STATISTICAL ANALYSIS PLAN

SEP380-201

A RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED, PARALLEL-GROUP STUDY OF SEP-4199 FOR THE TREATMENT OF MAJOR DEPRESSIVE EPISODE ASSOCIATED WITH BIPOLAR I DISORDER (BIPOLAR I DEPRESSION)

AUTHOR:

VERSION NUMBER AND DATE: V2.0, 21MAY2020

STATISTICAL ANALYSIS PLAN SIGNATURE PAGE

Statistical Analysis Plan V2.0 (Dated 21MAY2020) for Protocol SEP380-201.

	Name	Signature	Date
Author:			
Position:			
Company:			

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

	Name	Signature	Date
Approved By:			
Position:			
Company:			
Approved By:			
Position:			
Company:			
Approved By:			
Position:			
Company:			
Approved By:	Galluppi		
Position:			
Company:			
Approved By:			
Position:			
Company:			

MODIFICATION HISTORY

Unique Identifier for this Version	Date of the Document Version	Author	Significant Changes from Previous Authorized Version
V1.0	06SEP2018		Not Applicable – First Version
V2.0	15MAY2020		Addition of Physician Withdrawal Criteria in Section 17.6.6. Results of the population pharmacokinetic/pharmacodynamic analyses and the relationship between SEP-4199 PK and plasma prolactin will be reported separately from the CSR. (see Section 3.7). Medications that started the day of last study drug are considered concomitant as specified in Protocol. SAS code is appendix for MCP-Mod was changed. Addition of SAS code for MMRM and ANCOVA for all SEP-4199 results to be presented for safety questionnaires. Addition of sensitivity analyses due to COVID-19: impact of remote visits and overall impact. Addition of COVID-19 related adverse/pre-treatment event identification. Removal of appendix for Multiplicity and update in Section 7.5. Revision to the ITT population after the BDR Meeting to be consistent with prior Sunovion CNS studies.

TABLE OF CONTENTS

1. INTRODUCTION

This document describes the rules and conventions to be used in the presentation and analysis of efficacy and safety data for Protocol SEP380-201. It describes the data to be summarized and analyzed, including specifics of the statistical analyses to be performed.

Data and Safety Monitoring Board (DSMB) analysis plan will be described in separate documents. A Blind Data Review (BDR) Plan will be written to describe the process and the outputs to be delivered during the IPD/BDR meetings.

This statistical analysis plan (SAP) is based on protocol version 2.0, dated 04 June 2018. Specific amendments were developed for Japan (version 2.01A dated 29 January 2019) and Poland (version 2.00B dated 18 January 2019) Hereafter, this protocol version is referred to as the Clinical Study Protocol (CSP). This SAP covers both the United States/European Union (US/EU) sites analysis, the Japan sites analysis and their pooled analysis.

2. STUDY OBJECTIVES

2.1. PRIMARY OBJECTIVE

The primary objective is to evaluate the efficacy of SEP-4199 200 mg/day and 400 mg/day compared with placebo for major depressive episode associated with bipolar I depression (diagnosed by DSM-5 criteria) as measured by Montgomery-Asberg Depression Rating Scale (MADRS) total score.

2.2. SECONDARY OBJECTIVES

The secondary objective is to evaluate the effect of SEP-4199 200 mg/day and 400 mg/day compared with placebo on severity of illness as measured by the Clinical Global Impression Bipolar Version, Severity of Illness (CGI-BP-S) score (depression).

2.3. PHARMACOKINETIC OBJECTIVES:

The pharmacokinetic objectives are:

- To evaluate the therapeutic plasma concentration range of SEP-4199 200 mg/day and 400 mg/day for major depressive episode associated with bipolar I disorder.
- To determine the population PK of SEP-4199 200 mg/day and 400 mg/day.

2.4. OTHER OBJECTIVES

Other objectives for the study include:

- To determine the relationship between SEP-4199 PK and plasma prolactin for SEP-4199 200 mg/day and 400 mg/day.
- To characterize the exposure-response relationship of SEP-4199 200 mg/day and 400 mg/day and symptoms as measured by MADRS using population PK/PD.
- To evaluate the effect of SEP-4199 200 mg/day and 400 mg/day compared with placebo on anxiety symptoms, as measured by the Hamilton Rating Scale for Anxiety (HAM-A).
- To evaluate treatment response, defined as ≥ 50% reduction from baseline on the MADRS total score.
- To evaluate the effect of SEP-4199 200 mg/day and 400 mg/day on symptom remission, defined as a MADRS total score of ≤ 12 after 6 weeks of treatment.
- To evaluate the effect of SEP-4199 200 mg/day and 400 mg/day on functional impairment associated with bipolar depressive symptoms, as measured by the Sheehan Disability Scale (SDS) total score.
- To evaluate the effect of SEP-4199 200 mg/day and 400 mg/day on symptom severity as measured by the Quick Inventory of Depressive Symptomatology – Self Report (QIDS-SR16) total score.

2.5. SAFETY OBJECTIVES

The safety objectives are:

- To evaluate the effect of SEP-4199 200 mg/day and 400 mg/day on treatment-emergent mania, as assessed by the Young Mania Rating Scale (YMRS) or an adverse event (AE) of mania or hypomania.
- To evaluate safety and tolerability of SEP-4199 200 mg/day and 400 mg/day as measured by physical examinations, 12-lead electrocardiograms (ECG) parameters, vital signs, AE reports, clinical laboratory results, Columbia-Suicide Severity Rating Scale (C-SSRS), Abnormal Involuntary Movement Scale (AIMS), Barnes Akathisia Rating Scale (BARS), and the Modified Simpson-Angus Scale (SAS).

2.6. STUDY ENDPOINTS

2.6.1. PRIMARY EFFICACY ENDPOINT:

- Change from baseline in MADRS total score at Week 6.

2.6.2. SECONDARY EFFICACY ENDPOINT:

- Change from baseline in global severity assessed by the CGI-BP-S score (depression) at Week 6.

2.6.3. PHARMACOKINETIC ENDPOINTS:

- Plasma concentrations of R-amisulpride and S-amisulpride.
- Plasma concentrations of prolactin.

2.6.4. OTHER ENDPOINTS:

- Number and percentage of treatment responders, defined as ≥ 50% reduction from baseline in MADRS total score at Week 6.
- Time to treatment response, defined as ≥ 50% reduction from baseline MADRS total score.
- Change from baseline in anxiety symptoms based on the HAM-A total score at Week 6.
- Incidence of symptom remission, defined as a MADRS total score of ≤ 12 at Week 6.
- Time to remission, defined as MADRS total score of ≤ 12.
- Functional impairment assessed by change from baseline in the SDS total score at Week 6.
- Subject self-report of overall depressive symptom severity, assessed by change from baseline in the QIDS-SR16 total score at Week 6.

2.6.5. SAFETY ENDPOINTS:

- Incidence of treatment-emergent mania, defined as a YMRS score of ≥ 16 on any 2 consecutive visits or at the final assessment, or an AE of mania or hypomania.
- Incidence of AEs, discontinuation due to AEs, and serious AEs (SAEs).
- Changes in weight, clinical laboratory tests, vital signs, and ECG measurements.
- Change from baseline in AIMS, BARS, and Modified SAS.
- Frequency and severity of suicidal ideation and suicidal behavior using the C-SSRS.

3. STUDY DESIGN

3.1. GENERAL DESCRIPTION

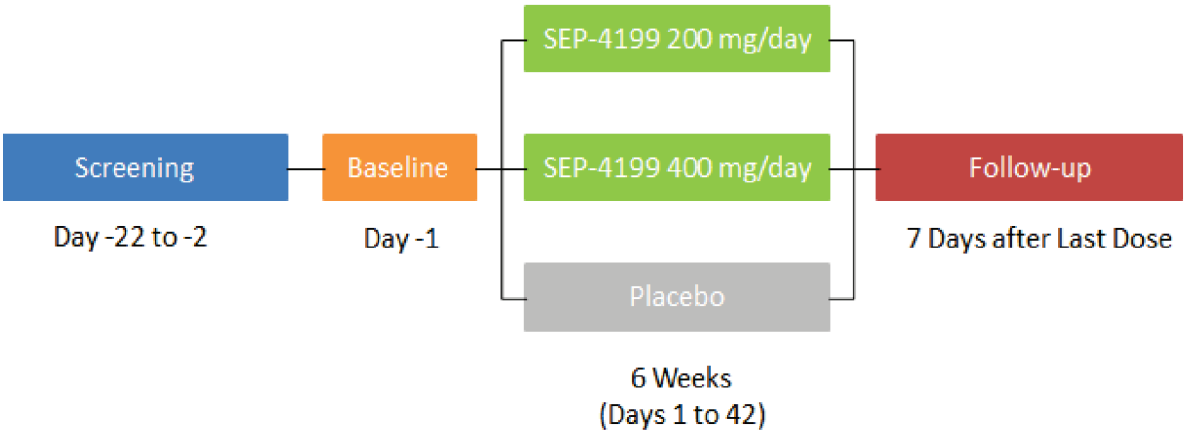
This is a randomized, double-blind, placebo-controlled, parallel-group, fixed-dose study designed to evaluate the efficacy, safety, and tolerability of treatment with SEP-4199 monotherapy given as 200 mg/day or 400 mg/day compared with placebo for the treatment of major depressive episode associated with bipolar I disorder (bipolar I depression). The study is projected to randomize approximately 294 subjects (98 subjects per treatment group) in the US and Europe to one of 3 treatment groups in a 1:1:1 ratio (SEP-4199 200 mg/day or 400 mg/day, or placebo).

In addition, a Japanese cohort will be included in the study to summarize efficacy and safety data for Japanese subjects with bipolar I depression. Approximately 45 subjects (15 subjects per treatment group) will be randomized in the Japanese cohort in a 1:1:1 ratio (SEP-4199 200 mg/day or 400 mg/day, or placebo).

The study will consist of a Screening period (up to 21 days), a 6-week treatment period (42 days), and a follow-up period (7 [±2] days after the last study drug dose). The screening visit will be defined as Visit 1, the Baseline/Week 0 visit will be defined as Visit 2, Weeks 1, 2, 4, and 6 will be defined as Visits 3-6, respectively, and the Follow-up/Week 7 visit will be defined as Visit 7 (Table 2).

A study schematic is presented in Figure 1. Details of the study assessments and other procedures to be performed at each visit are presented in Table 2 and Section 11 of the CSP. If necessary, subjects may return to the clinic at any time for an unscheduled visit.

Figure 1 Study Schematic – All Subjects



3.2. METHOD OF ASSIGNING SUBJECTS TO TREATMENT GROUPS

Randomization will be stratified by region. The treatment schedule will be generated by a non-study biostatistician. Once a subject is deemed eligible to be randomized at Day -1, an interactive voice/web-based system (IXRS) will perform the treatment assignment. Subjects will be randomized to 1 of 3 treatments in a 1:1:1 ratio:

- SEP-4199 200 mg/day
- SEP-4199 400 mg/day
- Placebo QD

3.3. BLINDING

Subjects, Investigators, clinical site staff, persons performing the assessments, clinical operations personnel (including the sponsor's bioanalytical manager), data analysts, and personnel at central laboratories (including imaging) will remain blinded to the identity of the treatment from the time of randomization until unblinding, using the following methods:

- Randomization data are kept strictly confidential until the time of unblinding, and will not be accessible by anyone involved in the study with the following exceptions: bioanalytical laboratory personnel involved in the analysis of PK samples, DSMB members involved in regular review of safety data, external statistical staff involved in preparing materials for DSMB reviews, and the Sponsor's clinical trials materials management.
- Prolactin levels will be blinded except for results from Visit 1 (Screening).

  The identity of the treatments will be concealed by the use of study drugs that are all identical in packaging, labeling, schedule of administration, and appearance.

Plasma concentrations of R-amisulpride, S-amisulpride, and total amisulpride will not be disclosed before unblinding. In the event that concentration data transfer to the sponsor becomes necessary prior to the database lock, then the concentration data can only be shared with the sponsor's bioanalytical project manager with dummy subject IDs.

In the case of a medical emergency where knowledge of study drug by the Investigator or an authorized delegate is essential for immediate medical management, a 24-hour code-break service will be available via the IXRS. The date and reason for unblinding are to be documented. Any subject for whom the treatment assignment was unblinded is to be discontinued from further study participation. The subject should return for a final study assessment. The identity and responsibility of those individuals at the study site who gain access to the unblinded treatment assignment must be documented. It is mandatory that all personnel who are involved in the unblinding, and who have access to the unblinded treatment assignment, maintain the confidentiality of the information and do not divulge the treatment assignment.

3.4. DETERMINATION OF SAMPLE SIZE

A total sample size of 279 evaluable subjects in US and Europe (93 per treatment group: SEP-4199 200 mg/day, SEP-4199 400 mg/day, and placebo) with a global 2-sided alpha of 0.05 will provide about 90% power to reject at least 1 truly significant primary comparison and about 75% power to reject both truly significant primary comparisons using the truncated Hochberg ($\gamma = 0.9$) procedure, assuming treatment effect sizes of 0.44 for both doses of SEP-4199. The effect sizes were selected considering the data from the completed Lurasidone bipolar depression monotherapy study (Loebel 2014). This effect size corresponds to a treatment difference of -3.85 for a SEP-4199 dose with a common standard deviation of 8.745 on the MADRS scale change from baseline at Week 6. An upward adjustment of approximately 5% is used to compensate for information lost due to subjects who are randomized and do not provide any postbaseline primary efficacy data. The total sample size will be approximately 294 randomized subjects (or N = 98 subjects per treatment group).

For the secondary efficacy endpoint, assuming the 93 subjects per treatment arm, the study will have 71% power to reject at least one truly significant comparison within the CGI-BP-S family using the regular Hochberg procedure, given a treatment difference of -0.47 for a SEP-4199 dose with a common standard deviation of 1.24 on the CGI-BP-S change from baseline at Week 6. The correlation between the primary and the secondary endpoint is assumed to be 0.8 for each treatment arm.

The sample size was calculated using the Mediana package version 1.05 within R version 3.4.2 (R Core Team 2017)

See APPENDIX 4 for additional power and Type I error rate calculations.

The sample size for the Japanese cohort (45 total [N = 15 subjects per group], including 9 dropouts) provides more than 80% probability to have a treatment difference point estimate greater than zero for the MADRS.

3.5. CHANGES IN THE CONDUCT OF THE STUDY

There are no anticipated changes from the protocol in the conduct of the study.

3.6. SCHEDULE OF EVENTS

Schedule of events can be found in Section 1, Table 2 Schedule of Assessments of the CSP.

3.7. CHANGES TO ANALYSIS FROM PROTOCOL

The following changes to analysis from protocol are conducted:

- During the BDR Meeting, it was determined that there were 3 subjects who were randomized but never received a dose and 4 more subjects who were randomized, received a dose, but did not have any postbaseline MADRS/CGI data. To be consistent with the previous Sunovion Central Nervous System (CNS) studies, the definition of the intention-to-treat (ITT) population was subsequently changed from “subjects who are randomized” to

"subjects who are randomized, received at least one dose of study medication, and have a baseline and at least one post-baseline efficacy measurement in MADRS Total Score or CGI-BP-S Depression Score".

- During the dry run 2 draft output review, some baseline characteristic tables (exposure, compliance, psychiatric and medical history) were decided to be only created for the Safety Population as the difference of subjects between the ITT and Safety populations does not affect the interpretation of the results.
- MedDRA dictionary version 19.1 was used for AE, MH and psychiatric disorders coding instead of version 20.1 (or higher) as indicated in protocol.
- WHO-DD dictionary version 01MAR2017 was used for medications coding instead of version 01SEP2017 (or more recent) as indicated in protocol.
- It was stated in the protocol that the methodology and results of the population pharmacokinetic/pharmacodynamic analyses and the relationship between SEP-4199 PK and plasma prolactin were to be described in a separate document from this SAP and reported as an appendix to the CSR. The analyses will be reported separately from the CSR.
- ECG markedly abnormal criteria Categories were updated in SAP to be more inclusive than in protocol by changing '>' to '≥'.

4. PLANNED ANALYSES

The following analysis will be performed for this study: Final analysis including EU/US sites analysis, Japan sites analysis, and pooled analysis of EU/US/Japan sites.

No interim analysis is planned.

4.1. DATA AND SAFETY MONITORING BOARD (DSMB)

An independent DSMB will monitor the study safety data approximately when 33%, 66%, and 100% of the data accrue. DSMB SAP and charter, describing the methodology and presentation of results and access to results are prepared as separate documents.

4.2. INTERIM ANALYSIS

Other than the DSMB safety reviews, there is no interim analysis of the US and EU subjects planned for this study.

Japan cohort enrollment ended earlier than anticipated, and there will only be one database lock for all 3 cohorts. All Type I error will be used for the primary US and EU subjects only analysis.

4.3. FINAL ANALYSIS

After the data base lock, all final planned analyses identified in this SAP will be performed by IQVIA Biostatistics following Sunovion authorization of this SAP, Sunovion authorization of analysis populations, and unblinding of treatment.

All population PK/PD analysis will be described in a separate document.

Some minor modifications may be necessary to the planned design of tables, figures, and listings to accommodate data collected during the actual study conduct.

4.4. COHORTS ANALYSIS

In general, all analyses of the Japanese cohort are descriptive in nature. The Japanese cohort will be reported both separately and pooled with the EU and US cohort.

5. ANALYSIS POPULATIONS

Agreement and authorization of subjects included/ excluded from each analysis population will be conducted prior to the unblinding of the study.

5.1. INTENTION-TO-TREAT [ITT] POPULATION

The intention-to-treat (ITT) population will consist of subjects who are randomized, received at least one dose of study medication, and have a baseline and at least one post-baseline efficacy measurement in MADRS Total Score or CGI-BP-S Depression Score. Subjects are included in the ITT population regardless of any protocol deviations. Subjects will be analyzed according to their randomized treatment group.

The ITT population is the primary population for all efficacy analyses.

5.2. PER PROTOCOL [PP] POPULATION

The per protocol (PP) population will consist of all ITT subjects who have no important protocol deviations (IPD), as determined during the IPD review meetings, that may affect the interpretation of the primary efficacy endpoint as defined in Section 2.6.1. See APPENDIX 7 for specifications of IPD. New IPD categories may be created through the review of investigator comments and/or protocol deviations log. The predefined and any new IPDs are defined during the Blinded Data Review (BDR) Process. See APPENDIX 6 for specifications of the BDR Process. Subjects in the PP population will be analyzed according to their randomized treatment group.

Analysis of the primary and the secondary efficacy endpoints will also be performed using the PP population and analyzed according to the randomized treatment group.

5.3. SAFETY [SAF] POPULATION

The safety (SAF) population will consist of all subjects who were randomized and received at least one dose of study drug. Subjects will be analyzed according to the predominant treatment received. The predominant treatment is defined as the treatment to which the subject is exposed for the greatest duration during the treatment period. In case of same duration for several treatments, the treatment corresponding to randomization will be considered as predominant.

Duration of exposure to a treatment is calculated as the number of days from first dose date to the last dose date of that treatment. The predominant treatment will generally be the same as the randomized treatment group, unless the subject takes incorrect study drug during their entire participation in the study.

If there is any doubt whether a subject was treated or not, they will be assumed to have been treated for the purposes of analysis.

5.4. PHARMACOKINETICS (PK) POPULATION

The PK population includes the SAF population subjects who have at least 1 valid postdose PK concentration value (including non-missing concentrations < LLOQ) without protocol deviations or events with the potential to affect PK.

The PK analysis will be based on the PK population.

6. GENERAL CONSIDERATIONS

6.1. REFERENCE START DATE AND STUDY DAY

Study Day will be calculated from the reference start date, and will be used to show start/ stop day of assessments and events.

Reference start date is defined as the day of the first dose of study drug, (Day 1 is the day of the first dose of study drug), and will appear in every listing where an assessment date or event date appears.

- If the date of the event is on or after the reference date then:

Study Day = (date of event – reference date) + 1.

- If the date of the event is prior to the reference date then:

Study Day = (date of event – reference date).

In the situation where the event date is partial or missing, Study Day, and any corresponding durations will appear partial or missing in the listings. Partial dates will be presented as is in the listings.

6.2. BASELINE

Unless otherwise specified, baseline is defined as the last non-missing measurement taken prior to the reference start date (including unscheduled assessments). In case of subjects randomized but not treated, baseline is defined as the last non-missing measurement taken prior to randomization date. In the case where time isn't available and the date of the last non-missing measurement and the reference start date coincide, that measurement will be considered pre-baseline. Adverse Events and medications commencing on the reference start date will be considered postbaseline.

6.3. DERIVED TIME POINTS

As per protocol, Visit 6 can be an End of Treatment (EOT) visit or an Early Termination (ET) visit. If a subject terminates early, his/her ET visit will be mapped to the next planned assessment visit if it is within the window of the next planned assessment visit. This applies to both efficacy and safety data. If assessments from an ET visit cannot be mapped to the next planned assessment visit, these assessments will be excluded from any mixed model for repeated measures (MMRM) or analysis of covariance (ANCOVA) analysis. No follow-up visit data will be included in any MMRM/ANCOVA analyses. In the case of the ET being out of window of the next planned visit, it will be mapped to an unscheduled visit as appropriate.

All applicable by-visit efficacy and safety tables will include descriptive statistics for the Follow-up visit and the "Last observation visit". The "Last observation visit" is defined as the post-baseline visit at which the last non-missing observation is collected for a subject for a given parameter, including unscheduled visits or unmapped ET visits. The "Last observation visit" will be excluded from any MMRM/ANCOVA analyses.

For remapping of the ET visits, see Table 1 below.

Table 1: Mapping of the ET visit.

Parameter	Early Termination Day Criteria	Week	Analysis Visit
ECG, Vital Sign, YMRS, C-SSRS, Concomitant Medication, Last dose of Benzodiazepines/ Sedatives/Hypnotics	1 ≤ study day ≤ 9	1	Visit 3/Week 1
	10 ≤ study day ≤ 16	2	Visit 4/ Week 2
	17 ≤ study day ≤ 30	4	Visit 5/Week 4
	31 ≤ study day ≤ 44	6	Visit 6/Week 6
MADRS, CGI-BP-S, HAM-A, QIDS-SR16, AIMS, BARS, SAS	1 ≤ study day ≤ 9	1	Visit 3/Week 1
	10 ≤ study day ≤ 16	2	Visit 4/Week 2
	17 ≤ study day ≤ 30	4	Visit 5/Week 4
	31 ≤ study day ≤ 44	6	Visit 6/Week 6
SDS, EQ-5D-5L, Weight, BMI, Waist Circumference, PWC, HbA1C, Lipid panel, Serum Insulin, HOMA-IR, CRP	1 ≤ study day ≤ 44	6	Visit 6/Week 6

Hematology, Serum Chemistry*, Urinalysis, Serum Prolactin, PK, Plasma Prolactin	1 ≤ study day ≤ 16	2	Visit 4/Week 2
	17 ≤ study day ≤ 44	6	Visit 6/Week 6

* Except for parameters listed above such as HbA1C, Lipid panel, Serum Insulin, HOMA-IR, CRP.
Abbreviations: AIMS = Abnormal Involuntary Movement Scale; BARS = Barnes Akathisia Rating Scale; BMI = Body Mass Index; CGI-BP-S = Clinical Global Impression Bipolar Version, Severity of Illness; C-SSRS = Columbia-Suicide Severity Rating Scale; ECG = 12-Lead Electrocardiogram; ET = Early Termination; EQ-5D = EuroQoL-5D; HAM-A = Hamilton Rating Scale for Anxiety; MADRS = Montgomery-Asberg Depression Rating Scale; PWC = Physician's Withdrawal Checklist; QIDS-SR16 = Quick Inventory of Depressive Symptomatology – Self-Report; SAS = Modified Simpson-Angus Scale; SDS = Sheehan Disability Scale; YMRS = Young Mania Rating Scale.

All planned visits will be mapped to weeks and analysis visits for summaries and statistical analyses where applicable (Table 2).

Table 2: Mapping of planned visits.

Planned Visit	Analysis Week	Analysis Visit
Visit 1	-1	Visit 1/Screening
Visit 2	0	Visit 2/Baseline
Visit 3	1	Visit 3/Week 1
Visit 4	2	Visit 4/Week 2
Visit 5	4	Visit 5/Week 4
Telephone Contact (TC)	5	TC
Visit 6	6	Visit 6/Week 6
Visit 7	7	Visit 7/Follow-up/Week 7

Original visit collected on the case report forms (CRFs) will be displayed in the listings.

6.4. RETESTS, UNSCHEDULED VISITS AND EARLY TERMINATION DATA

In general, for by-visit summaries, data recorded at the scheduled visits will be presented.

Unscheduled measurements will not be included in by-visit summaries. Unscheduled measurements collected prior to the first dose of study drug will contribute to the derivation of the baseline value. Unscheduled measurements collected postbaseline will contribute to the derivation of markedly abnormal postbaseline value, and best/worst case value where required (eg, shift tables).

Early termination data collected postbaseline will be assigned to the next planned visit for that assessment as explained in Section 6.3. This mapping will be implemented to all early termination data points used in the efficacy and safety analyses.

Listings will include scheduled, unscheduled (including retests), and early termination data with original dates and visits displayed.

6.5. WINDOWING CONVENTIONS

Apart from early termination data (see Section 6.3), no visit windowing will be performed during the analysis for this study. Data will be summarized by analysis visit and analyzed by week where applicable (see Section 6.3).

6.6. STATISTICAL TESTS

All statistical inference will be performed with 2-sided tests at the significance level of 0.05 and 2-sided 95% confidence intervals (CIs), unless otherwise specified in the description of the analyses.

6.7. COMMON CALCULATIONS

For quantitative measurements, change from baseline will be calculated as:

- Test Value at Visit X – Baseline Value

6.8. SOFTWARE VERSION

Apart for the supplementary analysis of the primary endpoint exploring the dose response relationship (generalized MCP-Mod using the DoseFinding package, version 0.9-16) which will use R software version 3.5.3 or higher (R Foundation for Statistical Computing, Vienna, Austria), all other analyses will be conducted using Statistical Analysis System (SAS®) version 9.4 or higher (SAS Institute, Inc., Cary, North Carolina).

7. STATISTICAL CONSIDERATIONS

7.1. ADJUSTMENTS FOR COVARIATES AND FACTORS TO BE INCLUDED IN ANALYSES

The following covariates are used in the analyses. For details of their inclusion in the models, see specific analysis sections.

- Baseline value of the variable to be analyzed

7.2. MULTICENTER STUDIES

This study will be conducted by multiple investigators at multiple centers in the US, Europe, and Japan.

When specified, statistical analysis will be adjusted for region with centers within the region pooled.

Region will be categorized as follows:

- US
- Europe
- Japan

7.3. MISSING DATA

For the MMRM models, missing observations are treated as missing at random (MAR) and no imputation for missing data will be applied.

Unless otherwise specified, any individual missing item in any scale will not be imputed. When calculating a total score, subscale score, or any assessment with more than one item, if one or more items are missing at a visit, then the associated score (ie, total score or subscale score) will be set to missing. For additional details, see the individual scale description sections.

Handling of missing efficacy data, if any, is described in Sections 16.1.2, 16.2.2, 16.3.2.

Handling of missing safety data, if any, is described in Sections 17.1.1.1, 17.1.1.2, and 17.6.

Handling of missing PK data if any is described in Section 18.

See APPENDIX 2 for details of incomplete/missing dates.

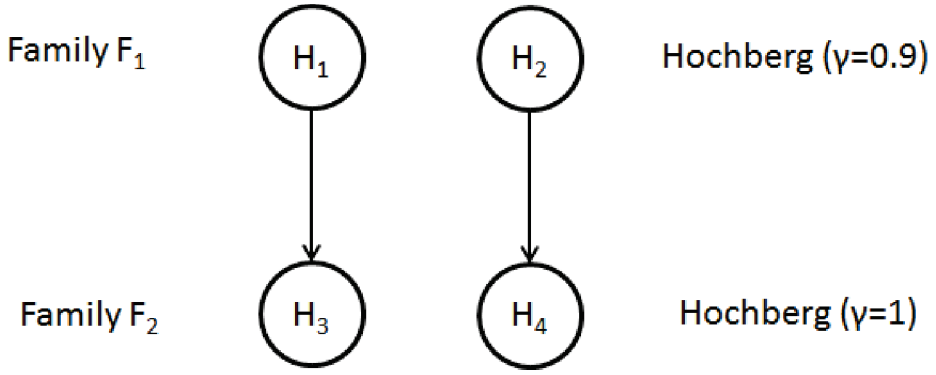
7.4. MULTIPLE COMPARISONS/ MULTIPLICITY

The truncated Hochberg gatekeeping (with the truncation parameter $\gamma = 0.9$) procedure constructed using the general mixture method will be used to control the global Family-Wise Error Rate (FWER) at 5% for the primary (E1) and secondary endpoint (E2) (Dmitrenko 2016). Using this method, a hypothesis is testable in a latter family only if the hypothesis in the former family associated with the same dose level is rejected.

The hypotheses associated with the primary and the secondary endpoints will be grouped into 2 hierarchical families as seen in Figure 2 below (See Sections 16.1.3 and 16.2.3 for the primary hypotheses [H1, H2] and secondary hypotheses [H3, H4]).

- Family F1: SEP-4199 200 mg/day versus placebo (H_1), and SEP-4199 400 mg/day versus placebo (H_2), based on change from baseline in MADRS total score at Week 6 (E1)
- Family F2: SEP-4199 200 mg/day versus placebo (H_3), and SEP-4199 400 mg/day versus placebo (H_4) based on change from baseline in CGI-BP-S depression score at Week 6 (E2)

Figure 2: Hierarchy of Testing for Primary and Secondary Endpoints



The mixture-based gatekeeping procedure will be performed in 2 steps:

- Step 1: The SEP-4199 vs placebo comparisons for E1 (hypotheses H1 and H2) will be performed using a truncated Hochberg (γ = 0.9) test.
- Step 2: The SEP-4199 vs placebo comparisons for E2 (hypotheses H3 and H4) will be performed using a regular Hochberg test only when the corresponding comparisons from Step 1 are rejected. For example, H4 is tested only if H2 is rejected.

The primary efficacy analysis will be repeated for the PP population to provide additional information on the robustness of the primary efficacy results. The Hochberg procedure will not be used for the PP population or for the analysis of Japan and pooled cohorts EU/US + Japan.

Adjusted p-values will be calculated using the SAS-based TreeGate macro and available in following website: http://multxpert.com/wiki/TreeGate_macro.

7.5. EXAMINATION OF SUBGROUPS

Subgroup analyses will be conducted as stated in the exploratory analysis sections. It should be noted that the study was not designed to detect treatment differences with high statistical power within subgroups.

The following subgroups will be assessed and described within the appropriate analysis sections:

- Region:
  - US
  - Europe
  - Japan
- Country
- Gender:

- Female
- Male

- Age (years):
  - < 55
  - ≥ 55

- Race in 2 categories:
  - Category 1 for efficacy modeling: White and Non-white (Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islanders, and Other including multiracials)
  - Category 2 for summary statistics for efficacy and safety: White, Non-white, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, and Other including multiracials

- Ethnicity
  - Hispanic or Latino
  - Not Hispanic or Latino

- Bipolar I diagnosis subtype
  - Rapid cycling (4-7 episodes of mood disturbance within the past 12 months)
  - Non-Rapid cycling

8. OUTPUT PRESENTATIONS

APPENDIX 1 shows conventions for presentation of data in outputs.

The templates provided with this SAP describe the presentations for this study and therefore the format and content of the summary tables, figures and listings to be provided by IQVIA.

All tables except efficacy and PK tables will present an overall All SEP-4199 column.

9. DISPOSITION AND WITHDRAWALS

All subjects who provide informed consent will be accounted for in this study.

Subject disposition will be presented for the screened subjects described by the randomized treatment group (where applicable) and overall for all subjects. The number and percentage of subjects who were screened, screen-failed, enrolled, randomized, randomized but not dosed, received study drug, completed or discontinued from the study treatment period, completed the treatment period and the follow-up visit, and completed the treatment period but not the follow-up visit will be presented. The number of subjects with non-missing change from baseline at Week 6 of MADRS total score and CGI-BP-S depression score will also be described. The number of subjects discontinued due to COVID-19 will be summarized and related reasons for discontinuations will be presented by relationship to COVID-

19.

With respect to the above, the following definitions apply:

- Screened Subjects: Any subject who signed the study specific informed consent and completed at least 1 study-related procedure.
- Enrolled Subjects: Any subject who was successfully screened and enrolled into the prerandomization period of the study.
- Randomized Subjects: Any subject who was randomized into the treatment period of the study and was assigned a randomization number.

COVID-19 related discontinuations are to be collected on the Disposition CRF page using the disposition reason = ”Other” or “Adverse Event”. For disposition reason = ”Other”, the sites are instructed to add the text “COVID-19” in the disposition reason detail. For disposition reason = ”Adverse Event” due to COVID-19, the adverse event preferred term will be identified through a predefined MedDRA preferred term search as detailed in Section 17.1.

The number of subjects randomized will also be described by site, country, and region by treatment group and overall. The number of subjects who complete (including remotely versus in-person, for Week 6 and the Follow-up visit only) and discontinue from the study drug before or at a given visit, along with reasons for study drug discontinuation, including due to COVID-19, will be described by the treatment group and overall using randomized subjects. Time to discontinuation from study drug will be plotted using a Kaplan-Meier (KM) curve.

The number of subjects in each population will also be described.

10. PROTOCOL DEVIATIONS

Important protocol deviations (IPDs) will be identified and documented based on blinded reviews of data listings. The IPD categories may include, but may not be limited to:

- Eligibility and Entry Criteria
- IP Compliance
- Concomitant Medication Criteria
- Laboratory Assessment Criteria

Further details on the identification of IPDs are provided in APPENDIX 7.

IPDs will be identified for all randomized subjects and presented in data listings. The number and percentage of subjects within each IPD category will be summarized by treatment group and overall for the ITT population.

In addition, the number and percentage of subjects with major and critical protocol deviations that were not considered to be IPDs will be also summarized by treatment group and overall for the ITT population.

In general, major and critical protocol deviations are defined as follows:

- Major: A deviation from protocol-related procedures that could affect the integrity of the data or adversely affect subjects.
- Critical: A deviation from protocol-related procedures that threatens integrity of the data, adversely affects subjects and/or could invalidate acceptability of a project (or part of it).

10.1. DEVIATIONS RELATED TO PK ANALYSIS

Changes to the procedures or events, which may impact the quality of the PK data, may be considered significant protocol deviations and will be described within the clinical study report body text. These changes or events will include any circumstances that will alter the evaluation of the PK. Examples include, but may not be limited to, vomiting following oral dosing occurring within the time frame of 2 times the median t_{max}, sample processing errors that lead to inaccurate bioanalytical results, and/or inaccurate dosing on the day of PK sampling. In the case of a significant protocol deviation or event, PK data collected for the affected dose group will be excluded from the study summary and inferential results but listed, if appropriate. Other changes to the procedures or events which do not impact the quality of the PK data will not be considered significant. A common example of a non-significant protocol deviation is a missed blood sample or deviations from blood collection times.

11. DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS

Demographic data and other baseline characteristics will be presented for the ITT population, SAF population, and PP population. For the ITT and PP populations, the data will be presented by randomized treatment groups. For the SAF population, the data will be presented by the predominant treatment received.

No statistical testing will be carried out for demographic or other baseline characteristics.

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

- Age (years) - calculated relative to date of informed consent, as a continuous variable and categorically:

  - < 55 years
  - >= 55 years

  and

  - < 65 years
  - >= 65 years

  and

  - <= 18 years
  - >18 and <65 years
  - >= 65 years

- Gender

- Race

  - American Indian or Alaska Native
  - Asian
  - Black or African American
  - Native Hawaiian or Other Pacific Islander
  - White
  - Multiracials as is (where more than one race is selected, ie, White/Asian, Black/Other)
  - Other
  - Other including multiracials
  - Non-White (Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islanders, and Other including multiracials)

- Ethnicity
  - Hispanic or Latino
  - Not Hispanic or Latino

- Country
  - Bulgaria
  - Poland
  - Russia
  - Serbia
  - Slovakia
  - Ukraine
  - USA
  - Japan

- Region
  - US
  - Europe
  - European Union (Bulgaria, Poland, Slovakia): for EudraCT
  - Japan

- Weight (kg)

- Height (cm)

- BMI (kg/m^2), as a continuous variable and categorically:
  - Underweight: < 18.5
  - Normal: 18.5 to < 25.0
  - Overweight: 25.0 to < 30.0
  - Obese: >= 30.0

- Waist circumference (cm)

- Baseline MADRS total score, as a continuous variable

- Baseline CGI-BP-S depression score, as a continuous variable

Demographic data collected for all screened subjects will be listed.

The following psychiatric history data will be summarized for the SAF population in a separate table:

- Age at onset of bipolar I disorder
- Time since initial onset of bipolar I disorder, in years, calculated relative to date of informed consent (see APPENDIX 2 for partial date imputation rules)
- Time since onset of current episode of major depression associated with bipolar I disorder symptoms, in months, calculated relative to date of informed consent (see APPENDIX 2 for partial date imputation rules)
- Frequency distribution of number of prior hospitalizations for Bipolar I depression (0, 1, 2, 3, and 4 or more)
- Frequency distribution of DSM-5 code in ascending numeric order
- Bipolar I diagnosis subtype (Bipolar I disorder without rapid Cycling and Bipolar I disorder with rapid cycling)
- Number and percentage of subjects with other psychiatric disorders
- Number and percentage of subjects with any other psychiatric disorders

Other psychiatric Disorder data will be summarized for SAF population in a separate table by summarizing Diagnosis and DSM-5 codes.

11.1. DERIVATIONS

- Time since initial onset of bipolar disorder (years):

(Date of informed consent – date of initial onset of bipolar disorder) / 365.25

- Time since initial onset of current episode of major depression (months):

(Date of informed consent – date of initial onset of current episode of major depression) / 30.25

12. MEDICAL HISTORY

Medical History information will be presented for the ITT and the SAF populations by treatment group. Medical and surgical history information, including both past and concomitant medical conditions and major surgical history, as collected on the Medical History CRF form, will be coded using MedDRA,

Version 19.1 or higher, and presented by SOC and PT. Medical History will be sorted alphabetically by SOC and by decreasing frequency of PT in the All SEP-4199 column.

13. MEDICATIONS

Medications will be presented for the SAF population and coded to indication-specific Anatomical Therapeutic Chemical (ATC) Level 3 and PT according to World Health Organization Drug (WHODRUG) drug dictionary, Version 01MAR2017 or higher. If ATC level 3 is missing then ATC level 2 will be used. Medication will be sorted alphabetically by ATC level3 and by decreasing frequency of PT in the All SEP-4199 column.

Whenever available, the time information should be accounted for in the derivation of prior, concomitant, and post-treatment medications, see APPENDIX 2 for the handling of partial dates for medications. In the case where it is not possible to define a medication as prior, concomitant, or post-treatment, the medication will be classified by the worst case; ie, concomitant.

- Prior medications are medications which stopped prior to the first dose of study drug.
- Concomitant medications are medications that:
  - Started prior to, at the time of, or after the first dose of study drug, AND
  - Started before or at the date (day) of the last dose of study drug, AND
  - Ended at the time of or after the first dose of study drug or were ongoing at the end of the study.
- Post-treatment medications are medications which started after the day of the last dose of study drug.

Prior and concomitant medication use will be summarized by treatment group by ATC Level 3 classification and PT using the SAF population. Subjects with multiple uses of a medication will be counted only once for a given drug class or PT. Number and percentage of subjects who take concomitant benzodiazepines, sedatives, or hypnotics will be summarized by visit (including follow-up visit) and overall. Prior, concomitant, and post-treatment medications will be provided in data listings.

14. STUDY DRUG EXPOSURE

Exposure to study drug will be calculated as the number of days from first dose date to last dose date. Duration of exposure to study drug will be summarized for the SAF population.

The date of first study drug administration will be taken from the CRF “Study Drug Administration / Drug Accountability” form.

The date of last study drug will be retrieved as the latest date between “Study Drug Administration / Drug Accountability”, “Blood PK Samples and 3 Prior Doses”, and “Subject Disposition - End of Study” forms. For subjects who are lost to follow-up, and for whom the last dose date is unknown, the last contact date will be used.

Interruptions and compliance are not taken into account for duration of exposure.

Overall number of tablets taken will be described by treatment group and overall as a continuous variable.

Duration (in days) of exposure will be summarized both as a continuous variable and categorically, and by planned visit days:

- Number and percentage of subjects with drug exposure ≥ 4, ≥ 7, ≥ 14, ≥ 21, ≥ 28, ≥ 35, and ≥ 42 days;
- Number and percentage of subjects with drug exposure for 1 - 3, 4 - 6, 7 - 13, 14 - 20, 21 - 27, 28 - 34, 35 - 41 and ≥ 42 days.

Person years of exposure will be described by treatment group and overall.

14.1. DERIVATIONS

- Duration of exposure (days)

  Last dose of study drug – First dose of study drug + 1.

- Person years of exposure

  Sum of total exposure (days) for all subjects within each treatment group / 365.25

- Overall number of tablets taken

  Sum of number of dispensed tablets – Sum of number of returned tablets, if any returned tablets is missing then overall number of tablets taken is set to missing

15. STUDY DRUG COMPLIANCE

Compliance to study drug will be presented for the SAF population.

Percent compliance will be calculated by visit and overall on the number of subjects with non-missing compliance data. For each postbaseline visit, if the number of tablets returned is missing, the corresponding compliance will be set as missing. Overall compliance will be set to missing if one or more blister cards are not returned to the site for accountability.

Non-compliance is defined as less than 75% or more than 125% non-missing compliance. Subjects with missing compliance will not be classified as non-compliant but described in the categorical analysis.

Compliance will be summarized both as a continuous variable (ie, mean percentage) and categorically (ie, number and percentage of subjects who are compliant vs. non-compliant, or with compliance < 75%, 75% - 125%, > 125%, and missing).

Subject level treatment compliance will be listed.

15.1. DERIVATIONS

Compliance with study drug will be calculated for each period defined by study visits (ie, per-visit compliance) and overall.

Per-visit compliance for the period defined by visit (V-1) and visit V will be calculated as:

$$\frac{\#\text{ Tablets dispensed at Visit (V-1)} - \#\text{ Tablets returned at Visit V}}{\#\text{ Tablets should be taken per day} \times (\text{Date of Visit V} - \text{Date of Visit (V-1)})} \times 100\%$$

Two tablets per day are supposed to be taken. If number tablets dispensed at Visit (V-1) and/or number tablets returned at Visit V are missing, per-visit compliance will not be calculated for the period impacted.

If a subject discontinued from the study in between Visit (V-1) and Visit V, then the above formula will be modified to:

$$\frac{\#\text{ Tablets dispensed at Visit (V-1)} - \#\text{ Tablets returned at Visit ET}}{\#\text{ Tablets should be taken per day} \times (\text{Date of Visit ET} - \text{Date of Visit (V-1)})} \times 100\%$$

The per-visit compliance calculated above will contribute to the summary of per-visit compliance for the period defined by visit (V-1) and visit V.

Overall compliance will be calculated as:

$$\frac{\text{Total \# tablets dispensed} - \text{Total \# tablets returned}}{\#\text{ Tablets should be taken per day} \times \text{Duration of Exposure}} \times 100\%$$

Duration of exposure is calculated as specified in Section 14.1.

If number tablets dispensed and/or number tablets returned are missing at 1 or more visits, overall compliance will not be calculated.

16. EFFICACY OUTCOMES

16.1. PRIMARY EFFICACY

Impact of the COVID-19 pandemic on efficacy will be investigated using additional sensitivity analyses (see Section 16.1.4.4, Section 16.1.4.5, Section 16.2.4, Section 16.3.3.5 and Section 16.3.3.6). Additional adhoc analysis may be conducted after DB lock to further investigate the COVID-19 impact.

16.1.1. PRIMARY EFFICACY VARIABLE & DERIVATION

The primary efficacy endpoint is the change from baseline in the MADRS total score at Week 6.

The MADRS is a clinician-rated assessment of the subject's level of depression. The measure contains 10 items that measure apparent and reported sadness, inner tension, reduced sleep and appetite, difficulty concentrating, lassitude, inability to feel, and pessimistic and suicidal thoughts. Each item is scored in a range of 0 to 6 points, with higher scores indicating increased depressive symptoms. Total

score will be equal to the sum of the 10 items (range between 0 and 60). MADRS is assessed at Visit 1/Screening, Visit 2/Baseline/Week 0, Visit 3/Week1, Visit 4/Week 2, Visit 5/Week 4, and Visit 6/Week 6.

16.1.2. Missing Data Methods for Primary Efficacy Variable

The MADRS total score at a visit will be set to missing at that visit if any one item is missing. The primary efficacy variable, change from baseline in MADRS score at Week 6, will be set to missing if MADRS score at Week 6 is missing or baseline value is missing. The same applies to all other visits.

The primary analysis of the primary efficacy variable will use an MMRM which makes an MAR assumption to the missing primary efficacy endpoints. Early termination data will be mapped as described in Section 6.3. Missing data will not be imputed.

16.1.3. Primary Analysis of Primary Efficacy Variable(s)

The primary efficacy estimand is defined as the efficacy of each SEP-4199 dose over placebo for 6 weeks, calculated as the difference in the primary efficacy endpoint between each SEP-4199 treatment group and the placebo treatment group in the ITT population. This estimand, which will only include those subjects who participate at sites located in the US and Europe, provides an estimate of the efficacy of each SEP-4199 dose over placebo, should a subject be able to tolerate and adhere to treatment for 6 weeks.

The 4 attributes of the primary efficacy estimand are as follows:

A. **Population of interest:** ITT subjects with major depressive episode associated with bipolar I disease (bipolar I disorder) at sites located in the US and Europe defined through the inclusion/exclusion criteria.

B. **The variable of interest (endpoint):** change from baseline in MADRS total score at Week 6

C. **Intercurrent events (and how to handle them)**

The estimand is based on the effect of treatment in the hypothetical absence of any intercurrent events, ie, if all subjects had adhered to treatment up to Week 6. This includes events where a subject continues the study regardless of the intercurrent event(s) (eg, subject received a prohibited medication during the treatment period and was not discontinued from the study) and/or events that cause the subject to discontinue from the study.

Any intercurrent event where subject continues the study regardless of the intercurrent event will be ignored for the primary analysis and the primary efficacy data will be analyzed as is.

The statistical analysis for intercurrent events that lead to treatment discontinuation rests on an assumption about measurements that would have been observed under the hypothetical setting where a subject does not discontinue from treatment. For example, a subject who discontinues treatment before Week 6 in the SEP-4199 treatment group is assumed to behave similar to a subject who completes the study up to Week 6 in the same SEP-4199 treatment group using the MMRM analysis based on the MAR assumption. The assumption for subjects in the placebo group is made similarly. However, as such assumptions cannot be verified, sensitivity analyses will be conducted to assess the

robustness of the conclusions, as indicated in Section 16.2.4.

D. **Population level summary of the variable:** Difference in means of change from baseline in MADRS total score at Week 6 comparing each SEP-4199 treatment group to placebo

The efficacy of SEP-4199 in terms of the MADRS total score will be evaluated using the following two null hypotheses:

H1: There is no difference in mean change from baseline at Week 6 on the MADRS total score in the SEP-4199 200 mg treatment arm compared to Placebo.

H2: There is no difference in mean change from baseline at Week 6 on the MADRS total score in the SEP-4199 400 mg treatment arm compared to Placebo.

The alternative hypothesis for each of the null hypotheses is that there is a difference. If at least 1 of the primary treatment group comparisons demonstrate a significant result, the study will be considered positive.

The primary efficacy endpoint will be analyzed for the ITT population using an MMRM analysis, with change from baseline in the MADRS total score as the response variable and with factors for treatment group, week, region, baseline MADRS total score, and the treatment-by-week interaction. This analysis method makes an MAR assumption to the missing primary efficacy endpoints. The Kenward Roger approximation will be used to estimate denominator degrees of freedom. An unstructured covariance matrix will be used for the within-subject correlation. A robust sandwich estimator for the standard error of the fixed effects (Diggle 2002) along with a spatial exponential or a spatial power covariance structures will be assumed sequentially in case the model fails to convergence. The first covariance structure to yield convergence will be used in the analysis.

The Least Square (LS) mean treatment differences (each SEP-4199 group minus placebo) of change from baseline at Week 6, their 2-sided 95% CIs, and the associated p-values will be calculated based on the MMRM. The LS mean of change from baseline over time will be plotted. Adjusted p-values will be calculated using a truncated Hochberg procedure with gamma parameter=0.9 (See Section 7.4).

At each postbaseline visit, within-group and between-group effect-size will be presented. Within-group effect size will be calculated as the absolute value of LS mean of change from baseline divided by SD, obtained as the SE of the LS Mean multiplied by the square root of the treatment group change from baseline sample size at each visit. Based on the MMRM, between-group effect size at a visit will be calculated as the absolute value of the LS means difference from placebo divided by the model estimate of the pooled SD at the visit, which is obtained from the square root of the diagonal element, associated at the visit, from the covariance matrix (R matrix of subjects with MADRS total score at all visits).

Multiplicity adjustment to control the global family wise Type I error rate at 5% is described in Section 7.4.

The normality and homoscedasticity assumptions underlying the primary MMRM model will be assessed graphically and included in the SAS outputs. Shapiro-Wilk statistic will be included in the SAS output for exploratory purposes. Marginal studentized and Pearson residuals will be plotted against the predicted values, respectively, and quantile-quantile (Q-Q) plots of these residuals versus the expected quantiles of the standard normal distribution will be included in the SAS outputs to provide a graphical view of similarity and difference in the 2 distributions. Additional analyses may be explored to better understand the potential impact of any deviations on the interpretation of the primary efficacy analysis.

16.1.4. SENSITIVITY ANALYSIS OF PRIMARY EFFICACY VARIABLE

To address early dropouts under the assumption of missing not at random (MNAR), a pattern-mixture model (PMM) using a placebo-based multiple imputation method (O'Kelly 2014; Ratitch 2011) and a PMM using multiple imputations with penalties (ie, tipping point analysis by deflating the individually estimated treatment effect size by known factors) (Levin 2016; Permutt 2015; Ratitch 2013) will be performed as sensitivity analyses to explore the robustness of the MMRM results for the primary analysis based on the ITT population.

To address the impact of COVID-19, sensitivity analyses will be conducted to investigate the effect of visits conducted remotely, and the effect of COVID-19 overall as a general psychosocial stressor using the 11 March 2020 global pandemic declaration by the World Health Organization as a cutoff date.

Sites are instructed to enter the text "COVID-19 remote" on the comments CRF page for any clinic visits conducted remotely. If a visit is conducted remotely, any scale assessments collected during that visit will be assumed to be collected "remotely". In addition, if a video is used during the remote visit, the text "video" will be added to the above mentioned comment text. Remote scale assessments to be collected are MADRS, CGI-BP-S, YMRS, C-SSRS, PWC, HAM-A, SDS, AIMS, BARS, SAS, EQ-5D-5L, and QIDS-SR16. QIDS-16 is a self-reported paper base assessment which will be mailed to the subject. Each remote assessment will be flagged in its individual listing. Video assessments information will be available in the comments listing.

16.1.4.1. Dropout Profiles

Patterns of missing data in the primary endpoint will be assessed through graphical tools based on the reason and timing of study withdrawal by subjects.

Change from baseline in MADRS total score by week will be described and plotted by regrouped reasons of early discontinuation.Table 3 shows the regrouping of the reasons. Similar reasons may be combined according to the following table, depending on the number of subjects under each reason:

Table 3. Regrouping of reasons of early discontinuation.

CRF Term	Grouped Term
Adverse event	Adverse event
Death (due to AE)	Adverse event
Lack of efficacy	Lack of efficacy
Lost to follow-up	Lost to follow-up or Withdrew consent
Withdrawal by subject	Lost to follow-up or Withdrew consent
Pregnancy	Noncompliance and other
Noncompliance with study drug	Noncompliance and other
Protocol deviation	Noncompliance and other

Other reasons	Noncompliance and other
Termination by sponsor	Noncompliance and other

Change from baseline in MADRS total score by week will also be described and plotted based on the timing of termination for Visit 3/Week 1 terminators, Visit 4/Week 2 terminators, Visit 5/Week 4 terminators, Visit 6/Week 6 terminators, and completers, separately for each treatment group.

- Visit 3/Week 1 terminators: subjects who discontinued after Visit 2/Baseline/Week 0 but before or on Visit 3/Week 1;
- Visit 4/Week 2 terminators: subjects who discontinued after Visit 3/Week 1 but before or on Visit 4/Week 2;
- Visit 5/Week 4 terminators: subjects who discontinued after Visit 4/Week 2 but before or on Visit 5/Week 4;
- Visit 6/Week 6 terminators: subjects who discontinued after Visit 5/Week 4 but before or on Visit 6/Week 6;
- Completers: subjects who completed the study treatment.

16.1.4.2. Pattern-mixture model with placebo-based multiple imputation

All ITT subjects are included in pattern-mixture model with placebo-based multiple imputation.

The MMRM model used in the primary analysis makes the assumption that data are MAR. However, the missing data mechanism may or may not be at random. Sensitivity to the missing data assumption will be tested by using the PMM with placebo-based multiple imputation method, exploring the robustness of the MMRM results of the primary efficacy analysis. In this analysis, missing values in the SEP-4199 (200 or 400 mg/day) treatment groups after remapping of the ET visits will be imputed based on data of the placebo group, assuming that after withdrawal, subjects from the SEP-4199 (200 or 400 mg/day) group will exhibit the same future evolution of depressive episode as subjects from the placebo group, and that subjects who discontinue from the placebo group will exhibit the same future evolution of depressive episode as subjects in the placebo group remaining in the study. This approach does not assume a sustained benefit of experimental treatment after discontinuation.

Two separate imputation procedures are used to impute missing values. Firstly, the Markov chain Monte Carlo (MCMC) method is used to perform partial imputation to obtain datasets with monotone missing patterns. Then a sequential regression multiple imputation method is used to impute the monotone missing values.

Under the assumption that the MADRS total scores have a multivariate normal distribution, the MCMC method is used to impute only intermittent missing values (using the SAS MI procedure with MCMC statement), by using a data augmentation algorithm, with each iteration n consisting of an imputation step and a posterior step. The imputation step uses a random draw of $\theta^{(n)}$, parameter of the joint imputation model, to sample missing values from a conditional distribution $P(Y_{mis}|x, y_{obs}, \theta^{(n)})$, obtaining $y_{mis}^{(n)}$, the subset of missing values that need to be filled in to achieve monotone missingness. The posterior step simulates a new draw of the parameter $\theta^{(n+1)}$ from the posterior distribution given the current monotone missing data $P(\theta|x, y_{obs}, y_{mis}^{(n)})$ with a non-informative Jeffreys prior. Treatment group will be taken into account for this imputation (ie, missing data at intermediate visits will be imputed for each treatment group using no-missing data from all subjects within the treatment group). These steps are repeated to obtain 1000 datasets with monotone missingness. The random seed number is 12345.

The remaining monotone missing data will be imputed using sequential regression multiple imputation, where a separate regression model is estimated for imputation of each variable (ie, MADRS total score at each time point). Imputation of values in the placebo group will assume MAR. Imputation of values in the SEP-4199 (200 or 400 mg/day) group will be done as if the subject had been a member of the placebo group. Missing values in the SEP-4199 (200 or 400 mg/day) groups will be imputed using the imputation model of the placebo group, ie, conditional on subject values observed at time points prior to discontinuation. Each sequential regression model (ie, for imputation of values at a given time point) will include explanatory variables for all previous (Baseline, Week 1, Week 2, and Week 4) values of MADRS total score. Missing values at a given time point in placebo and SEP-4199 (200 or 400 mg/day) treatment groups will be imputed from the same imputation model, conditional on subject values observed or imputed at previous time points. The SAS MI procedure with the MONOTONE REG statement is used to specify that the regression method will be used for the imputation, and the MNAR statement with MODEL option will be used for the MADRS total score at each postbaseline visit to specify that only observations from the placebo group should be used to estimate the imputation model. The random seed number is 56789.

No rounding restriction will be applied to imputed MADRS total scores. The imputed MADRS total scores must be within the range of 0 to 60.

Each of the 1000 imputed datasets will be analyzed using the same MMRM model as the primary efficacy analysis. Results from the analysis of each imputed dataset, ie, the LS means of each treatment group, the LS mean treatment difference, and their standard errors, will be combined using Rubin's imputation rules (using the SAS MIANALYZE procedure) to produce pooled LS mean estimates, their standard errors and 95% CI, and a pooled p-value for the test of null hypothesis of no treatment effect.

16.1.4.3. Tipping point analysis

All ITT subjects are included in tipping point analysis.

Sensitivity to departures from the MAR assumption will also be investigated using a tipping point analysis. In this analysis, departures from MAR in the SEP-4199 (200 mg and 400 mg) group will be assessed assuming that subjects who discontinue the study have, on average, efficacy outcomes after discontinuation that are worse by some amount δ (ie, a percentage of the LS mean treatment difference) compared to other similar subjects with observed data at the same time point (ie, compared to a value which would have been assumed under a MAR model).

A series of analyses will be performed with increasing values of δ until the analysis conclusion of a statistically significant treatment effect no longer holds. The value of δ that overturns the primary results will represent a tipping point. An interpretation of clinical plausibility of the assumption underlying the tipping point will be provided. After one treatment group comparison reaches the tipping point, the analyses will continue for the other treatment group comparison until it reaches the tipping point.

Change from baseline in MADRS total score will be analyzed based on data observed while the subject remains on study as well as data imputed using multiple imputation methodology for the time points at which no value is observed. Intermittent (non-monotone) missing data will be imputed first based on the MAR assumption and a multivariate joint Gaussian imputation model using the MCMC method within each treatment group, as described above for the pattern-mixture model with placebo-based multiple imputation.

The remaining monotone missing data will be imputed using sequential regression multiple imputation, where a separate regression model is estimated for imputation of each variable (ie, MADRS total score at each time point). Each regression model will include explanatory variables for treatment and all previous (Baseline, Week 1, Week 2, and Week 4) values of MADRS total score.

After the MAR-based imputations have been generated for MADRS total score at each time point, a value of δ will be added to all imputed change from baseline values in the SEP-4199 (200 mg or 400 mg) group. This approach assumes that the marginal mean of unobserved subject measurements is worse by δ at each time point after discontinuation compared to the marginal mean of subjects with observed data at the same time point.

No rounding restriction will be applied to imputed continuous values. The imputed MADRS total scores must be within the range of 0 to 60.

A total of 1000 imputed datasets will be generated. The random seed number is 12345 for the partial imputation step and 56789 for the sequential regression multiple imputation step.

Each of the 1000 imputed and δ-adjusted datasets will be analyzed using the same MMRM model as the primary efficacy analysis. Results from the analysis of each imputed dataset, ie, the LS mean treatment difference and its standard error, will be combined using Rubin's imputation rules (using the SAS MIANALYZE procedure) to produce a pooled LS mean estimate of treatment difference, its standard error and 95% CI, and a pooled p-value for the test of null hypothesis of no treatment effect.

Analyses will be conducted with different values of δ at each visit, which represents a percentage of the LS mean treatment difference at that visit, starting at 5% with 5% increments, until either the tipping point is identified or the 100% penalty is applied. The actual δ for each percent increment will be presented in the table.

16.1.4.4. Impact of remote visits

To assess the impact of remote visits, primary MMRM analysis (see Section 16.1.3) will be repeated for MADRS Total Score after removal of all assessments conducted during remote visits.

16.1.4.5. Impact of COVID-19 Overall

To assess the potential impact of COVID-19 overall, primary MMRM analysis (see Section 16.1.3) will be repeated for MADRS Total Score by only keeping the records reported before 11 March 2020, corresponding to the WHO global pandemic declaration date.

16.1.5. Supportive Analysis of Primary Efficacy Variable(s)

16.1.5.1. Analysis of covariance

The primary efficacy endpoint will be analyzed using an ANCOVA on the ITT population as supportive analysis for each postdose visit. The model will include terms for treatment, region, and baseline MADRS total score as covariate. The LS mean of treatment differences (each SEP-4199 group minus placebo), their 2-sided 95% CIs, and the associated p-values will be obtained from the model.

Based on the ANCOVA, within group effect size is calculated as the absolute value of the LS Mean at

each visit divided by the standard deviation, obtained as the SE of the LS Mean multiplied by the square root of the treatment group change from baseline sample size at that visit. Between-group effect size at a visit will be calculated as the absolute value of the LS means difference from Placebo divided by the model estimate of the pooled standard error (the standard error of the LS mean difference divided by the square root of the sum of inverse treatment group change from baseline sample sizes).

16.1.5.2. Analysis on the per protocol population

The primary MMRM analysis will be repeated for the PP population to examine the impact of protocol deviations.

16.1.5.3. Analysis of dose-response

A supplementary analysis of the primary efficacy of SEP-4199 will be conducted on the ITT population to explore the dose response relationship using the generalized Multiple Comparison Procedure – Modeling (gMCP-Mod) methodology (Pinheiro 2014).

The gMCP-Mod approach to dose response will be implemented using R software, in particular the DoseFinding package (APPENDIX 3).

First, to obtain estimates of the dose response, an MMRM model on the ITT population with factors for treatment dose level, region, week, and baseline MADRS score as a continuous covariate will be used to model the change from baseline in MADRS score over time from the placebo and SEP-4199 groups. The LS mean of change from baseline in the MADRS total score at Week 6 for each SEP-4199 group and placebo and the corresponding variance-covariance matrix will be obtained from the model.

Next, the gMCP-Mod approach will use the estimates obtained from the MMRM model to analyze the dose response relationship at Week 6. The following set of candidate models for the dose response relationship will be considered (APPENDIX 3), but in case of inability to correctly fit data, a model could be removed from the analyses:

- Linear model (no parameters)
- Quadratic model (δ)
- E_{max} model (ED_{50})
- Sigmoidal E_{max} (ED_{50}, h)

Once a set of candidates are identified, the selection will be made using the MCPMod function from DoseFinding R package. The null hypothesis of a constant dose-response curve for the primary endpoint will be tested at the significance level of 5% against the one-sided alternative hypothesis of a non-constant dose response curve using the gMCP-Mod methodology.

For each of the dose-response models in the candidate set, the null hypothesis that the sum of the components of the optimal contrast vector is equal to zero will be tested. To do so, the contrast coefficients of each model are chosen so they maximize the power to detect the absence of dose response.

Under the null hypothesis, the vector of the optimal contrasts for the candidate models follows

asymptotically normal distribution. If the maximum asymptotic z statistic for the model contrast tests exceeds the critical value (corresponding to a unilateral test with alpha = 0.05), then the null hypothesis will be rejected and the proof of concept of a non null dose-response will be demonstrated.

The final detection of a significant dose-response signal is based on the adjusted p-value of the contrast test statistic. If no model contrast is significant (could be due to small sample size, high variance within the data, choice of candidate models, etc.), no dose response relationship will be further explored. Otherwise, the models with significant adjusted p-values will be selected for further investigation. The best/optimal model will be the one with the minimum Akaike Information Criterion (AIC) criterion. AIC for each model will be displayed in the table and the optimal model with the minimum AIC will be indicated.

The target dose of interest is the minimum effective dose (MED) level that will achieve a target effect of delta over placebo effect. The target dose of interest will be estimated using each significant model using inverse regression techniques. The target dose will be computed on both continuous and discrete scales using R software. Effective dose (ED_p), the dose level that achieves a certain percentage p of the maximum effect over placebo (ie, ED_{10}, ED_{25}, ED_{50}, ED_{75}, ED_{90}) will also be calculated on both continuous and discrete scales.

Alternatively, a weighted average of the significant model contrasts will also be used to calculate a weighted average of the estimated target dose.

16.1.6. Subgroup Analysis of Primary Efficacy Variable

For each of the subgroup factors listed in Section 7.5, change from baseline in MADRS score at Week 6 will be analyzed using the same MMRM method with additional terms for the subgroup, subgroup-by-treatment interaction, subgroup-by-visit interaction, and subgroup-by-treatment-by-visit interaction using the ITT population. The MMRM with the additional by region subgroup interaction terms will be repeated after removal of all assessments conducted during remote visits and also by only keeping the records reported before 11 March 2020 separately. For country analysis, region will be replaced by country in the model. Summary statistics, LS mean treatment differences (each SEP-4199 group minus placebo), and effect size of change from baseline by week, their 2-sided 95% CIs, and the associated p-values will be provided as well as the p-value for the treatment-by-subgroup interaction. Its statistical significance will be assessed at the 0.10 level for homogeneity of the treatment effect across the different categories of a subgroup factor. In case a significant interaction effect is detected, estimates by subgroup will be examined to determine the nature of the interaction (qualitative or quantitative).

For these subgroups, a forest plot will be produced.

16.2. Secondary Efficacy

The secondary efficacy analyses will be performed for the ITT and PP population.

16.2.1. Secondary Efficacy Variable & Derivations

16.2.1.1. Change from baseline in Clinical Global Impression Bipolar Version, Severity of Illness (CGI-BP-S) Score (Depression) at Week 6

The CGI-BP-S is a clinician-rated assessment of the subject's current illness state on a 7-point scale, where a higher score is associated with greater illness severity. The CGI-BP-S score takes one of the following values: 1 (normal, not at all ill), 2 (borderline mentally ill), 3 (mildly ill), 4 (moderately ill), 5 (markedly ill), 6 (severely ill), 7 (among the most extremely ill patients). CGI-BP-S is assessed at Visit 2/Baseline/Week 0, Visit 3/Week 1, Visit 4/Week 2, Visit 5/Week 4, and Visit 6/Week 6.

16.2.2. Missing Data Methods for Secondary Efficacy Variable(s)

The secondary efficacy variable, change from baseline in CGI-BP-S depression score at Week 6, will be set to missing if the CGI-BP-S depression score at Week 6 is missing or baseline value is missing. The same applies to all other visits. The analysis of the secondary efficacy variable will use a MMRM which makes an MAR assumption to the missing secondary efficacy endpoints. Early termination data will be mapped as described in Section 6.3. Missing data will not be imputed.

16.2.3. Analysis of Secondary Efficacy Variable

The secondary efficacy estimand for this study is the difference in means of change from baseline in the CGI-BP-S depression score at Week 6 comparing each SEP-4199 treatment group to placebo in ITT population with major depressive episode associated with bipolar I disease (bipolar I disorder) at sites located in US and Europe defined through the inclusion/exclusion criteria, should a subject be able to tolerate and adhere to treatment up to Week 6.

The efficacy of SEP-4199 in terms of the CGI-BP-S depression score will be evaluated using the following two null hypotheses:

H3: There is no difference in mean change from baseline at Week 6 on the CGI-BP-S depression score in the SEP-4199 200 mg treatment arm compared to Placebo.

H4: There is no difference in mean change from baseline at Week 6 on the CGI-BP-S depression score in the SEP-4199 400 mg treatment arm compared to Placebo.

The alternative hypothesis for each of the null hypotheses is that there is a difference.

The CGI-BP-S depression score will be analyzed by the MMRM method described for the primary efficacy endpoint with baseline CGI-BP-S depression score used as covariate (see Section 16.1.3) for the ITT population. LS mean of change from baseline over time will be plotted including p-value.

The CGI-BP-S mania and overall bipolar illness scores will only be presented in listings.

16.2.4. Sensitivity Analysis of Secondary Efficacy Variable

Sensitivity analysis similar to the one described in Section 16.1.4 will also be produced for the secondary endpoint for the ITT population. See APPENDIX 3 for details.

16.2.5. Supportive Analysis of Secondary Efficacy Variable

ANCOVA and analysis on PP population as described in Section 16.1.5 will also be produced for the secondary endpoint.

16.2.6. Subgroup Analysis of Secondary Efficacy Variable

The same subgroup analysis as the one performed for primary analysis will be performed with the secondary endpoint (see Section 16.1.6) for the ITT population.

16.3. Other Efficacy Variables

All other efficacy endpoints will be analyzed using the ITT population.

16.3.1. Other Efficacy Variable Derivations

16.3.1.1. Hamilton Rating Scale for Anxiety (HAM-A)

The HAM-A is a rating scale developed to quantify the severity of anxiety symptomatology. The 14 items are: anxious mood, tension, fears, insomnia, intellectual, depressed mood, somatic (muscular), somatic (sensory), cardiovascular symptoms, respiratory symptoms, gastrointestinal symptoms, genitourinary symptoms, autonomic symptoms, and behavior at interview. Each of the 14 items is rated on a 5-point scale, ranging from 0 (not present) to 4 (very severe). The HAM-A total score is the sum of the 14 individual items and ranges from 0 to 56. A higher score is associated with a greater degree of anxiety. A score of ≥14 has been suggested to indicate clinically significant anxiety (Rush 2000). If 1 or more items are missing at a visit, the total score will be set to missing. HAM-A is assessed at Visit 2/Baseline/Week 0, Visit 3/Week 1, Visit 4/Week 2, Visit 5/Week 4, and Visit 6/Week 6.

16.3.1.2. Quick Inventory of Depressive Symptomatology – Self-Report 16-Item (QIDS-SR16)

The QIDS-SR16 is a 16-item self-report measure of depressive symptomatology which uses a computerized assessment interface for administration. The scoring system for the QIDS-SR16 converts responses from 16 separate items into nine DSM-IV symptom criterion domains. Each item is rated 0 to 3. For symptom domains that consist of more than one item, the highest score among the items relevant to the given domain is taken. The total score equals the sum of the nine individual domain scores and ranges from 0 to 27. Higher score indicates worse depression symptoms. The nine domains comprise: depressed mood (Item 5); concentration/decision making (Item 10); self outlook (Item 11); suicidal ideation (Item 12); decreased interest (Item 13); decreased energy (Item 14); sleep disturbance (initial, middle, and late insomnia or hypersomnia) (Items 1 to 4); appetite/weight disturbance (Items 6 to 9); and psychomotor disturbance (Items 15 and 16). The total score will be set to missing if one or more domain scores are missing at a visit. The appetite/weight disturbance symptom domain score will be set to missing if three or more items are missing at a visit, given the planned skip pattern of items 6 to 9, and the remaining symptom domain scores will be set to missing if one or more items within the

respective domain are missing. QIDS-SR16 is assessed at Visit 2/Baseline/Week 0, Visit 3/Week 1, Visit 4/Week 2, Visit 5/Week 4, and Visit 6/Week 6.

16.3.1.3. Sheehan Disability Scale (SDS)

Three items are self-rated using an 11-point visual analog scale ranging from 0 to 10 to assess disability across three domains: work/school, social life, and family life; the three items will be summarized individually in addition to the SDS total score. The SDS total score is calculated as the sum of the 3 items and ranges from 0 (unimpaired) to 30 (highly impaired). If one or more items are missing at a visit, as can occur when a subject opts out of the work/school item because it does not apply, the authors of the scale recommend setting the total score to missing (Sheehan 2008). Number of days lost due to symptoms, and number of days underproductive due to symptoms are also collected. SDS is assessed at Visit 2/Baseline/Week 0 and Visit 6/Week 6.

16.3.1.4. EuroQoL 5-D (EQ-5D-5L)

The EQ-5D-5L is a standardized instrument developed by the EuroQol Group as a measure of health-related quality of life that can be used in a wide range of health conditions and treatments. The EQ-5D consists of a descriptive system and the EQ VAS.

The descriptive system comprises five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems and extreme problems, corresponding to scores of 1 to 5, respectively. The patient is asked to indicate his/her health state by ticking the box next to the most appropriate statement in each of the five dimensions. This decision results in a 1-digit number that expresses the level selected for that dimension. The digits for the five dimensions can be combined into a 5-digit number that describes the patient's health state. An index value (a weighted scoring of the 5 dimension scores with a possible range from 0 to 1) will be assigned to each observed health state using the US value set as defined in APPENDIX 8.

The EQ VAS records the patient's self-rated health on a vertical visual analogue scale, where the endpoints are labeled 'The best health you can imagine' and 'The worst health you can imagine'. The VAS can be used as a quantitative measure of health outcome that reflect the patient's own judgment. EQ-5D-L is assessed at Visit 2/Baseline/Week 0 and Visit 6/Week 6.

16.3.1.5. MADRS responders

A MADRS responder is defined as achieving a ≥ 50% reduction from the baseline total score at Week 6. The MADRS total score percentage change will be defined as (value at postbaseline visit – baseline value) x 100 / (baseline value). Subjects having a negative percentage change indicate improvement in MADRS total score.

Any ITT subject with missing response at Week 6 will be classified as non-responder. As a sensitivity analysis, the same analysis will be repeated by excluding subjects with missing responses, where applicable.

In addition, the following endpoints will be explored through graphing:

- Subjects achieving a ≥ 40% reduction from the baseline total score at Week 6;

- Subjects achieving a ≥ 30% reduction from the baseline total score at Week 6;
- Subjects achieving a ≥ 20% reduction from the baseline total score at Week 6.

16.3.1.6. MADRS remission

A MADRS remission is defined as a MADRS total score of ≤ 12 at Week 6.

Any ITT subject with missing MADRS total score at Week 6 will be classified as non-remitter. As a sensitivity analysis, the same analysis will be repeated by excluding subjects with missing MADRS total score at Week 6, where applicable.

Additionally, subjects with MADRS total scores ≤ 10 and ≤ 8 at Week 6 are of interest.

16.3.1.7. Young Mania Rating Scale (YMRS)

The YMRS is an 11-item instrument used to assess the severity of mania in patients with a diagnosis of bipolar disorder. The 11 items are: Elevated Mood, Increased Motor Activity Energy, Sexual Interest, Sleep, Irritability, Speech (Rate and Amount), Language-Thought Disorder, Content, Disruptive-Aggressive Behaviour, Appearance and Insight. The YMRS is a clinician-rated assessment. Ratings are based on patient self-reporting, combined with clinician observation (accorded greater score). Seven items are rated on a 5-point scale, ranging from 0 to 4, and four items are rated on a 9-point scale, ranging from 0 to 8. The YMRS total score is the sum of the 11 individual items and ranges from 0 to 60. A higher score is associated with a greater severity of mania. If one or more items are missing at a visit, the total score will be set to missing. YMRS is assessed at Visit 1/Screening, Visit 2/Baseline/Week 0, Visit 3/Week 1, Visit 4/Week 2, Visit 5/Week 4, Visit 6/Week 6, and Visit 7/Follow-up/Week 7.

16.3.2. MISSING DATA METHODS FOR OTHER EFFICACY VARIABLE(S)

Any individual missing item in any scale will not be imputed.

For HAM-A, SDS, and YMRS, if 1 or more items are missing at a visit, the total score will be set to missing. Similarly, if an individual domain score for EQ-5D-5L is missing, the index value cannot be calculated.

For QIDS-SR16, the total score will be set to missing if 1 or more domain scores are missing at a visit. The appetite/weight disturbance symptom domain score will be set to missing if 3 or more items are missing at a visit, given the planned skip pattern of items 6 to 9, and the remaining symptom domain scores will be set to missing if 1 or more items within the respective domain are missing.

For MADRS responder, any ITT subject with a missing response at Week 6 will be classified as a non-responder.

For MADRS remission, any ITT subject with a missing MADRS total score at Week 6 will be classified as a non-remitter.

16.3.3. ANALYSIS OF OTHER EFFICACY VARIABLES

16.3.3.1. Change from baseline in HAM-A total score at Week 6

The HAM-A total score and change from baseline will be summarized by time point for each treatment group. Change from baseline at Week 6 will be analyzed by the MMRM method described for primary endpoint with baseline HAM-A total score as covariate (see Section 16.1.3).

Change from baseline at Week 6 will be analyzed by the ANCOVA model described in Section 16.1.5.1 with baseline HAM-A total score as covariate.

16.3.3.2. Change from baseline in QIDS-SR16 total score at Week 6

The QIDS-SR16 total score and change from baseline will be summarized by timepoint for each treatment group. Change from baseline at Week 6 will be analyzed by the MMRM method described for primary endpoint with baseline QIDS-SR16 total score as covariate (see Section 16.1.3).

Change from baseline at Week 6 will be analyzed by the ANCOVA model described in Section 16.1.5.1 with baseline QIDS-SR16 total score as covariate.

16.3.3.3. Change from baseline in SDS total score at Week 6

The SDS total score and change from baseline will be summarized by time point for each treatment group. Change from baseline at Week 6 will be analyzed by the ANCOVA model described in Section 16.1.5.1 with baseline SDS total score as covariate.

In addition, the following composite variables will be summarized: days = 0 at baseline, days at Week 6 ≥ 1; days ≥ 1 at baseline, days at Week 6 = 0; days = 0 at baseline, days at Week 6 ≥ 3; days ≥ 3 at baseline, days at Week 6 = 0. Results for these composite variables will be evaluated using logistic regression utilizing a 0-1 indicator (1 = met the criteria, 0 = did not meet the criteria) as the dependent variable, treatment and region as categorical factors, and baseline SDS total score as a covariate. If the model fails to converge, the model will be attempted without region first, and if still not converging then without baseline SDS total score. Contrasts will be evaluated for each treatment group compared to placebo. Odds Ratios, 95% CIs and Wald chi-square p-values will be presented for each contrast. Only n and percent will be presented in cases where the number of subjects meeting the above composite variable criteria is too small for an analysis to be performed. A Cochran-Armitage trend test will also be conducted on this outcome across placebo and the treatment groups.

16.3.3.4. Change from baseline in EQ-5D-5L Scores at Week 6

The EQ dimension scores, VAS score, and index score and change from baseline will be summarized by time point for each treatment group. Change from baseline of EQ VAS and index score at Week 6 will be analyzed by the ANCOVA model described in Section 16.1.5.1 with baseline EQ VAS and baseline index score as a covariate, respectively.

16.3.3.5. MADRS responders at Week 6

For MADRS responder, any ITT subject with missing response at Week 6 will be classified as Non-responder. As a sensitivity analysis, the same analysis will be repeated by only using observed values and not including subjects with missing responses. This sensitivity analysis will be repeated by removing remote visits due to COVID-19.

The MADRS total score responder proportions at Week 6 will be compared among the 3 treatment groups using logistic regression. This model will include treatment and region as fixed effects and baseline MADRS score as a covariate. If the model fails to converge, the model will be attempted without region first, and if still not converging then without baseline MADRS total score. Effect size and Number Needed to Treat (NNT) will be provided.

Between group effect size will be defined as: Natural Log(odds ratio)xsqrt(3)/pi

The NNT will be derived as 1/Risk Reduction where Risk Reduction (RR) = (SEP-4199 trt group – placebo response rate). The NNT results will be provided in whole numbers with any fractional values rounded up to the nearest whole number.

The 95% CI will also be presented (only when the lower and upper limits are of the same sign) and are computed by taking the reciprocal of the 95% lower and upper bound of the RR. The lower confidence limit will be rounded down to the largest integer value that is less than the computed estimate, and the upper confidence limit will be rounded up to the smallest integer value that is greater than the computed estimate.

Cochran-Armitage test for trend will be used to test for association between MADRS response and the three treatments.

The log rank testing method will be used to test for differences among the treatments for time to MADRS response. ITT subjects discontinuing the study and those completing the study without meeting the criteria for MADRS response will be censored at the date of last dose of study drug or, if unknown, the end of study date. The analysis will be supplemented with Kaplan-Meier curves to illustrate the differences among the three treatments. A Cox proportional hazards model will also be used with treatment, region, and baseline value as parameters. Wald Chi-square p-value will be included.

A cumulative distribution curve for all treatment groups will be presented, plotting the cumulative percent of MADRS responders versus the percent reduction at Week 6 from baseline.

16.3.3.6. MADRS remission at Week 6

Remission incidence rate, defined as MADRS total score ≤12, will be analyzed similarly to the MADRS responder analysis, including the sensitivity analysis of observed values and of assessing the removal of remote visits (see Section 16.3.3.5).

Subjects with MADRS total scores ≤ 10 and ≤ 8 at Week 6 will be summarized using descriptive statistics.

A cumulative distribution curve for all treatment groups will be presented, plotting the cumulative percent of MADRS remissions versus the MADRS total score at Week 6.

16.3.3.7. Change from baseline in YMRS total score over time

The YMRS total score and change from baseline will be summarized by time point for each treatment group. Change from baseline over time will be analyzed by the MMRM method described for primary endpoint with baseline YMRS total score as covariate (see Section 16.1.3).

Change from baseline over time will also be analyzed by the ANCOVA model described in Section 16.1.5.1 with baseline YMRS total score as covariate.

17. SAFETY OUTCOMES

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

There will be no statistical comparisons between the treatment groups for safety data, unless otherwise specified within the relevant section.

17.1. PRETREATMENT AND ADVERSE EVENTS

Pretreatment events and AEs will be coded using the MedDRA central coding dictionary, Version 19.1 or higher.

Adverse events are untoward medical occurrences that occurred on or after the first dose of study medication, with a missing start date and a stop date on, or with both a missing start and stop date. Untoward medical occurrences that started between informed consent and prior to the first dose of study medication are pre-treatment events.

Any COVID-19 related Adverse Events and pretreatment events will be identified using a pre-defined search.

For the COVID-19 related terms, see Table 4 below.

Table 4. COVID-19 related adverse/pre-treatment event identification.

Verbatim text	Lower Level Term (LLT)	Preferred Term	Search
	Coronavirus test positive	Coronavirus test positive	Using LLT
Includes the term “COVID-19”	Virus test positive	Virus test positive	Using LLT and text “COVID-19” in the verbatim text.

Whenever available, the time information should be accounted for in the derivation of AEs vs. pretreatment events. In the case where time isn’t available, untoward medical occurrences that started on or after the day of the first dose of study drug will be considered AEs; those that started before the day of the first dose of study drug will be considered pretreatment.

See APPENDIX 2 for handling of partial dates for AEs. In the case where it is not possible to define an untoward medical occurrence as an AE or pretreatment event, it will be classified by the worst case, ie, AE.

An overall summary of the incidence of AEs within each of the categories described in the following sections will be provided as specified in the templates, including summary of COVID-19 related AEs and SAEs. This summary will also be repeated by region, race (category 2 in Section 7.5), sex, and age (< 55 and ≥ 55). The overall incidence summary will also be provided for AEs related to study drug.

Listings will be provided for all AEs, AEs leading to discontinuation of study drug, AEs leading to discontinuation from the study, serious adverse events (SAEs), and AEs leading to death. A listing for pretreatment events will also be presented. Any COVID-19 related AE will be flagged in the AE listings.

For incidence summaries, each subject will be counted only once within each SOC and PT. Adverse Events will be sorted alphabetically by SOC and then by decreasing frequency of PT within each SOC based on the All SEP-4199 treatment group. For the incidence summary table by SOC, High Level Term (HLT), and PT, each subject will be counted only once within each SOC, HLT and PT, and AEs will be sorted alphabetically by SOC and HLT, and then by decreasing frequency of PT within each SOC and HLT based on the All SEP-4199 group. If not otherwise specified, all summaries will present incidence (number of subjects and percentages) and number of events.

17.1.1. ALL AEs

Adverse events will be presented by SOC and PT for AE incidence and number of events. A listing of all AEs will be presented.

An AE summary table by SOC, HLT, and PT will also be provided.

Adverse events reported by ≥ 2% of subjects who received SEP-4199 and that occurred at greater incidence than in the placebo group will be summarized by SOC and PT.

Non-serious AEs occurring with a PT frequency exceeding 5% in any treatment group will be summarized by PT.

Adverse events will also be presented by maximum severity and by strongest relationship to the study drug as specified in the sections below.

17.1.1.1. Severity

Severity is classified as mild/ moderate/severe (increasing severity). Adverse events with a missing severity will be summarized as severe. If a subject reports an AE more than once within the same SOC/PT, the AE with the worst severity will be used in the corresponding severity summaries. For this summary, AEs will be presented by SOC and PT.

17.1.1.2. Relationship to Study Drug

Relationship to study drug, as indicated by the investigator, is classed as “not related”/ “possible”/ “probable”/ “definite” (increasing strength of relationship). A “related” AE is defined as an AE with a relationship to the study drug of “possible”, “probable” or “definite”. A “not related” AE is defined as an AE with a relationship to the study drug of “not related”. Adverse events with a missing relationship to the study drug will be regarded as “related” to the study drug. If a subject reports the same AE more than once within the same SOC/PT, the AE with the strongest relationship to study drug will be used in the corresponding relationship summaries. For this summary, AEs will be presented in 2 categories, related and not related, by SOC and PT.

17.1.2. AEs LEADING TO DISCONTINUATION FROM STUDY TREATMENT PERIOD

Adverse events leading to discontinuation from the study treatment period are AEs with “Caused Study Discontinuation” = “Yes” on the AE CRF page for subjects with reason for discontinuation as AE on the Study Disposition CRF page or AE action taken=”Drug Withdrawn”. A summary of AEs leading to

discontinuation from the study treatment period by SOC and PT will be presented. A listing of AEs leading to discontinuation from the study treatment period will be presented.

17.1.3. SERIOUS ADVERSE EVENTS

Serious AEs are those AEs recorded as “Serious” on the AE CRF page. Summaries of SAEs by SOC and PT will be prepared. Related SAEs will be summarized. A listing of SAEs will be presented.

17.1.4. ADVERSE EVENTS LEADING TO DEATH

AEs leading to death are those AEs which are recorded as having an outcome of “Fatal” on the AE CRF page. A summary of AEs leading to death by SOC and PT will be prepared. Related AEs leading to death will be summarized. Overall deaths resulting from all causes will also be presented. A listing of deaths will be presented.

17.1.5. ADVERSE EVENTS BY SUBGROUP

As stated above, the overall incidence summary table and the AEs by SOC and PT summary table will also be presented by the subgroups of region as appropriate, race (category 2 in Section 7.5), gender, and age.

17.2. LABORATORY EVALUATIONS

All summaries will be based on the SAF population. Laboratory data to be reported for this study include Hematology, Serum Chemistry (including lipid panel and thyroid panel), Urinalysis, Urine drug screening, Hemoglobin (HbA1c), serum follicle stimulating hormone (FSH), Serum Prolactin, Serology, Serum Insulin, C-Reactive Protein (CRP), and urine pregnancy test. A list of laboratory assessments to be included in the outputs is included in section 21 of the CSP. Urine drug screening, serum FSH, serology, and urine pregnancy test will only be listed.

Presentations will use international system of units (SI).

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

The following summaries will be provided for laboratory data:

- By-visit summary of observed values and changes from baseline for continuous data in hematology, chemistry, and urinalysis.
- By visit summary of the number and percentage of subjects in each outcome category for categorical data in urinalysis (if applicable).

- Shift in laboratory results (chemistry, hematology, urinalysis) from baseline to Week 6 according to the reference range criteria. The normal reference ranges from the central laboratory will be used to determine whether the laboratory test value is below, within, or above the normal range. Urinalysis will be presented as normal or abnormal (including low, high, abnormal) in the shift table.

- By planned visit (after mapping of the ET visit) and overall (including unscheduled visits) summary of markedly abnormal postbaseline value (MAPLV, see Section 17.2.1 and APPENDIX 5). Subjects will be represented in the count of a particular MAPLV (potentially clinically significant [PCS] low or high) if they have experienced that MAPLV at least once postbaseline, regardless of baseline value. The number and percentage of subjects with MAPLV will be presented by treatment group.

- Shift in markedly abnormal laboratory results from baseline to postbaseline according to the markedly abnormal criteria as presented in APPENDIX 5.

- Serum prolactin results will be summarized overall and separately by gender. Glucose, insulin, and lipid panel results, along with a derived variable for homeostasis model assessment of insulin resistance (HOMA-IR) will be summarized for fasting only and overall (fasting, non-fasting, or fasting status unknown combined) status. HOMA-IR will be derived based on glucose and insulin results as: HOMA-IR = Glucose (mg/dL) x Insulin (mU/L) / 405 (Matthews 1985). HbA1c will also be summarized.

- The change from baseline values at Week 6 for selected laboratory parameters will be evaluated using a nonparametric rank ANCOVA analysis with adjustments for the corresponding baseline values for comparison between each SEP-4199 group and placebo. Using the regression residuals from the rank ANCOVA analysis as scores, the Mantel-Haenszel row mean scores test will be used to compare between the treatment groups. This analysis will be conducted for CRP, Glucose (Fasting and Overall), Insulin (Fasting and Overall), HOMA-IR (Fasting and Overall), Cholesterol Total (Fasting and overall), HDL (Fasting and Overall), LDL (Fasting and Overall), Triglycerides (Fasting and Overall), Prolactin (Overall, Male, and Female), and HbA1c (Fraction of 1).

All laboratory data will be provided in data listings, with the values outside the reference ranges flagged.

17.2.1. Laboratory Reference Ranges and Markedly Abnormal Criteria

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

- Low: Below the lower limit of the laboratory reference range.

- Normal: Within the laboratory reference range (upper and lower limit included).

- High: Above the upper limit of the laboratory reference range.

For laboratory parameters with categorical outcomes, measurements will be compared with the relevant laboratory reference values and categorized as normal or abnormal.

In addition to the high and low quantitative laboratory assignments (as identified by means of the laboratory reference ranges), markedly abnormal quantitative safety (and other) laboratory assessments will also be identified in accordance with the predefined markedly abnormal criteria as presented in APPENDIX 5.

17.3. ECG EVALUATIONS

Results from the central ECG will be included in the reporting of this study.

The following ECG parameters will be reported for this study:

- PR Interval (msec)
- RR Interval (msec)
- QRS Interval (msec)
- QT Interval (msec)
- QTcF Interval (msec) [derived]
- QTcB Interval (msec) [derived]
- Heart rate (HR) (beats/min)
- ECG findings
- Overall assessment of ECG (investigator's judgment):
    - Normal
    - Abnormal, clinically Significant (CS)
    - Abnormal, not clinically significant (NCS)

The following summaries will be provided for ECG data for the SAF population:

- By visit summary of observed values and changes from baseline (for quantitative measurements)
- By visit and overall (including unscheduled visits) summary of ECG overall assessment results. Percentage of subjects will be based on the number of subjects with ECG overall assessment available at the given time point. Any unscheduled ECG that occurs after first dose will be included in the overall post-treatment summaries.
- Shift in ECG overall assessments from baseline to Week 6.

- Number and percentage of subjects with QTc levels in each of the QTc categories

The number and percentage of subjects with QTc values in the following categories will be identified, same criteria apply to both QTcF and QTcB:

- >= 450 msec for males \ >=470 msec for females at any postbaseline time point (including unscheduled visits) not present at baseline
- >= 480 msec for male or females at any postbaseline time point (including unscheduled visits) not present at baseline
- >= 500 msec for males or females at any postbaseline time point (including unscheduled visits) not present at baseline
- ≥ 30 msec and < 60 msec increase from baseline for at least one postbaseline measurement (including unscheduled visits)
- ≥ 60 msec increase from baseline for at least one postbaseline measurement (including unscheduled visits)

All ECG parameters, overall interpretation, and findings will be provided in data listings. A listing displaying ECG values for subjects with at least 1 prolonged QTc will be produced.

17.3.1. ECG Specific Derivations

The following 3 measures are provided by the ECG vendor and those provided values will be analyzed. If one of these values is missing and derivation in the course of analysis is needed, the following derivations may be used:

- Bazett's Correction of QT interval (msec)

$$QTcB\ (\text{msec}) = \frac{QT\ (ms)}{\sqrt{RR\ (ms)/1000}}$$

- Fridericia's Correction of QT interval (msec)

$$QTcF\ (\text{msec}) = \frac{QT\ (ms)}{\sqrt[3]{RR\ (ms)/1000}}$$

- RR Interval – If RR Interval is not available it will be derived from HR as follows, for the derivation of the QTc corrections

$$RR\ (\text{msec}) = 1000 * \frac{60}{HR\ (bpm)}$$

17.3.2. ECG Markedly Abnormal Criteria

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

- Absolute values for QT interval, QTcB interval and QTcF will be classified as:
  - >= 450 msec for males \ >= 470 msec for females
  - >= 480 msec for males or females
  - >= 500 msec for males or females
- Change from Baseline for QT interval, QTcB interval and QTcF will be classified as:
  - >= 30 msec and < 60 msec increase from baseline
  - >= 60 msec increase from baseline

17.4. VITAL SIGNS

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

- Supine Systolic Blood Pressure (mmHg)
- Standing Systolic Blood Pressure (mmHg)
- Supine Diastolic Blood Pressure (mmHg)
- Standing Diastolic Blood Pressure (mmHg)
- Supine Pulse Rate (beats/min)
- Standing Pulse Rate (beats/min)
- Respiratory Rate (breaths/min)
- Temperature (C)
- Weight (kg)
- BMI (kg/m^2)
- Waist circumference (cm)

The following summaries will be provided for vital signs data for the SAF population:

- Observed value and change from baseline by visit including the follow-up visit.
- BMI Categories will be described by visit.

- Mantel-Haenszel row mean scores test will be used to compare each SEP-4199 group and placebo similar to the laboratory parameters.

- By visit and overall summary of Markedly Abnormal Postbaseline Vital Signs (MAPVS, see Section 17.4.1). This analysis will include only subjects having experienced MAPVS at least once postbaseline.

- The number and percentage of subjects with orthostatic hypotension and orthostatic tachycardia will be summarized by treatment for baseline and the overall postbaseline period, as well as by visit. As specified in Section 6.3, any orthostatic hypotension or tachycardia events that occurred at the early termination visit will be assigned to the next planned visit. Orthostatic hypotension is defined as a decrease of ≥ 20 mmHg in systolic blood pressure or ≥ 10 mmHg in diastolic blood pressure after the subject had been standing for at least 2 to 4 minutes, compared to the systolic and diastolic blood pressures measured in the supine position, respectively. Orthostatic tachycardia is defined as a heart rate increase of at least 20 beats per minute (bpm) after the subject was standing for at least 2 to 4 minutes, compared to the heart rate measured in the supine position and a heart rate > 100 bpm after the subject was standing for at least 2 to 4 minutes. Standing vital signs will be used as collected on the CRFs.

All vital signs data will be provided in a data listing, including flagging of MAPVS.

17.4.1. Vital Signs Markedly Abnormal Criteria

MAPVS measurements will be identified in accordance with the following predefined markedly abnormal criteria:

Table 5. Markedly abnormal vital sign criteria.

Parameter Name	Low	Decrease from Baseline	High	Increase from Baseline
Systolic BP	≤ 90 mmHg	≥ 20 mmHg	≥ 180 mmHg	≥ 20 mmHg
Diastolic BP	≤ 50 mmHg	≥ 15 mmHg	≥ 105 mmHg	≥ 15 mmHg
Pulse rate	≤ 50 bpm	≥ 15 bpm	≥ 120 bpm	≥ 15 bpm
Respiration Rate	≤ 10 breaths/min	≥ 50%	≥ 25 breaths/min	≥ 50%
Temperature	N/A		≥ 38.3 °C	≥ 0.8°C increase from baseline

Note: A vital sign measurement is MAPVS Low if it is below the specified lower limit and decreased from baseline or MAPVS High if it is above the specified upper limit and increased from baseline.

17.5. PHYSICAL AND NEUROLOGICAL EXAMINATION

As all physical and neurological findings will be recoded as medical history or AEs, no specific analysis of physical and neurological examination will be performed.

17.6. OTHER SAFETY ASSESSMENTS

17.6.1. TREATMENT-EMERGENT MANIA – YOUNG MANIA RATING SCALE (YMRS)

The YMRS is an 11-item instrument used to assess the severity of mania in patients with a diagnosis of bipolar disorder (see Section 16.3.1.7 for details). Number and percentage of patients with treatment-emergent mania, as assessed by the YMRS total score of ≥ 16 on any 2 consecutive visits or at the last YMRS assessment, or an AE of mania or hypomania as assessed by medical review will be summarized by treatment group for the SAF population. For incidence summaries, each subject will be counted only once if multiple criteria are met. Mania incidence rates will be analyzed for the SAF population similarly to the MADRS responder analysis. The mania indicator will be set to 1 if the subject exhibits postbaseline mania, 0 if the subject does not experience post-baseline mania and has at least 1 non-missing postbaseline YMRS total score, and to missing if there is no postbaseline YMRS total score.

The change from baseline over time in the YMRS total score will also be analyzed for the ITT population using the ANCOVA and MMRM method as described in the efficacy section (Sections 16.3.1.7 and 16.3.3.7).

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

The C-SSRS is a tool designed to systematically assess and track suicidal behavior and suicidal ideation for life time, past 3 months (90 days) prior to the screening visit, and throughout the study. The strength of this suicide classification system is in its ability to comprehensively identify suicidal events while limiting the over-identification of suicidal behavior. The C-SSRS Baseline/Screening Version is used at the screening visit and the C-SSRS Since Last Visit Version is used from Visit 2 (lead-in) onward. Subjects with Type 4 (active suicidal ideation with some intent to act, without specific plan) or Type 5 (active suicidal ideation with specific plan and intent) suicidal ideation during the study will be discontinued from the study and referred to a mental health professional. C-SSRS is assessed at Visit 1/Screening, Visit 2/Baseline/Week 0, Visit 3/Week 1, Visit 4/Week 2, Visit 5/Week 4, Visit 6/Week 6, and Visit 7/Follow-up/Week 7.

C-SSRS includes four sections: Suicidal Ideation, Intensity of Ideation, Suicidal Behavior, and Answer for Actual Suicide Attempts.

The following outcomes are C-SSRS categories and have binary responses (yes/no). The categories are re-ordered from the scale to facilitate the definitions of the C-SSRS endpoints, and to provide clarity in the presentation of the results.

Suicidal ideation is measured by 5 categories, representing 5 subtypes of suicidal ideation with increasing severity:

- Category 1: Wish to be Dead
- Category 2: Non-specific Active Suicidal Thoughts
- Category 3: Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act
- Category 4: Active Suicidal Ideation with Some Intent to Act, without Specific Plan
- Category 5: Active Suicidal Ideation with Specific Plan and Intent

Suicidal behavior is measured by 5 categories, representing 5 subtypes of suicidal behavior:

- Category 6: Preparatory Acts or Behavior
- Category 7: Aborted Attempt
- Category 8: Interrupted Attempt
- Category 9: Actual Attempt (non-fatal)
- Category 10: Completed Suicide

The 10 categories above are not mutually exclusive. Subjects will be counted in each category for which they have an event.

Self-injurious behavior without suicidal intent is a non-suicide-related C-SSRS outcome, and also has a binary response (yes/no).

For the purpose of C-SSRS analysis, “baseline” and “postbaseline” are defined as follows.

Time point	Study Visit	C-SSRS Version	Derivation Rule
Baseline	Screening/Visit 1	Baseline/Screening – Past 3 Months	Most severe outcome
	Visit 2/Baseline*	Since Last Visit	
Post-baseline	All postbaseline visits up to and including Visit 6/Week 6 and Visit 7/follow-up visit, including unscheduled visits	Since Last Visit	Most severe outcome

* Note: The Visit 2/Baseline C-SSRS assessment must be administered prior to the first dose of study medication in order to be used in the C-SSRS Baseline derivation.

C-SSRS composite endpoints will be derived for each time point of interest (ie, baseline, postbaseline, and each study visit) as follows:

- Any suicidal ideation: A “yes” answer to any one of the 5 suicidal ideation questions on C-SSRS (Categories 1-5).
- Any suicidal behavior: A “yes” answer to any one of the 5 suicidal behavior questions on the C-SSRS (Categories 6-10).
- Any Suicidal Ideation or Behavior: A “yes” answer to any one of the 10 suicidal ideation and behavior questions on the C-SSRS (Categories 1-10).

For each subject, the suicidal ideation score at each time point of interest (ie, baseline, postbaseline, each study visit including follow-up visit) is defined as the maximum suicidal ideation category (1-5) present for the time of interest. If no ideation is present a score of 0 is assigned. A suicidal ideation score of 4 or 5 is considered serious.

Intensity of ideation for the most severe ideation subtype is measured in terms of frequency, duration, controllability, deterrents, and reasons for ideation. Each is measured with responses ranging from 1 to 5 for frequency and duration, and from 0 to 5 for controllability, deterrents, and reasons for ideation. The ideation intensity total score is the sum of responses to the five items and can range from 2 to 25 for subjects with endorsed suicidal ideation. If one or more of these five items are missing at an assessment, the total score will be set to missing. If a subject did not endorse any suicidal ideation, a score of 0 for the ideation intensity total score will be given.

The number and percentage of subjects with any suicidality, any suicidal ideation, and any suicidal behavior in the SAF population will be presented for:

- Baseline (as defined above)
- Postbaseline (as defined above)
- Each scheduled study visit: Screening (lifetime; past 3 months), Visit 2/Baseline/Week 0, Visit 3/Week 1, Visit 4/Week 2, Visit 5/Week 4, Visit 6/Week 6, and Visit 7/Follow-up/Week 7.

Responses to each question will be listed.

17.6.3. Abnormal Involuntary Movement Scale (AIMS)

The AIMS is a clinician-rated assessment of abnormal movements consisting of unobtrusive observation of the subject at rest (with shoes removed) and several questions or instructions directed toward the subject. It contains seven items related to: facial, lip, jaw, and tongue movements, upper and lower extremity movements, and trunk movements. Three other items assess the subject at a global level, and two items assess dental status. The individual items on the AIMS are rated from 0 = ‘None’ to 4 = ‘Severe’. The (non-global) total AIMS score is the sum of items 1 through 7. The possible range for the total AIMS score is 0 to 28. Items 8 through 12 are not included in the total AIMS score. Higher values of the total AIMS score indicate increased severity in abnormal movement. If 1 or more of the AIMS total score items are missing at a visit, the score will be set to missing. Item 8, representing global severity score, will be summarized separately. AIMS is assessed at Visit 2/Baseline/Week 0, Visit 3/Week 1, Visit 4/Week 2, Visit 5/Week 4, and Visit 6/Week 6.

All summaries and analyses will be based on the SAF population. Descriptive statistics for absolute value and change from baseline will be displayed at each visit by treatment group. Each total score will be analyzed using MMRM and ANCOVA based on the change from baseline to Week 6 similar to the primary efficacy endpoint, with the respective baseline values as covariate (see Sections 16.1.3 and 16.1.5.1).

The AIMS total score at each visit will also be classified as 'abnormal' if either: at least two items have a response of 'mild' or higher; or at least one item has a response of 'moderate' or higher. Otherwise, non-missing total scores will be classified as 'normal'. This is a modification of the Schooler-Kane Criteria for Tardive Dyskinesia (Schooler 1982). Shifts from baseline will be summarized by visit, postbaseline overall (based on maximum severity), and treatment group.

Frequency distribution of AIMS global severity score will be provided by visit and treatment group. Postbaseline AIMS global severity scores will be classified as 'worsened', 'unchanged', or 'improved', relative to a subject's baseline score. A higher score than that of baseline would be classified as 'worsened'. Conversely, a lower score would be classified as 'improved'. These postbaseline changes will be summarized by visit and treatment group.

17.6.4. BARNES AKATHISIA RATING SCALE (BARS)

The BARS is a rating scale geared toward assessment of neuroleptic-induced akathisia, though it can be used to measure akathisia associated with other drugs as well. The BARS consists of four items, including one item assessing objective restlessness, two items targeting subjective restlessness (awareness and related distress), and one global clinical assessment item. All items are anchored and utilize a 4-point scale, except for the global rating which has a 6-point scale (from absence of akathisia through severe akathisia). The subjective and objective items are summed to yield a total score. The BARS total score is the sum of items 1 through 3 and ranges from 0 to 9. Higher values of the BARS total score indicate higher severity of akathisia. If one or more of items 1 to 3 at a visit are missing the total will not be calculated. BARS is assessed at Visit 2/Baseline, Visit 3/Week 1, Visit 4/Week 2, Visit 5/Week 4, and Visit 6/Week 6.

All summaries and analyses will be based on the SAF population. Descriptive statistics for absolute value and change from baseline in BARS total score and global clinical assessment score will be displayed at each visit by treatment group. The BARS total score and global clinical assessment score will be analyzed using MMRM and ANCOVA based on the change from baseline to Week 6 similar to the primary efficacy endpoint, with the respective baseline values as covariate (see Sections 16.1.3 and 16.1.5.1).

Categorical scores for the four BARS items will be summarized with number and percentage of observations by visit and treatment group.

The postbaseline BARS Global Clinical Assessment of Akathisia responses will be classified as 'worsened', 'unchanged', or 'improved', relative to a subject's baseline response. A higher score than that of baseline would be classified as 'worsened'. Conversely, a lower score would be classified as 'improved'. Shifts from baseline will be summarized by visit and treatment group.

17.6.5. MODIFIED SIMPSON-ANGUS SCALE (SAS)

The Modified SAS is a clinician-rated assessment of neuroleptic-induced Parkinsonism consisting of 10 items. Items are anchor-based, rated on a 5-point scale, and address rigidity, gait (bradykinesia), tremor, glabellar tap, and salivation (Simpson 1970). The SAS mean score is defined as the average of all 10 items and ranges between 0 and 4. Lower values of the SAS mean score indicate milder symptoms. If one or more items are missing at a visit the SAS mean score will be set to missing. SAS is assessed at Visit 2/Baseline, Visit 3/Week 1, Visit 4/Week 2, Visit 5/Week 4, and Visit 6/Week 6.

All summaries and analyses will be based on the SAF population. Descriptive statistics for absolute value and change from baseline in SAS mean score will be displayed at each visit by treatment group. The SAS mean score will be analyzed using MMRM and ANCOVA based on the change from baseline to Week 6 similar to the primary efficacy endpoint, with the baseline SAS mean score value as covariate (see Sections 16.1.3 and 16.1.5.1).

The SAS mean score at each visit will also be classified as 'abnormal' if it exceeds 0.3 (Rush 2000). Otherwise, non-missing mean scores will be classified as 'normal.' Shifts from baseline will be summarized by visit and treatment group.

17.6.6. Physician Withdrawal Checklist (PWC)

The Physician Withdrawal Checklist (PWC) is used to evaluate symptoms of withdrawal after discontinuation of study drug. The scale includes 20 symptoms and each symptom is assessed on a 4 point scale using the following: 0=Not Present, 1=Mild, 2=Moderate and 3=Severe. The score for each question is summed to compute a total score ranging from 0 to 60. If the response to any question is missing, then the total score will be missing.

This checklist will be assessed at the end of treatment visit (Visit 6/Week 6) and Visit 7/Follow-up/Week 7 visit.

All summaries and analyses will be based on the SAF population. The PWC will be summarized using descriptive statistics by visit for the total score. The number and percentage of subjects with a Present response (mild, moderate, or severe) will be provided for each of the 20 PWC items. In addition, responses of mild, moderate and severe will be summarized separately at each visit.

17.7. Subgroup Analysis of Safety Variables

Adverse events will be presented by the subgroups of region, race, gender, and age (See Section 17.1.5).

18. Pharmacokinetic Analysis

All PK analysis will be performed using PK population.

Blood sample for plasma concentrations of R-amisulpride, S-amisulpride, total amisulpride, and/or plasma prolactin measurement will be collected at Visit 1/Screening (only for plasma prolactin), Visit 2/Baseline/Day -1, Visit 4/Week 2, Visit 6/Week 6, and Visit 7/Follow-up/Week 7. The PK collection times from Predose on Day -1 to Follow-up and deviations and/or any values excluded from analysis will be presented in data listings. Any plasma concentration or summary statistics below the lower limit of

quantification (LLOQ) will be represented by “BLQ” (below the limit of quantification) in tables and listings. Amisulpride plasma concentration will not be presented for the placebo group.

R-amisulpride, S-amisulpride, total amisulpride, and plasma prolactin concentrations at each scheduled sample collection time point (including ET and Follow-up, see Section 6.3 for mapping of the ET visit) will be summarized descriptively (n, mean, median, minimum, maximum, coefficient of variation [CV] and if appropriate, geometric mean and geometric CV [GCV]). In addition, if there is at least 1 concentration < LLOQ within a treatment group at a time point, it will be set to ½ LLOQ for summary statistics calculations. A second approach will be to assume a truncated normal distribution for the specific concentration data if at least 10% of the concentrations at a time point are < LLOQ within a treatment group. The mean and variance of truncated normal will be calculated (APPENDIX 3). Number and percentage of concentrations that are below the LLOQ will be provided for each visit. All PK summaries will be presented by dose.

18.1. DERIVATION

LLOQ is 0.0500 ng/mL for R-amisulpride and S-amisulpride and LLOQ is 1.56 ng/mL for plasma prolactin.

Below derivations apply after concentrations lower then LLOQ are set to ½ LLOQ.

- Coefficient of variation

  100*Standard Deviation/Mean

- Geometric Mean

  Exponential (mean of $\log_e$ transformed data)

- Geometric CV

  Square Root (Exponential (Variance ($\log_e$ transformed data)-1))*100

19. DATA NOT SUMMARIZED OR PRESENTED

Disposition, demographics, and pretreatment events will be listed for screened subjects including the screen failures. A randomized subject’s any premedication data from his/her screen failed period will be included in ADAM and in Tables/Listings. Any pre-treatment medical occurrences from the subject’s screen failed period will be excluded from ADAE and all other screen failed period data will be excluded from any baseline calculations for the later randomized subject. The information on scales collected on paper is not included in ADAM and in Tables/Listings. Other data that are collected on screen failures will not be presented, but will be available in the clinical study database and SDTM domains.

20. CHANGES IN THE ANALYSIS SPECIFIED IN THE STATISTICAL ANALYSIS PLAN

Any changes or deviations during the analysis and reporting process from the statistical analysis plan designed will be described and justified in the final report.

21. REFERENCES

Bornkamp B, Pinheiro JC, Bretz F. MCPMod: An R package for the design and analysis of dose-finding studies. Journal of Statistical Software. 2009;29(7):1–23.

Buckland ST, Burnham KP, Augustin NH. Model selection an integral part of inference. Biometrics. 1997;53:603–618.

Dmitrienko A, Koch GG. Analysis of clinical trials using SAS: A practical Guide. Second Edition. SAS Press: Cary, NC. 2017.

Dmitrienko A, Kordzakhia G, Brechenmacher T. Mixture-based gatekeeping procedures for multiplicity problems with multiple sequences of hypotheses. Journal of Biopharmaceutical Statistics. 2016;26(4):758-780.

Dmitrienko A, Molenberghs G, Chuang-Stein C, Offen W. Analysis of Clinical Trials Using SAS: A Practical Guide. SAS Press: Cary, NC. 2005.

Dmitrienko A, Tamhane A, Liu L. Mixtures of multiple testing procedures. Northwestern University. Department of Industrial Engineering and Management Sciences. Working Paper 08-04. 2008.

Diggle PJ, Heagerty P, Liang KY, Zeger SL. The analysis of Longitudinal Data. Oxford, England: 2nd Edition. Oxford University Press. 2002.

Genz A, Bretz F. Computation of Multivariate Normal and t Probabilities. Springer, 2009.

Gregory L, FDA. “Sensitivity Analysis Framework and A Novel Tipping Point Approach” Regulatory-Industry Statistics Workshop. Washington D.C. 30 Sept 2016.

Hochberg Y, Tamhane AC. Multiple Comparison Procedures. Wiley, New York. 1987.

Loebel A, Cucchiaro J, Silva R, Kroger H, Hsu J, Sachs G. Lurasidone monotherapy in the treatment of bipolar I depression: a randomized, double-blind, placebo-controlled study. Am J Psychiatry. 2014 Feb;171(2):160 8.

Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia. 1985;28 (7):412–9.

O’Kelly M, Ratitch B. Clinical Trials with Missing Data: A Guide for Practitioners. John Wiley & Sons. 2014.

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical. Computing, Vienna, Austria. URL https://www.R-project.org/. 2017.

Ratitch B, O’Kelly M. Implementation of pattern-mixture models using standard SAS/STAT procedures. SUGI. 2011.

Ratitch B, O’Kelly M, Tosiello R. Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. Pharmaceutical Statistics. 2013;12:337–347.

Rush A, Pincus H, First M, Blacker D, Endicott J, Keith S, Phillips K, Ryan N, Smith G, Tsuang M, Widiger T, Zarin D. Handbook of Psychiatric Measures, Washington D.C: American Psychiatric Association. 2000.

Rush J, Pincus H, First M, Blacker D, Endicott J, Keith S, Phillips K, Ryan N, Smith G, Tsuang M, Widiger T, Zarin D. Handbook of Psychiatric Measures. American Psychiatric Association. 2000;554-556.

Schooler NR, Kane JM. Research Diagnoses for Tardive Dyskinesia. Arch. Gen. Psychiatry. 1982;39:486-487.

Sheehan KH, Sheehan DV, Assessing treatment effects in clinical trials with the Discan metric of the Sheehan Disability Scale. International Clinical Psychopharmacology. 2008;23:70-83.

Permutt T. Sensitivity analysis for missing data in regulatory submissions. Statistics in Medicine. 2015.

Pinheiro JC, Bornkamp B, Glimm E and Bretz F. Model-based dose finding under model uncertainty using general parametric models. Statistics in Medicine. 2014;33:1646–1661.

APPENDIX 1. PROGRAMMING CONVENTIONS FOR OUTPUTS

OUTPUT CONVENTIONS

Where applicable, the Appendix_Compilation_Working_Guidelines_Final 07May2014.pdf document – provided by Sunovion – will be followed.

In addition, the following output conventions are to be followed:

- The first row in the body of the table or listing should be blank
- The left-hand column should start in column 1. No indenting or centering of the output should occur.
- Rounding should be done with the SAS function ROUND.
- Numbers in tables should be rounded, not truncated.
- Alphanumeric output should be left aligned.
- Numbers should be decimal point aligned.
- Whole numbers should be right aligned.
- Text values should be left aligned.
- The first letter of a text entry should be capitalized.
- The width of the entire output should match the linesize (134)

- Univariate Statistics:

  If the raw data has N decimal places, then the summary statistics should have the following decimal places:

  - Minimum and maximum: N
  - Mean, median, Q1, and Q3: N + 1
  - SD: N + 2

- Frequencies and percentages (n and %):

  - Percent values should be reported inside parentheses, with one space between the count (n) and the left parenthesis of the percentage. Parentheses should be justified to accept a maximum of 100.0 as a value and padded with blank space if the percent is less than 100.0.
  - Percentages will be reported to one decimal place, except cases where percent <100.0% but >99.9% will be presented as '>99.9%' (eg, 99.99% is presented as >99.9%); and cases where percent < 0.1% will be presented as '<0.1%' (eg, 0.08% is presented as <0.1%). Rounding will be applied after the <0.1% and >99.9% rule.
  - Where counts are zero, no percentage should appear in the output.

- Confidence Intervals:

  - Confidence intervals and estimates are presented to one place more than the raw data, and standard errors to two places more than the raw data.
  - Confidence intervals should be justified so that parentheses displayed on consecutive lines of a table "line up".
  - Boundary values of confidence intervals should be separated by a comma.

  - Boundary values should be padded as necessary to accept negative values and to allow alignment of the decimal place.

- P-values:
  - P-values should be reported to three decimal places, except values <1.000 but >0.999 will be presented as '>0.999' (eg, 0.9998 is presented as >0.999); and values <0.001 will be presented as '<0.001' (eg, 0.0009 is presented as <0.001). Rounding will be applied after the <0.001 and >0.999 rule.
- Ratios:
  - Ratios should be reported to one more decimal place than the raw data.
- Spacing:
  - There must be a minimum of 1 blank space between columns (preferably 2).
- Missing values:
  - A "0" should be used to indicate a zero frequency.
  - A blank will be used to indicate missing data in an end-of-text table or subject listing.
- Figures:
  - Figures should be provided in RTF files using the SAS Output Delivery System (ODS), as Computer Graphics Metafile (CGM) formatted graphical output generated by SAS.
  - The CGM file itself should contain the title or footer.
  - The image should be clear and of high quality when viewed in the Word document, and when printed.
  - In general, boxes around the figures should be used.
- Footers should be defined as follows:
  - A continuous line of underscores ('_') will follow the body of the table or listing prior to any footnotes at the bottom of the page.
  - Table footnotes should be defined using compute statements in the proc report, and should appear directly after the body of the table.
  - If text wraps across more than one line (for a note), the first letter for all lines of text after the first one will be indented to align beneath the first letter of the text in the first line.

DATES & TIMES

Depending on data available, dates and times will take the form DDMMMYYYYThh:mm:ss.

SPELLING FORMAT

English US.

PRESENTATION OF TREATMENT GROUPS

For outputs, treatment groups will be represented as follows and in that order:

Treatment Group	For Tables, Graphs and Listings
Placebo	Placebo
SEP-4199 200 mg	SEP-4199 200 mg
SEP-4199 400 mg	SEP-4199 400 mg
All SEP-4199	All SEP-4199 (only applies to tables)

LISTINGS

All listings will be sorted by the following (unless otherwise indicated in the template):

- Randomized/Actual treatment received as applicable, displaying SEP-4199 first and then placebo.
- Subject ID,
- Original date/time (where applicable) - listings of AEs, concomitant medications, medical histories etc should be sorted in chronological order, with earliest AE, medication or history coming first,
- For listings where non-randomized subjects are included, these will appear in a category after the randomized treatment groups labelled ‘Not Randomized’.

APPENDIX 2. PARTIAL DATE CONVENTIONS

Imputed dates will not be presented in the listings. In the algorithms for date, SEP-4199 study med start date the date of first dose of study drug.

ALGORITHM FOR ADVERSE EVENTS:

The concept of "date" below should also include time information whenever available.

START DATE	STOP DATE	ACTION
Known	Known	If start date < SEP-4199 study med start date, then pre-treatment events If start date >= SEP-4199 study med start date, then SEP-4199 AE
Known	Partial	If start date < SEP-4199 study med start date, then pre-treatment events If start date >= SEP-4199 study med start date, then SEP-4199 AE
Known	Missing	If start date < SEP-4199 study med start date, then pre-treatment events If start date >= SEP-4199 study med start date, then SEP-4199 AE
Partial, but known components show that it cannot be on or after SEP-4199 study med start date	Known	Pre-treatment events
Partial, but known components show that it cannot be on or after SEP-4199 study med start date	Partial	Pre-treatment events
Partial, but known components show that it cannot be on or after SEP-4199 study med start date	Missing	Pre-treatment events
Partial	Partial	Impute stop date as latest possible date (ie, last day of month if day unknown or 31st December if day and month are unknown). Then: If stop date < SEP-4199 study med start date, then pre-treatment events If stop date >= SEP-4199 study med start date, then SEP-4199 AE

START DATE	STOP DATE	ACTION
Partial, could be on or after SEP-4199 study med start date	Known	If stop date < SEP-4199 study med start date, then pre-treatment events If stop date >= SEP-4199 study med start date, then SEP-4199 AE
Partial, could be on or after SEP-4199 study med start date	Partial	Impute stop date as latest possible date (ie, last day of month if day unknown or 31st December if day and month are unknown). Then: If stop date < SEP-4199 study med start date, then pre-treatment events If stop date >= SEP-4199 study med start date, then SEP-4199 AE
Partial	Missing	Assumed SEP-4199 AE
Missing	Known	If stop date < SEP-4199 study med start date, then pre-treatment events If stop date >= SEP-4199 study med start date, then SEP-4199 AE
Missing	Partial	Impute stop date as latest possible date (ie, last day of month if day unknown or 31st December if day and month are unknown). Then: If stop date < SEP-4199 study med start date, then pre-treatment events If stop date >= SEP-4199 study med start date, then SEP-4199 AE
Missing	Missing	Assumed SEP-4199 AE

ALGORITHM FOR PRIOR / CONCOMITANT / POST-TREATMENT MEDICATIONS:

The concept of "date" below should also include time information whenever available. Time will not be imputed.

START DATE	STOP DATE	ACTION
Known	Known	If stop date < SEP-4199 study med start date, assign as prior. If stop date >= SEP-4199 study med start date and start date <= end of treatment date of last dose of study drug), assign as concomitant. If stop date >= SEP-4199 study med start date and start date > end of treatment (date of last dose of study drug), assign as post treatment.
Known	Partial	Impute stop date as latest possible date: • If only day unknown, impute as the earlier of (last day of the month; date of the last study visit). • If month and day unknown, impute as the earlier of (31st December; date of the last study visit). Then: If stop date < SEP-4199 study med start date, assign as prior. If stop date >= SEP-4199 study med start date and start date <= end of treatment (date of last dose of study drug), assign as concomitant. If stop date >= SEP-4199 study med start date and start date > end of treatment (date of last dose of study drug), assign as post treatment.
Known	Missing	If stop date is missing could never be assumed a prior medication. If start date <= end of treatment (date of last dose of study drug), assign as concomitant. If start date > end of treatment (date of last dose of study drug), assign as post treatment.
Partial	Known	Impute start date as earliest possible date: *CRF questions: 'Started prior to study?' = Yes; 'Started after last dose of study medication?' = No.* • If only day unknown, impute as the later of (first day of the month; date of birth). • If month and day unknown, impute as the later of (1st January; date of birth). *CRF questions: 'Started prior to study?' = No; 'Started after last dose of study medication?' = Yes.* • If only day unknown, impute as the later of (first day of the month; end of treatment + 1). • If month and day unknown, impute as the later of (1st January; end of treatment + 1). *CRF questions: 'Started prior to study?' = No; 'Started after last dose of study medication?' = No.* • If only day unknown, impute as the later of (first day of the month; date of ICF). • If month and day unknown, impute as the later of (1st January; date of ICF). Then: If stop date < SEP-4199 study med start date, assign as prior. If stop date >= SEP-4199 study med start date and start date <= end of treatment (date of last dose of study drug), assign as concomitant.

		If stop date >= SEP-4199 study med start date and start date > end of treatment (date of last dose of study drug), assign as post treatment.
Partial	Partial	Impute start date as earliest possible date: *CRF questions: 'Started prior to study?' = Yes; 'Started after last dose of study medication?' = No.* • If only day unknown, impute as the later of (first day of the month; date of birth). • If month and day unknown, impute as the later of (1st January; date of birth). *CRF questions: 'Started prior to study?' = No; 'Started after last dose of study medication?' = Yes.* • If only day unknown, impute as the later of (first day of the month; end of treatment + 1). • If month and day unknown, impute as the later of (1st January; end of treatment + 1). *CRF questions: 'Started prior to study?' = No; 'Started after last dose of study medication?' = No.* • If only day unknown, impute as the later of (first day of the month; date of ICF). • If month and day unknown, impute as the later of (1st January; date of ICF). Impute stop date as latest possible date: • If only day unknown, impute as the earlier of (last day of the month; date of the last study visit). • If month and day unknown, impute as the earlier of (31st December; date of the last study visit). Then: If stop date < SEP-4199 study med start date, assign as prior. If stop date >= SEP-4199 study med start date and start date <= end of treatment (date of last dose of study drug), assign as concomitant. If stop date >= SEP-4199 study med start date and start date > end of treatment (date of last dose of study drug), assign as post treatment.
Partial	Missing	Impute start date as earliest possible date: *CRF questions: 'Started prior to study?' = Yes; 'Started after last dose of study medication?' = No.* • If only day unknown, impute as the later of (first day of the month; date of birth). • If month and day unknown, impute as the later of (1st January; date of birth). *CRF questions: 'Started prior to study?' = No; 'Started after last dose of study medication?' = Yes.* • If only day unknown, impute as the later of (first day of the month; end of treatment + 1). • If month and day unknown, impute as the later of (1st January; end of treatment + 1). *CRF questions: 'Started prior to study?' = No; 'Started after last dose of study medication?' = No.* • If only day unknown, impute as the later of (first day of the month; date of ICF). • If month and day unknown, impute as the later of (1st January; date of ICF). Then: If stop date is missing could never be assumed a prior medication. If start date <= end of treatment (date of last dose of study drug), assign as concomitant. If start date > end of treatment (date of last dose of study drug), assign as post

		treatment.
Missing	Known	If stop date < SEP-4199 study med start date, assign as prior. If stop date >= SEP-4199 study med start date, assign as concomitant
Missing	Partial	Impute stop date as latest possible date: • If only day unknown, impute as the earlier of (last day of the month; date of the last study visit). • If month and day unknown, impute as the earlier of (31st December; date of the last study visit). Then: If stop date < SEP-4199 study med start date, assign as prior. If stop date >= SEP-4199 study med start date, assign as concomitant.
Missing	Missing	Assign as concomitant

PARTIAL DATE IMPUTATION RULES FOR INITIAL ONSET OF BIPOLAR I DISORDER/CURRENT EPISODE OF MAJOR DEPRESSION ASSOCIATED WITH BIPOLAR I DISORDER SYMPTOMS:

For subjects with partial onset dates of bipolar I disorder/current episode, impute the onset date using the following rules:

- If only day unknown, impute as the earlier of: last day of the month, or date of ICF.
- If both month and day unknown, impute as the earlier of: 31st December of the year, or date of ICF.

APPENDIX 3. STATISTICAL MODEL SPECIFICATIONS

MIXED MODEL FOR REPEATED MEASURES (MMRM)

```
Ods graphics on;
PROC MIXED DATA=XXX plots=(studentpanel(marginal)  pearsonpanel(marginal));
        CLASS SUBJID REGION TRTPN WEEK;
        MODEL TOTCFB = TOTBL REGION TRTPN WEEK TRTPN*WEEK / DDFM=KR RESIDUAL OUTPM=OUTRESIDUAL;
        REPEATED WEEK/TYPE=UN SUBJECT=SUBJID;
        LSMEANS TRTPN*WEEK /PDIFF CL;
RUN;
*Where WEEK = Week 1, Week 2, Week 4, Week 6;
Ods graphics off;
```

In case the MMRM model assuming the unstructured covariance variance structure fails to converge, the spatial exponential or spatial power covariance structure will be assumed sequentially along with a robust sandwich estimator. The first covariance structure to yield convergence will be used in the analysis.

```
PROC MIXED EMPIRICAL;
        CLASS SUBJID REGION TRTPN WEEK;
        MODEL TOTCFB = TOTBL REGION TRTPN WEEK TRTPN*WEEK / RESIDUAL OUTP=OUTRESIDUAL;
        REPEATED WEEK/TYPE=SP(EXP) (WEEK) SUBJECT=SUBJID;
        REPEATED WEEK/TYPE=SP(POW) (WEEK) SUBJECT=SUBJID;
        LSMEANS TRTPN*WEEK /PDIFF CL COV;
RUN;
*EMPIRICAL IS FOR THE SANDWICH ESTIMATOR;
```

Within-group effect size will be calculated as the absolute value of LS mean of change from baseline divided by SD, obtained as the SE of the LS Mean multiplied by the square root of the treatment group change from baseline sample size at each visit. Between-group effect size at a visit will be calculated as the absolute value of the LS means difference from Placebo divided by the model estimate of the pooled SD at the visit, which is obtained from the square root of the diagonal element, associated at the visit, from the covariance matrix (R matrix of subjects with MADRS total score at all visits). Although the variance-covariance estimates are calculated differently depending on the covariance structure being applied for the model, the diagonal of the R matrix will contain the variance estimate at each visit that can be used for the effect size calculations regardless of the covariance structure.

For All SEP-4199 column for safety questionnaires:

- LS Means estimate, 95%CI and SE is obtained using following statements:

```
LSMESTIMATE TRT01PN*AVISITN "MEAN(200MG+400MG) at Week 1" [1, 1 1] [1, 2 1]  / CL divisor=2;
LSMESTIMATE TRT01PN*AVISITN "MEAN(200MG+400MG) at Week 2" [1, 1 2] [1, 2 2]  / CL divisor=2;
LSMESTIMATE TRT01PN*AVISITN "MEAN(200MG+400MG) at Week 4" [1, 1 3] [1, 2 3]  / CL divisor=2;
LSMESTIMATE TRT01PN*AVISITN "MEAN(200MG+400MG) at Week 6" [1, 1 4] [1, 2 4]  / CL divisor=2;
```

- LS Means Difference estimate (vs Placebo), 95%CI and SE is obtained using following statements:

```
LSMESTIMATE TRT01PN*AVISITN "MEAN(200MG+400MG) vs  Placebo at Week 1" [0.5, 1 1] [0.5, 2 1] [-1, 3 1]  / CL ;
LSMESTIMATE TRT01PN*AVISITN "MEAN(200MG+400MG) vs  Placebo at Week 2" [0.5, 1 2] [0.5, 2 2] [-1, 3 2]  / CL ;
LSMESTIMATE TRT01PN*AVISITN "MEAN(200MG+400MG) vs  Placebo at Week 4" [0.5, 1 3] [0.5, 2 3] [-1, 3 3]  / CL ;
LSMESTIMATE TRT01PN*AVISITN "MEAN(200MG+400MG) vs  Placebo at Week 6" [0.5, 1 4] [0.5, 2 4] [-1, 3 4]  / CL ;
```

NORMALITY CHECK

```
PROC UNIVARIATE DATA=OUTRESIDUAL NORMAL;
        VAR STUDENTRESID;
        VAR PEARSONRESID;
RUN;
```

ANALYSIS OF COVARIANCE (ANCOVA)

```
PROC MIXED;
        BY WEEK;
        CLASS TRTPN REGION;
        MODEL TOTCFB = TRTPN REGION TOTBL;
        LSMEANS TRTPN / PDIFF CL;
RUN;
```

For All SEP-4199 column for safety questionnaires:

- LS Means estimate, 95%CI and SE is obtained using

  ```
  LSMESTIMATE  TRT01AN "MEAN(200 mg,400mg)" 1 1 0 / CL DIVISOR=2;
  ```

- LS Means Difference estimate (versus Placebo) , 95%CI and SE is obtained using

  ```
  ESTIMATE "MEAN(200+400) VS PLACEBO" TRT01AN 0.5 0.5 -1 / CL;
  ```

MH TEST (FOR SELECTED SAFETY PARAMETERS)

```
*Conducted for each separate SEP-4199 group versus Placebo;
PROC RANK OUT=OUT1 NPLUS1 TIES=MEAN;
        VAR VSCFB VSBASE;
        RANKS RKVSCFB RKVSBASE;
RUN;

PROC GLM DATA=OUT1;
        CLASS TRTPN;
        MODEL RKVSCFB = TRTPN RKVSBASE;
        OUTPUT OUT=RESIDUAL R=RESID;
RUN;

PROC FREQ DATA=RESIDUAL;
        TABLES TRTAN*RESID / NOPRINT CMH2;
RUN;
```

For All SEP-4199 column, please select all data and create a new treatment code equal to 1 if subject was treated by SEP-4199 (either 200mg or 400mg) and equal to 0 if he received Placebo. Then use same code as above with this new treatment code.

TIME TO RESPONSE ANALYSIS

```
* Select Raw Log-rank p-value in SURVDIFF ods dataset;
* Response probability are obtained in PRODUCTLIMITESTIMATES ods dataset;
ODS OUTPUT SURVDIFF=SURVDIFF PRODUCTLIMITESTIMATES=PRODUCTLIMITESTIMATES;
PROC LIFETEST OUTSURV=PROBRESPONSE TIMELIST=(9 16 30 44);
        TIME RSPTIME*RSPSTAT(1);
        STRATA TRTPN / DIFF=CONTROL ('Placebo');
RUN;

ODS OUTPUT HAZARDRATIOS=HAZARDRATIOS PARAMETERESTIMATES=PARAMETERESTIMATES;PROC PHREG;
        CLASS REGION TRTP (REF='Placebo') / PARAM=REF;
        MODEL RSPTIME*RSPSTAT(1)=TRTP REGION TOTBL;
        HAZARDRATIO TRTP / DIFF=REF;
       CONTRAST '200 MG VS. PLACEBO' TRTP 1 0 / ESTIMATE;
        CONTRAST '400 MG VS. PLACEBO' TRTP 0 1 / ESTIMATE;
RUN;
```

LOGISTIC REGRESSION

```
* Placebo is reference;
PROC LOGISTIC DESCENDING;
        CLASS TRTPN REGION / PARAM=REF;
        MODEL RESP = TRTPN REGION TOTBL;
        CONTRAST '200 MG vs. PLACEBO' TRTPN 1 0  / ESTIMATE=EXP;
        CONTRAST '400 MG vs. PLACEBO' TRTPN 0 1  / ESTIMATE=EXP;
        CONTRAST '200 MG vs. 400 MG'  TRTPN -1 1 / ESTIMATE=EXP;
CONTRAST '(200 MG + 400 MG) vs Placebo'  TRTPN 0.5 0.5 / ESTIMATE=EXP;RUN;
```

COCHRAN-ARMITAGE TEST

```
PROC FREQ;
       TABLE TRTP*RESP/CMH TREND;
RUN;
```

MULTIPLICITY-ADJUSTED P-VALUES

```
*Example raw p-values;
DATA STUDY;
        INPUT HYP $ FAMILY PARALLEL $ SERIAL $ RAWP;
        DATALINES;
        H1 1 0000 0000 0.1403
        H2 1 0000 0000 0.0095
        H3 2 0000 1000 0.0559
        H4 2 0000 0100 0.0051
        ;
RUN;

%TREEGATE(TEST=HOCHBERG, GAMMA=0.9);
```

PATTERN MIXTURE MODEL (PMM) WITH PLACEBO BASED MI

The steps to implement this sensitivity analysis are as follows. For Japan cohort analysis, code must be adapted by removing REGION variable.

Step 1: Intermittent missing MADRS total scores will be imputed for each treatment using non-missing data from all ITT subjects within the treatment group by a Markov Chain Monte Carlo (MCMC) imputation model which assumes a multivariate normal distribution over all variables included in the imputation model. This imputation is based on the MAR assumption, ie, the missing data are assumed to follow the same model as the other patients in their respective treatment group.

```
PROC MI DATA=xxx OUT=yyy NIMPUTE=1000 SEED=12345 MINIMUM=. . 0 0 0 0 MAXIMUM=. . 60 60 60 60
MINMAXITER=100;
        BY TRTP;
        VAR REGION MADRS_BASE MADRS_Wk1 MADRS_Wk2 MADRS_Wk4 MADRS_Wk6;
        MCMC CHAIN=MULTIPLE IMPUTE=MONOTONE NBITER=5000 NITER=200;
RUN;
```

As a result, each of the 1000 datasets will only have a monotone missing data pattern.

There is no missing value expected at baseline for MADRS total score since it is related to inclusion/exclusion criteria. For the secondary endpoint, there could be missing values at baseline for CGI-BP-S scores. To impute CGI-BP-S scores, intermittent imputation to obtain monotone missing pattern has two steps, 1a and 1b.

In Step 1a, missing baseline CGI-BP-S total scores will be imputed. In Step 1b, the intermittent missing data will be filled in, assuming MAR, to obtain only monotone missing data.

Step 1a: If there are any ITT subjects who do not have a baseline CGI-BP-S total score, their baseline data needs to be imputed first. Missing baseline data will be imputed using a monotone regression on 100 imputations with a model that includes region, and baseline CGI-BP-S depression score. Additional covariates may be added as appropriate.

```
PROC MI DATA=xxx OUT=aaa NIMPUTE=100 SEED=1234 MINIMUM=. 1 MAXIMUM=. 7 MINMAXITER=100;
        BY TRTP;
        VAR REGION CGI_BASE;
        MCMC CHAIN=MULTIPLE IMPUTE=MONOTONE NBITER=5000 NITER=200;
RUN;
```

Step 1b: Intermittent missing CGI-BP-S total scores will be imputed for each TRTP using non-missing data from all subjects within the treatment group by MCMC imputation model which assumes a multivariate normal distribution over all variables included in the imputation model. This imputation is based on the MAR assumption, ie, the missing data are assumed to follow the same model as the other patients in their respective treatment group.

```
PROC MI DATA=aaa OUT=yyy NIMPUTE=10 SEED=4567 MIN=. . 1 1 1 1 MAX=. . 7 7 7 7 MINMAXITER=100;
        BY _IMPUTATION_;
        BY TRTP;
        VAR REGION CGI_BASE CGI_WK1 CGI_WK2 CGI_WK4 CGI_WK6;
        MCMC CHAIN=MULTIPLE IMPUTE=MONOTONE NBITER=5000 NITER=200;
RUN;
```

As a result, each of the 1000 datasets will only have a monotone missing data pattern.

Step 2: This step will use the MNAR assumption (placebo group based imputation). The assumption is that the trajectories of all the patients (including the placebo group subjects who discontinue) are assumed to follow the placebo group after their discontinuation.

```
PROC MI DATA = yyy OUT = zzz NIMPUTE = 1 SEED = 56789 MINIMUM=. . 0 0 0 0 MAXIMUM=. . 60 60 60 60;
        BY _IMPUTATION_;
        VAR REGION MADRS_BASE MADRS_WK1 MADRS_WK2 MADRS_WK4 MADRS_WK6;
        CLASS TRTP REGION;
        MONOTONE REGRESSION(/ DETAILS);
        MNAR MODEL(MADRS_WK1 MADRS_WK2 MADRS_WK4 MADRS_WK6/ MODELOBS = (TRTP = 'Placebo'));
RUN;
```

The MODEL option specifies that only observations where TRT = "Placebo" are used to derive the imputation model for the missing postbaseline MADRS scores. Thus, the TRT variable is not specified in the VAR list. This approach is also known as jump to reference (jump to placebo in our case) where it is assumed that the discontinuations from the SEP-4199 treatment groups immediately change to have the distribution of the placebo group. This imputation model is also the imputation model that is used to impute missing postbaseline observations in the placebo group.

Step 3: Fit the primary MMRM model using each of the 1000 imputed datasets (BY _IMPUTATION_) and calculate the inferential statistics on each imputed dataset.

```
PROC MIXED DATA=zzz;
        BY _IMPUTATION_;
        CLASS SUBJID REGION TRTP WEEK;
        MODEL TOTCFB = MADRS_BASE REGION TRTP WEEK TRTP*WEEK / DDFM=KR;
        REPEATED WEEK/TYPE=UN SUBJECT=SUBJID;
        LSMEANS TRTP*WEEK /PDIFF CL;
        ODS OUTPUT DIFFS=CHGDIFF LSMEANS=CHGLSMN;
RUN;
```

Step 4. Use proc MIANALYZE to combine the estimates from Step 3 based on Rubin's imputation rules. See example for LS Means below.

```
PROC MIANALYZE DATA = CHGLSMN;
        BY TRTP WEEK;
        MODELEFFECTS ESTIMATE;
```

```
        STDERR STDERR;
        ODS OUTPUT PARAMETERESTIMATES = LSMN (DROP = PARM);
RUN;
```

TIPPING POINT ANALYSIS

This analysis assumes that the data in the SEP-4199 treatment groups is not missing at random. The tipping point based assumption will be used, ie, the trajectories of the subjects in the SEP-4199 treatment groups after withdrawal are assumed to be worse by an amount of delta, after the MI using the MAR assumption.

The steps to implement this sensitivity analysis are as follows. For Japan cohort analysis, code must be adapted by removing REGION variable.

Step 1: Run the primary MMRM analysis for the ITT population. Obtain LS mean difference between each SEP-4199 group and the placebo group at each visit. Suppose these values are stored in variable LSMDIFF in dataset DIFF (LSMDIFF_200_PLACEBO, LSMDIFF_400_PLACEBO). This dataset should also contain the variable WEEK. These differences will be used when applying penalty on the imputed values.

Step 2: Similar to above, intermittent missing data will be filled in, assuming MAR, to obtain only monotone missing data.

```
PROC MI DATA=xxx OUT=yyy NIMPUTE=1000 SEED=12345 MINIMUM=. . 0 0 0 0 MAXIMUM=. . 60 60 60 60
MINMAXITER=100;
        BY TRTP;
        VAR REGION MADRS_BASE MADRS_Wk1 MADRS_Wk2 MADRS_Wk4 MADRS_Wk6;
        MCMC CHAIN=MULTIPLE IMPUTE=MONOTONE NBITER=5000 NITER=200;
RUN;
```

Step 3: Impute the remaining missing data using an MAR-based regression model.

```
PROC MI DATA = yyy OUT = zzz NIMPUTE = 1 SEED = 56789 MINIMUM=. . . 0 0 0 0 MAXIMUM=. . . 60 60 60 60
MINMAXITER=100;
        BY _IMPUTATION_;
        VAR TRTP REGION MADRS_BASE MADRS_WK1 MADRS_WK2 MADRS_WK4 MADRS_WK6;
        CLASS TRTP REGION;
        MONOTONE REGRESSION;
RUN;
```

Step 4: For a range of penalty values (ie, 5%, 10%, etc of the LS mean difference obtained in Step 1), apply penalty to the change from baseline values calculated from imputed MADRS total scores in each SEP-4199 group at each postbaseline visit.

Suppose ANALYSIS is the dataset reformatted from zzz that contains each subject's data in long format. It should contain the missing value flag MISSVAL (ie, whether the MADRS total score for a given visit is missing in the original data): 1 = missing; 0 = available. Other variables in analysis should

include: SUBJID, TRTP, REGION, WEEK, TOTBL, AVAL (MADRS total score at each postbaseline visit), TOTCFB, LSMDIFF, and _IMPUTATION_.

```
DATA ANALYSIS_PENALTY;
        SET ANALYSIS;
        /* store the original TOTCFB values in ORIGCHG */
        ORIGCHG = TOTCFB;
        /* if a MADRS total score is missing in the original dataset, apply a penalty to the TOTCFB value in each SEP-
        4199 group only */
        DELTA = XXX;
        IF MISSVAL=1 AND TRTP=”200 MG” THEN TOTCFB = ORIGCHG – ((DELTA/100)* LSMDIFF_200_PLACEBO);
        IF MISSVAL=1 AND TRTP=”400 MG” THEN TOTCFB = ORIGCHG – ((DELTA/100)* LSMDIFF_400_PLACEBO);
RUN;
```

Repeat the above step for all delta values, ie, 5, 10, …, 100%.

Step 5: Fit the primary MMRM model for each penalty and imputation to obtain the estimates at each postbaseline visit.

Suppose the primary MMRM model adopted the unstructured covariance model:

```
ODS OUTPUT DIFFS=CHGDIFF_TIPPING LSMEANS=CHGLSMN_TIPPING;
PROC MIXED DATA=ANALYSIS_PENALTY;
        BY DELTA _IMPUTATION_;
        CLASS SUBJID REGION TRTP WEEK;
        MODEL TOTCFB = TOTBL REGION TRTP WEEK TRTP*WEEK / DDFM=KR;
        REPEATED WEEK / SUB=SUBJID TYPE=UN;
        LSMEANS TRTP*WEEK / PDIFF CL;
RUN;
```

Step 6: Use proc MIANALYZE similar to above to combine the estimates from each imputation for each delta using Rubin’s rules. See example for LS Means below.

```
ODS OUTPUT PARAMETERESTIMATES=LSMEST_TIPPING_MIAN;
PROC MIANALYZE DATA = CHGLSMN _TIPPING;
        BY DELTA TRTP WEEK;
        MODELEFFECTS ESTIMATE;
        STDERR STDERR;
RUN;
```

This analysis is repeated for a range of deltas, starting at 0% of the treatment difference at each postbaseline visit, and increasing in increments of 5% until either the tipping point (where the statistical significance of the treatment effect is lost) is identified or the 100% penalty is applied. Tipping point could be reached at different deltas for each SEP-4199 treatment group. Once the tipping point %delta is identified, smaller %delta increments (eg 1%) may be selected for further analysis.

MCP-MOD ANALYSIS

The primary efficacy endpoint, change from baseline in MADRS score at Week 6, will also be analyzed using the gMCP-Mod approach for the ITT population. The analysis will be done in two steps.

Step 1: MCP step

The MCP step consists of specifying a set of candidate models for the dose response relationship. First, the following four candidate model families will be selected to describe the potential dose-response curve *f(d,θ)*.

- Linear model: $f(d, \theta) = E_0 + \delta\, d$

  For the linear model, E_0 is the drug effect at d≈0 and δ is the slope associated with d. There is no parameter to be estimated for the linear model.

- Quadratic model: $f(d, \theta) = d + \delta\, d^2$

  The quadratic model is intended to capture a possible non-monotonic dose-response relationship. For the quadratic model, there is one parameter δ to be estimated.

- E_{max} model: $f(d, \theta) = E_0 + E_{max}\, d /(ED_{50} + d)$

  The Emax model is used to represent monotone, concave dose-response shapes. To distinguish it from the more general sigmoid emax model it is sometimes also called hyperbolic emax model. For the E_{max} model, there is one parameter ED_{50} to be estimated.

- Sigmoidal E_{max}: $f(d, \theta) = E_0 + E_{max}\, d^h /(ED_{50}^h + d^h)$

  The sigmoid Emax model is an extension of the (hyperbolic) Emax model by introducing an additional parameter h, that determines the steepness of the curve at the ED_{50} value. The sigmoid Emax model describes monotonic, sigmoid dose-response relationships. For the SigmEmax model, there are two parameters ED_{50} and h to be estimated.

The next step is to determine the initial values of model parameters for the candidate models. Initial values for standardized model parameters can be calculated based on observed value at Week 6..

```
fitlinear<- fitMod(dose=dose, resp= resp, data = xxx, model="linear")
fitquadratic<- fitMod(dose=dose, resp= resp, data = xxx, model="quadratic")
fitemax<- fitMod(dose=dose, resp=resp, data = xxx, model="emax")
fitsigemax<- fitMod(dose=dose, resp=resp, data = xxx, model="sigEmax")
```

Note: Dose is 0 for placebo, 200 for SEP-4199 200mg/day and 400 for SEP-4199 400mg/day. Resp is the response so change from baseline at Week 6 in MADRS Total Score.

Based on an estimate of the dose-response vector and its associate variance/covariance matrix, optimal contrast will be computed for each model to maximize the power to test the null hypothesis of a constant dose-response curve. To obtain this estimate, an MMRM model with factors for treatment dose level, region, week, and baseline MADRS score as a continuous covariate will be used to model the change from baseline in MADRS score over time from the placebo and SEP-4199 groups for the ITT population. This analysis can be performed using the SAS software (Dmitrienko, 2017).

```
PROC MIXED;
        CLASS SUBJID REGION TRTPN WEEK;
        MODEL TOTCFB = TOTBL REGION TRTPN WEEK TRTP*WEEK / DDFM=KR RESIDUAL;
        REPEATED WEEK/TYPE=UN SUBJECT=SUBJID;
        LSMEANS TRTPN*WEEK /PDIFF CL COV;
        *ODS OUTPUT LSMEANS=MUH (KEEP=TRTPN WEEK ESTIMATE);
        *ODS OUTPUT LSMEANS=COVH (KEEP=TRTPN WEEK COVx COVy COVz) ;
RUN;
```

where COVx COVy COVz are representing the columns associated to LS Means estimated for Placebo, SEP-4199 200mg/day and 400 for SEP-4199 400mg/day at Week 6.

The vector of the optimal contrast estimates corresponding to the set of candidate models is asymptotically normally distributed. The test statistic used for establishing an overall dose response signal is the maximum t-statistics of the individual model test statistics. Critical values for tests with (asymptotically) exact level α can be derived from the joint distribution of the t statistics, which can be obtained from the joint distribution of the contrast estimates (Pinheiro 2014). Multiplicity adjusted p-values for the individual model contrast tests can be derived similarly. The mvtnorm package in R includes functions to calculate quantiles and probabilities for the underlying multivariate normal distributions (Pinheiro 2014, Genz 2009).

The optimal contrasts corresponding to the candidate models are calculated using the optContr function in the DoseFinding package in R (Pinheiro 2014). Model specific contrast coefficients and contrast correlation matrix will be displayed in R outputs.

```
doses <- c(0, 200, 400)
Delta=coef(fitquadratic)["b2"]/abs(coef(fitquadratic)["b1"])

mod <- Mods(emax     = coef(fitemax)["ed50"],
        quadratic = Delta,
        sigEmax   = coef(fitsigemax)[c("ed50","h")],
        linear    = NULL,
        doses     = doses, direction = "decreasing")
plot(mod, xlab= "Doses" , ylab= "Model  Means" )

contMat <- optContr(mod, S=covH)
summary(contMat)
plot(contMat, xlab= "Doses" , ylab= "Optimal Contrasts" )
```

The optimal model contrast tests will be performed using the MCTtest function in the DoseFinding package in R, based on the multiple comparison approach described above (Pinheiro 2014).

```
MCTtest(doses, muH, S=covH, type = "general", critV = TRUE, pVal = TRUE, alpha=0.05, alternative='one.sided', contMat=contMat)
```

Step 2: Mod step

After a dose response signal is established in the first step, one proceeds to the Mod step, fitting the dose response profile and estimating target doses based on the models identified in the MCP step (Pinheiro 2014). The candidate models will be retained to fit the data. The Generalized AIC (gAIC) will be used to select the model for which the maximum likelihood fit provides the smallest gAIC value. The fitMod function in DoseFinding can be used for this analysis.

If at least one significant model is found in Step 1 then use following code:

```
ModStep=MCPMod(doses, muH, S=covH, models=mod, type = "general", doseType="TD", Delta=3.5, pVal = T)
```

If no significant model is found in Step 1 then use following code:

```
ModStepEM <- fitMod(doses, muH, S=covH, model="emax", type = "general")
gAIC(ModStepEM)
coef(ModStepEM)
plot(ModStepEM)

ModStepQ <- fitMod(doses, muH, S=covH, model="quadratic", type = "general")
gAIC(ModStepQ)
coef(ModStepQ)
plot(ModStepQ)

ModStepLin <- fitMod(doses, muH, S=covH, model="linear", type = "general")
gAIC(ModStepLin)
coef(ModStepLin)
plot(ModStepLin)

ModStepsigEM <- fitMod(doses, muH, S=covH, model="sigEmax", type = "general")
gAIC(ModStepsigEM)
coef(ModStepsigEM)
plot(ModStepsigEM)
```

Estimation of target doses is done based on the selected fitted model for the dose response parameter (Bornkamp 2009).

```
TD(ModStep, Delta=xxx, direction="decreasing")

TD(ModStep, Delta=xxx, TDtype = "discrete", doses=doses, direction="decreasing")
```

Note: Appropriate Delta will be explored based on results

```
ED(ModStep, p=0.5)

ED(ModStep, p=0.5, EDtype = "discrete", doses=doses)
```

Note: Repeat for 10%, 25%, 50%, 75% and 90%

Alternatively, model averaging approaches can be used to avoid the need to select a single model. The individual AIC values for the candidate models with significant contrast test statistics determine the model averaging weights (Buckland 1997). This applies both to dose response and target dose estimation (Pinheiro 2014).

```
MM2 <- MCPMod(doses, muH, S=covH, type = "general", models=mod, alpha = 0.05, Delta=xxx, selModel =
"aveAIC", doseType = "TD", alternative = "one.sided")
```

Note: Appropriate Delta will be explored based on results

```
TDEst <- MM2$doseEst%*%MM2$selMod

TDEst
```

PK Analysis

If there is at least one PK concentration <LLOQ within a treatment group at a time point, two approaches will be used to handle the BLQ data. The first approach is to set PK concentration < LLOQ to ½ LLOQ for summary statistics calculations.

A second approach is to assume a truncated normal distribution for the specific concentration data if at least 10% of the concentrations at a time point are <LLOQ within a treatment group. The likelihood of the BLQ sample assumes that the value is less than the LLOQ, and the BLQ data will be censored with values less than LLOQ as LLOQ. Proc Lifereg in SAS will be used to perform a Tobit analysis. The Tobit model is a regression model for left-censored data assuming a normally distributed error term. The model parameters are estimated by maximum likelihood. Mean and SD will be provided by visit and treatment group.

```
DATA XXX;
        SET XXX;
        IF AVAL<AAA THEN DO;
          LOWER=.;
          UPPER=AAA;
        END;
        ELSE DO;
```

```
          LOWER=AVAL; UPPER=AVAL;
        END;
RUN;

PROC LIFEREG DATA=XXX;
        BY TRT01A WEEK;
        MODEL (LOWER, UPPER) = / D=NORMAL;
RUN;
```

APPENDIX 4. SAMPLE SIZE: POWER CALCULATION

The optimal value of the truncation parameter is chosen to maximize an appropriate power function. For this study, definition 1, the probability to reject at least one truly significant comparison for endpoint E1 in Family 1 was considered as the primary. In addition, the following power functions were also investigated:

- Definition 2: Probability of rejecting at least one truly significant comparison for endpoint E2 (Family 2)
- Definition 3: Probability of rejecting at least one truly significant comparison among all 4 comparisons for endpoint E1 and endpoint E2 in Family 1 and Family 2
- Definition 4: Probability of rejecting both truly significant comparisons for endpoint E1 (Family 1)
- Definition 5: Probability of rejecting both truly significant comparisons for endpoint E2 (Family 2)
- Definition 6: Probability of rejecting all four truly significant comparisons for endpoint E1 and endpoint E2 in Family 1 and Family 2
- Definition 7: Probability of rejecting at least one truly significant comparison for endpoint E1 (Family 1) and at least one truly significant comparison for endpoint E2 (Family 2)
- Definition 8: Probability of rejecting both truly significant comparisons for endpoint E1 (Family 1) and at least one truly significant comparison for endpoint E2 (Family 2)
- Definition 9: Probability of rejecting H_2 (400 mg/day versus Placebo in Family 1) and H_4 (400 mg/day versus Placebo in Family 2)
- Definition 10: Probability of rejecting H_1(200 mg/day versus Placebo in Family 1) and H_3 (200 mg/day versus Placebo in Family 1).

Table 6 below shows two scenarios considered for effect size assumptions.

Table 6. True effect sizes used in the power assessment

	200 mg/day vs Placebo	400 mg/day vs Placebo
Scenario 1		
Family 1 (Endpoint E1)	0.44	0.44
Family 2 (Endpoint E2)	0.38	0.38
Scenario 2		
Family 1 (Endpoint E1)	0.22 (half the 400 mg/day dose)	0.44

Family 2 (Endpoint E2)	0.19 (half the 400 mg/day dose)	0.38

Table 7 shows the correlation assumption between the 2 endpoints within a treatment.

Table 7. Correlation assumption used in the power assessment

	Endpoint 1	Endpoint 2
Endpoint 1	1	0.8
Endpoint 2	0.8	1

Table 8 and Table 9 below show the power calculations based on 93 subjects/group sample size for a range of gamma, γ, parameters for scenario 1 and scenario 2, respectively based on 10,000 simulation runs with a 1-sided 0.025 alpha level. The larger the value of γ, the higher the power of the Family 1 test and the lower the power of the Family 2 test. In the extreme case, when the truncation parameter is set to 1, the power of the Family 1 test will be maximized (the hypotheses in Family 1 would be tested using the regular Hochberg test) and the power of the Family 2 test would be minimized (it would be reduced to 0 and it would be impossible to reject any hypotheses in Family 2 unless all hypotheses in Family 1 are rejected). On the other hand, if the truncation parameter is set to 0, the power of the Family 1 test would be minimized (the truncated Hochberg test would be reduced to the Bonferroni test) but the power of the Family 2 test would be maximized.

Table 8. Power simulations for Scenario 1

Gamma	Defn 1	Defn 2	Defn 3	Defn 4	Defn 5	Defn 6	Defn 7	Defn 8	Defn 9	Defn 10
0	0.8939	0.7645	0.8939	0.6538	0.5221	0.5221	0.7645	0.6048	0.6409	0.6457
0.1	0.8939	0.7606	0.8939	0.6677	0.5296	0.5296	0.7606	0.6164	0.6434	0.6468
0.2	0.8942	0.7578	0.8942	0.6816	0.5357	0.5357	0.7578	0.6274	0.6448	0.6487
0.3	0.8950	0.7549	0.8950	0.6939	0.5415	0.5415	0.7549	0.6370	0.6456	0.6508
0.4	0.8964	0.7513	0.8964	0.7041	0.5453	0.5453	0.7513	0.6443	0.6466	0.6500
0.5	0.8980	0.7469	0.8980	0.7152	0.5492	0.5492	0.7469	0.6518	0.6452	0.6509
0.6	0.8992	0.7411	0.8992	0.7247	0.5526	0.5526	0.7411	0.6585	0.6441	0.6496
0.7	0.9004	0.7338	0.9004	0.7341	0.5559	0.5559	0.7338	0.6652	0.6424	0.6473
0.8	0.9016	0.7265	0.9016	0.7426	0.5592	0.5592	0.7265	0.6713	0.6402	0.6455
0.9	0.9027	0.7173	0.9027	0.7517	0.5633	0.5633	0.7173	0.6782	0.6391	0.6415
1	0.9034	0.6847	0.9034	0.7605	0.5663	0.5663	0.6847	0.6847	0.6236	0.6274

The following definitions are the same by definition: defn1=defn 3, Defn2=defn7, Defn5=defn6.

Table 9. Power simulations for Scenario 2

Gamma	Defn 1	Defn 2	Defn 3	Defn 4	Defn 5	Defn 6	Defn 7	Defn 8	Defn 9	Defn 10
0	0.7856	0.6124	0.7856	0.2143	0.1445	0.1445	0.6124	0.1962	0.6002	0.1567
0.1	0.7859	0.6029	0.7859	0.2273	0.1502	0.1502	0.6029	0.2061	0.5912	0.1619
0.2	0.7858	0.5892	0.7858	0.2344	0.1536	0.1536	0.5892	0.2118	0.5780	0.1648
0.3	0.7863	0.5810	0.7863	0.2476	0.1581	0.1581	0.5810	0.2226	0.5706	0.1685
0.4	0.7868	0.5687	0.7868	0.2554	0.1604	0.1604	0.5687	0.2280	0.5586	0.1686
0.5	0.7871	0.5525	0.7871	0.2630	0.1631	0.1631	0.5525	0.2341	0.5426	0.1730
0.6	0.7875	0.5347	0.7875	0.2726	0.1666	0.1666	0.5347	0.2417	0.5254	0.1759
0.7	0.7877	0.5135	0.7877	0.2811	0.1692	0.1692	0.5135	0.2483	0.5045	0.1782
0.8	0.7879	0.4830	0.7879	0.2904	0.1728	0.1728	0.4830	0.2563	0.4746	0.1812
0.9	0.7885	0.4339	0.7885	0.2996	0.1760	0.1760	0.4339	0.2633	0.4263	0.1836

1	0.7890	0.2709	0.7890	0.3092	0.1788	0.1788	0.2709	0.2709	0.2637	0.1860

The following definitions are the same by definition: defn1=defn 3, Defn2=defn7, Defn5=defn6.

Type I error rate assessment:

Table 10 displays the Type I error rate (probability of rejecting at least one null hypothesis out the four null hypotheses) for various values of γ assuming a zero effect size for both endpoints and a correlation of 0.8 between endpoints based on 10,000 simulation runs.

Table 10. Type 1 error rate: probability of rejecting at least one null hypothesis out of 4 comparisons

Gamma	Zero effect size
0	0.0237
0.1	0.0237
0.2	0.0237
0.3	0.0237
0.4	0.0239
0.5	0.0239
0.6	0.0240
0.7	0.0240
0.8	0.0240
0.9	0.0240
1	0.0242

APPENDIX 5. PREDEFINED MARKEDLY ABNORMAL CRITERIA IN SI UNIT

Category Parameter Name Age/Gender Restriction, if any	PCS Low	PCS High
HEMATOLOGY		
WBC	≤ 2.8 x 10^9/L	≥ 16 x 10^9/L
Neutrophils (abs)	< 0.5 x 10^9/L	> 13.5 x 10^9/L
Lymphocytes (abs)	N/A	> 12 x 10^9/L
Monocytes (abs)	N/A	> 2.5 x 10^9/L
Eosinophils (abs)	N/A	> 1.6 x 10^9/L
Basophils (abs)	N/A	> 1.6 x 10^9/L
Neutrophils (relative)	≤ 0.15	> 0.85
Lymphocytes (relative)	N/A	≥ 0.75
Monocytes (relative)	N/A	≥ 0.15
Eosinophils (relative)	N/A	≥ 0.10
Basophils (relative)	N/A	≥ 0.10
Hemoglobin Male Female	 ≤ 115 g/L ≤ 95 g/L	 ≥ 190 g/L ≥ 175 g/L
Hematocrit Male Female	 ≤ 0.37 ≤ 0.32	 ≥ 0.60 ≥ 0.54
RBC	≤ 3.5 x 10^{12}/L	≥ 6.4 x 10^{12}/L

Category Parameter Name Age/Gender Restriction, if any	PCS Low	PCS High
Platelet Count	≤ 75 x 10^9/L	≥ 700 x 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
Phosphate	< 0.65 mmol/L	> 1.65 mmol/L
Bicarbonate	< 15.1 mmol/L	> 34.9 mmol/L
Magnesium	< 0.4 mmol/L	> 1.23 mmol/L
AST （IU/L）	N/A	≥ 3 x ULN
ALT （IU/L）	N/A	≥ 3 x ULN
Alkaline Phosphatase (IU/L）	N/A	≥ 1.5 x ULN
GGT (Gamma-Glutamyl Transferase （IU/L）	N/A	≥ 2.5 x ULN
LDH （IU/L）	N/A	≥ 3 x ULN
CK（IU/L）	N/A	> 2.5 x ULN
Creatinine	N/A	≥ 177 umol/L
Creatinine Clearance	< 0.48343 mL/s	N/A
BUN	N/A	≥ 10.7 mmol/L
Total bilirubin	N/A	≥ 34.2 umol/L OR > 2 x ULN

Category Parameter Name Age/Gender Restriction, if any	PCS Low	PCS High
Total protein	≤ 45 g/L	≥ 100 g/L
Albumin	≤ 25 g/L	N/A
Total Cholesterol	N/A	> 7.76 mmol/L
HDL-Cholesterol	< 0.78 mmol/L	N/A
LDL-Cholesterol	N/A	> 4.14 mmol/L
Triglycerides	N/A	> 3.42 mmol/L
Uric acid Male Female	 N/A N/A	 > 595 umol/L > 476 umol/L
Glucose	< 2.78 mmol/L	> 13.9 mmol/L
HbA1c	N/A	≥ 0.075
Prolactin	N/A	≥ 5 x ULN
COAGULATION		
aPTT (sec)	N/A	> 1.5 x ULN
INR (ratio)	N/A	> 1.5 x ULN
URINALYSIS		
RBC	N/A	> 25 hpf
WBC	N/A	> 25 hpf

APPENDIX 6. BLINDED DATA REVIEW (BDR) PROCESS

The blinded review of the SEP3800-201 data will be conducted prior to unblinding in order to permit the checking and assessment of data for the purposes of identifying additional information for analyses specified in the SAP. This process will be completed prior to unblinding.

This review will be performed by appropriate personnel including, but not limited to a biostatistician and an MD. Relevant outputs will be generated by IQVIA shortly before/after database soft lock in order to facilitate the blinded review process.

Any changes to the planned analysis based on a blinded review of data will be documented with detailed explanation for deviations from the pre-defined plans. Decisions made at this meeting will be documented prior to unblinding to be incorporated into the analysis.

The BDR process will include review of analysis populations, checking distribution assumptions, identifying any outliers, defining possible transformations for the efficacy analyses, checking stratification, and reviewing the extent of missing data for the primary and the secondary efficacy endpoints. In addition, the IPD review process detailed in APPENDIX 7 will help identify any IPDs.

APPENDIX 7. SPECIFICATIONS OF IMPORTANT PROTOCOL DEVIATIONS

IPD REVIEW PROCESS

Only randomized subjects will be included in the IPD listings. The IPD criteria defined below will be applied consistently to all subjects regardless of the protocol version under which a subject was randomized. The IPD criteria given may be updated during the IPD review meetings, or anytime during the study, but will be final before the database lock.

For study SEP380-201, IPD review will be conducted periodically. In additional to the interim IPD reviews (around 33% and 66% of randomized subjects), there will be one final review. Interim IPD reviews will be conducted following the completion of each batch cleaning. The final IPD review will take place after database soft lock. The first 2 IPD review meetings will concentrate on identifying additional data issues that might prevent the team from determining IPDs. The last IPD review meeting will concentrate on identifying any data issues associated with the last batch and overall and also identifying the IPDs for all randomized subjects.

For each IPD review, data listings (in EXCEL spreadsheets) will be generated and distributed to the reviewers one week before the IPD review meeting. (This timeline will be shortened at the final review.) These listings, together with the protocol deviations log provided by the clinical monitor, will be reviewed to identify IPDs.

Programmable IPDs will be identified programmatically and presented in data listings at each interim IPD review and the final review. The IPD review team will consider all cases and decide if any updates to the programming criteria are necessary. These programming criteria will be finalized and approved by the IPD review team prior to database hard lock.

For IPDs that cannot be determined programmatically, manual review of relevant data listings will be the primary method of determination. The determinations will be documented in a tracking document containing at least subject ID, deviation date noted by Clinical Manager, deviation start date, deviation type, deviation description and deviation severity (major, critical). This tracking document will be reviewed and approved by the IPD review team after all data issues discovered during the IPD review have been resolved but prior to database hard lock.

It should be noted that the identification of IPDs related to receiving wrong treatment requires unblinded data. Therefore, no data listing will be provided during the IPD review; these IPDs will be identified programmatically using the approved criteria after database hard lock and treatment unblinding. Similarly, only potential IPDs related to study drug overdose will be identified during the IPD review. IPDs related to SEP-4199 overdose can only be ascertained after treatment unblinding.

The approved manual IPD determinations and programmable IPD criteria will be provided to the IQVIA Biostatistics team. Programmable IPDs will be identified based on the final, unblinded study data using the approved programming criteria. Both programmable and manually determined IPDs will be included in the analysis dataset and used in downstream derivations.

The entire IPD review process will be completed prior to database hard lock.

IPD SPECIFICATIONS

0.1 Did not satisfy important inclusion or exclusion criteria

Subjects who did not satisfy one or more of the following eligibility criteria at screening and/or at baseline will be considered IPDs whether or not an exemption was granted during the study. Subjects who violated one or more of the below criteria will be identified programmatically based on the data recorded on the Inclusion/Exclusion Criteria CRF pages, and presented in a data listing.

In addition, supporting data are available to facilitate programmatic or manual identification of violations of the above criteria. These supporting data will also be presented in data listings.

INCLUSION CRITERIA

I.1 Subject is 18 to 65 years of age, inclusive, at the time of informed consent with bipolar I disorder, current episode depressed with or without rapid cycling disease course (≥ 4 episodes of mood disturbance but < 8 episodes in the previous 12 months) with or without psychotic features (diagnosed by DSM-5 criteria, and confirmed by the SCID-5-CT). The current episode of major depression associated with bipolar I disorder must be confirmed by the Investigator and noted in the source records.

I.2 Subject provides written informed consent and is willing and able to comply with the protocol in the opinion of the investigator.

I.3 Subject or legally acceptable representative must possess an educational level and degree of understanding of English or the local language that enables them to communicate suitably with the Investigator and the study coordinator.

I.4 Subject must have a lifetime history of at least one bipolar manic or mixed manic episode. It is strongly recommended that a reliable informant (eg, family member or caregiver) be available to confirm this history.

I.5 Subject's current major depressive episode is ≥ 4 weeks and less than 12 months in duration at Screening.

I.6 Subject has a MADRS total score ≥ 22 at both Screening and Baseline.

I.7 Subject has an YMRS total score ≤ 12 at Screening.

I.8 Female subjects of childbearing potential must have a negative serum ß-hCG test at Screening.

I.9 Females who participate in this study must be one of the following:

- Unable to become pregnant (eg, postmenopausal, surgically sterile, etc)
- Practicing abstinence or part of an abstinent lifestyle

- Using and will continue to use a highly effective form of birth control for at least 28 days prior to administration of the first dose of study drug, during the treatment period, and 2 months after completion or premature discontinuation from the study drug.

I.10 Male subjects with partners of child bearing potential must be practicing abstinence, part of an abstinent life style or using protocol-specified methods of birth control. See Section 10.4 for further information on acceptable methods of birth control.

I.11 Subject is in good physical health on the basis of medical history, physical examination, and laboratory screening.

I.12 Subjects with type 2 diabetes are eligible for study inclusion only if all of the following conditions are met within 30 days prior to Screening:

- Subject's random screening glucose is < 200 mg/dL (11.1 mmol/L).
- Subject's Hemoglobin A1c (HbA1c) ≤ 7.0%.
- If a subject is currently being treated with oral anti-diabetic medication(s), the dose must have been stable for at least 30 days prior to screening. Such medication may be adjusted or discontinued during the study, as clinically indicated.
- Subject has not required hospitalization for diabetes or related complications in the past 12 months.
- Note: Subjects with type 2 diabetes that is newly diagnosed during screening are ineligible for the study.

I.13 Subject who requires concomitant medication treatment with the following agents may be included if they have been on stable doses for the specified times: 1) oral hypoglycemics must be stabilized for at least 30 days prior to baseline; 2) thyroid hormone replacement must be stable for at least 90 days prior to baseline; 3) anti-hypertensive agents must be stable for at least 30 days prior to baseline. The subject's medical condition should be deemed clinically stable following consultation with the Medical Monitor as needed.

INCLUSION CRITERIA – JAPAN SPECIFIC

I.3 Subject provides written informed consent and is willing and able to comply with the protocol in the opinion of the investigator. If the subject is considered a minor per local regulations at the time of collection of the informed consent, written consent will be obtained from a legally acceptable representative (guardian) in addition to that obtained from the subject. If the subject is hospitalized involuntarily, written consent will be obtained from a legally acceptable representative (guardian) in addition to that obtained from the subject.

I.12 Subjects with type 2 diabetes are eligible for study inclusion only if all of the following conditions are met:

- Subject's random screening glucose is < 200 mg/dL (11.1 mmol/L) at Screening.
- Subject's Hemoglobin A1c (HbA1c) <= 7.0% at Screening.
- If a subject is currently being treated with oral anti diabetic medication(s), the dose must have been stable for at least 30 days prior to Screening. Such medication may be adjusted or discontinued during the study, as clinically indicated.
- Subject has not required hospitalization for diabetes or related complications in the 12 months prior to Screening.
- Subjects with type 2 diabetes that is newly diagnosed during screening are ineligible for the study.

I.13 Subject who requires concomitant medication treatment with the following agents may be included if they have been on stable doses for the specified times: 1) Oral hypoglycemics must be stabilized for at least 30 days prior to Screening. 2) Thyroid hormone replacement must be stable for at least 90 days prior to Screening. 3) Anti hypertensive agents must be stable for at least 30 days prior to Screening. The subject's medical condition should be deemed clinically stable following consultation with the Medical Monitor as needed.

INCLUSION CRITERIA – POLAND SPECIFIC

I.9 Females who participate in this study must be one of the following:

- Unable to become pregnant (eg, postmenopausal, surgically sterile, etc)
- Using and will continue to use a highly effective form of birth control for at least 28 days prior to administration of the first dose of study drug, during the treatment period, and 2 months after completion or premature discontinuation from the study drug

EXCLUSION CRITERIA

E.1 Subject has a lifelong history or presence of symptoms consistent with a major psychiatric disorder other than bipolar I disorder as defined by DSM-5. Exclusionary disorders include but are not limited to moderate to severe alcohol use disorder (within past 12 months), substance use disorder (other than nicotine or caffeine) within past 12 months, bipolar II disorder, schizoaffective disorder, obsessive compulsive disorder, posttraumatic stress disorder.

E.2 Subject demonstrates a decrease (improvement) of ≥ 25% in MADRS total score from Screening to Baseline, or subject's MADRS total score is < 22 at Baseline.

E.3 Subject has received treatment with antidepressants within 3 days of randomization, fluoxetine at any time within 28 days, an MAO inhibitor within 21 days or clozapine within 120 days. All other psychotropic medications with the exceptions of lorazepam, temazepam, eszopiclone, zopiclone, zolpidem, and zolpidem CR require 3 days minimum washout. Depot neuroleptics must be discontinued at least one treatment cycle prior to randomization.

E.4 Subject has suspected/confirmed Borderline Personality Disorder.

E.5 Subject currently has a clinically significant neurological, metabolic (including type 1 diabetes), hepatic, renal, hematological, pulmonary, cardiovascular, gastrointestinal, and/or urological disorder such as unstable angina, congestive heart failure (uncontrolled), or central nervous system (CNS) infection that would pose a risk to the subject if they were to participate in the study or that might confound the results of the study. Subjects with a known history of HIV seropositivity will be excluded.

E.6 Subject has evidence of any chronic organic disease of the CNS such as tumors, inflammation, active (or history of) seizure disorder, vascular disorder, Parkinson's disease, Alzheimer's disease or other forms of dementia, myasthenia gravis, or other degenerative processes. In addition, subjects must not have a history of intellectual disability or persistent neurological symptoms attributable to serious head injury. Past history of febrile seizure, is not exclusionary.

E.7 Subject has a history of malignancy < 5 years prior to signing the informed consent, except for adequately treated basal cell or squamous cell skin cancer or in situ cervical cancer. Subjects with pituitary tumors of any duration are excluded.

E.8 Subject demonstrates evidence of acute hepatitis, clinically significant chronic hepatitis, or evidence of clinically significant impaired hepatic function through clinical and laboratory evaluation (use screening values for laboratory evaluation). Subject has a history of stomach or intestinal surgery or any other condition that could interfere with absorption, distribution, metabolism, or excretion of medications.

E.9 Subject has knowledge of any kind of cardiovascular disorder/condition known to increase the possibility of QT prolongation or history of additional risk factors for torsade de pointes (eg, heart failure, hypokalemia, family history of Long QT Syndrome or Brugada Syndrome) or cardiac conduction disorders, or requires treatment with an antiarrhythmic medication.

E.10 Subject has family history of QTc prolongation or of unexplainable sudden death at < 50 years of age.

E.11 Abnormal 12-lead ECG at Screening, including:

- QTcF > 450 ms (male subjects) or > 470 ms (female subjects)
- QRS > 110 ms
- PR > 200 ms
- Second- or third-degree atrioventricular block

- Any rhythm other than sinus rhythm, which is interpreted by the Investigator to be clinically significant

E.12 Subject has a history of neuroleptic malignant syndrome (NMS).

E.13 Subject exhibits evidence of severe tardive dyskinesia, severe dystonia, or any other severe movement disorder. Severity is to be determined by the investigator.

E.14 Subject has been diagnosed with type 1 diabetes, or insulin-dependent diabetics.

E.15 Subject who has any abnormal laboratory parameter at screening that indicates a clinically significant medical condition as determined by the investigator.

- Subjects with fasting blood glucose at screening ≥ 126 mg/dL (7.0 mmol/L) will be excluded from the study.
- Subjects with fasting blood glucose from 100-125 mg/dL (5.6-6.9 mmol/L) may enter the study based on the approval of the Medical Monitor.
- Subjects who are found to have been non-fasting at Screening may be allowed if their blood glucose is < 200 mg/dL. Subjects with random (nonfasting) blood glucose at screening ≥ 200 mg/dL (11.1 mmol/L) must be retested in a fasted state.
- Subjects with HbA1c > 7.0% will be excluded.

E.16 Subject has a prolactin concentration > 100 ng/mL at screening or have a history of pituitary adenoma.

E.17 Subject has a body mass index (BMI) ≥ 40 or < 18 kg/m2 at Screening.

E.18 Subject has a history of non-response to an adequate (6-week) trial of three or more antidepressants (with or without mood stabilizers) during the current episode.

E.19 Subject is considered by the Investigator to be at imminent risk of suicide or injury to self, others, or answers “yes” to “Suicidal Ideation” item 4 (active suicidal ideation with some intent to act, without specific plan) or item 5 (active suicidal ideation with specific plan and intent) on the C-SSRS assessment at the Screening visit (in the past month [30 days]) or Baseline.

E.20 Subject tests positive for drugs of abuse at screening or baseline. In the event a subject tests positive for cannabinoids (tetrahydrocannabinol), the investigator will evaluate the subject’s ability to abstain from cannabis during the study. This information will be discussed with the Medical Monitor for study enrollment consideration.

E.21 Subject has a history of hypersensitivity to more than two distinct chemical classes of drug (eg, sulfas and penicillins).

E.22 Subjects have received depot neuroleptics unless the last injection was at least one treatment cycle before randomization.

E.23 Subject requires treatment with a drug that consistently prolongs the QTc interval.

E.24 Subject has received ECT within 90 days prior to randomization or is expected to require ECT during the study course.

E.25 Subject is currently participating, or has participated in a study with an investigational or marketed compound or device within 6 months prior to signing the informed consent, or has participated in 3 or more studies within 18 months prior to signing the informed consent.

EXCLUSION CRITERIA – JAPAN SPECIFIC

E.3 Subject has received treatment with antidepressants within 3 days of randomization, fluoxetine at any time within 28 days, an MAO inhibitor within 21 days or clozapine within 120 days. All other psychotropic medications with the exceptions of lorazepam, eszopiclone, zopiclone, and zolpidem require 3 days minimum washout (see Section 10.3.3 for lorazepam, eszopiclone, zopiclone, and zolpidem treatment restrictions). Depot neuroleptics must be discontinued at least one treatment cycle prior to randomization.

E.19 Subject is considered by the Investigator to be at imminent risk of suicide or injury to self, others, or answers “yes” to “Suicidal Ideation” item 4 (active suicidal ideation with some intent to act, without specific plan) or item 5 (active suicidal ideation with specific plan and intent) on the C SSRS assessment at the Screening visit (in the past 3 months) or Baseline.

Partial or completely missing data will be handled as follows:

- I-6 and I-7: If the MADRS or YMRS score is missing at screening or baseline, the subject will be considered to have violated the eligibility criteria and hence an IPD.

- I-5: If the date of initial onset of current major depressive episode, the date will be imputed and the time since onset will be calculated using the algorithms described in APPENDIX 2.If the date is completely missing, the subject will be considered to have violated the eligibility criteria and hence an IPD.

0.2 Received disallowed concomitant medication

Concomitant medications are defined as any medications taken during the course of the study, with a start date on or after the date of the first dose of study drug and on or before the date (day) of the last dose of study drug; or with a start date prior to, and an end date on or after, the date of the first dose of study drug, or marked as ongoing. Medications that ended prior to the date of the first dose of study drug will be considered prior medications. Medications that started after the date of the last dose of study drug will not be considered concomitant, but will be considered post-treatment.

Protocol section 10.3 discusses the requirements on prior and concomitant medications.

All medications taken by the randomized subjects under review will be provided in a data listing. To facilitate the review, the listing will present the ATC text of *all* ATC levels. In addition, preferred drug name, reported drug name (verbatim), indication, dose level, frequency, route, the start and end dates, and a flag for the psychotropic medications *based on investigator assessment* will be included. The listing will also include flags for whether a medication is prior, concomitant, or post-treatment, and whether or not a medication started prior to the first dose of study drug.

The disposition data of all randomized subjects under review will also be provided in a data listing.

The reviewers will manually review both listings to identify any IPDs as a result of the use of disallowed

concomitant medication(s) as specified in the protocol.

0.3 Received wrong treatment or was overdosed (only after unblinding)

A subject is considered to have received wrong treatment if they were randomized to study drug **A** but instead received study drug B at some point during the study. Any subjects who received wrong treatment on 1 or more days during the double-blind treatment period will be considered IPDs. These subjects will be identified programmatically after study data unblinding.

The drug dispensing record based on the Study Drug Administration / Drug Accountability CRF page will be presented in a data listing. The number of tablets taken from each blister card will be calculated, as well as the number of tablets should have been taken from the card. If for a given card the number taken is larger than the number should have been taken, that card will be flagged for study drug overdose. Additional 'study drug overdose' incidences may be identified from the protocol deviations log and/or investigator comments.

The IPD review team will determine if any of the incidences identified above constitutes a potential IPD. These determinations are made without knowledge of which treatment (placebo or SEP-4199) the subject was actually overdosed with.

After unblinding, those incidences that were determined to be a potential IPD with the treatment involved being SEP-4199 will be considered IPDs.

0.4 Overall double-blind compliance rate < 75% or > 125%

The overall compliance rate of the double-blind treatment period is calculated as: (number of tablets taken / number of tablets should have been taken) × 100%. Subjects whose overall compliance rate is < 75% or > 125% will be identified programmatically. These subjects will be considered IPDs. Subjects whose overall compliance rate cannot be calculated will not be considered IPDs.

0.5 Received benzodiazepines, sedatives or hypnotics within 8 hours of a subsequent efficacy assessment

Protocol section 10.3.3 specifies: "Benzodiazepines, sedatives or hypnotics should not be administered within 8 hours prior to any psychiatric assessments."

For a given study visit, the date and time of the last dose of benzodiazepines, sedatives or hypnotics taken immediately before the visit, and the collection start date and time of efficacy assessments at the visit, will be used to identify subjects who took any benzodiazepines, sedatives or hypnotics medications within 8 hours of a subsequent efficacy assessment. Such cases will be identified programmatically and flagged in a data listing. The IPD review team will determine whether any of these cases constitutes an IPD.

0.6 Treatment unblinded during double-blind period

Subjects whose treatment was unblinded during the double-blind period and who completed additional assessments on MADRS or CGI-BP-S after unblinding will be considered IPDs.

Subjects whose treatment was unblinded will be identified based on the Subject Disposition – End of Study CRF page and presented in a data listing. The date of unblinding will be based on the IRT data transfer. Those subjects who have MADRS or CGI-BP-S data collected on or after the date of unblinding will be identified programmatically and flagged in the listing. The complete MADRS and CGI-

S assessment records for all unblinded subjects will be provided in a separate listing.

0.7 Tested positive for urine drug screen

All urine drug screen results will be presented in a data listing. The IPD review team will determine for each positive result whether this constitutes an IPD.

0.8 Missing laboratory results before randomization

All subjects with a complete missing laboratory category evaluation (either Chemistry, Hematology, or Toxicology) will be presented in a data listing. The IPD review team will determine for each missing assessment before randomization whether this constitutes an IPD.

APPENDIX 8. **EQ-5D-5L INDEX VALUE**

EQ-5D-5L health states, defined by the EQ-5D-5L descriptive system, will be converted into a single index value using an appropriate value set. For international trials, EuroQol recommends applying a single value set or crosswalk to all study sites. US value set will be used for US, Europe, and Japan for this study. Please refer the following EuroQol website for EQ-5D-5L value sets and further information.

https://euroqol.org/publications/key-euroqol-references/value-sets/

Document Approval Signature(s)

This document has been signed electronically in compliance with IQVIA and IQVIA Companies' electronic signature policies and procedures. This page is a manifestation of the electronic signatures, which may be confirmed within the IQVIA and IQVIA Companies' source system in accordance with IQVIA and IQVIA Companies' records retention policies. Note: the Title field in the signature block has been intentionally left blank.

UserName: Sansas, Benoit (q1010574)
Title:
Date: Friday, 22 May 2020, 03:29 AM Eastern Time
Meaning: I Approve the Document
==========