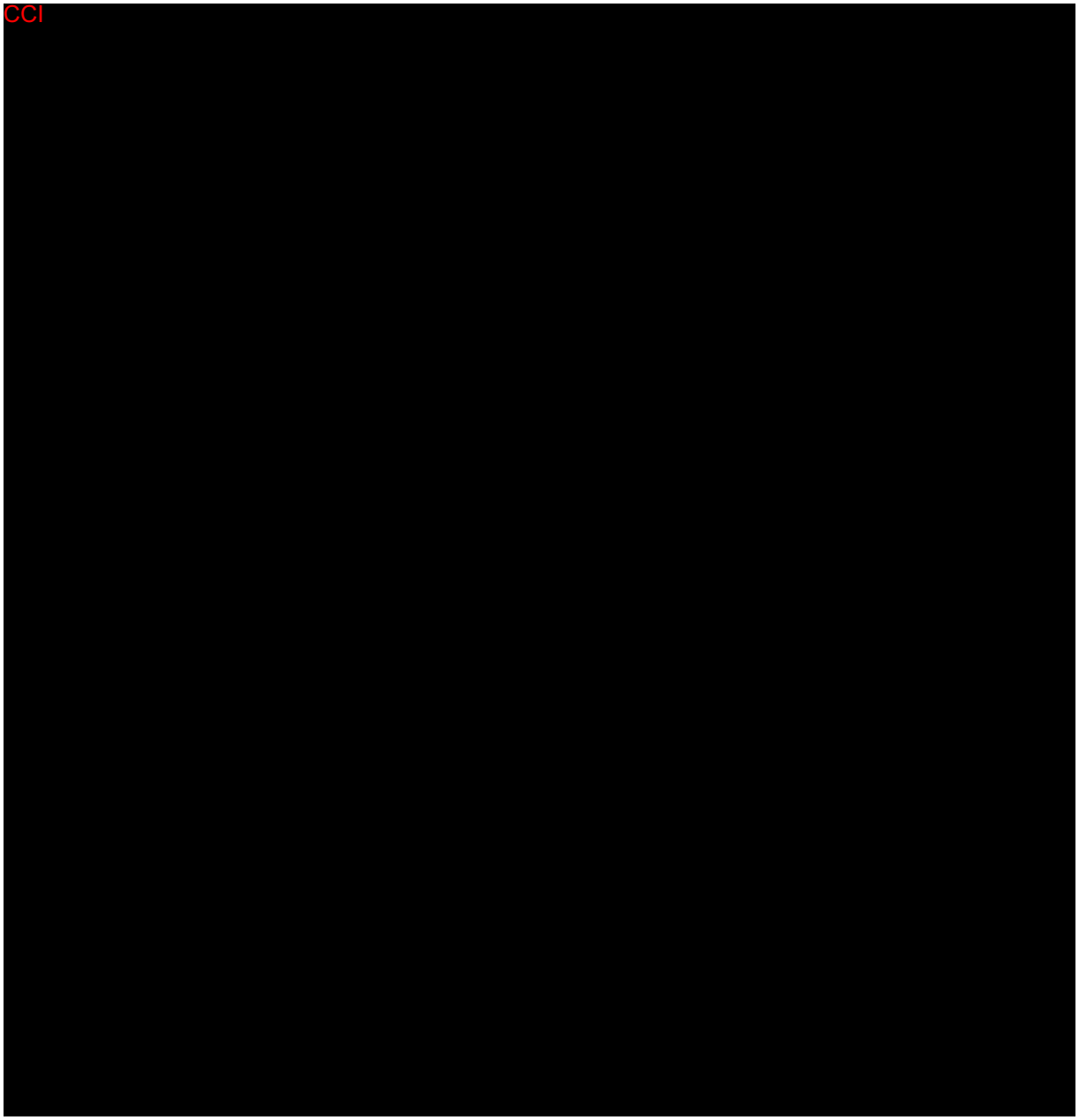
6.1 Primary Outcome Variables

Primary efficacy endpoint is change from the end of Phase A (Week 8 visit) to the end of Phase B (Week 14 visit) in the MADRS total score for subjects who randomized to Phase B.

6.2 Secondary Outcome Variables

Key secondary efficacy endpoint is change from end of Phase A (Week 8 visit) to end of Phase B (Week 14 visit) in the CGI-S.

CCI



Appropriate summaries of efficacy endpoints will also be provided for Phase A/A+.

6.4 Safety Outcome Variables

Secondary safety endpoints include assessments in adverse events (AEs), physical examinations (including body weight and waist circumference), vital signs (blood pressure, heart rate, and body temperature), 12 lead electrocardiograms (ECGs), and clinical laboratory assessments (hematology, serum chemistry, urinalysis, prolactin, glycosylated hemoglobin, and coagulation parameters). In addition to the standard safety variables, other safety variables including the Columbia-Suicide Severity Rating Scale (C-SSRS), SAS, AIMS, BARS, CSFQ, and PSQI will also be evaluated.

7 Disposition and Demographic Analysis

7.1 Subject Disposition

Subject disposition is summarized for the ADT analysis set, Phase A+ Analysis Set and randomized analysis set. For the randomized analysis set, disposition is summarized by treatment group, by treatment group and ADT, and by center.

Subject completion rate is presented for subjects in Phase A, Phase A+ and Phase B. Subject completion rate for Phase B (randomized analysis set) will be summarized by treatment group and ADT, and by week.

Reasons for discontinuation are summarized for ADT analysis set, Phase A+ Analysis Set by ADT and for randomized analysis set by treatment group.

7.2 Demographic and Baseline Characteristics

Baseline demographic characteristics including age, race, ethnicity, height, weight, waist circumference, and body mass index (BMI) will be tabulated by ADT and overall for the ADT analysis set and Phase A+ analysis set using the baseline assessments for all parameters.

For the randomized analysis set, demographic characteristics will be summarized by treatment group. Age, race, ethnicity, height, weight, waist circumference, and BMI will be tabulated by sex at birth, where for weight, waist circumference, and BMI the assessment at the end of Phase A will be used, and for the other parameters the baseline assessments will be used.

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 Treatment Compliance

Based on the Study Medication panel of the eSource Form, compliance in taking study medication is calculated by dividing the number of tablets taken by the total number of tablets the subjects were scheduled to take during the study period. Compliance is calculated on Placebo for Phase A and Phase A+, and on double-blind study medication for Phase B. For lost-to-follow up subjects, last study medication end date record will be used as the treatment end date.

Antidepressant medication compliance will be calculated similarly based on the Antidepressant Therapy Medication panel of the eSource Form.

7.4 Prior and Concomitant Medications

Number and proportion of subjects taking concomitant medications prior to Phase A, during Phase A, during Phase B, during Phase A+, and after study therapy are tabulated by drug classification using the World Health Organization (WHO) drug dictionary. For Phase B **Safety Analysis Set**, data will be presented by treatment group and overall.

7.5 Protocol Deviations

Protocol deviations are summarized by center and type of deviation for randomized subjects by treatment group, and non-randomized subjects by ADT group. Listing of protocol deviation will list the treatment phases during which the deviations occurred.

7.6 Medical and Psychiatric History and Antidepressant Treatment History

A summary of medical and psychiatric history will be presented for ADT analysis set and Phase A+ analysis set (by ADT and overall) and for the randomized analysis set (by treatment group and overall). Summarized will be the duration of the current depressive episode, number of lifetime episodes, number and percentage of subjects who have a single current major depressive episode or recurrent major depressive episodes.

If subjects have only 1 MDE listed then categorize them as having single MDE and if subjects had more than 1 MDE then categorize them as having recurrent MDE.

7.7 Baseline Psychiatric Evaluation

For the randomized analysis set, baseline and end of Phase A 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: MADRS Total Score, CGI-S, CCI [redacted].

For the ADT analysis set and Phase A+ analysis set, baseline psychiatric scale evaluation will be summarized by ADT and overall.

8 Efficacy Analysis

8.1 Estimands

8.1.1 Primary Estimand

Based on historical trials in Brexpiprazole aMDD indication (Trials 331-13-214[8] 331-10-227[9] 331-10-228,[10] 331-12-282[11]) as an adjunctive therapy to antidepressants for the treatment of MDD in adults, we expect the response profiles to be similar between dropouts and completers. The drop-out rate should be low based on the historical data, and we will closely monitor dropout rate with blind data. The primary estimand defining the treatment effect of interest uses the hypothetical strategy specified in the International Council for Harmonisation (ICH) E9 Addendum. The estimand, or target of estimation, following the hypothetical strategy is the treatment effect seen, assuming no withdrawals occurred. Subjects who withdraw from IMP treatment either could have lost their treatment effect, had the subjects not taken any other treatment after withdrawal, or could have their treatment effect been masked, had the subjects taken other treatment after withdrawal. This means that any observations made after subjects stop IMP will most likely not contribute relevant information about the treatment effect of the drug. Due to this strategy, the last efficacy assessment after premature trial discontinuation will be performed only at the ET visit. Every effort will be made to complete all of the ET evaluations prior to administering any additional medications for the treatment of MDD or other prohibited medications. In the case of terminal or lost to follow-up events, no ET evaluations would be expected, and only scheduled assessments would be performed before such an event occurred.

The primary estimand for this trial is defined by the following components:

- Population: all subjects with MDD who meet the inclusion/exclusion criteria. For the efficacy analyses, the FAS will be used to represent the population of interest. Further details can be found in Section 5.
- Endpoint: the change from the end of Phase A (Week 8 visit) to the end of Phase B (Week 14 visit) in MADRS total score.
- Intercurrent Events: Intercurrent events refer to premature treatment discontinuation (ie, early dropout) prior to end of Phase B (Week 14 visit) attributable to AEs, lack of efficacy, withdrawal of consent, or any other causes.
- Treatment condition: one of the randomized treatment groups of the following per protocol: SEP-363856 (50 to 75 mg/day) + ADT or placebo + ADT.
- Population-level summary: difference in mean change of MADRS total score between two arms.

For the primary analysis of the primary endpoint, unobserved future data after participant discontinuation due to lack of efficacy and AE will be imputed using control-based multiple imputation under the MNAR assumption. Unobserved data after participant discontinuation due to other reasons will be analyzed under missing at random assumption. The resulting multiple imputed datasets will then be analyzed using a MMRM and the results will be combined using Rubin's rules. The MMRM model will include fixed factors for treatment, visit, center, and treatment-by-visit interaction, and include end of Phase A (Week 8 visit) baseline MADRS total score by visit as a covariate. An unstructured covariance matrix will be used to model the within-subject correlation. Kenward-Roger approximation will be used to calculate the denominator degrees of freedom. The main estimator of the primary efficacy estimand is the least squares mean difference in MADRS total score from the end of Phase A (Week 8 visit) to the end of Phase B (Week 14 visit) derived from the primary analysis.

For randomized participants who have baseline but do not have any post-baseline MADRS total score, missing post-baseline data will be imputed using a standard MAR-based multiple imputation approach from PROC MI, except for participants who discontinue due to lack of efficacy, and AE. For those participants missing due to lack of efficacy and AE, missing post-baseline data will be imputed using control-based multiple imputation under the MNAR assumption.

8.1.2 Secondary and Other Estimands

The key secondary estimand has similar attributes to the primary estimand except that key secondary endpoint is change from the end of Phase A (Week 8 visit) to end of

Phase B in CGI-S score. The estimands for other efficacy endpoints have similar attributes to the primary and key secondary estimands.

8.2 Primary Efficacy Endpoint

The primary efficacy endpoint is the change from the end of Phase A (Week 8 visit) to the end of Phase B (Week 14 visit) in the MADRS total score.

8.2.1 Primary Efficacy Analysis

The null hypothesis of the statistical test comparing SEP-363856 + ADT and placebo + ADT, based on the primary efficacy endpoint, is that the change in MADRS total score using SEP-363856 + ADT is equal to the change in MADRS using the placebo + ADT.

$$H0: \Delta\mu_{\text{MADRS SEP-363856 + ADT}} = \Delta\mu_{\text{MADRS placebo + ADT}}$$

$$HA: \Delta\mu_{\text{MADRS SEP-363856 + ADT}} \neq \Delta\mu_{\text{MADRS placebo + ADT}}$$

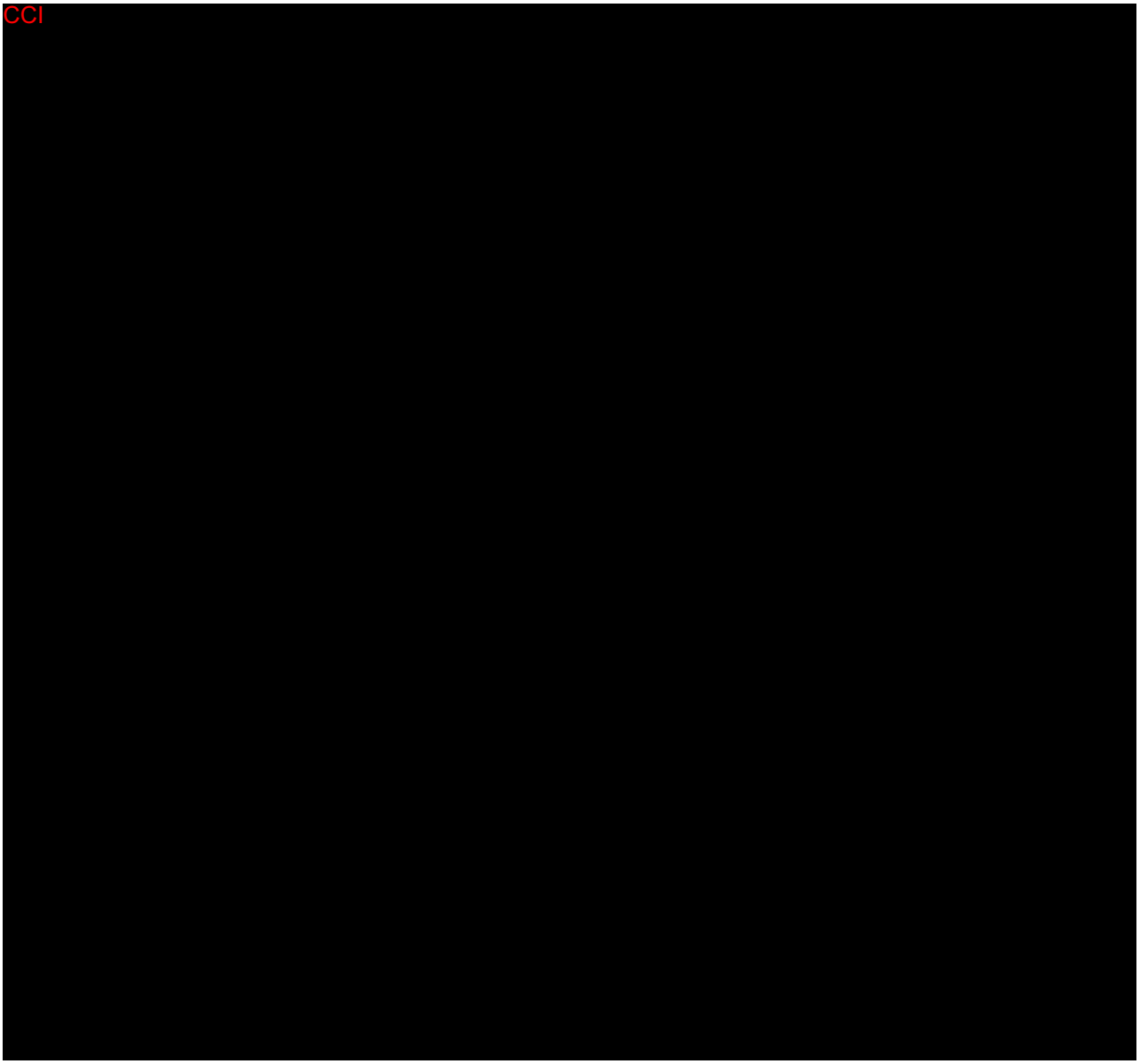
The objective of the primary analysis is to compare the efficacy of SEP-363856 (50 to 75 mg) to placebo as adjunctive therapy to an assigned open label marketed ADT in subjects with MDD who demonstrate an inadequate response to a prospective 8-week trial of the same assigned open-label ADT in Phase A. The primary efficacy endpoint is the change from the end of Phase A (Week 8 visit) to the end of Phase B (Week 14 visit) in MADRS total score. For analysis of Phase B data, the end of Phase A measurement is considered as baseline which is defined as the last available measurement prior to the first dose of double-blind IMP in Phase B.

The primary analysis will be performed on the FAS which includes all randomized subjects who take at least 1 dose of IMP in Phase B and who have an end of Phase A (Week 8) MADRS total score. Unobserved future data after participant discontinuation due to lack of efficacy and AE will be imputed using control-based multiple imputation under the MNAR assumption. Unobserved data after participant discontinuation due to other reasons will be analyzed under missing at random assumption. The resulting multiple imputed datasets will then be analyzed using a MMRM model with an unstructured variance covariance structure in which the change from the end of Phase A (Week 8 visit) in MADRS total score (at Weeks 9 to 14) will be the dependent variable. The model will include fixed class effect terms for treatment, pooled trial site, visit week, and an interaction term of treatment by visit week. The model will also include the interaction term of end of Phase A (Week 8 visit) values of MADRS total score by visit week as covariates. The primary comparison between the SEP-363856 + ADT group and the placebo + ADT group at Week 14 will be estimated as the overall difference between

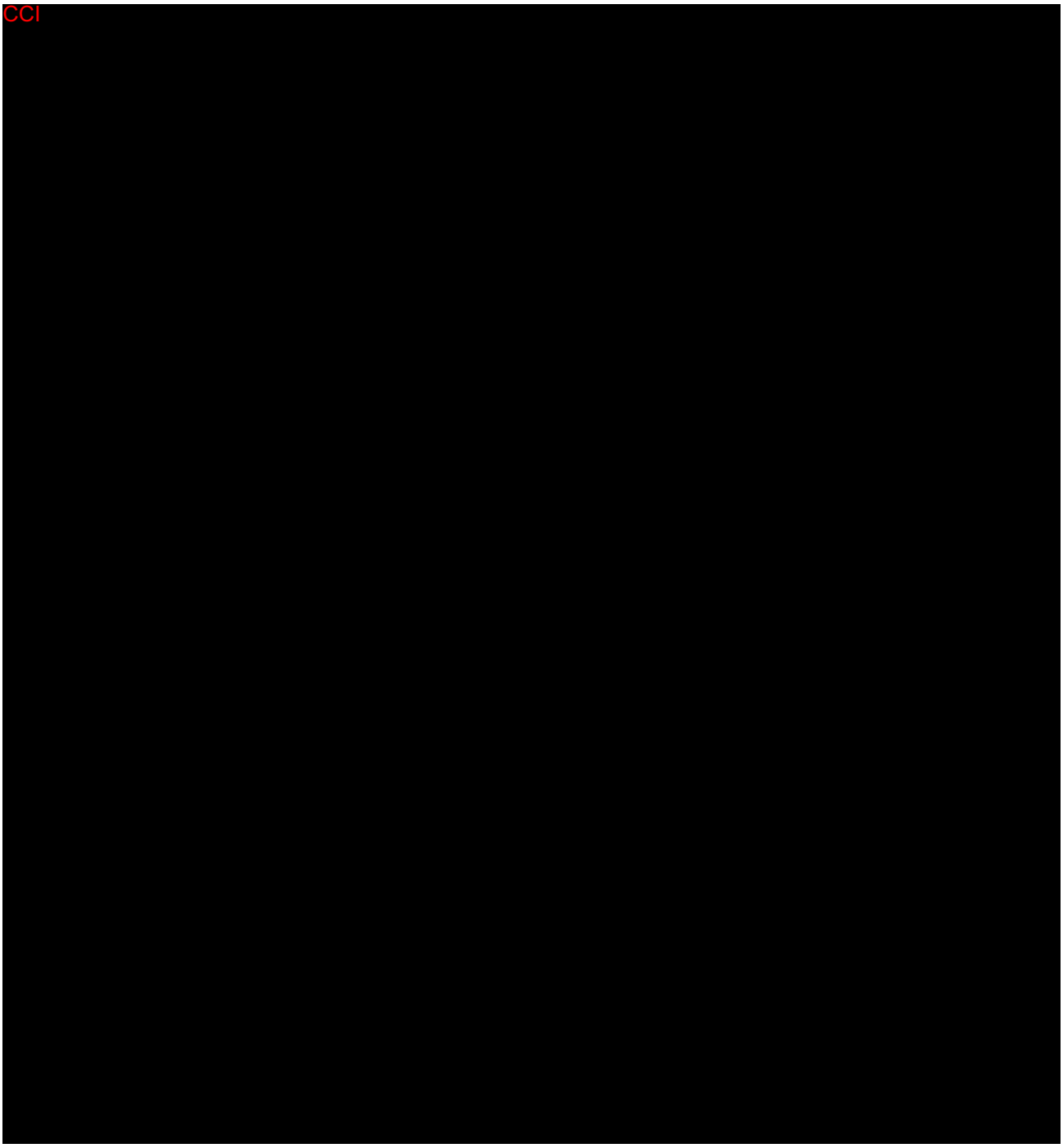
least squares (LS) means utilizing the computing software SAS procedure PROC MIXED using Rubin's rules. The comparison between the SEP-363856 + ADT group and placebo + ADT group will be tested at a significance level of 0.05 (2-sided).

In case there is a convergence problem with MMRM model with the unstructured variance covariance matrix, the following structures other than unstructured will be used in order of 1) heterogeneous toeplitz, 2) heterogeneous autoregressive of order 1, and 3) heterogeneous compound symmetry and the first (co)variance structure converging to the fit will be used as the primary analysis. If a structured covariance has to be used, the "sandwich" estimator of the standard error of the fixed effects parameters will be used in order to deal with possible model misspecification of the covariance matrix.

CCI



CCI



CCI



CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

8.3 Subgroup Analyses

Subgroup analyses of change from end of Phase A in MADRS Total Score to every study week in Phase B will be performed by the following factors:

- Sex at Birth
- Race (White and All Other Races)
- Age group (Age < 55 and Age ≥55)
- ADT (escitalopram, fluoxetine, paroxetine, sertraline, duloxetine, venlafaxine)
- Region (North America and Europe)
- Country (US and Non-US)
- Antidepressant class (Selective Serotonin Reuptake Inhibitors vs Serotonin-norepinephrine Reuptake Inhibitors)
- Subgroups based on percentages of improvement from baseline of Phase A in MADRS Total Score (< 25% improvement and ≥ 25% improvement)
- CCI [redacted]

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. If the sample size is too small (< 5 subjects) for the subgroup, then MMRM analysis will not be performed.

Interaction effects of treatment-by-subgroup will be assessed at Week 14 for the subgroups identified in the previous paragraph. MMRM analyses will be performed by adding interaction of terms for subgroup-by-week and treatment-by-subgroup-by-week. These treatment-by-subgroup interaction analyses will be presented in statistical documentation.

8.4 Key Secondary Efficacy Endpoint

The key secondary efficacy variable is the change from end of Phase A (Week 8 visit) to end of Phase B (Week 14 visit) in CGI-S.

This key secondary efficacy variable will be analyzed by fitting the same MMRM model described in the primary analysis. For the key secondary efficacy variable, a hierarchical testing procedure will be used so that the overall experiment-wise type I error rate is maintained at 0.05.

8.5 Control of Experiment-wise Type 1 Error

The overall type I error rate for the primary and key secondary endpoint will be maintained at 0.05. To control the overall experiment-wise type I error at 0.05 level, a fixed sequence testing approach will be applied. The statistical test will be performed in the following order:

1) Change from the end of Phase A (Week 8 visit) to the end of Phase B (Week 14 visit) in MADRS total score.
2) Change from end of Phase A (Week 8 visit) to end of Phase B (Week 14 visit) in CGI-S.

The test for the subsequent hypothesis will be performed only if the test for the previous hypothesis is statistically significant.

CCI



CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

8.8 Efficacy Outcome Measures in Phase A and A+

Summary statistics will be provided for efficacy outcome measures in Phase A and Phase A+. No inferential statistical analysis will be performed.

9 Safety Analyses

Standard safety variables to be analyzed include AEs, physical examinations (including body weight and waist circumference), vital signs (blood pressure, heart rate, and body temperature), 12-lead ECGs, and clinical laboratory assessments (hematology, serum chemistry, urinalysis, prolactin, glycosylated hemoglobin, and coagulation parameters). In addition to the standard safety variables, other safety variables including C-SSRS, SAS, AIMS, BARS, CSFQ, and PSQI will also be evaluated.

Analyses of Phase B safety data will be performed on the safety analysis set unless indicated otherwise.

The Safety Sample will also be analyzed for Phase A safety data, as applicable, by treatment groups for the subjects eventually being assigned in Phase B.

In addition, Phase A safety data are analyzed for the ADT analysis set, by ADT and overall. Phase A+ safety data are analyzed for the Phase A+ analysis set, by ADT and overall.

9.1 Extent of Exposure

9.1.1 Exposure to SEP-363856

The start date of double-blind study therapy will be the first day of double-blind dosing. The number and percentage of subjects who receive double-blind study medication, will be presented by week and by treatment group and by ADT. The first day of double-blind dosing is defined as “Day 1”. Each dosing week will be based on the actual week; ie,

1 week (Day 1 to 7), 2 weeks (Day 8 to 14), etc. This summary will be performed on the safety analysis set.

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 subjects receiving double-blind study medication, and the mean and range of the mean daily dose for each week.

9.1.2 Exposure to ADT

The number and percent of subjects who receive an ADT in Phase B will be presented by week, by ADT treatment group, and by double-blind treatment group for the safety analysis set. A similar summary will be prepared for Phase A. For each ADT the mean daily dose will be calculated in the same way as for double-blind study medication.

Summaries will be presented for ADT administered to subjects in the ADT analysis set during Phase A and Phase A+ analysis set during Phase A+.

In addition, the mean daily dosage of ADT will be summarized for the ADT analysis set during Phase A and the Phase A+ analysis set during Phase A+ by week and by ADT group using descriptive statistics.

9.2 Adverse Events

All AEs will be coded by Primary System Organ Class (SOC) and Preferred Term (PT) according to the Medical Dictionary for Regulatory Activities (MedDRA). The incidence of the following events will be summarized:

- Treatment-emergent AEs (TEAEs) by severity
- Potentially drug-related TEAEs
- TEAEs with an outcome of death
- Serious TEAEs
- Discontinuations due to TEAEs
- Abuse-potential TEAEs
- Psychiatric TEAEs that are serious or led to study discontinuation.

Treatment-emergent adverse events will be summarized by double-blind treatment group for Phase B, ADT groups for ADT analysis set and Phase A+ analysis set. Incidence of TEAEs by SOC and MedDRA preferred term will be summarized for sex at birth, race, and age subgroups.

9.2.1 Adverse Events in Phase B

Treatment-emergent adverse events in Phase B are defined as AEs with an onset date on or after the start of double-blind treatment. In more detail, TEAEs are all AEs which started after start of double-blind study drug treatment; or if the event was continuous from end of Phase A and was worsening, serious, study drug related, or resulted in death, discontinuation, interruption or reduction of study therapy. Adverse events occurring up to 7 days after the last day of double-blind dosing will be included in the summary tables.

The incidence of AEs in Phase B will be tabulated by treatment group and overall using the safety analysis set. Summary by ADT groups will also be provided. Incidence of TEAE during Phase B of at least 5% in the SEP-363856 + ADT group and also greater than placebo + ADT group by SOC and MedDRA preferred term will be provided.

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

The incidence of psychiatric TEAEs leading to discontinuation and serious psychiatric TEAEs will be summarized. TEAEs related to abuse potential will be provided.

9.2.2 Adverse Events in Phase A and A+

Treatment-emergent adverse events in Phase A are defined as AEs with an onset date in Phase A on or after the start of ADT treatment as entered on the open-label dosing form. Adverse events in Phase A will be summarized for subjects in the safety analysis set, ie, subjects who receive double-blind study medication. Adverse events occurring up to 7 days after the last day of ADT medication, but prior to the start of Phase B, will be included in these summary tables. The incidence of AEs in Phase A will be tabulated by the double-blind treatment subjects receive in Phase B.

Adverse events in Phase A will also be summarized by ADT for subjects in the ADT analysis set. Adverse events occurring up to 7 days after the last day of ADT medication, but prior to the start of Phase B or Phase A+ will be included in these summary tables.

The incidence of AEs in Phase A+ will be tabulated by ADT for subjects in the Phase A+ analysis set. Adverse events with an onset date up to 7 days after the last day of ADT medication will be included in these summary tables.

The incidence of psychiatric TEAEs leading to discontinuation and serious psychiatric TEAEs will be summarized. Treatment-emergent adverse events related to abuse potential will be provided.

9.3 Vital Sign Data

Vital signs include body temperature, heart rate, systolic blood pressure, and diastolic blood pressure. In addition, body weight, waist circumference, and BMI are measured. Summary statistics for vital signs will be provided. For Phase B vital signs, change from end of Phase A will be summarized for the safety analysis set by treatment group. Potentially clinically relevant results in vital signs will also be summarized. Similar summaries will be provided for the safety analysis set during Phase A, for the ADT analysis set during Phase A, and the Phase A+ analysis set during Phase A+.

9.3.1 Vital Signs in Phase B

Descriptive statistics will be provided by treatment group, for change from baseline in vital signs parameters.

The shift in BMI from baseline will be summarized by treatment group.

To assess whether results are potentially clinically relevant in Phase B, the end of Phase A assessments are used as "baseline." Potentially clinically relevant vital signs measurements identified in Phase B for the safety analysis set will be summarized by treatment group. Where end of Phase A values are missing and an end of Phase A value is necessary to identify a potentially clinically relevant event, then only the criterion values in Appendix 1 will be used to identify such events. All potentially clinically relevant events or changes will be listed and included in summary tables.

9.3.2 Vital Signs in Phase A and A+

Descriptive statistics will be provided by treatment group, for change from baseline in vital signs parameters.

Potentially clinically relevant vital signs measurements identified in Phase A for the safety analysis set will be summarized by treatment group.

Potentially clinically relevant vital signs measurement in Phase A (baseline through Week 8) will be listed and summarized for the ADT analysis set by ADT and overall.

Potentially clinically relevant vital signs measurements identified in Phase A+ (Week 9 through Week 14) for the Phase A+ analysis set will be listed and summarized by ADT and overall.

9.4 Clinical Laboratory Data

Summary statistics for routine clinical laboratory measurements, prolactin concentrations, coagulation parameters, HbA1c, and thyroid-stimulating hormone (TSH) will be provided. For Phase B laboratory tests, change from end of Phase A will be summarized by treatment group. Potentially clinically relevant results in laboratory tests will also be summarized.

9.4.1 Clinical Laboratory Tests in Phase B

Descriptive statistics will be provided by treatment group, for change from baseline in clinical laboratory parameters. The shift in laboratory parameters from baseline will be summarized by treatment group. The metabolic parameters will be summarized by fasting and not fasting.

Potentially clinically relevant laboratory measurement test results in Phase B will be identified for the safety analysis set and will be summarized by treatment group and listed. Criteria for identifying laboratory values of potential clinical relevance are provided in Appendix 2.

9.4.2 Drug Induced Liver Injury

Total bilirubin level should be checked for any subject with increased alanine aminotransferase (ALT) or aspartate aminotransferase (AST) levels ≥ 3 × upper limit of normal (ULN) or baseline.

Reporting all drug induced liver injuries (DILI) as serious adverse events (SAEs) to the US Food and Drug Administration (FDA) based on Hy's Law:

- AST or ALT > 3 × ULN or baseline and
- T_Bili > 2 × ULN or baseline

A separate incidence table will be provided for DILI cases, and the corresponding listing will be provided for safety analysis set during Phase B and Phase A, for the ADT analysis set during Phase A, and the Phase A+ analysis set during Phase A+.

9.4.3 Clinical Laboratory Tests in Phase A and A+

Descriptive statistics will be provided by treatment group, for change from baseline in clinical laboratory parameters.

Potentially clinically relevant laboratory measurement test results in Phase A will be summarized for the safety analysis set by treatment group and overall as well as listed by subject and by laboratory test.

Potentially clinically relevant laboratory measurement test results in Phase A (baseline through Week 8) will be listed by subject and by laboratory test and summarized for the ADT analysis set by ADT and overall.

Potentially clinically relevant laboratory measurement test results in Phase A+ (Week 9 through Week 14) will be listed and summarized for the Phase A+ analysis set by ADT and overall.

9.5 Physical Examination Data

By-subject listings will be provided for physical examination.

9.5.1 Body Weight, Waist Circumference, and Body Mass Index

9.5.1.1 Body Weight, Waist Circumference, and Body Mass Index in Phase B

Analyses of body weight, waist circumference, and BMI will be performed for the safety analysis set. The mean change from end of Phase A to Week 14 (OC) and last visit in Phase B in body weight will be tabulated and analyzed using ANCOVA. The ANCOVA models for both the OC and last visit analyses will include the end of Phase A body weight and the treatment group.

Percentages of subjects showing significant weight gain (≥ 7 % increase in weight), as well as percentages of subjects showing significant weight loss (≥ 7 % decrease in weight) from end of Phase A to Week 14 (OC and LOCF) will be analyzed using Cochran-Mantel-Haenszel (CMH) General Association Test.

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

9.5.1.2 Body Weight, Waist Circumference, and Body Mass Index in Phase A and A Plus

Body weight, waist circumference, and BMI in Phase A of the study will be summarized for the safety analysis set by treatment group and overall, and for the ADT Sample by ADT and overall. Number of subjects, sample mean, sample standard deviation, and range will be presented for change from baseline in weight, in waist circumference, and in BMI at Week 8. These summaries will be performed on the OC data set.

Body Weight, waist circumference, and BMI from subjects in the Phase A+ Sample will be summarized per study week. Number of subjects, sample mean, sample standard deviation, and range will be presented for change from baseline in weight, waist circumference, and BMI at study weeks 8 and 14. These summaries will be performed on the OC data and last visit.

9.6 12-Lead Electrocardiograms

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

For the analysis of QT, data from three consecutive complexes (representing 3 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:

QTcF is the length of the QT interval corrected for heart rate by the Fridericia formula:

$$QTcF=QT/(RR)^{0.33}$$

9.6.1 12-Lead Electrocardiograms in Phase B

Descriptive statistics will be provided by treatment group, for change from baseline in ECG.

To assess whether results are potentially clinically relevant in Phase B, the end of Phase A assessments are used as "baseline." Potentially clinically relevant changes in the 12-lead ECG identified in Phase B for the safety analysis set will be listed and summarized by treatment group. If end of Phase A measures are missing and the end of Phase A ECG assessment is necessary to identify clinically relevant events, then the criterion values in Appendix 3 only will be used to identify the relevant events.

9.6.2 12-Lead Electrocardiograms in Phase A and A+

Descriptive statistics will be provided by treatment group, for change from baseline in ECG.

Potentially clinically relevant changes in the 12-lead ECG identified in Phase A for the safety analysis set will be listed and summarized by treatment group.

Potentially clinically relevant ECG measurement results in Phase A (baseline through Week 8) will be listed and summarized for the ADT analysis set by ADT and overall.

Potentially clinically relevant ECG measurements identified in Phase A+ (Week 9 through Week 14) for the Phase A+ analysis set will be listed and summarized by ADT and overall.

9.7 Suicidality

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, ideation by type, and treatment-emergent suicidal behavior and ideation. Summary will be provided for Phase A (safety analysis set and ADT analysis set), Phase B (safety analysis set), and Phase A+ (Phase A+ analysis set).

Suicidality is defined as report of at least one occurrence of any type of suicidal ideation or at least 1 occurrence of any type of suicidal behavior during assessment period (count each person only once).

Treatment-emergent suicidal behavior and ideation is summarized by 4 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.

For Phase B analyses, End of Phase A is being used as “baseline.”

9.8 Simpson-Angus Scale, Abnormal Involuntary Movement Scale, and Barnes Akathisia Rating Scale

9.8.1 Simpson-Angus Scale, Abnormal Involuntary Movement Scale, and Barnes Akathisia Rating Scale in Phase B

The mean change from end of Phase A to every study week in Phase B obtained from the SAS Total Score, AIMS Total Score (total of the first 7 item scores), and the BARS Global Clinical Assessment will be tabulated and analyzed using ANCOVA. Analyses will be performed on the OC data set. In addition, analyses will be performed using the maximum (ie, the worst) value observed during Phase B and the last visit data to determine the change from end of Phase A score. The ANCOVA model for the OC data set will include the end of Phase A measure and the treatment group. The ANCOVA model for change at the last visit and for change to the maximum value will include the end of Phase A measure, study center and treatment group. The same analyses will be performed on the AIMS individual item scores 8, 9, and 10. In addition, incidence BARS

Global Clinical Assessment of Akathisia during Phase B by severity category will be provided. Analyses of these extrapyramidal symptoms (EPS) rating scales will be performed for the safety analysis set.

9.8.2 SAS, AIMS, and BARS in Phase A and A+

SAS Total Score, AIMS Total Score (total of the first 7 item scores), and the BARS Global Clinical measurements from subjects in the Phase A+ analysis set will be summarized by ADT per study week. Number of subjects, sample mean, sample standard deviation, and range will be presented for change from baseline in: SAS Total Score, AIMS Total Score, AIMS individual item scores 8, 9, and 10, and the Barnes Akathisia Scale Global Clinical Assessment. These summaries will be performed on the OC data and last visit.

9.9 Sexual Functioning Questionnaire

9.9.1 Sexual Functioning Questionnaire in Phase B

Change from end of Phase A (Week 8 visit) to the end of Phase B (Week 14 visit) in scores for the CSFQ scale will be evaluated separately by sex at birth using ANCOVA. The ANCOVA model for the OC data set will include the end of Phase A score as covariate and treatment group as main effects. The ANCOVA model for last visit will include the end of Phase A score as covariate, and study center, and treatment group as main effects.

9.9.2 Sexual Functioning Questionnaire in Phase A and A+

The CSFQ in Phase A of the study will be summarized for the safety analysis set by treatment group and overall, and for the ADT analysis set by ADT and overall. Number of subjects, sample mean, sample standard deviation, and range will be presented for change from baseline in CSFQ scale at Week 8. These summaries will be performed on the OC data set.

CSFQ from subjects in the Phase A+ analysis set will be summarized per study week. Number of subjects, sample mean, sample standard deviation, and range will be presented for change from baseline in CSFQ at study Weeks 8 and 14. These summaries will be performed on the OC data and last visit.

9.10 Pittsburgh Sleep Quality Index

9.10.1 Pittsburgh Sleep Quality Index in Phase B

Change from end of Phase A (Week 8 visit) to the end of Phase B (Week 14 visit) in scores for the PSQI scale will be evaluated using ANCOVA. The ANCOVA model for the OC data set will include the end of Phase A item score as covariate and treatment group as main effects. The ANCOVA model for last visit will include the end of Phase A item score as covariate, and study center and treatment group as main effects.

9.10.2 Pittsburgh Sleep Quality Index in Phase A and A+

The PSQI will be summarized for the safety analysis set by treatment group and overall, and for the ADT analysis set by ADT and overall. Number of subjects, sample mean, sample standard deviation, and range will be presented for change from baseline in PSQI at Week 8. These summaries will be performed on the OC data set.

Pittsburgh Sleep Quality Indexes from subjects in the Phase A+ analysis set will be summarized per study week. Number of subjects, sample mean, sample standard deviation, and range will be presented for change from baseline in PSQI at study Weeks 8 and 14. These summaries will be performed on the OC data and last visit.

10 Rules for Scoring

10.1 Montgomery-Asberg Depression Rating Scale

The MADRS is utilized as the primary efficacy assessment of a patient's level of depression. The MADRS consists of 10 items, all rated on a 0 to 6 scale with zero being the "best" rating and 6 being the "worst" rating. A higher score on the MADRS represents a higher severity of the level of depression. The MADRS Total Score is the sum of ratings for all 10 items. The possible Total scores are from 0 to 60. The MADRS Total Score will be un-evaluable if less than 8 of the 10 items are recorded. If 8 or 9 of the 10 items are recorded, the MADRS Total Score will be the mean of the recorded items multiplied by 10 and then rounded to the first decimal place.

10.2 Clinical Global Impression

Clinical Global Impression consists of two scales: CGI-S and CGI-C.

Clinical Global Impression Severity items are: 1 = normal, not at all ill; 2 = borderline ill; 3 = mildly ill; 4 = moderately ill; 5 = markedly ill; 6 = severely ill; and 7 = among the most extremely ill patients. The CGI-S is therefore a 7-point scale from 1 through 7.

Clinical Global Impression - Change items are: 1 = very much improved, 2 = much improved, 3 = minimally improved, 4 = no change, 5 = minimally worse, 6 = much worse, and 7 = very much worse. During Phase A and Phase A+, the response at a given week will be compared to the subject's condition at baseline prior to the first dose of IMP in Phase A. For subjects who are randomized into Phase B, the subject's condition at a given week will be compared to the subject's condition at randomization (ie, Week 8 of Phase A).

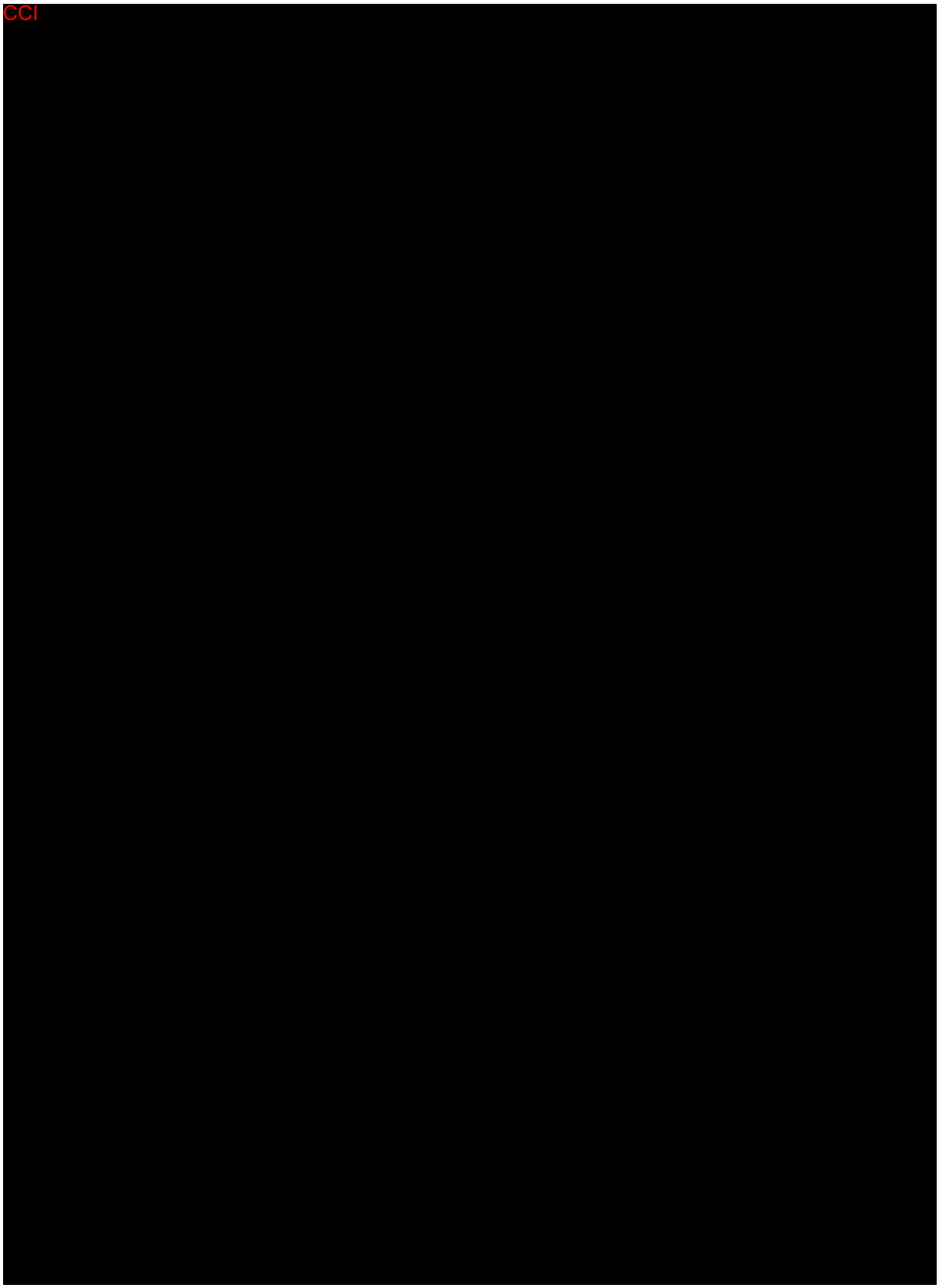
CCI



10.6 Columbia-Suicide Severity Rating Scale

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. The "Since Last Visit" version will be completed at all visits after screening.

CCI



and S2 and rounding this score to the first decimal place. The possible total scores are from 0 to 56.

10.9 Changes in Sexual Functioning Questionnaire

The CSFQ is a 14-item questionnaire that produces an overall measure of sexual functioning. The total score will be calculated as the sum of the scores for all 14 items. If three or more of the 14 items contributing to CSFQ total score calculation are missing at a visit, the total score will be set to missing for that visit. If 12 or 13 of the 14 items are recorded, the CSFQ total score will be the mean of the recorded items multiplied by 14 and then rounded to the first decimal place.

10.10 Simpson-Angus Scale

The SAS is a rating scale used to measure EPS. The SAS is a 10-item scale, with each item rated from 0 to 4, with a score of zero representing absence of symptoms, and a score of 4 representing a severe condition. The SAS Total score is the sum of ratings for all 10 items, with possible Total scores from 0 to 40. The SAS Total score will be un-evaluable if less than 8 of the 10 items are recorded. If 8 or 9 of the 10 items are recorded, the Total score will be the mean of the recorded items multiplied by 10 and then rounded to the first decimal place.

10.11 Abnormal Involuntary Movement Scale

The AIMS assessment consists of 10 items describing symptoms of dyskinesia. Facial and oral movements (items 1 through 4), extremity movements (items 5 and 6), and trunk movements (item 7) will be observed unobtrusively while the subject is at rest (eg, in the waiting room), and the investigator will also make global judgments on the subject's dyskinesias (items 8 through 10). Each item will be rated on a 5-point scale, with a score of zero representing absence of symptoms (for item 10, no awareness), and a score of 4 indicating a severe condition (for item 10, awareness, severe distress). For this scale, the subject is to be sitting on a hard, firm chair. In addition, the AIMS includes 2 yes/no questions that address the subject's dental status. The AIMS Total Score is the sum of the ratings for the first 7 items. The possible total scores are from 0 to 28. The AIMS Total Score will be un-evaluable if less than 6 of the first 7 items are recorded. If only 1 of the first 7 items is missing, then the total score will be the mean of the recorded items multiplied by 7 and then rounded to the first decimal place.

10.12 Pittsburgh Sleep Quality Index

The PSQI contains 19 self-rated questions and 5 questions rated by the bed partner or roommate (if one is available). Only self-rated questions are included in the scoring. The

19 self-rated items are combined to form 7 “component” scores, each of which has a range of 0 to 3 points. In all cases, a score of “0” indicates no difficulty, while a score of “3” indicates severe difficulty. The 7 component scores are then added to yield 1 “global” score, with a range of 0 to 21 points, “0” indicating no difficulty and “21” indicating severe difficulties in all areas.

10.13 Barnes Akathisia Rating Scale

The Barnes Akathisia Rating Scale will be used to assess the presence and severity of akathisia. The first 3 items will be rated on a 4-point scale, with a score of zero representing absence of symptoms and a score of 3 representing a severe condition. The global clinical evaluation will be made on a 6-point scale, with zero representing absence of symptoms and a score of 5 representing severe akathisia.

CCI

[redacted]

CCI

[redacted]

12 References

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

[2] Diggle P, Kenward MG. Informative drop-out in longitudinal data analysis. Applied Statistics. 1994;43(1):49-93.

[3] Little RJA. Pattern-mixture models for multivariate incomplete data. J Am Stat Assoc. 1993;88(421):125-134.

[4] Little RJA. Modeling the drop-out mechanism in repeated measures studies. J Am Stat Assoc. 1995;90(431):1112-1121.

[5] Hedeker D, Gibbons RD. Application of random effects pattern-mixture models for missing data in longitudinal studies. Psychological Methods. 1997;2(1):64-78.

[6] Ali MW, Siddiqui O. Multiple imputation compared with some informative dropout procedures in the estimation and comparison of rates of change in longitudinal clinical trials with dropouts. J Biopharm Stat. 2000;10(2):165-181.

[7] Wu MC, Bailey KR. Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. Biometrics. 1989;45(3):939-955.

[8] Hobart M. A phase 3, multicenter, randomized, double-blind, placebo-controlled trial of the safety and efficacy of fixed-dose brexpiprazole (OPC-34712) as adjunctive therapy in the treatment of adults with major depressive disorder with and without anxious distress. Otsuka Clinical Study Report for Protocol 331-13-214, issued 14 Oct 2016.

[9] Skuban A. A phase 3, multicenter, randomized, double-blind, placebo-controlled trial of the safety and efficacy of two fixed-doses of OPC-34712 as adjunctive therapy in the treatment of adults with major depressive disorder, the Polaris trial. Otsuka Clinical Study Report for Protocol 331-10-227, issued 04 Jun 2014

[10] Youakim JM. A phase 3, multicenter, randomized, double-blind, placebo-controlled trial of the safety and efficacy of fixed-dose OPC-34712 as adjunctive therapy in the treatment of adults with major depressive disorder, the Pyxis Trial. Otsuka Clinical Study Report for Protocol 331-10-228, issued 18 Sep 2013.

[11] Hobart M. A phase 3, multicenter, randomized, double-blind, placebo- and active comparator-controlled trial of flexible-dose brexpiprazole (OPC-34712_ as adjunctive therapy in the treatment of adults with major depressive disorder, the Delphinus trial. Otsuka Clinical Study Report for Protocol 331-12-282, issued 22 Mar 2017.

[12] Van Elteren PH. On the combination of independent two sample tests of Wilcoxon. Bull Inst Intern Staist. 1960;37:351-361.

[13] Mehrotra DV, 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.

[14] Cortina JM, Nouri H. Effect size for ANOVA designs. Thousand Oaks, CA. Sage Publications; 2000.

[15] Hopkins SC, Tomioka S, Szabo ST, Koblan KS. A clinical trial inclusion criteria to enrich for patients presenting with canonical symptom structure in bipolar depression. Contemp Clin Trials. 2024;145:107644.

13 Appendices

Appendix 1 Criteria for Identifying Vital Signs and Weight 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	Decrease of ≥ 20 mmHg in systolic blood pressure and/or a decrease of ≥ 10 mmHg in diastolic blood pressure after at least 2 to 4 minutes standing compared to the previous supine blood pressure.	-
Weight	-	≥ 7% increase ≥ 7% decrease
Temperature (°C)	≥ 38.3°C	≥ 0.8°C increase

FDA = US Food and Drug Administration; NDA = new drug application.

[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 (27 Feb 1987).

Appendix 2 Criteria for Identifying Laboratory Values of Potential Clinical Relevance

Category Parameter Name Age/Sex Restriction, if any	Low	High
HEMATOLOGY		
WBC	≤ 2.8 × 10^9/L	≥ 16 × 10^9/L
Neutrophils (abs)	< 0.5 × 10^9/L	> 13.5 × 10^9/L
Lymphocytes (abs)	N/A	> 12 × 10^9/L
Monocytes (abs)	N/A	> 2.5 × 10^9/L
Eosinophils (abs)	N/A	> 1.6 × 10^9/L
Basophils (abs)	N/A	> 1.6 × 10^9/L
Neutrophils (relative)	≤ 15%	> 85%
Lymphocytes (relative)	N/A	≥ 75%
Monocytes (relative)	N/A	≥ 15%
Eosinophils (relative)	N/A	≥ 10%
Basophils (relative)	N/A	≥ 10%
Hemoglobin		
Male	≤ 115 g/L	≥ 190 g/L
Female	≤ 95 g/L	≥ 175 g/L
Hematocrit		
Male	37%	60%
Female	32%	54%
RBC	≤ 3.5 × 10^{12}/L	≥ 6.4 × 10^{12}/L
Platelet Count	≤ 75 × 10^9/L	≥ 700 × 10^9/L
SERUM CHEMISTRY		
Sodium	< 130 mmol/L	> 150 mmol/L
Potassium	< 3 mmol/L	> 5.5 mmol/L
Chloride	≤ 90 mmol/L	≥ 118 mmol/L
Calcium	< 1.75 mmol/L	≥ 3.1 mmol/L
Inorganic phosphorus	< 0.65 mmol/L	> 1.65 mmol/L
Bicarbonate	< 15.1 mmol/L	> 34.9 mmol/L
Magnesium	< 0.33 mmol/L	> 1.01 mmol/L
AST	N/A	≥ 3 × ULN
ALT	N/A	≥ 3 × ULN
Alkaline Phosphatase	N/A	≥ 1.5 × ULN
CK	N/A	> 2.5 × ULN
Creatinine	N/A	≥ 176.8 umol/L
Blood Urea Nitrogen (BUN)	N/A	≥ 10.7 mmol/L
Total bilirubin	N/A	≥ 34.2 umol/L OR > 2 × ULN
Total protein	≤ 45.0 g/L	≥ 100 g/L
Albumin	≤ 25 g/L	N/A
Total Cholesterol	N/A	> 7.758 mmol/L
HDL-Cholesterol	< 0.78 mmol/L	N/A
LDL-Cholesterol	N/A	> 4.91 mmol/L
Triglycerides	N/A	> 3.39 mmol/L
Uric Acid		
Male	N/A	> 594.8 umol/L
Female	N/A	> 475.8 umol/L

Category **Parameter Name** **Age/Sex Restriction, if any**	**Low**	**High**
Glucose		
Fasting	-	≥ 5.551 mmol/L
Non-Fasting	-	≥ 11.102 mmol/L
Insulin	< 18.595 pmol/L	> 178.719 pmol/L
Lactate dehydrogenase		
Male	< 135 U/L	> 281 U/L
Female	< 135 U/L	> 330 U/L
C-reactive protein	N/A	> 47.6 nmol/L
Gamma glutamyl transferase		
Male	< 10 U/L	> 49 U/L
Female	< 5 U/L	> 32 U/L
Glycosylated hemoglobin (HbA1c)	N/A	≥ 7.5%
Prolactin	N/A	≥ 5 × ULN
Coagulation Parameters		
PT (Prothrombin Time)	< 9.4 s	> 12.5 s
aPTT (activated PTT)	< 25.1 s	> 36.5 s
INR	< 0.8	> 1.2
URINALYSIS		
Blood	N/A	test positive
Protein	N/A	test positive
Glucose	N/A	test positive
Ketones	N/A	test positive
pH	< 4.6	> 8.0
Specific gravity	< 1.002 g/mL	> 1.035 g/mL

abs = absolute; ALT = alanine transaminase; aPTT = activated PTT; AST = aspartate aminotransferase; BUN = blood urea nitrogen; HbA1C = glycosylated hemoglobin; HDL = high-density lipoprotein; INR = international normalized ratio; LDL = low-density lipoprotein; N/A = not applicable; PT = prothrombin time; RBC = red blood cell; TSH = thyroid-stimulating hormone; ULN = upper limit of normal; WBC = white blood cell.

Appendix 3 Criteria for Identifying Electrocardiogram 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 QTcF	QTcF > 450 msec (male) QTcF > 470 msec (female)	-

[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.

Note: Potential Clinical Relevance will be identified by post-baseline diagnosis or the criterion value and change relative to baseline based on measurements. "No current diagnosis" in footnotes b, c, and d means that the condition already existed and diagnosed at baseline.

This page is a manifestation of an electronically captured signature

SIGNATURE PAGE

Document Name: 382-201-00001_SAP

Document Number: 1000236384

Signed by	Meaning of Signature	Server Date (GMT) dd-MM-yyyy hh:mm:ss
PPD	Biostatistics Approval	06-Nov-2025 17:01:36
PPD	Clinical Approval	06-Nov-2025 20:11:04