Official Title: A Randomized, Double-Blind, Placebo-Controlled Study Evaluating the Efficacy and Safety of Sage-217 in the Treatment of Adults With Severe Postpartum Depression

NCT Number: NCT04442503

Document Dates: SAP Version 1.0: 13 April 2022

9. STATISTICAL METHODS INTERIM ANALYSIS PLAN

The following statistical analysis plan is provided:

Statistical analysis plan Version 1.0, 11 April 2022

SAGE THERAPEUTICS INCORPORATED

Statistical Analysis Plan

Methods

Protocol Number 217-PPD-301

A RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED STUDY EVALUATING THE EFFICACY AND SAFETY OF SAGE-217 IN THE TREATMENT OF ADULTS WITH SEVERE POSTPARTUM DEPRESSION

Author of SAP: [redacted]

Version: Final Version 1.0

Version Date of SAP: 11 April 2022

Sponsor: Sage Therapeutics, Inc.
215 First Street
Cambridge, Massachusetts 02142

Authorization Signature Page

A Randomized, Double-Blind, Placebo-controlled Study Evaluating the Efficacy and Safety of SAGE-217 in the Treatment of Adults with Severe Postpartum Depression

Author:

DocuSigned by:
[redacted]
Signer Name: [redacted]
Signing Reason: I approve this document
Signing Time: 12-Apr-2022 | 10:05 EDT
[redacted]

12-Apr-2022 | 10:05 EDT

[redacted], M.S., [redacted]
Sage Therapeutics, Inc.

Date

Approved by:

DocuSigned by:
[redacted]
Signer Name: [redacted]
Signing Reason: I approve this document
Signing Time: 12-Apr-2022 | 10:51 EDT
[redacted]

12-Apr-2022 | 10:51 EDT

[redacted], MD, MBA, [redacted]
Sage Therapeutics, Inc.

Date

DocuSigned by:
[redacted]
Signer Name: [redacted]
Signing Reason: I approve this document
Signing Time: 13-Apr-2022 | 11:54 EDT
[redacted]

13-Apr-2022 | 11:54 EDT

[redacted]
Sage Therapeutics, Inc.

Date

DocuSigned by:
[redacted]
Signer Name: [redacted]
Signing Reason: I approve this document
Signing Time: 12-Apr-2022 | 09:24 EDT
[redacted]

12-Apr-2022 | 09:25 EDT

[redacted], PhD, [redacted]
Sage Therapeutics, Inc.

Date

TABLE OF CONTENTS

TABLE OF CONTENTS 3
LIST OF TABLES 5
1. LIST OF ABBREVIATIONS 6
2. INTRODUCTION 7
3. STUDY OBJECTIVES 8
3.1. Primary Objective 8
3.2. Secondary Objective 8
[redacted] [redacted] 8
4. STUDY ENDPOINTS 8
4.1. Efficacy Endpoints 8
4.1.1. Primary Efficacy Endpoint 8
4.1.2. Secondary Efficacy Endpoints 8
4.1.2.1. Key Secondary Efficacy Endpoints 8
4.1.2.2. Other Secondary Endpoints 8
[redacted] [redacted] 9
5. STUDY DESIGN 9
5.1. Overall Design 9
5.2. Sample Size and Power 10
5.3. Randomization 10
5.4. Blinding and Unblinding 11
6. MODIFICATIONS 11
6.1. Modifications to the Approved Clinical Study Protocol 11
6.2. Modifications to the Approved Statistical Analysis Plan 11
6.3. Modifications to the Approved DMC Charter 12
7. ANALYSIS SETS 12
7.1. Randomized Set 12
7.2. Full Analysis Set 12
7.3. Safety Set 12
7.4. Per Protocol Set 12
[redacted] [redacted] 12
8. STATISTICAL ANALYSIS 12

8.1. General Considerations 12
8.1.1. Study Day Definition 14
8.1.2. Missing Data 14
8.2. Background Characteristics 14
8.2.1. Participant Disposition 14
8.2.2. Protocol Deviations 15
8.2.3. Demographics and Baseline Characteristics 15
8.2.4. Medical/Surgical History 17
8.2.5. Prior and Concomitant Medications 17
8.2.6. Investigational Product Exposure 18
8.2.7. Investigational Product Adherence 19
8.3. Efficacy Analysis 20
8.3.1. Definition of Efficacy Variables 20
8.3.1.1. Hamilton Rating Scale for Depression (HAM-D) 20
8.3.1.2. Clinical Global Impression – Improvement (CGI-I) 21
8.3.1.3. Clinical Global Impression – Severity (CGI-S) 22
8.3.1.4. Hamilton Anxiety Rating Scale (HAM-A) 22
8.3.1.5. Montgomery-Åsberg Depression Rating Scale (MADRS) 22
8.3.1.6. Edinburgh Postnatal Depression Scale (EPDS) 22
8.3.1.7. Patient Health Questionnaire (PHQ-9) 23
8.3.2. Visit Windows 23
8.3.3. Analysis of Efficacy Variable(s) 24
8.3.3.1. Mixed Effects Model for Repeated Measures 26
8.3.3.2. Generalized Estimating Equation (GEE) Models 28
8.3.3.3. Sensitivity Analysis 28
8.3.3.4. Analysis of Time to First HAM-D Response/Remission 31
8.3.3.5. Multiplicity Adjustment for Key Secondary Endpoints 32
8.3.4. Characterization of Durability of SAGE-217 Treatment Effect 32
8.3.4.1. Durability of Clinically Meaningful Treatment Effect for SAGE-217 .. 32
8.3.4.2. Durability of Treatment Effect at Day 45 via Statistical Comparison of SAGE-217 Versus Placebo 33
8.3.5. Anxious Subgroup Analysis 34

8.4. Safety Analysis 34
8.4.1. Adverse Events 35
[redacted] 37
[redacted] 40
[redacted] 42
8.4.5. Physical Examination 42
[redacted] 42
[redacted] 44
[redacted] 44
[redacted] 44
9. SUMMARY OF INTERIM AND DMC ANALYSES 44
10. REFERENCES 44
11. LIST OF APPENDICES 45
11.1. Appendix A: Schedule of Assessments 45
11.2. Appendix B: Details of Statistical Methodology 48
11.3. Appendix C: Handling of Missing Dates 50
11.4. Appendix D: List of Displays 53

LIST OF TABLES

Table 1: Diagnostic Screening Test 16
Table 2: HAM-D Subscale Calculation 20
Table 3: Visit Windows for Efficacy Analysis 24
Table 4: Safety Endpoints and Variables in the Summary Tables 35
[redacted] 37
[redacted] 39
[redacted] 41
[redacted] 42

1. LIST OF ABBREVIATIONS

Abbreviation or specialist term	Explanation
AE	adverse event
AR	Autoregressive
ATC	anatomical therapeutic chemical
BMI	body mass index
CGI-I	Clinical Global Impression scale for improvement
CGI-S	Clinical Global Impression scale for severity
COVID-19	A syndrome caused by infection with the SARS-CoV-2 virus
CS	Compound Symmetry
[redacted]	
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
[redacted]	
eCRF	electronic case report form
EOT	end of treatment
EPDS	Edinburgh Postnatal Depression Scale
ET	early termination
FAS	Full Analysis Set
FSH	follicle stimulating hormone
GEE	generalized estimating equation
HAM-A	Hamilton Rating Scale for Anxiety
HAM-D	Hamilton Depression Rating Scale
hCG	human chorionic gonadotropin
ICF	informed consent form
IRT	interactive response technology
IP	Investigational product
[redacted]	
MADRS	Montgomery-Åsberg Depression Rating Scale
MMRM	mixed effects model for repeated measures
MedDRA	Medical Dictionary for Regulatory Activities

Abbreviation or specialist term	Explanation
NCS	not clinically significant
PCS	Potentially clinically significant
[redacted]	
PHQ-9	Patient health Questionnaire
[redacted]	
PPD	Postpartum depression
PRO	patient-reported outcome
PT	preferred term
[redacted]	
SAE	serious adverse event
SAP	statistical analysis plan
SCID-5-CT	Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition for clinical trials
SD	standard deviation
SE	standard error
SI	International System of Units
SOC	System Organ Class
TEAE	treatment-emergent adverse event
UN	Unstructured
WHO	World Health Organization

2. INTRODUCTION

This statistical analysis plan (SAP) is for the final analysis of data from 217-PPD-301 study, and is based on clinical study protocol, version 2.0, dated 29 JAN 2021.

The purpose of the SAP is to describe in detail the statistical methodology and analyses to be conducted for 217-PPD-301 protocol. The SAP will be approved and finalized before database lock.

Any changes made to the SAP after the clinical database lock has occurred but before treatment unblinding will be documented and discussed in the clinical study report for this study. No changes can be made to the SAP after treatment unblinding.

3. STUDY OBJECTIVES

3.1. Primary Objective

The primary objective of Study 217-PPD-301 is to determine if treatment with SAGE-217 reduces depressive symptoms in adults with severe postpartum depression (PPD) compared to placebo.

3.2. Secondary Objective

The secondary objectives of Study 217-PPD-301 are:

- To determine if treatment with SAGE-217 reduces anxiety symptoms compared to placebo
- To assess self-report of depressive symptoms
- To evaluate the safety and tolerability of SAGE-217

[redacted]

4. STUDY ENDPOINTS

4.1. Efficacy Endpoints

4.1.1. Primary Efficacy Endpoint

The primary efficacy endpoint is the change from baseline in the 17-item Hamilton Rating Scale for Depression (HAM-D) total score at Day 15. Details of the estimand are provided in Section 8.3.3.1 .

4.1.2. Secondary Efficacy Endpoints

4.1.2.1. Key Secondary Efficacy Endpoints

- Change from baseline in HAM-D total score at Day 3
- Change from baseline in HAM-D total score at Day 28
- Change from baseline in HAM-D total score at Day 45
- Change from baseline in Clinical Global Impressions – Severity (CGI-S) score at Day 15

4.1.2.2. Other Secondary Endpoints

- HAM-D response at Day 15 and Day 45

- HAM-D remission at Day 15 and Day 45
- Clinical Global Impressions – Improvement (CGI-I) response at Day 15
- Change from baseline in Hamilton Anxiety Rating Scale (HAM-A) total score at Day 15
- Change from baseline in Montgomery-Åsberg Depression Rating Scale (MADRS) total score at Day 15
- Change from baseline in HAM-D subscale at Day 15
- Change from baseline in self-reported measures of depressive symptoms, as assessed by the Edinburgh Postnatal Depression Scale (EPDS) total score and 9-item Patient Health Questionnaire (PHQ-9)
- Incidence of treatment-emergent adverse events (TEAEs)

[redacted]

- [redacted]
- [redacted]

5. STUDY DESIGN

5.1. Overall Design

Study 217-PPD-301 is a randomized, double-blind, placebo-controlled, parallel group study of the efficacy and safety of SAGE-217 in adults diagnosed with severe PPD. Participants, site staff, and sponsor personnel will be masked to treatment allocation.

This study will consist of a Screening Period of up to 28 days, a 14-day double-blind Treatment Period, and a Follow-up Period through Day 45.

The diagnosis of depression must be determined according to the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders-5 (DSM-5) Axis I Disorders (SCID-5-CT) performed by a qualified healthcare professional. Participants will undergo preliminary screening procedures to determine eligibility. A full personal medical history will be taken including documentation of all major depression episodes, other Axis I, and any prior PPD episodes. In addition, PPD episodes experienced by immediate female family members, including siblings, parents, and grandparents will be documented.

Psychotropic medications are permitted provided participants are on a stable dose for at least 30 days prior to Day 1 and agree to continue a stable dose through the Day 45 assessments. Initiation of new psychotropic medications that may potentially have an impact on efficacy

and/or safety endpoints will not be allowed within 30 days prior to Day 1 through completion of the Day 45 assessments.

On Day 1, eligible participants will be stratified based on antidepressant treatment use at baseline (Yes = current and stable; No = not treated or withdrawn (≥30 days or >5 half-lives)) and randomized within each stratum to 1 of 2 treatment groups (SAGE-217 50 mg or matching placebo) in a 1:1 ratio.

Participants will self-administer investigational product (IP) once daily at approximately 8 PM with fat-containing food (eg, within 1 hour of an evening meal which contains fat, or with a fat-containing snack), on an outpatient basis, for 14 days. As local regulations permit, IP administration will be monitored via a medication adherence monitoring platform used on smartphones to visually confirm IP ingestion.

During the treatment period, participants will be able to receive 50 mg IP as long as there are no dose limiting safety/tolerability concerns. Participants who cannot tolerate 50 mg will receive 40 mg for the remainder of the treatment period. At the discretion of the investigator, participants who cannot tolerate the 40-mg dose will be discontinued from IP.

Participants who discontinue IP early should return to the site for the remaining study visits as scheduled (ie, relative to the first dose of IP) unless the participant withdraws consent or loses the capacity to grant consent. If a participant withdraws from the study/stops study participation early, an early termination visit should be conducted, if possible.

5.2. Sample Size and Power

Using a 2-sided test at an alpha level of 0.05, a sample size of approximately 86 evaluable participants per treatment group would provide 90% power to detect a placebo-adjusted treatment difference of approximately 4 points in the primary endpoint, change from baseline in HAM-D total score at Day 15 assuming standard deviation of 8 points.

Assuming a 10% dropout and a 1:1 randomization ratio within each stratum (antidepressant treatment use at baseline, yes or no), approximately 192 randomized participants (96 per treatment group) will be required to obtain 86 evaluable participants per treatment group. Evaluable participants are defined as those randomized participants who receive IP and have a valid baseline and at least 1 postbaseline HAM-D assessment. Additional participants may be randomized if the dropout rate is higher than 10%.

5.3. Randomization

Participants who meet the eligibility criteria will be randomized in a stratified manner based on antidepressant treatment use at baseline (Yes = current and stable; No = not treated or withdrawn (≥30 days or > 5 half-lives)) at baseline; Randomization will be done within each stratum in a 1:1 ratio to receive SAGE-217 50 mg or matched placebo.

Randomization schedules will be generated by an independent statistician. The allocation to treatment group (SAGE-217 50 mg, or placebo) will be based on the randomization schedule. Randomization will be performed centrally via an interactive response technology (IRT) system.

5.4. Blinding and Unblinding

The Sponsor, site personnel and participants will remain blinded until the database lock after all participants have completed the Day 45 visit.

During the study, the blind is to be broken only when the safety of a participant is at risk and the treatment plan is dependent on the study treatment received. Unless a participant is at immediate risk, the Investigator should make diligent attempts to contact Sage prior to unblinding the study treatment administered to a participant. Requests from the Investigator about the treatment administered to study participants should be discussed with the Sage Medical Monitor. If the unblinding occurs without Sage's knowledge, the Investigator must notify Sage within 24 hours of breaking the blind.

All circumstances surrounding a premature unblinding must be clearly documented in the source records.

In all cases where the IP allocation for a participant is unblinded, pertinent information (including the reason for unblinding) must be documented in the participant's records and on the eCRF. At the time of withdrawal from the study/stopping participation, if possible, an EOT and/or ET visit should be conducted.

If a participant becomes unblinded to the participant's treatment assignment before database lock, the participant will be excluded from the Per Protocol Set (see Section 7.4), but will be included in the Full Analysis Set (FAS) (see Section 7.2).

6. MODIFICATIONS

6.1. Modifications to the Approved Clinical Study Protocol

The following statistical considerations have been added to the analysis plan compared to the statistical section of the Clinical Study Protocol, version 2.0, dated 29 Jan 2021

1) Change from baseline in HAM-D total score at Day 3, 28, and 45, and change from baseline in CGI-S at Day 15 prespecified as key secondary endpoints and testing sequence defined
2) Multiplicity adjustment procedure for strong control of type I error rate described
3) The FAS definition is updated as 'all randomized participants administered IP with valid baseline and at least 1 post-baseline total score in at least one of HAM-D, HAM-A, MADRS, CGI-S, EPDS and PHQ-9, or at least 1 post-baseline value of CGI-I. In the protocol, it is at least 1 post-baseline HAM-D assessment
4) MADRS response and remission were added
5) Analyses of durability of treatment effect were added

6.2. Modifications to the Approved Statistical Analysis Plan

This is first version of the SAP.

6.3. Modifications to the Approved DMC Charter

Not applicable

7. ANALYSIS SETS

7.1. Randomized Set

The Randomized Set is defined as all participants who are randomized.

7.2. Full Analysis Set

The FAS is defined as all randomized participants who are administered IP with valid baseline total score and at least 1 post-baseline total score in at least one of HAM-D, HAM-A, MADRS, CGI-S, EPDS and PHQ-9, or at least 1 post-baseline value of CGI-I.

7.3. Safety Set

The Safety Set is defined as all participants who are administered IP.

7.4. Per Protocol Set

The Per Protocol Set is defined as all participants in the FAS without any major protocol deviations that could affect efficacy. For further details, see Section 8.2.2 .

In addition, the Per Protocol Set will also exclude FAS participant with any of the following conditions:

1. Participants who consumed <22 capsules (ie, <80% of assigned number of capsules)
2. Participants who consumed incorrect IP (ie, IP other than that to which they were randomized to receive) at any time during the study
3. Participants or study personnel who were unblinded to participant's treatment assignment before database lock

[redacted]

8. STATISTICAL ANALYSIS

8.1. General Considerations

Unless otherwise specified, continuous endpoints will be summarized with n, mean standard deviation (SD), median, minimum (min) and maximum (max). If the measurements in the source (raw) data are integers, then the corresponding mean and median will be presented to 1 decimal place and the SD to 2 decimal places; If the raw scores are integers and the difference between two treatment groups is small, in result the rounded values for two

treatment groups are equal (for instance, if mean is 0.07 and 0.14 respectively, both groups would have reported as 0.1 even though one number is doubling of another) then two decimal places will be present. If the measurements are obtained to 1 decimal place, then the mean and median will be presented to 2 decimal places and the SD to 3 decimal places and so forth. Minimum and maximum will be displayed as reported in the source (raw) data. LS mean will be presented to 1 decimal place, and standard error (SE) will be presented to 2 decimal places. P-value will be presented to 4 decimal places with p-values less than 0.0001 reported as "<0.0001". P-values larger than 0.9999 will be reported as ">0.9999". In addition, change from baseline values (visit value – baseline value) will be calculated at each time point and summarized descriptively.

For categorical endpoints, descriptive summaries will include counts and percentages. Percentages will be presented to 1 decimal place unless otherwise specified; the denominator of percentages will be the number of participants in the analysis set used unless specified otherwise.

All analyses and summary outputs will be generated using SAS® 9.4 or higher.

All summaries and figures will be provided by treatment group: placebo and SAGE-217. Participants who are randomized to 50 mg treatment and received 40 mg dose due to dose reduction will be summarized under one treatment group. For efficacy data analysis, participants' data are analyzed by as randomized. For safety data analysis, participants' data are analyzed per the actual treatment received, and this is determined as follows: if a participant received any dose of SAGE-217 at any point of time, the participant is assigned to actual treatment of SAGE-217. All participant data, including those derived, to support tables and figures will be presented in the participant data listings. In general, the participant data listings will be sorted by participant number and assessment visit and date (and time, if applicable). The treatment will be identified for each participant.

For the purpose of all safety and efficacy analyses, baseline is defined as the last non-missing measurement including unscheduled visits prior to the first dose of investigational product, unless stated otherwise. If the time of an assessment is collected, baseline will be the latest assessment prior to first dose of IP administration time; if the time of an assessment is not collected, the assessment on Day 1 is assumed to be prior to dosing if the protocol mentions that this assessment needs to be before dosing or it is collected as "pre-dose".

Antidepressant treatment use at baseline (yes=current and stable; no=not treated or withdrawn (≥30 days or > 5 half-lives)) identified by ATC level 3 code of N06A will be used in the subgroup analysis and mixed effects model for repeated measures (MMRM).

[redacted]

8.1.1. Study Day Definition

It is to be noted that the investigational product is administered in the evening with food. The assessments at the clinic on Day 1 are hence before the first dose of investigational product.

Study day will be defined as follows:

- The day of participant receiving the first dose of investigational product is designated as Day 1.
- For visit days after Day 1, study day = visit date – Day 1 date + 1.
- For visit days prior to Day 1, study day = visit date – Day 1 date. Thus, study days for screening visit are negative numbers. There is no "Day 0".

8.1.2. Missing Data

All participants will be used in the analyses, as per the analysis populations, using all non-missing data available. Missing scheduled visit values may be replaced by windowing unscheduled visit values (see Section 8.3.2). Efficacy analyses will use sensitivity analyses to assess the impact of the assumptions made about missing data. Safety analyses will not impute missing data, unless specified. Imputation of missing data in scoring of questionnaires is discussed in respective sections below. Handling of missing or incomplete dates are described in Section 11.3 Appendix C.

8.2. Background Characteristics

All displays in this section will be presented by the treatment arms of placebo, SAGE-217 and overall.

8.2.1. Participant Disposition

The analyses of participant disposition will use all participants who provided written informed consent to the study.

The summaries of participant disposition will include the number of participants who were screened, number (%) screen failed, randomized, did and did not receive investigational product (IP), the number (%) of participants who completed the study, prematurely withdrew from the study, primary reasons for not completing the study, completed treatment, discontinued IP prematurely, and primary reasons for discontinuing IP.

All percentages will be calculated based on the participants who were randomized and received IP. Treatment arm assignment will be according to the planned treatment. If a participant is rescreened because the participant has been a screen failure the first time, the status of the participant will be determined from the second screening. In the count of screened participants, this participant will be counted only once.

A completer for the study is defined as one who completed the last follow up visit based on the study completion CRF page with the completion question answered Yes.

A participant is marked as completing the treatment if the complete treatment question on the study treatment completion CRF page is answered Yes.

The number of participants in each analysis set will be provided. Using Randomized Set, the reason for not being included in other analysis sets will be summarized.

A separate data listing will be provided for all participants who prematurely discontinued IP or prematurely withdrew from the study with reasons, number of days on investigational product, date of withdrawal from the study, using the Safety Set.

8.2.2. Protocol Deviations

Protocol deviations identified during the study will be captured and categorized by the study review team as major and minor deviations without any unblinding information on an ongoing basis.

All major protocol deviations will be summarized by randomized treatment using the FAS.

The study team will identify the major protocol deviations related to efficacy or that may have impact on efficacy to determine the participants in the FAS to be excluded for the Per Protocol Set prior to database lock in a blinded fashion (some major protocol deviations may not lead to participants' data to be excluded from the Per Protocol Set).

In addition, COVID-19 related protocol deviations such as remoted telephone/video visit/assessment, home healthcare visit, missed visit/assessment, out of window visit/assessment, safety reporting, investigational product administration, and other will be documented and provided as a separate listing.

All protocol deviations (major and minor) will be included in a data listing for randomized participants. Randomized treatment will be used.

Any violation of inclusion/exclusion criteria will be presented with the randomized treatment in a separate data listing using the Randomized Set.

8.2.3. Demographics and Baseline Characteristics

The following analyses will use the Safety Set (using actual treatment received), and the FAS (using randomized treatment).

Demographic data (age at informed consent [years]), race, ethnicity, employment status, highest education level, marital/civil status) and baseline characteristics, such as height (cm), weight (kg), and body mass index (BMI) (kg/m^2), will be summarized by treatment group (randomized for the FAS and actual for the Safety Set) and overall.

Baseline subgroups will be summarized for the following categories:

- Race (Black or African American, White, Other)
- Age (18-24, 25-45 years)
- Baseline antidepressant use (Yes, No) (for definition of antidepressant, see Section 8.2.5)
- BMI ($\leq$18.4, 18.5-24.9, 25-29.9, $\geq$30 kg/m^2)

- Baseline HAM-D total score (<median of baseline HAM-D total score, >= median of baseline HAM-D total score, median will be calculated on FAS)
- Onset of PPD (3rd trimester vs. postpartum)
- History of PPD (1st episode vs recurrent PPD episode)
- COVID-19 History (not impacted, hospitalization, SARS-CoV-2 test positive, patient isolation, suspected COVID, COVID-19) (refer to Section 8.2.4 for details)
- Depression with elevated anxiety based on baseline HAMD anxiety subscale (>=7 vs. <7 or normalized score ≥39 vs.<39).
- Depression with elevated anxiety based on baseline HAMA total score (>=20, <20).
- Country (US, Rest of World)

A separate listing for participants who were randomized under incorrect stratification of antidepressant use at baseline will be provided. The antidepressant medications are identified from concomitant medication records with Anatomic Therapeutic Classification (ATC) level 3 = N06A. The stratum for antidepressant use at baseline is determined from the coding of antidepressant, comparing the start/end dates of the medication versus the date of first dose of investigational product. Any deviation from the antidepressant stratum recorded at IRT system are part of protocol deviations and will be included in protocol deviation listing.

Diagnostic labs are part of screening. The results of the diagnostic screening test in Table 1 will be provided in a data listing using the Safety Set.

Table 1: Diagnostic Screening Test

	Diagnostic	
Serum	**Urine**	**Breathalyzer**
Participants that are not surgically sterile at screen: serum hCG	Drug screen: including amphetamines, barbiturates, benzodiazepines, cannabinoids, cocaine, opiates, phencyclidine	Alcohol
Participants, if not surgically sterile at Screen: FSH	Participants that are not surgically sterile at all other visits: urine hCG	

Abbreviations: FSH=follicle stimulating hormone; hCG=human chorionic gonadotropin;

8.2.4. Medical/Surgical History

The following analyses will use the Safety Set.

Time since initial diagnosis of PPD, antidepressant usage, any past PPD episodes, and onset of current PPD (3rd trimester, within 4 week of delivery) will be summarized. Time since initial diagnosis of PPD and days since start of current episode will be calculated using: First dose date of the investigational product – date of interest. For imputation of incomplete dates in medical history, please see Section 11.3.

Disease history such as the diagnosis of PPD, ICD-10 code, current or past episode, date of delivery, start date of current PPD episode and Family history of PPD will be listed.

Medical/surgical history collected at screening will be coded using the Medical Dictionary for Regulatory Activities (MedDRA), Version 23.0 or higher. Medical/surgical history data will be summarized by system organ class (SOC) and preferred term (PT). A summary of medical/surgical history that are ongoing at the time of screening will be provided separately. A Participant will be counted once within each level of summarization. Medical/surgical history will also be listed.

In addition, participant's COVID-19 history will be collected via medical history eCRF page and will be coded by coded by MedDRA. COVID-19 history will be summarized by COVID-19 related preferred term (PT), low level term, high level term and reported term, in medical history data. For participants without COVID-19 related PT, the participants are categorized as 'Not impacted'. If a participant had multiple PTs related to COVID-19, then the participant will be counted once under the worst case in the following order: Hospitalization > SARS-CoV-2 test positive > Patient isolation > Suspected COVID-19 > COVID-19.

A participant history of psychiatric disorders and family history of postpartum depression episodes will be summarized and listed.

8.2.5. Prior and Concomitant Medications

The following analyses will use the Safety Set.

All medications taken and procedures undergone during the study will be recorded; in addition, psychotropic medications taken within 6 months prior to screening, and non-psychotropic medications taken within 30 days prior to screening will also be collected. All medications will be coded using World Health Organization-Drug (WHO) Global B3 March 2020 or later.

Medications will be presented according to whether they are being taken prior to and/or during the study (concomitant). Prior medications are defined as those taken prior to the initiation of the start of investigational product. Concomitant medications are defined as those with a start date on or after the first dose of investigational product or those with a start date before the first dose of investigational product that are ongoing or with a stop date on or after the first dose of investigational product. If medication dates are incomplete and it is not clear whether the medication was concomitant, it will be assumed concomitant. For

imputation of missing concomitant medication dates, please refer to Appendix C, Section 11.3. Note that medication taken before the initial dosing of investigational product and continued after the initial dosing will be categorized as a prior medication and separately as a concomitant medication.

Concomitant medications will be further divided by usage period as follows:

- On-treatment concomitant medications are those that have been used any time from start of first dose to the last dose of investigational product.
- Post-treatment concomitant medications are those that have been started after the last dose of investigational product.

Prior and concomitant non-psychotropic medication use will be summarized by anatomical therapeutic chemical (ATC) level 1 and Standard Medication Name. Similar summary tables will be provided for psychotropic medications. Separate but similar summaries will be provided for concomitant medication use for on-treatment and post-treatment periods as defined above. A Participant will be counted once within each level of summarization.

In addition, for prior and concomitant psychotropic medication separate summaries will be provided by ATC level 1 and ATC level 4.

Antidepressants that have been taken at the same dose for at least 30 days prior to the first dose of investigational product are permitted if the participant intends to continue the stable dose through the follow-up period (Day 45). Initiation of new antidepressants or any other medications that may potentially have an impact on efficacy or safety endpoints is prohibited from at least the screening visit and through completion of the Day 45 assessments. Antidepressant medications are identified by ATC level 3 code of N06A. A summary of antidepressant use at baseline and any change in these medications post-baseline (including the follow up period) will be provided.

Concomitant procedures are recorded on a separate eCRF page; this will be presented in a listing by participant. The study day for the end date of the procedure will be provided when a complete end date is available.

Prohibited medications are reviewed by the medical monitor in the study team on an ongoing basis in blinded fashion; any medication identified as prohibited medication intake is captured in the protocol deviations list.

8.2.6. Investigational Product Exposure

The following analyses will use the Safety Set.

Total drug exposure (in mg) is defined as the total investigational product in mg for SAGE-217 that was taken during the study. Total drug exposure for participants randomized to placebo is zero, unless the participant has taken SAGE-217 by mistake due to being provided wrong investigational product kits, in which case the total exposure comes from SAGE-217 exposure. If the participant skips the dose on any of the days, the dose taken is 0 mg.

The kit for 50 mg dose contains two capsules – one for 30 mg, the other for 20 mg. For participants who took only 1 capsule inadvertently for a dosing day of 50 mg, it is assumed that the participant took the higher dose (ie, 30 mg capsule) for the day and will be calculated as such for the total drug exposure and percent of the planned exposure received.

The kit for 40 mg dose contains two capsules, 20 mg each; therefore, taking one capsule will unambiguously be assigned to 20 mg.

Total exposure duration to investigational product (in days) is defined as: Date of last dose – date of first dose + 1. Note that this does not exclude days when the dose has been missed.

Percent of the planned exposure received is defined as the total drug exposure, divided by planned exposure, times 100. For participants who complete the treatment period without dose reduction, planned exposure is 14 days of treatment planned, times 50 mg for participants randomized to SAGE-217. For participants who reduce dose to 40 mg, the planned exposure is the number of days (last date of 50 mg – first dose date + 1), times 50 mg plus the number of days (date of study day 14 – first date of 40 mg + 1), times 40 mg for participants randomized to SAGE-217. For participants who discontinue the treatment early and reduce dose, the planned exposure is the number of days (last date of 50 mg – first dose date + 1), times 50 mg plus the number of days (last date of dose - first date of 40 mg + 1), times 40 mg for participants randomized to SAGE-217.

For participants who discontinued the treatment early without dose reduction, the planned exposure is the number of days (last date of 50 mg – first dose date + 1), times 50 mg.

For participants randomized to placebo, this measure is not applicable.

Total drug exposure, total exposure duration and percent of the planned exposure received will be summarized descriptively. Number and percentage of participants with less than 22 capsules consumed will be provided.

The number and percent of participants who had dose adjustment (ie, reduced, interrupted, withdrawn) and dose modification (ie, missed dose, took an extra day of dosing, took more than the planned dose for dosing day, partial dose, other) will also be summarized descriptively.

A listing of IP administration & IP exposure will be provided. Whether the participant has eaten fat-containing food within the last 1 hour of the IP administration will be included in the IP exposure data listing.

8.2.7. Investigational Product Adherence

The following analyses will use the Full Analysis Set. Investigational product adherence (%) is defined as the number of capsules taken, divided by the number of capsules planned to be taken, times 100.

The schedule of investigational product is two capsules per day. The number of capsules planned to be taken is defined as follows:

1. If the participant discontinues treatment within Day 1 and Day 14 (both inclusive), the number of capsules planned to be taken is the last dose day of investigational product × 2.

2. If the participant does not discontinue treatment, the number of capsules planned to be taken is 28.

Investigational product adherence will be summarized descriptively and listed. Number and percentage of participants with investigational product adherence in categories (<80%, 80-100%, >100%) will be provided. Investigational product accountability will be listed.

8.3. Efficacy Analysis

8.3.1. Definition of Efficacy Variables

The efficacy variables are defined below:

8.3.1.1. Hamilton Rating Scale for Depression (HAM-D)

The 17-item HAM-D will be used to rate the severity of depression in participants already diagnosed as depressed. HAM-D is collected on screening and the clinic visit on Days 1, 3, 8, 15, 21, 28 and 45. The 17-item HAM-D comprises of individual ratings related to the following symptoms: depressed mood (sadness, hopeless, helpless, worthless), feelings of guilt, suicide, insomnia (early, middle, late), work and activities, retardation (slowness of thought and speech; impaired ability to concentrate; decreased motor activity), agitation, anxiety (psychic and somatic), somatic symptoms (gastrointestinal and general), genital symptoms, hypochondriasis, loss of weight, and insight. Each item is scored in a range of 0 to 2 or 0 to 4, with higher scores indicating a greater degree of depression. The score for each item will be summed to compute a total score, which ranges from 0 to 52. If more than 3 individual items are missing a response, the HAM-D total score will not be calculated and will be left as missing. If less than or equal to 3 individual item scores are missing, the missing item scores will be imputed by the mean of all other available item scores, or the maximum possible values for the missing responses, whichever is smaller, to calculate the HAM-D total score.

Four HAM-D subscale scores will be calculated as the sum of the individual rating scores related to each subscale, divided by the total possible score within the subscale, multiplied by 100, and rounded to a whole number. If more than one item is missing or HAM-D total score is missing, the subscale score is left as missing; if one item on a particular subscale is missing but has been imputed for the calculation of total score, the imputed value from total score calculation will be used in subscale score calculation for that item. Following Table 2 describes the subscale score calculation:

Table 2: HAM-D Subscale Calculation

HAM-D Subscales	Items	Calculation
Core	Depressed mood Feeling of guilt Suicide Work and activities Retardation	Sum of the 5-item responses/20 x 100. If more than one item responses are missing or HAM-D total score is missing, leave as missing; otherwise, use the imputed item score used to

HAM-D Subscales	Items	Calculation
		calculate HAM-D total score to calculate the subscale.
Anxiety	Anxiety psychic Anxiety somatic Somatic symptoms gastrointestinal Somatic symptoms general Hypochondriasis Insight	Sum of the 6-item responses/18 x 100. If more than one item responses are missing or HAM-D total score is missing, leave as missing; otherwise, use the imputed item score used to calculate HAM-D total score to calculate the subscale.
Bech-6	Depressed mood Feeling of guilt Work and activities Retardation Anxiety psychic Somatic symptoms general	Sum of the 6-item responses/22 x 100. If more than one item responses are missing or HAM-D total score is missing, leave as missing; otherwise, use the imputed item score used to calculate HAM-D total score to calculate the subscale.
Maier	Depressed mood Feeling of guilt Work and activities Retardation Agitation Anxiety psychic	Sum of the 6-item responses/24 x 100. If more than one item responses are missing or HAM-D total score is missing, leave as missing; otherwise, use the imputed item score used to calculate HAM-D total score to calculate the subscale.

HAM-D response will be defined as having a 50% or greater reduction from baseline in HAM-D total score; only participants who have a non-missing total score of HAM-D at baseline as well as the visit will be considered in HAM-D response evaluations.

HAM-D remission will be defined as having a HAM-D total score of ≤7; if HAM-D total score is missing, remission will not be defined.

For a sensitivity analysis the worst-case scenario imputation will be used, ie, missing values for HAM-D response (remission) will be considered as “No response” (“No remission”).

8.3.1.2. Clinical Global Impression – Improvement (CGI-I)

The Clinical Global Impression - Improvement (CGI-I) employs a 7-point Likert scale to measure the overall improvement in the participant’s condition post-treatment. The Investigator will rate the participant’s total improvement. Response choices include: 1=very much improved, 2=much improved, 3=minimally improved, 4=no change, 5=minimally worse, 6=much worse, and 7=very much worse. The CGI-I is only rated at post-treatment assessments on Days 3, 8, 15, 21, 28 and 45. By definition, all CGI-I assessments are evaluated against baseline conditions. CGI-I response will be defined as having a CGI-I score of “very much improved” or “much improved.” Missing CGI-I at the visit will not be evaluated for response. For a sensitivity analysis the worst-case scenario imputation will be used, ie, missing values for CGI-I response will be considered as “No response”

8.3.1.3. Clinical Global Impression – Severity (CGI-S)

The Clinical Global Impression - Severity (CGI-S) uses a 7-point Likert scale to rate the severity of the participant's illness at the time of assessment, relative to the clinician's past experience with participants who have the same diagnosis. Considering total clinical experience, a participant is assessed on severity of mental illness at the time of rating as 1=normal, not at all ill; 2=borderline mentally ill; 3=mildly ill; 4=moderately ill; 5=markedly ill; 6=severely ill; and 7=extremely ill. CGI-S is collected at Screening, Days 1, 3, 8, 15, 21, 28 and 45.

8.3.1.4. Hamilton Anxiety Rating Scale (HAM-A)

The 14-item HAM-A will be used to rate the severity of symptoms of anxiety. HAM-A is collected during the clinic visit at Days 1, 3, 8, 15, 21, 28 and 45. Each of the 14 items is defined by a series of symptoms, and measures both psychic anxiety (mental agitation and psychological distress) and somatic anxiety (physical complaints related to anxiety). Scoring for HAM-A is calculated by assigning scores of 0 (not present) to 4 (very severe), with a total score range of 0 to 56, where <17 indicates mild severity, 18 to 24, mild to moderate severity, and 25 to 30, moderate to severe severity. The HAM-A total score will be calculated as the sum of the 14 individual item scores. If more than 3 individual items are missing a response, the HAM-A total score will not be calculated and will be left as missing. If less than or equal to 3 individual item scores are missing, the missing item scores will be imputed by the mean of all other available item scores to calculate the HAM-A total score.

8.3.1.5. Montgomery-Åsberg Depression Rating Scale (MADRS)

The MADRS is a 10-item diagnostic questionnaire used to measure the severity of depressive episodes in participants with mood disorders. MADRS is collected during the clinic visit on Days 1, 8, 15, 28 and 45.

Each MADRS item range from 0 to 6; higher MADRS scores indicate more severe depression. The MADRS total score will be calculated as the sum of the 10 individual item scores, which ranges from 0 to 60. If more than 2 individual items are missing, the MADRS total score will not be calculated and will be left as missing. If less than or equal to 2 individual item scores are missing, the missing item scores will be imputed by the mean of all other available item scores to calculate the MADRS total score.

MADRS response will be defined as having a 50% or greater reduction from baseline in MADRS total score; only participants who have a non-missing total score of MADRS at baseline as well as the visit will be considered in MADRS response evaluations. MADRS remission will be defined as having a MADRS total score of ≤10; if MADRS total score is missing, remission will not be defined.

For a sensitivity analysis the worst-case scenario imputation will be used, ie, missing values for MADRS response (remission) will be considered as "No response" ("No remission").

8.3.1.6. Edinburgh Postnatal Depression Scale (EPDS)

The EPDS is a patient-rated depressive symptom severity scale specific to the perinatal period (Cox 1987). The EPDS will be collected in the clinic at Screening and Days 1, 3, 8, 15, 21, 28 and 45.
The scale consists of 10 questions. Items 1, 2, and 4 are scored from 0 to 3. Items 3, 5, 6, 7, 8, 9, and 10 are reverse scored from 3 to 0. After the reverse scoring is done, higher scores indicate more severe symptoms. The EPDS total score will be calculated as the sum of the 10 individual item scores. If more than 2 individual items are missing, the EPDS total score will not be calculated and will be left as missing. If 1 or 2 individual item scores are missing, the missing item scores will be imputed by the mean of all other available item scores. Imputed individual scores will be rounded to the nearest integer.

8.3.1.7. Patient Health Questionnaire (PHQ-9)

The PHQ-9 is a 9-item participant-rated depressive symptom severity scale. It is collected during the clinic visit at Screening and Days 1, 3, 8, 15, 21, 28 and 45. Scoring is based on responses to specific questions, as follows: 0=not at all; 1=several days; 2=more than half the days; and 3=nearly every day.

The PHQ-9 total score will be calculated as the sum of the 9 individual item scores. If more than 1 individual item is missing, the PHQ-9 total score will not be calculated and will be left as missing. If 1 individual item score is missing, the missing item score will be imputed by the mean of all other available item scores to calculate the PHQ-9 total score. The PHQ-9 total score will be categorized as follows: 1 to 4=minimal depression, 5 to 9=mild depression, 10 to 14=moderate depression, 15 to 19=moderately severe depression; and 20 to 27=severe depression.

8.3.2. Visit Windows

The scheduled visits will not be windowed and will be used at nominal visit value for treatment period (Days 3, 8 and 15). Post-treatment period visits (Days 21, 28 and 45) and unscheduled, end-of-treatment (EOT) and early termination (ET) visit will be mapped to a scheduled visit for analysis. In order to accommodate as much data as possible into analysis, the windows in the table below have been widened compared to protocol-specified operational window, to have no gap between them; these windows are used for analysis purposes only. Once analysis visit windows are assigned, all visits, including scheduled visits, and windowed visits will be eligible for being flagged as the "analyzed record" within the analysis window; a participant's individual analysis visit window could potentially contain more than 1 visit. In the event of multiple visits falling within an analysis window or in case of a tie, the following rules will be used in sequence to determine the "analyzed record" for the analysis visit window:

For treatment period visits:

- If the data from the scheduled visit is available, then the scheduled visit data will be used.
- If there is no data from the scheduled visit available, the windowed visit data will be used.

For post-treatment period visits:

- If more than one windowed visits available in the same window, the data closest to the targeted study day will be used.
- If there is a tie with distance from the targeted day, the later data will be used.

Table 3 displays windows for efficacy analysis.

Table 3: Visit Windows for Efficacy Analysis

Scheduled Visit (+/1 window days) in protocol	**Target Study Day**	**Study Day Window for Visit in Analysis**
Day 1	Day 1 (predose)	Day 1 (predose) or last non-missing assessment before pre-dose
Day 3 (±1 day)	Day 3	Day 2 - Day 5
Day 8 (+1 day)	Day 8	Day 6 - Day 11
Day 15 (+1 day)	Day 15	Day 12 - Day 17
Day 21 (±1 day)	Day 21	Day 18 - Day 23
Day 28 (±3 day)	Day 28	Day 24 - Day 31
Day 45 (±3 days)	Day 45	≥Day 32

The summary by visit will use the "analyzed records" only – at most one per participant. The data not flagged as the "analyzed record" will be included in listings. A windowed visit that does not fall under any analysis window will remain in the database and will be included in the listings.

8.3.3. Analysis of Efficacy Variable(s)

The FAS will be used for all efficacy summary tables. Participants will be analyzed according to randomized treatment.

The following efficacy endpoints will be summarized descriptively by scheduled assessment time point:

- HAM-D total score – observed, change from baseline, percent change from baseline
- HAM-D subscale scores – observed, change from baseline, percent change from baseline
- HAM-D individual item score – observed, change from baseline
- HAM-D response – missing response not accounted
- HAM-D response – missing response counted as No response
- HAM-D remission – missing remission not accounted

- HAM-D remission – missing remission counted as No remission
- CGI-I score -observed
- CGI-I response - missing response not accounted
- CGI-I response – missing response counted as No response
- CGI-S score – observed and change from baseline
- HAM-A total score – observed, change from baseline, percent change from baseline
- HAM-A individual item score -– observed, change from baseline
- MADRS total score – observed, change from baseline, percent change from baseline
- MADRS individual item score – observed, change from baseline
- MADRS response – missing response not accounted
- MADRS remission – missing remission not accounted
- MADRS response – missing response counted as No response

  MADRS remission – missing remission counted as No remission
- EPDS total score– observed, change from baseline, percent change from baseline
- PHQ-9 score – observed (including categories), change from baseline

The HAM-D change from baseline in total score will also be presented by the following subgroups:

- Antidepressant use at baseline: yes, no
- Age group: 18-24, 25-45 years
- Race: White, Black or African American, Other
- BMI (≤18.4, 18.5-24.9, 25-29.9, ≥30 kg/m^2)
- HAM-D total score at baseline: < median of baseline HAM-D total score, ≥median of baseline HAM-D total score
- Onset of PPD (3^{rd} trimester vs. postpartum)
- History of PPD (1^{st} episode vs recurrent PPD episode)
- Depression with elevated anxiety, by Baseline HAMA total score (>=20, <20)
- Country (US vs. Rest of World)

No model-based analysis will be performed for the subgroup of "Depression with elevated anxiety, by baseline HAM-D anxiety subscale" since over 95% of participants had depression with elevated anxiety categorized by baseline HAM-D anxiety subscale (>=7) or the subgroup of COVID-19 history as majority of participants were 'Not impacted';

similarly for the subgroup of country as only handful of participants were enrolled outside the US.

In addition, post-baseline percentage improvement in HAM-D total score will be presented in histogram over scheduled visits by treatment group under the following categories: < 0% (worsened), ≥0% but < 25%, ≥25% but < 50%, ≥50% but <75%, ≥75%. Post baseline HAM-D total score will also be presented in histogram over time: ≤7, >7 but ≤15, >15 but ≤23, >23 but <26, >=26. Supporting data will be presented in summary tables.

Bar charts over scheduled visits by treatment for HAM-D response, HAM-D remission and CGI-I response will be provided. In addition, line plots for LS means for HAM-D total score and bar charts for HAM-D response and remission will be provided by antidepressant use at baseline.

As part of supportive analyses, summary statistics and Model-based estimates for the HAM-D total scores will be provided for Per Protocol Set.

All above efficacy variables will be listed in a separate listing.

8.3.3.1. Mixed Effects Model for Repeated Measures

The estimand for the primary efficacy analysis is defined as follows:

1. The treatment regimens for participants are: placebo or SAGE-217 for 14 days.
2. The target population is adult participants with a diagnosis with severe PPD (baseline HAM-D total score ≥26).
3. The variable of interest is the change from baseline in HAM-D total score at Day 15.
4. The population summary level is the model-based estimate of the difference between SAGE-217 and placebo treatments in mean change from baseline in HAM-D total score or in CGI-S total score.
    a. The population summary level for the primary endpoint is the model-based estimate of the difference between SAGE-217 and placebo in change from baseline in HAM-D total score at Day 15
    b. The population summary level for the key secondary endpoints is the model-based estimate of the difference between SAGE-217 and placebo in change from baseline in HAM-D total score at Day 3, Day 28, Day 45 and in change from baseline in CGI-S total score at Day 15 respectively
5. The intercurrent events could be:
    a. The premature discontinuation of treatment for any reason. Treatment policy strategy will be used.
    b. Certain medications including, but not limited to, new antidepressants or benzodiazepines are prohibited in the protocol until Day 45 follow-up; however, the treatment policy strategy requires that data following initiation of such prohibited medications are included in analysis. The protocol does not specify any rescue process, hence there is no rescue medication.

Data from SAGE-217 group versus placebo group will be analyzed using a mixed effects model for repeated measures (MMRM); the model will include treatment (SAGE-217 or placebo), baseline HAM-D total score, antidepressant use at baseline (Yes or No), assessment time point, and time point-by-treatment interaction as explanatory variables. All explanatory variables will be treated as fixed effects. All post-baseline time points will be included in the model. The main comparison will be between SAGE-217 and placebo at Day 15. Model-based point estimates ie, treatment difference in least squares [LS] mean is the estimate of the effect and will be reported where applicable along with 95% confidence intervals, and p-values. An unstructured (UN) covariance structure with the default Newton-Raphson algorithm used by SAS PROC MIXED will be used to model the within-participant errors. If there is a convergence issue with the unstructured covariance model, the Fisher Scoring algorithm (via the SCORING option of the PROC MIXED statement), the no-diagonal factor analytic structure (via the TYPE=FA0(*T*) option of the REPEATED statement, where *T* is the total number of time points), Toeplitz, Autoregressive (1) [AR (1)] covariance structure will be used, following this sequence until convergence is achieved. If the model still does not converge with AR (1) structure, no results will be reported. When the covariance structure is not UN, the sandwich estimator for the variance-covariance matrix will be derived, using the EMPIRICAL option in the PROC MIXED statement in SAS. Observed Margins (OM) option will be used in SAS with LSMEAN statement. The p-value will be interpreted at 5% 2-sided level of significance.

If the comparison of SAGE-217 versus placebo is significant at 0.05 level, the hypotheses testing for the key secondary endpoints will proceed in the prespecified fixed sequence as described in Section 8.3.3.5 .

Similar to the methods described above for the primary endpoint, MMRM will be used for the analysis of the change from baseline in other time points in HAM-D total score, all time points in HAM-D subscale scores (Core Subscale score, Anxiety Subscale score, Bech-6 Subscale score and Maier Subscale score), HAM-D individual item scores, CGI-S score, HAM-A total score, HAM-D individual item scores, MADRS total score, MADRS individual item scores, EPDS total score, and PHQ-9 total score. PHQ-9 total score category will be summarized.

The MMRM analyses will also be used for change from baseline in HAM-D total score within each baseline subgroup level defined in Section 8.2.3 separately. If any treatment group for any level of subgroup has ≤15 participants, the subgroup level will not be used in the analysis.

For each model, the comparison of interest will be between SAGE-217 and placebo at Day 15. In addition, model-based point estimates (ie, LS means), 95% confidence intervals, and p-values will be reported for all time points.

Line plot of model-based LS Mean and standard error (SE) over time will be prepared for change from baseline in HAM-D total score, CGI-S score, HAM-A total score, MADRS total score. Forest plot for subgroup analysis for change from baseline in HAM-D total score at Day 15 – LS mean, LS mean difference, 95% confidence interval, and P-value – will be

provided. In addition, a Forest plot for individual items in HAM-D and for subscales of HAM-D will also be provided.

Summary of HAM-D total score along with model-based estimates will be provided for Per Protocol Set as a supportive analysis for the primary endpoint.

8.3.3.2. Generalized Estimating Equation (GEE) Models

Generalized estimating equation (GEE) methods will be used for the analysis of HAM-D response (missing response not accounted or missing response counted as No response respectively) and HAM-D remission (missing remission not accounted or missing remission counted No remission respectively). GEE models will include terms for treatment (SAGE-217, or Placebo), baseline HAM-D score, antidepressant use at baseline (Yes or No), assessment time point, and time point-by-treatment as explanatory variables.

An unstructured (UN) covariance structure will be used to model the within-participant errors. If there is a convergence issue with the unstructured covariance model, then exchangeable covariance structure will be used. If the model still does not converge with exchangeable structure, no results will be reported.

Model-based point estimates (ie, odds ratios), 95% confidence intervals, and p-values will be reported.

A GEE method will also be used for the analysis of MADRS response (missing response not accounted or missing response counted as No response respectively) and remission (missing remission not accounted or missing remission counted No remission respectively) including terms for treatment, baseline MADRS score, antidepressant use at baseline, assessment time point, and time point-by-treatment as explanatory variables.

A GEE method will also be used for the analysis of CGI-I response (missing response not accounted or missing response counted as No response respectively) including terms for treatment, baseline CGI-S score, antidepressant use at baseline, assessment time point, and time point-by-treatment as explanatory variables.

Model diagnostics to access the goodness of fit for the GEE models will be examined. In the event of poor goodness of fit, logistic regression will be performed separately for each visit, including baseline score and antidepressant use at baseline as explanatory variables.

Bar charts over scheduled visits by treatment for HAM-D response/remission, MADRA response/remission and CGI-I response will be provided for FAS.

8.3.3.3. Sensitivity Analysis

8.3.3.3.1. Imputation Based on Study Withdrawal Reasons for Assessing MAR Assumption for Missing HAM-D Total Scores

A sensitivity analysis will be used to investigate the impact of missing data if ≥ 5% of participants have missing HAM-D total score at timepoints (ie, HAM-D total score at Days 3, 8, 15, 21, 28 or 45).

Imputation based on study withdrawal reason will be used. The missing HAM-D total scores will be imputed using multiple imputation methods. The MMRM model will use the imputed

dataset (all observed and imputed values included) to estimate the treatment difference. The FAS will be used for sensitivity analyses. Sample SAS codes for MI imputation is provided in Section 11.2, Appendix B.

Imputation distribution:
The imputation distribution for the missing HAM-D total score at each timepoint will be assumed to have a normal distribution. All randomized participants will be classified as non-missing category, missing category 1, or missing category 2, based on the following rules:

- Non-missing category: Participant with non-missing change from baseline in HAM-D total score at Day X
- Missing category 1: Participant discontinued due to adverse events, physician decision, protocol deviation, non-compliance with investigational product, or other, and is missing change from baseline in HAM-D total score at Day X, lost to follow up, withdrawal by participant, or participant completed study but is missing change from baseline in HAM-D total score at Day X or participants discontinued due to sponsor decision,
- Missing category 2: Participant discontinued due to pregnancy.

Imputation algorithm:
Missing values of HAM-D total score at all visits will be imputed using the fully conditional specification (FCS) models. Two different imputation models will be used based on reasons of missing:

- **Missing category 1:** simulate missing values of HAM-D total score using an imputation model based on the non-missing HAM-D total scores for placebo group (Jump to Reference). This represents a conservative approach as it tends to reduce the difference between treatment and placebo group since higher values of change from baseline represents worse outcome, and placebo is supposed to provide a higher value.
- **Missing category 2:** simulate missing values of HAM-D total score using an imputation model based on the non-missing HAM-D total scores within the same treatment group.

Analysis model:
The complete MI method is described below:

- Impute missing values using the normal distribution specified in the above algorithm to form a complete dataset (imputed dataset). After imputation, all FAS participants will have non-missing change from baseline in HAM-D total score at all scheduled assessment time points.
- Repeat the process K (K=100) times, using the procedure described above to form K imputed complete datasets with the same variance structure.

- Fit the MMRM model including treatment, baseline antidepressant use, and baseline HAM-D total score, assessment time point, and time point-by-treatment, to each imputed dataset, to estimate the treatment effect and its variance.
- Combine the results from the K imputed datasets using the SAS procedure PROCMIANALYZE, to derive the MI estimator.

We fit the analysis model (MMRM model specified before) to the kth completed dataset, denoting the estimate of the treatment difference θ by θ_k from the kth completed dataset, and denoting the corresponding estimate of the variance V_k. The MI estimator of θ, $\tilde{\theta}_{MI}$, is the average of the K individual estimators:

$$\tilde{\theta}_{MI} = \frac{1}{K}\sum_{k=1}^{K} \theta_k$$

The estimated variance of $\tilde{\theta}_{MI}$ is a combination of the between- and within-imputation variability as follows:

$$V_{MI} = W + \left(1 + \frac{1}{K}\right) B$$

where $W = \frac{1}{K}\sum_{k=1}^{K} V_K$ is the within-imputation variability and $B = \frac{1}{K-1}\sum_{k=1}^{K}\left(\theta_k - \tilde{\theta}_{MI}\right)^2$ is the between-imputation variance.

It has been shown that the statistic

$$T = \frac{\tilde{\theta}_{MI} - \theta}{\sqrt{V_{MI}}}$$

has an approximate t distribution where $V = (K-1)\left(1 + \frac{W}{B}\right)^2$.

8.3.3.3.2. Tipping Point Analysis for Assessing Missing HAM-D total score

The purpose of a tipping point analysis is to evaluate the sensitivity of results in missing data assumed missing at random (MAR) by finding out the size of the change assumed missing not a random (MNAR) that tips statistically significant results to become not statistically significant. This analysis will use FAS.

A tipping point analysis, missing HAM-D total score from each study visit will be imputed using multiple imputation in the following steps:

1. For treatment group (i.e., SAGE-217 50 mg) a shift from 0 to 4 with increment of 0.2 will be added to the imputed HAM-D total score values at each study visit.

Assumed arbitrary missing pattern, a multiple imputation step with SAS MI FCS method will be used to impute the missing values of HAM-D total score with "mnar" procedure to make the shift. One hundred datasets will be generated for each study visit.

2. For each of the 100 complete datasets after imputation, the change from baseline will be calculated so that the MMRM model will be fitted to estimate treatment differences and corresponding p-values.

3. The 100 sets of MMRM results will be combined with SAS PROC MIANALYZE, which combines estimates as specified in step 2.

4. Steps 1 to 3 will be repeated with different values of the shift parameter until the treatment effect is no longer significant at 0.05 level at each study visit. The range of shift may be adjusted at the time of the final analysis if it will not reach a tipping point properly at majority of study visits.

The tipping point will be provided by study visit along with the main estimates in change from baseline by study visit from MMRM in Section 8.3.3.1 . Sample SAS code for MI is provided in Section 11.2 Appendix B.

8.3.3.3.3. Weighted Generalized Estimating Equation

The standard GEE method in Section 8.3.3.2 is valid if the data are missing completely at random (MCAR), but it can lead to biased results if the data are missing at random (MAR). The weighted GEE implements the inverse probability-weighted method to account for dropouts under the MAR assumption. This analysis will be performed using the binary secondary endpoints, ie. HAM-D response, HAM-D remission, MADRA response and MADRA remission and CGI-I response. This analysis will use FAS.

A weighted GEE method does not apply for the data with intermittent missing from any of the participants and the procedure does not converge in SAS. A SAS procedure with MI Markov Chain Monte Carlo (MCMC) with 100 imputations will be used to fill out intermittent missing values so that the data will have monotone missing pattern. Therefore, a weighted GEE on each binary endpoint will be run as a sensitivity analysis. Since the endpoints are binary outcomes, the imputation to fill out intermittent missing will be performed based on the raw scores. The imputed values will be rounded and converted as each binary endpoint as defined in Section 8.3.1.1 and Section 8.3.1.2 .

Therefore, a weighted GEE model including baseline HAM-D score, antidepressant use at baseline (Yes or No), assessment time point, and time point-by-treatment as explanatory variables will be used to estimate point estimates (ie, odds ratios), 95% confidence intervals, and p-values. The results from the 100 imputed datasets will be combined using the SAS procedure PROC MIANALYZE.

Sample SAS code for weighted GEE is provided in Section 11.2 Appendix B.

8.3.3.4. Analysis of Time to First HAM-D Response/Remission

Using the FAS, Kaplan-Meier (KM) survival method will be provided for time to first HAM-D response; the median time (Days) to first response, min, max, 25th and 75th percentiles will be estimated from KM analysis. A participant will be censored at the participant's last day of HAM-D evaluation in the database if the participant did not have a response. Similar analysis will be done for first HAM-D remission.

These analyses will be provided on FAS for supportive purposes only.

8.3.3.5. Multiplicity Adjustment for Key Secondary Endpoints

Multiplicity adjustment to statistical testing of hypotheses of the key secondary endpoints are conducted by using fixed sequence strategy. Only if the primary endpoint is statistically significant at 0.05 level, the key secondary endpoints will be tested sequentially, testing each endpoint at 5% level of significance only if the previous endpoint in the below sequence has been significant at 5% level. If an endpoint is not significant at 5% level, the next endpoint in the sequence will be interpreted only with nominal p-value.

The sequence of testing key secondary endpoints is as follows:

- Change from Baseline in HAM-D total score at Day 3
- Change from Baseline in HAM-D total score at Day 28
- Change from Baseline in HAM-D total score at Day 45
- Change from Baseline in CGI-S at Day 15

A secondary endpoint not included above is not adjusted for multiplicity, and hence will be interpreted with nominal p-value.

8.3.4. Characterization of Durability of SAGE-217 Treatment Effect

After completion of 14 days of treatment, the study participant is followed for 31 days, without any further treatment with study drug. During this 31-day follow-up period, clinic visits are scheduled for Days 21, 28 and 45. Day 45 change from baseline in HAM-D total score is pre-specified as a key secondary endpoint in this study.

Demonstration of SAGE-217 durable effect will be characterized via 2 complementary approaches:
1) clinically meaningful durable SAGE-217 treatment effect at Day 45 (see Section 8.3.4.1 below), and
2) statistically significant change from baseline at Day 45 in SAGE-217 versus placebo (for statistical demonstration of durable SAGE-217 treatment effect, see Section 8.3.4.2 below).

8.3.4.1. Durability of Clinically Meaningful Treatment Effect for SAGE-217

Durability of treatment effect is assessed over the post-treatment period in the SAGE-217 group, based on the efficacy observed at Day 15. The primary endpoint for clinically durable effect will be examined by the percent retention of the Day 15 reduction from baseline in HAM-D total score (referred to as "percent retention of D15 change from baseline (CFB)" going forward).

Percent retention of D15 CFB is defined as follows: Let X_b be baseline HAM-D total score, X_{15} be Day 15 HAM-D total score, X_y be Day Y (Y>15) HAM-D total score. Then percent retention (%) for Day Y is $\frac{X_y - X_b}{X_{15} - X_b} \times 100$. It will be calculated for the scheduled visits after Day 15 for the participants by the randomized treatment groups. For example, for a

participant with a baseline HAM-D of 27 (X_b) and a HAM-D score of 13 (X_{15}) at Day 15, percent retention at Day 45 with HAM-D of 16 (X_y) would be 79%. If D45 HAM-D score is 18, percent retention would be 64%.

A summary of percent retention of D15 CFB will be presented for participants at Day 15 for post-Day 15 visits by treatment group, based on FAS. This will serve as the primary approach for assessing clinically durable effect; mean percent retention of D15 CFB at Day 45 is considered clinically meaningful durability if it is >=65%. The mean percent retention (±SE) over time by treatment group will be presented in a line plot. In addition, the number and percent of participants with at least 65% retention of D15 CFB at each of post-Day 15 visit will also be provided by treatment group.

Following analysis will be presented by treatment group for supportive purposes to assist with further understanding of clinical durability of effect of SAGE-217. These analyses will be conducted using FAS.

A. A summary of percent retention of D15 CFB will be presented for HAM-D responders at Day 15 for post-Day 15 visits. The mean percent retention (±SD) over time will be presented in a line plot. In addition, the number and percent of HAM-D responders at Day 15 with at least 65% retention of D15 CFB at each of post-Day 15 visit will also be provided.

B. The number and percent of HAM-D responders at post-Day 15 visits among HAM-D responders at Day 15 will be provided. A bar chart will be provided.

C. The number and percent of HAM-D remitters (HAM-D total score <=7) at post-Day 15 visits among the HAM-D remitters at Day 15 will be provided. A bar chart will be provided.

D. Relapse: A relapse is defined for HAM-D responders at Day 15. A relapse is defined as having at least 2 consecutive HAM-D total score >=20 after Day 15. The number and percent of participants with relapse will be provided.

E. Rebound: A rebound is defined for HAM-D responders at Day 15. A rebound is any HAM-D total score (after Day 15) >=Baseline HAM-D total score. The number and percent of participants with rebound will be provided.

F. The number and percent of CGI-I responders (CGI-IR) at post-Day 15 visits among the CGI-I responders at Day 15 will be provided. A bar chart will be provided base.

Post Day 15 HAM-D results supporting the above analyses will be listed separately for the participants based on FAS.

8.3.4.2. Durability of Treatment Effect at Day 45 via Statistical Comparison of SAGE-217 Versus Placebo

These analyses have been discussed in Section 8.3.3.1 and Section 8.3.3.2 as part of the efficacy analyses with Day 45 as one of key secondary efficacy endpoints:

A. MMRM analysis of change from baseline in HAM-D total score at Day 45 (key secondary efficacy endpoint), comparing SAGE-217 versus placebo (LS mean, p-value and 95% CI) will be provided.
B. GEE analysis of HAM-D response/remission rates comparing SAGE-217 versus placebo at Day 45 (Odds Ratio, p-value and 95% CI) will be provided.

8.3.5. Anxious Subgroup Analysis

Efficacy analysis will be repeated for the anxiety subgroup.

- Depression with elevated anxiety, by Baseline HAMA total score (>=20, <20)

HAM-D total score and CGI-S score observed, change from baseline, percent change from baseline - will be summarized descriptively by scheduled assessment time point for the subgroup, respectively.

A MMRM analysis will be provided for the change from baseline in HAM-D total score and CGI-S score by scheduled assessment time point for each group, respectively.

A GEE method will also be used for the analysis of HAM-D response, HAM-D remission, and CGI-I response by scheduled assessment time point for each subgroup, respectively.

8.4. Safety Analysis

A secondary objective is to evaluate the safety and tolerability of SAGE-217 as assessed by the incidence and severity of adverse events; [redacted]. Safety analyses will be conducted using the Safety Set, unless specified otherwise. The data will be presented by the actual treatment received rather than the treatment to which the participant has been randomized; for definition of actual treatment assignment, please see Section 8.1

The safety endpoints evaluated at scheduled visits within treatment period are taken as done in nominal visit, without any windowing. If a value is available for a nominal scheduled visit, that value will be used in summary by visit. If scheduled visit value is not available, a value from the specific visit window will be included in summary. For Post-treatment period visits, the choice of the visit will follow the same rule as described in Section 8.3.2.

All parameters will be converted to consistent units according to the International System of Units (SI) before analysis.

Any time during treatment, last value on treatment and last value on study will be included in the summaries whenever indicated in the relevant sections below. Any time during treatment is defined as measurement on or after first dose, on or before last dose + 1 day. Last value on treatment is defined as the last post-baseline value between first dose of investigational product (exclusive) and up to last dose of investigational product + 1 days (inclusive). Last value on study is defined as the last post-baseline value after the first dose of investigational product.

The safety endpoints and variables considered in the summary tables for this study are summarized in Table 4.

Table 4: Safety Endpoints and Variables in the Summary Tables

Safety Evaluation	Incidence	Observed Value	Change from Baseline	Abnormality/Clinical Significance (CS)	Potentially Clinical Significance (PCS)
AEs	X				
[redacted]					

Note: PCS criteria are outlined in sections 8.4.2-8.4.4

X = to be summarized in tables
Z = to be presented in listings only

8.4.1. Adverse Events

Adverse events (AEs) are collected starting at the time of informed consent and throughout the duration of the participant's participation in the study. A treatment-emergent adverse event (TEAE) is defined as an adverse event with onset on or after the start of investigational product. The TEAEs will be further categorized by the phase of occurrence as follows:

Adverse events are assigned an AE period based on the onset date/time. AE periods are defined as follows:

- Pre-treatment AE: AE onset date before first investigational product dosing date
- TEAE: AE onset date/time on or after first investigational product dose date/time (If an AE start date same as investigational product first dose date, but no time either in AE start or treatment start, then consider this AE to be in treatment period TEAE.)
- On-treatment TEAE: AE onset date/time on or after first investigational product dose date/time and on or before investigational product last dose date + 1 day (Note that time does not matter for the end of this period.)
- Post-treatment TEAE: AE onset date after investigational product last dose date +1 day (Typically, Day 16 through Day 45 – time does not matter)

If the date of an adverse event is incomplete and an unambiguous determination could not be made with respect to its onset time versus the first dose of investigational product and/or last dose of investigational product, the adverse event will be assumed to be a TEAE and a

treatment period AE. For imputation of missing AE dates, please refer to Appendix C, Section 11.3.

All adverse events will be coded using MedDRA version 23.0 or higher.

An overview summary table of TEAEs will present the number and percentage of participants as well as the number of events for the following:

- TEAE
  - On-treatment TEAE
  - Post-treatment TEAE
- TEAEs by maximum severity (severe>moderate>mild)
- TEAE leading to investigational product dose reduction
- TEAE leading to discontinuation of investigational product
- TEAE leading to withdrawal from the study
- Treatment-emergent Serious Adverse Event (SAE)
- Death

Incidence of TEAEs in following categories will be provided by SOC and PT. A participant is counted only once under each SOC and PT in case of multiple occurrences of the same AE. These tables will be sorted by decreasing frequency of System Organ Class (SOC) in SAGE-217 group, then in placebo group, then alphabetically first within SOC then within preferred term.

- TEAE
  - On-treatment TEAE
  - Post-treatment TEAE
- TEAEs by maximum Severity
- TEAEs by relationship
- Serious TEAEs
- TEAEs leading to discontinuation of investigational product
- TEAEs leading to withdrawal from the study
- TEAEs leading to investigational product dose reduction

Listing of AEs with onset prior to first dose of investigational product will be provided. All listings on TEAEs will provide the period designation for each AE.

A summary of most common study period TEAE just by preferred term where the incidence is more than 2% in any treatment group will be provided, sorted by decreasing frequency of system organ class (SOC) first by SAGE-217 group, then by placebo, then alphabetically first within SOC then within preferred term.

For maximum severity, participants will be counted only once within each SOC and PT at the maximum severity in the following order: severe> moderate> mild; an AE with missing severity will be omitted from severity presentation. For relationship to investigational product, participant will be counted only once within each SOC and PT at the strongest relationship to investigational product in the following order: related > not related, an AE with relationship missing is treated as related. The incidences will be presented by descending frequency of SOC and then, within a SOC, by descending frequency of PT based on the participant count, and in alphabetical order of PT if the incidence within a PT is a tie. Adverse events with onset before the first dose of investigational product will be provided in a separate listing. Separate data listing for deaths and SAEs will be provided. A listing of all adverse events experienced by participants who had investigational product dose reduction due to adverse event will be generated as well. A listing of participants with COVID-19 related AEs will be provided.

In addition, TEAE summary by SOC/PT will also be presented by the following subgroups:

- Antidepressant use at baseline: yes, no
- Age group: 18-24, 25-45 years
- Race: White, Black or African American, Other
- BMI ($\leq$18.4, 18.5-24.9, 25-29.9, $\geq$30 kg/m^2)
- Baseline HAM-D total score: < median of baseline HAM-D total score, $\geq$ median of baseline HAM-D total score
- Country (US, Rest of World)

[redacted]

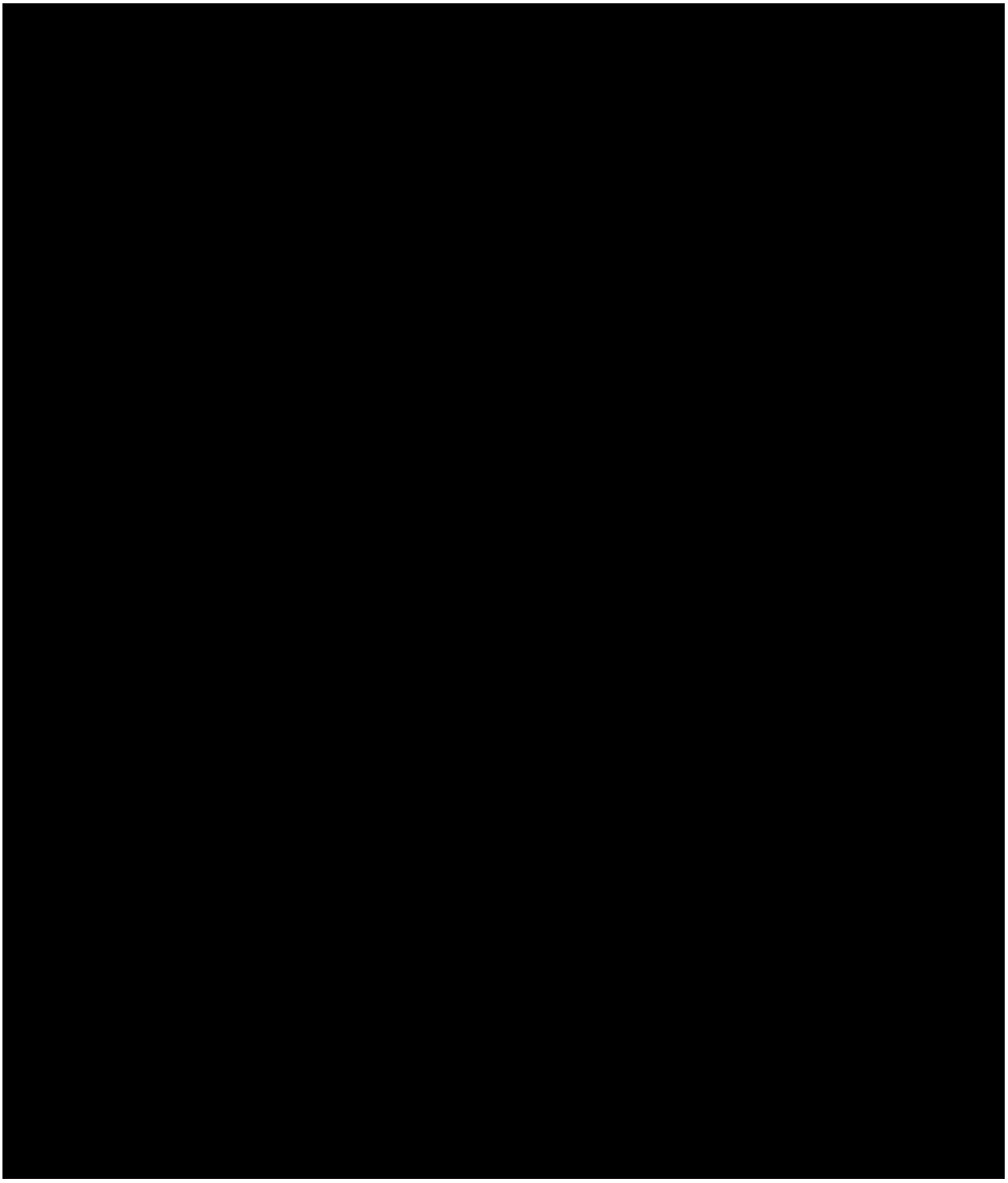
[redacted]

[redacted]

[redacted]

[redacted]

[redacted]



[redacted]

[redacted]

[redacted].

[redacted].

[redacted].

[redacted]

[redacted]

8.4.5. Physical Examination

Physical examination is scheduled on screening, Days 8 and 21. Only clinically significant abnormalities are captured in the database – for post-baseline observations, these will be reported as adverse events, hence these will be included in AE displays; for pre-baseline observations, these will be reported as medical history, hence these will be included in Medical History displays. The dates of physical examination will be listed to confirm that the examination was done.

[redacted]

[redacted].

[redacted]

[redacted]

9. SUMMARY OF INTERIM AND DMC ANALYSES

Not applicable

10. REFERENCES

Clinical study protocol, version 2.0, 29 JAN 2021, Company: Sage Therapeutics Inc.

11. LIST OF APPENDICES

11.1. Appendix A: Schedule of Assessments

	Study Period / Visit Day								
	Screening Period	Treatment Period[a]				Follow-up Period			
Study Procedure	D-28 to -1	D1 (assessments completed pre-dose)	D3 (±1 d)	D8 (+1 d)	D15 (+1 d)	D18 (±1 d)	D21 (±1 d)	D28 (±3 d)	D45/ET (±3 d)
Informed Consent	X								
Duplicate Participant Check [b]	X								
Inclusion/Exclusion	X	X							
SCID-5-CT	X								
Demographics	X								
Medical/Family History [c]	X								
Participant training [d]	X	X							
Physical Examination [e]	X			X	X		X		X[f]
Body Weight/Height	X				X (weight only)		X (weight only)		X (weight only) [f]
[redacted]									
Drug & Alcohol Screen [h]	X	X	X	X	X	X	X	X	X
Pregnancy Test [i]	X	X			X				X
Hormone Sample [j]	O			O	O				
Genetic Sample [k]	O								
[redacted]									

Study Procedure	Study Period / Visit Day								
	Screening Period	Treatment Period[a]				Follow-up Period			
	D-28 to -1	D1 (assessments completed pre-dose)	D3 (±1 d)	D8 (+1 d)	D15 (+1 d)	D18 (±1 d)	D21 (±1 d)	D28 (±3 d)	D45/ET (±3 d)
[redacted]									
CGI-I			X	X	X		X	X	X
CGI-S	X	X	X	X	X		X	X	X
HAM-A [o]		X	X	X	X		X	X	X
HAM-D [o, p]	X	X	X	X	X		X	X	X
MADRS		X		X	X			X	X
EPDS	X	X	X	X	X		X	X	X
PHQ-9	X	X	X	X	X		X	X	X
[redacted]									
Randomization		X							
Dispense IP		X		X					
IP Administration		X (once daily in the evening through Day 14 - inclusive)							
IP Adherence [r]		X							
IP Accountability/Return				X	X				X [f]
Adverse Events [c]	X (from time of ICF throughout the duration of participation)								
Prior/Concomitant Medications [c, s]	X								

CGI-I = Clinical Global Impressions - Improvement; CGI-S = Clinical Global Impressions - Severity; [redacted] D = day; [redacted]; EPDS = Edinburgh Postnatal Depression Scale; ET = early termination; HAM-A = Hamilton Anxiety Rating Scale; HAM-D = Hamilton Rating

Scale for Depression, 17-item; IP = investigational product; MADRS = Montgomery-Åsberg Depression Rating Scale; O = optional; PHQ-9 = Patient Health Questionnaire;

Note: When , and/or blood draws are scheduled to occur at the same time, the order of assessments should be as follows: , blood draw.

[a] An unscheduled visit may be needed if a dose adjustment is deemed necessary by the investigator. Any remaining current dose should be returned, and the adjusted dose should be dispensed at this visit.

[b] Participants in the US only will be asked to authorize that their unique participant identifiers be entered into a registry with the intent of identifying participants who may meet exclusion criteria for participation in another clinical study.

[c] Information regarding diagnosis, isolation, and/or hospitalization due to COVID-19 will be documented as part of medical history, AE collection, and prior/concomitant medication/procedure collection at screening and throughout the study.

[d] Participants will be trained on the use of software applications and devices necessary for the conduct of the study by site personnel.

[e] A full physical examination will be conducted at screening and abbreviated physical examinations will be conducted thereafter. A full physical examination includes assessment of body systems (eg; head; eye; ear, nose, and throat; heart; lungs; abdomen; and extremities). An abbreviated physical examination includes a brief medical history followed by targeted physical examination.

[f] To be completed at ET only for participants who withdraw from the study/stop study participation before the Day 15 visit.

[h] Urine toxicology for selected drugs of abuse and breath test for alcohol

[i] Urine and serum pregnancy test to be conducted at screening; the urine test is recommended to precede other screening assessments. Qualifying criteria will be based on serum test results. Urine pregnancy test to be conducted at all other visits indicated. A urine pregnancy test will also be collected as part of the ET assessments for participants who discontinue the study early.

[j] An optional blood sample for stress hormone levels, kynurenine biochemistry, and markers of inflammation, where consent is given.

[k] An optional genetic sample for biomarker testing, where consent is given.

.

.

[o] The assessment timeframe for HAM-D scales will refer to the past 7 days (1 week) at Screening and “Since Last Visit” for all other visits. The assessment timeframe for HAM-A scales will refer to the past 7 days (1 week) at all visits.

[p] HAM-D is to be completed as early during the visit as possible.

.

[r] As local regulations permit, IP administration will be monitored via a medication adherence monitoring platform used on smartphones to visually confirm IP ingestion. IP adherence will not be captured after participants discontinue IP. Optional consent may be required in regions with restricted regulations.

[s] Prior medications will be collected at screening and concomitant medications and/or procedures will be collected at each subsequent visit (see Section 9.2.1 from protoocol).

11.2. Appendix B: Details of Statistical Methodology

Sample SAS code for Mixed Effects Model for Repeated Measures (MMRM):

- If type=un:

```
ods output lsmeans=estimates diffs=diffs;
proc mixed data=&data;
class trtpn avisitn usubjid antidep;
model chg=base trtpn avisitn trtpn*avisitn antidep / ddfm=kr s;
repeated avisitn / subject=usubjid type=un;
lsmeans trtpn*avisitn /diff=all cl alpha=0.05 OM;
** assuming trtpn= 1 for the Placebo, trtpn=2 for SAGE-217
run;
```

if above model doesn't converge, use following model:

```
ods output lsmeans=estimates diffs=diffs;
proc mixed data=&data scoring=x;
class trtpn avisitn usubjid antidep;
model chg=base trtpn avisitn trtpn*avisitn antidep / ddfm=kr s;
repeated avisitn / subject=usubjid type=un;
lsmeans trtpn*avisitn /diff=all cl alpha=0.05 OM;
** assuming trtpn= 1 for the Placebo, trtpn=2 for SAGE-217
run;
```

if above 'scoring' option doesn't converge, use following model:

```
ods output lsmeans=estimates diffs=diffs;
proc mixed data=&data empirical;
class trtpn avisitn usubjid antidep;
model chg=base trtpn avisitn trtpn*avisitn antidep / ddfm=kr s;
repeated avisitn / subject=usubjid type=fa0(6);
lsmeans trtpn*avisitn /diff=all cl alpha=0.05 OM;
** assuming trtpn= 1 for the Placebo, trtpn=2 for SAGE-217
run;
```

- If type=un does not converge, use type=TOEP or AR(1);

```
ods output lsmeans=estimates diffs=diffs;
proc mixed data=&data empirical;
class trtpn avisitn usubjid antidep;
model chg=base trtpn avisitn trtpn*avisitn antidep / s;
```

```
repeated avisitn / subject=usubjid type=TOEP;
**If type=TOEP does not converge, use type=AR(1);
lsmeans trtpn*avisitn /diff=all cl alpha=0.05 OM;
** assuming trtpn= 1 for the Placebo, trtpn=2 for SAGE-217
run;
```

Sample SAS code for Generalized Estimating Equation (GEE):

```
 ods output lsmeans=lsmeans diffs=diffs ObStats=dd;

 proc genmod data=&data desc  plots=all;
class usubjid trtpn antidep avisitn;
 model aval=base trtpn avisitn trtpn*avisitn antidep /dist=bin link=logit OBSTATS;
repeated subject=usubjid / within=avisitn type=un; * if convergence not met, use type=exch;
lsmeans trtpn*avisitn / diff exp cl;
run;
```

Sample SAS code for Multiple Imputation (MI):

```
** Missing category 1, trtp=A represent the PLACEBO group
proc mi data=in seed=xxxx nimpute=20 round=. . . . 1 1 1 1 1 1 1 1 1 1 1 1 1 out= fcs_reg1;
class trtp antidep;
FCS reg;
mnar model (base day3 day8 day15 day21 day28 day45/ modelobs= (trtp='A'));
var antidep base day3 day8  day15 day21 day28  day45;
run;
** Missing category 2
proc mi data=&data seed=xxxx nimpute=20 round=. . . . 1 1 1 1 1 1 1 1 1 1 1 1 1 out=fcs_reg2;
class trtp antidep;
fcs nbiter=20 reg (base day3 day8  day15  day21 day28  day45/details);
var trtp antidep base day3 day8 day15 day21 day28  day45;
run;
```

A sample of the SAS code for sensitivity analysis using the tipping-point approach is provided as below:

```
proc mi data=&data out=temp seed=xxxxxx nimpute=100;
class trtpn antidep;
var trtpn antidep base day3 day8 day15 day21 day28  day45 ;
fcs reg;
mnar adjust(day_xx/shift=xx adjustobs=(trtpn='2'))
    adjust(day_xx/shift=xx adjustobs=(trtpn='2'))
    adjust(day_xx/shift=xx adjustobs=(trtpn='2'))
    adjust(day  xx/shift=xx adjustobs=(trtpn='2'))
    adjust(day_xx/shift=xx adjustobs=(trtpn='2'))
```

```
    adjust(day_xx/shift=xx adjustobs=(trtpn='2'))

run;
```

Sample SAS code for Weighted GEE analysis:

```
*Chage the missing pattern from arbitrary to monotone;
proc mi data=work1 out=im1 seed= 12345 nimpute=100;
mcmc chain=multiple impute= monotone ;
var base antidep day3  day8  day15 day21 day28  day42 ;
run;

proc gee data=cgiim desc ;
by _imputation_;
class usubjid trtpn antidep avisitn ;
missmodel BASE trtpn avisitn trtpn*avisitn antidep / type=obslevel;
model dep = BASE trtpn avisitn trtpn*avisitn antidep  /dist=bin link=logit ;
repeated subject=usubjid / type=un ; * if convergence not met, use type=exch;
lsmeans trtpn*avisitn / diff exp cl;
run;
ods output close;
```

11.3. Appendix C: Handling of Missing Dates

Dates missing the day, or both the day and month of the year will adhere to the following conventions in order to classify TEAEs and to classify prior and concomitant medications.

In general, listings will present the actual partial or missing values rather than the imputed values that may be used in derivation. In instances where imputed values will be presented, imputed values will be flagged.

Adverse Events

If the AE start date is completely missing, do not impute a date but consider it as TEAE, unless the AE end date is before the initiation of treatment, in which case the AE will be considered prior.

For partial AE start dates:

- When the year is known, but the month and day is unknown, then:
    - If the year matches the year of first dose date and the end date (if present) is after first dose date, or AE is ongoing, then impute as the month and day of earlier date of (the first dose date + 1 day, last dose date).
    - If the year of AE onset < year of initiation of the treatment, then the month and day will be set to December 31st.
    - If the year of AE onset > the year of initiation of treatment, then the month and day will be set to January 1st.
- If the year and month are known, but the day is unknown, then:

- If the year of AE onset = the year of initiation of the treatment and:
  - the month of AE onset = the month of initiation of the treatment, then the day will be set to the day of initiation of the treatment.
  - the month of AE onset < the month of initiation of the treatment, then the day will be set to the last day of month of the particular year.
  - if the month of AE onset > the month of initiation of the treatment, then the day will be set to the 1st day of month.
- If the year of AE onset < the year of initiation of the treatment, then the day will be set to the last day of month of the particular year.
- If the year of AE onset > the year of initiation of the treatment, then the day will be set to the 1st day of month.

If the imputed AE onset date is after the AE stop date, then the onset date will be set to the stop date.

- When the year and day are present and the month is missing, treat it as if the day is missing, and only year is present. Follow the imputation rules for "year is known, but the month and day is unknown".
- When the year is missing, but the month and/or day is known, treat this date as missing; do not impute.

Dates in Disease History (Dates of diagnosis, current episode, first episode)

- If the year is present and the month and day are missing, then the month and day will be set to January 1.
- If the year and day are present and the month is missing, then the month will be set to January.
- If the year and month are present and the day is missing, then the day will be set to the 1st day of month

Prior and Concomitant Medications

For the partial start date of medication:

- If the year is present and the month and day are missing, then the month and day will be set to January 1.
- If the year and day are present and the month is missing, then the month will be set to January.
- If the year and month are present and the day is missing, then the day will be set to the 1st day of month.
- If the imputed start date of medication is after the end date (imputed date if applicable) of medication, then the start date will be set to the end date of medication.

For the partial end date of medication:

- If the year is present and the month and day are missing, then the month and day will be set to either December 31 or date of death whichever is earlier.

- If the year and month are present and the day is missing, then the day will be set to the last day of the month or month of death.
- If the year and day are present and the month is missing, then treat it as if the day is also missing. Set the month and day to be either December 31 or date of death whichever is earlier.

11.4. Appendix D: List of Displays

Tables

Table Number	Title	Analysis Set
Table 14.1.1.1	Summary of Participant Disposition	All Participants
Table 14.1.1.2	Summary of Analysis Sets	All Participants
Table 14.1.2.1	Summary of Major Protocol Deviations	Full Analysis Set
Table 14.1.2.2	Reasons for Exclusion from Analysis Sets	Randomized Set
Table 14.1.3.1.1	Summary of Demographics and Baseline Characteristics	Safety Set
Table 14.1.3.1.2	Summary of Demographics and Baseline Characteristics	Full Analysis Set
Table 14.1.3.1.3	Summary of Baseline Subgroups	Safety Set
Table 14.1.3.1.4	Summary of Baseline Subgroups	Full Analysis Set
Table 14.1.3.2.1	Summary of Disease History	Safety Set
Table 14.1.3.3.1	Summary of Medical and Surgical History	Safety Set
Table 14.1.3.3.2	Summary of Medical and Surgical History ongoing at Screening	Safety Set
Table 14.1.3.3.3	Summary of Participant History of Psychiatric Disorder	Safety Set
Table 14.1.3.3.4	Summary of Family History of Postpartum Depression	Safety Set
Table 14.1.4.1	Summary of Prior Non-Psychotropic Medications	Safety Set
Table 14.1.4.2	Summary of Concomitant Non-Psychotropic Medications	Safety Set
Table 14.1.4.3	Summary of On-treatment Concomitant Non-Psychotropic Medications	Safety Set
Table 14.1.4.4	Summary of Post-treatment Concomitant Non-Psychotropic Medications	Safety Set
Table 14.1.4.5.1	Summary of Prior Psychotropic Medications	Safety Set
Table 14.1.4.5.2	Summary of Prior Psychotropic Medications by ATC Level 1 and Level 4	Safety Set
Table 14.1.4.6.1	Summary of Concomitant Psychotropic Medications	Safety Set
Table 14.1.4.6.2	Summary of Concomitant Psychotropic Medications by ATC Level 1 and Level 4	Safety Set
Table 14.1.4.7	Summary of On-treatment Psychotropic Medications	Safety Set

Table Number	Title	Analysis Set
Table 14.1.4.8	Summary of Post-treatment Psychotropic Medications	Safety Set
Table 14.1.4.9	Summary of Use of Concomitant Antidepressant Medications	Safety Set
Table 14.1.4.10	Summary of Investigational Product Exposure	Safety Set
Table 14.1.4.11	Summary of Investigational Product Adherence	Full Analysis Set
Table 14.2.1.1.1	Summary of Hamilton Rating Scale for Depression (HAM-D) Total Score by Study Visit	Full Analysis Set
Table 14.2.1.1.2	Model-based Results on Change from Baseline in Hamilton Rating Scale for Depression (HAM-D) Total Score by Study Visit	Full Analysis Set
Table 14.2.1.1.3	Model-based Sensitivity Analysis on Change from Baseline in Hamilton Rating Scale for Depression (HAM-D) Total Score at Day 15 (Missing Data Imputed Algorithmically Based on Reason of Missing)	Full Analysis Set
Table 14.2.1.1.4	Summary of Hamilton Rating Scale for Depression (HAM-D) Total Score by Study Visit	Per Protocol Set
Table 14.2.1.1.5	Model-based Results on Change from Baseline in Hamilton Rating Scale for Depression (HAM-D) Total Score by Study Visit	Per Protocol Set
Table 14.2.1.1.6	Imputation Based on Study Withdrawal Reasons for Assessing MAR Assumption for Missing HAM-D Total Scores	Full Analysis Set
Table 14.2.1.1.7	Tipping Point Analysis for Assessing Missing HAM-D total score	Full Analysis Set
Table 14.2.1.2.1	Summary of Hamilton Rating Scale for Depression (HAM-D) Subscale Scores by Study Visit	Full Analysis Set
Table 14.2.1.2.2	Model-based Results on Change from Baseline in Hamilton Rating Scale for Depression (HAM-D) Subscale Score by Study Visit	Full Analysis Set
Table 14.2.1.3.1	Summary of Hamilton Rating Scale for Depression (HAM-D) Individual Item Score by Study Visit	Full Analysis Set
Table 14.2.1.3.2	Model-based Results on Hamilton Rating Scale for Depression (HAM-D) Individual Item Score, Change from Baseline by Study Visit	Full Analysis Set
Table 14.2.1.4.1	Summary of Hamilton Rating Scale for Depression (HAM-D) Response by Study Visit	Full Analysis Set
Table 14.2.1.4.2	Model-based Results on Hamilton Rating Scale for Depression (HAM-D) Response by Study Visit	Full Analysis Set

Table Number	Title	Analysis Set
Table 14.2.1.4.3	Sensitivity Analysis: Summary of Hamilton Rating Scale for Depression (HAM-D) Response by Study Visit	Full Analysis Set
Table 14.2.1.4.4	Summary of Hamilton Rating Scale for Depression (HAM-D) Total Score - Percent Improvement - by Study Visit	Full Analysis Set
Table 14.2.1.4.5	Summary of Time to First Hamilton Rating Scale for Depression (HAM-D) Response - Kaplan-Meier Analysis	Full Analysis Set
Table 14.2.1.5.1	Summary of Hamilton Rating Scale for Depression (HAM-D) Remission by Study Visit	Full Analysis Set
Table 14.2.1.5.2	Model-based Results on Hamilton Rating Scale for Depression (HAM-D) Remission by Study Visit	Full Analysis Set
Table 14.2.1.5.3	Sensitivity Analysis: Summary of Hamilton Rating Scale for Depression (HAM-D) Remission by Study Visit	Full Analysis Set
Table 14.2.1.5.4	Summary of Hamilton Rating Scale for Depression (HAM-D) Total Score in Categories, by Study Visit	Full Analysis Set
Table 14.2.1.5.5	Summary of Time to First Hamilton Rating Scale for Depression (HAM-D) Remission - Kaplan-Meier Analysis	Full Analysis Set
Table 14.2.1.6.1	Summary of Change from Baseline in HAM-D Total Score by Study Visit and Antidepressant Use at Baseline	Full Analysis Set
Table 14.2.1.6.2	Model-based Results on Change from Baseline in Hamilton Rating Scale for Depression (HAM-D) Total Score by Study Visit and Antidepressant Use at Baseline	Full Analysis Set
Table 14.2.1.6.3	Summary of Change from Baseline in HAM-D Total Score by Study Visit and Age Group	Full Analysis Set
Table 14.2.1.6.4	Model-based Results on Change from Baseline in Hamilton Rating Scale for Depression (HAM-D) Total Score by Study Visit and Age Group	Full Analysis Set
Table 14.2.1.6.5	Summary of Change from Baseline in HAM-D Total Score by Study Visit and Race	Full Analysis Set
Table 14.2.1.6.6	Model-based Results on Change from Baseline in Hamilton Rating Scale for Depression (HAM-D) Total Score by Study Visit and Race	Full Analysis Set
Table 14.2.1.6.7	Summary of Change from Baseline in HAM-D Total Score by Study Visit and BMI (kg/m^2) Group	Full Analysis Set

Table Number	Title	Analysis Set
Table 14.2.1.6.8	Model-based Results on Change from Baseline in Hamilton Rating Scale for Depression (HAM-D) Total Score by Study Visit and BMI (kg/m^2) Group	Full Analysis Set
Table 14.2.1.6.9	Summary of Change from Baseline in HAM-D Total Score by Study Visit and Baseline HAM-D Total Score Category	Full Analysis Set
Table 14.2.1.6.10	Model-based Results on Change from Baseline in Hamilton Rating Scale for Depression (HAM-D) Total Score by Study Visit and Baseline HAM-D Total Score Category	Full Analysis Set
Table 14.2.1.6.11	Summary of Change from Baseline in HAM-D Total Score by Study Visit and Country	Full Analysis Set
Table 14.2.1.6.12	Model-based Results on Change from Baseline in Hamilton Rating Scale for Depression (HAM-D) Total Score by Study Visit and Country	Full Analysis Set
Table 14.2.1.6.13	Summary of Change from Baseline in HAM-D Total Score by Study Visit and Onset of PPD	Full Analysis Set
Table 14.2.1.6.14	Model-based Results on Change from Baseline in Hamilton Rating Scale for Depression (HAM-D) Total Score by Study Visit and Onset of PPD	Full Analysis Set
Table 14.2.1.6.15	Summary of Change from Baseline in HAM-D Total Score by Study Visit and History of PPD	Full Analysis Set
Table 14.2.1.6.16	Model-based Results on Change from Baseline in Hamilton Rating Scale for Depression (HAM-D) Total Score by Study Visit and History of PPD	Full Analysis Set
Table 14.2.1.6.17	Summary of Change from Baseline in HAM-D Total Score by Study Visit and Depression with elevated anxiety (by Baseline HAMA Total Score)	Full Analysis Set
Table 14.2.1.6.18	Model-based Results on Change from Baseline in Hamilton Rating Scale for Depression (HAM-D) Total Score by Study Visit and Depression with elevated anxiety (by Baseline HAMA Total Score)	Full Analysis Set
Table 14.2.1.7.1	Summary of HAM-D Response by Study Visit and Antidepressant Use at Baseline	Full Analysis Set
Table 14.2.1.7.2.1	Model-based Results on Hamilton Rating Scale for Depression (HAM-D) Response by Study Visit and Antidepressant Use at Baseline	Full Analysis Set

Table Number	Title	Analysis Set
Table 14.2.1.7.2.2	Sensitivity Analysis: Model-based Results on Hamilton Rating Scale for Depression (HAM-D) Response by Study Visit and Antidepressant Use at Baseline	Full Analysis Set
Table 14.2.1.7.3	Summary of Hamilton Rating Scale for Depression (HAM-D) Response by Study Visit and Depression with elevated anxiety (by Baseline HAMA Total Score)	Full Analysis Set
Table 14.2.1.7.4	Model-based Results on Hamilton Rating Scale for Depression (HAM-D) Response by Study Visit and Depression with elevated anxiety (by Baseline HAMA Total Score)	Full Analysis Set
Table 14.2.1.7.5	Weighted GEE Model-Based Results on Hamilton Rating Scale for Depression (HAM-D) Response by Study Visit	Full Analysis Set
Table 14.2.1.8.1	Summary of HAM-D Remission by Study Visit and Antidepressant Use at Baseline	Full Analysis Set
Table 14.2.1.8.2.1	Model-based Results on Hamilton Rating Scale for Depression (HAM-D) Remission by Study Visit and Antidepressant Use at Baseline	Full Analysis Set
Table 14.2.1.8.2.2	Sensitivity Analysis: Model-based Results on Hamilton Rating Scale for Depression (HAM-D) Remission by Study Visit and Antidepressant Use at Baseline	Full Analysis Set
Table 14.2.1.8.3	Summary of Hamilton Rating Scale for Depression (HAM-D) Remission by Study Visit and Depression with elevated anxiety (by Baseline HAMA Total Score)	Full Analysis Set
Table 14.2.1.8.4	Model-based Results on Hamilton Rating Scale for Depression (HAM-D) Remission by Study Visit and Depression with elevated anxiety (by Baseline HAMA Total Score)	Full Analysis Set
Table 14.2.1.8.5	Weighted GEE Model-Based Results on Hamilton Rating Scale for Depression (HAM-D) Remission by Study Visit	Full Analysis Set
Table 14.2.2.1.1	Summary of Clinical Global Impression – Improvement (CGI-I) by Study Visit	Full Analysis Set
Table 14.2.2.1.2	Summary of Clinical Global Impression – Improvement (CGI-I) by Study Visit – Categorical Results	Full Analysis Set

Table Number	Title	Analysis Set
Table 14.2.2.2.1	Summary of Clinical Global Impression – Improvement (CGI-I) Response by Study Visit	Full Analysis Set
Table 14.2.2.2.2	Model-based Results on Clinical Global Impression – Improvement (CGI-I) Response by Study Visit	Full Analysis Set
Table 14.2.2.2.3	Sensitivity Analysis: Summary of Clinical Global Impression - Improvement (CGI-I) Response by Study Visit	Full Analysis Set
Table 14.2.2.2.4	Weighted GEE Model-Based Results on Clinical Global Impression - Improvement (CGI-I) Response by Study Visit	Full Analysis Set
Table 14.2.2.2.5	Summary of Clinical Global Impression - Improvement (CGI-I) Response by Study Visit and Depression with elevated anxiety (by Baseline HAMA Total Score)	Full Analysis Set
Table 14.2.2.2.6	Model-based Results on Clinical Global Impression - Improvement (CGI-I) Response by Study Visit and Depression with Elevated Anxiety (by Baseline HAMA Total Score)	Full Analysis Set
Table 14.2.3.1.1	Summary of Clinical Global Impression – Severity (CGI-S) Score by Study Visit	Full Analysis Set
Table 14.2.3.1.2	Model-based Results on Change from Baseline in Clinical Global Impression - Severity (CGI-S) by Study Visit	Full Analysis Set
Table 14.2.3.1.3	Summary of Clinical Global Impression – Severity (CGI-S) by Study Visit – Categorical Results	Full Analysis Set
Table 14.2.3.1.4	Summary of Clinical Global Impression – Severity (CGI-S) Score by Study Visit and Depression with Elevated Anxiety (by Baseline HAMA Total Score)	Full Analysis Set
Table 14.2.3.1.5	Model-based Results on Change from Baseline in Clinical Global Impression - Severity (CGI-S) by Study Visit and Depression with Elevated Anxiety (by Baseline HAMA Total Score)	Full Analysis Set
Table 14.2.4.1.1	Summary of Hamilton Rating Scale for Anxiety (HAM-A) Total Score by Visit	Full Analysis Set
Table 14.2.4.1.2	Model-based Results on Change from Baseline in Hamilton Rating Scale for Depression (HAM-A) Total Score by Study Visit	Full Analysis Set
Table 14.2.4.2.1	Summary of Hamilton Rating Scale for Anxiety (HAM-A) Individual Item Score by Study Visit	Full Analysis Set

Table Number	Title	Analysis Set
Table 14.2.4.2.2	Model-based on Results on Change from Baseline in Hamilton Rating Scale for Anxiety (HAM-A) Individual Item Score by Study Visit	Full Analysis Set
Table 14.2.5.1.1	Summary of Montgomery-Asberg Depression Rating Scale (MADRS) Total Score by Study Visit	Full Analysis Set
Table 14.2.5.1.2	Model-based Results on Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Total Score by Study Visit	Full Analysis Set
Table 14.2.5.2.1	Summary of Montgomery-Asberg Depression Rating Scale (MADRS) Individual Item Score by Study Visit	Full Analysis Set
Table 14.2.5.2.2	Model-based Results on Change from Baseline in Montgomery-Asberg Depression Rating Scale (MADRS) Individual Item Score by Study Visit	Full Analysis Set
Table 14.2.5.3.1	Summary of Montgomery-Asberg Depression Rating Scale (MADRS) Response by Study Visit	Full Analysis Set
Table 14.2.5.3.2	Model-based Results on Montgomery-Asberg Depression Rating Scale (MADRS) Response by Study Visit	Full Analysis Set
Table 14.2.5.3.3	Sensitivity Analysis: Montgomery-Asberg Depression Rating Scale (MADRS) Response by Study Visit	Full Analysis Set
Table 14.2.5.3.4	Weighted GEE Model-Based Results on Montgomery-Asberg Depression Rating Scale (MADRS) Response by Study Visit	Full Analysis Set
Table 14.2.5.4.1	Summary of Montgomery-Asberg Depression Rating Scale (MADRS) Remission by Study Visit	Full Analysis Set
Table 14.2.5.4.2	Model-based Results on Montgomery-Asberg Depression Rating Scale (MADRS) Remission by Study Visit	Full Analysis Set
Table 14.2.5.4.3	Sensitivity Analysis: Montgomery-Asberg Depression Rating Scale (MADRS) Remission by Study Visit	Full Analysis Set
Table 14.2.5.4.4	Weighted GEE Model-Based Results on Montgomery-Asberg Depression Rating Scale (MADRS) Remission by Study Visit	Full Analysis Set
Table 14.2.6.1.1	Summary of Patient Health Questionnaire (PHQ-9) Total Score by Study Visit	Full Analysis Set
Table 14.2.6.1.2	Summary of Patient Health Questionnaire (PHQ-9) Total Score Category by Study Visit	Full Analysis Set

Table Number	Title	Analysis Set
Table 14.2.6.1.3	Model-based Results on Change from Baseline in Patient Health Questionnaire (PHQ-9) Total Score by Study Visit	Full Analysis Set
Table 14.2.7.1.1	Summary of Edinburgh Postpartum Depression Scale (EPDS) Total Score by Study Visit	Full Analysis Set
Table 14.2.7.1.2	Model-based Results on Change from Baseline in Edinburgh Postpartum Depression Scale (EPDS) by Study Visit	Full Analysis Set
Table 14.2.8.1	Summary of Hamilton Rating Scale for Depression (HAM-D) Percent Retention (%) of Day 15 Change from Baseline at Post-Day 15 Visits	Full Analysis Set – Participants who Had HAM-D Improvement at Day 15
Table 14.2.8.2	Summary of Hamilton Rating Scale for Depression (HAM-D) Percent Retention (%) of Day 15 Change from Baseline at Post-Day 15 Visits	Full Analysis Set - Day 15 HAM-D Responders Only
Table 14.2.8.3	Summary of Hamilton Rating Scale for Depression (HAM-D) Response at Post-Day 15 Visits, for HAM-D Responders at Day 15	Full Analysis Set - Day 15 HAM-D Responders Only
Table 14.2.8.4	Summary of Hamilton Rating Scale for Depression (HAM-D) Remission at Post-Day 15 Visits, for HAM-D Remitters at Day 15	Full Analysis Set - Day 15 HAM-D Remitters Only
Table 14.2.8.5	Summary of Hamilton Rating Scale for Depression (HAM-D) Relapse and Rebound Post-Day 15 Visits, for HAM-D Responders at Day 15	Full Analysis Set - Day 15 HAM-D Responders Only
Table 14.2.8.6	Summary of Clinical Global Impression-Improvement (CGI-I) Response at Post-Day 15 Visits, for CGI Responders at Day 15	Full Analysis Set – Day 15 CGI-I Responders Only
Table 14.3.1.1	Overview of Treatment-Emergent Adverse Events	Safety Set
Table 14.3.1.2	Summary of Treatment-Emergent Adverse Events by System Organ Class and Preferred Term	Safety Set
Table 14.3.1.3.1	Summary of On-Treatment Treatment Emergent Adverse Events by System Organ Class and Preferred Term	Safety Set
Table 14.3.1.3.2	Summary of Post-Treatment-Emergent Adverse Events by System Organ Class and Preferred Term	Safety Set
Table 14.3.1.4.1	Summary of Treatment-Emergent Adverse Events by System Organ Class, Preferred Term, and Use of Antidepressant at Baseline	Safety Set

Table Number	Title	Analysis Set
Table 14.3.1.4.2	Summary of Treatment-Emergent Adverse Events by System Organ Class, Preferred Term, and Age Group	Safety Set
Table 14.3.1.4.3	Summary of Treatment-Emergent Adverse Events by System Organ Class, Preferred Term, and Race Group	Safety Set
Table 14.3.1.4.4	Summary of Treatment-Emergent Adverse Events by System Organ Class, Preferred Term, and BMI at Baseline	Safety Set
Table 14.3.1.4.5	Summary of Treatment-Emergent Adverse Events by System Organ Class, Preferred Term, and Baseline HAM-D Total Score	Safety Set
Table 14.3.1.4.6	Summary of Treatment-Emergent Adverse Events by System Organ Class, Preferred Term, and Country	Safety Set
Table 14.3.1.5	Summary of Treatment-Emergent Adverse Events by System Organ Class, Preferred Term, and Maximum Severity	Safety Set
Table 14.3.1.6	Summary of Treatment-Emergent Adverse Events by System Organ Class, Preferred Term, and Relationship to Investigational Product	Safety Set
Table 14.3.1.7	Summary of Treatment Period Treatment-Emergent Adverse Events Leading to Discontinuation of Investigational Product by System Organ Class and Preferred Term	Safety Set
Table 14.3.1.8	Summary of Treatment-Emergent Adverse Events Leading to Withdrawal from the Study by System Organ Class and Preferred Term	Safety Set
Table 14.3.1.9	Summary of Treatment Emergent Adverse Events leading to Investigational Product Dose Reduction by System Organ Class and Preferred Term	Safety Set
Table 14.3.1.9	Summary of Treatment-Emergent Serious Adverse Events by System Organ Class and Preferred Term	Safety Set
Table 14.3.1.10	Summary of Most Common (>2%) Treatment-Emergent Adverse Events by Preferred Term	Safety Set
[redacted]		

Table Number	Title	Analysis Set
[redacted]		

Figures

Figure Number	Title	Analysis Set
Figure 14.2.1.1	Line Plot of LS Mean (±SE) Change from Baseline in Hamilton Rating Scale for Depression (HAM-D) Total Score Over Time by Treatment Group	Full Analysis Set
Figure 14.2.1.1.1	Line Plot of LS Mean (±SE) Change from Baseline in Hamilton Rating Scale for Depression (HAM-D) Total Score over Time by Treatment Group and Antidepressant Use at Baseline	Full Analysis Set
Figure 14.2.1.2	Forest Plot of LS Mean (95% CI) Change from Baseline in Hamilton Rating Scale for Depression (HAM-D) Total Score by Treatment Group and Subgroups on Day 15	Full Analysis Set
Figure 14.2.1.3.1	Bar Chart of Hamilton Rating Scale for Depression (HAM-D) Response over Time	Full Analysis Set
Figure 14.2.1.3.2	Sensitivity Analysis: Bar Chart of Hamilton Rating Scale for Depression (HAM-D) Response over Time	Full Analysis Set
Figure 14.2.1.3.3	Bar Chart of Hamilton Rating Scale for Depression (HAM-D) Response over Time by Antidepressant Use at Baseline	Full Analysis Set
Figure 14.2.1.3.4	Sensitivity Analysis: Bar Chart of Hamilton Rating Scale for Depression (HAM-D) Response over Time by Antidepressant Use at Baseline	Full Analysis Set
Figure 14.2.1.4	Histogram of Hamilton Rating Scale for Depression (HAM-D) Percentage Improvement over Time	Full Analysis Set
Figure 14.2.1.5.1	Bar Chart of Hamilton Rating Scale for Depression (HAM-D) Remission over Time	Full Analysis Set
Figure 14.2.1.5.2	Sensitivity Analysis: Bar Chart of Hamilton Rating Scale for Depression (HAM-D) Remission over Time	Full Analysis Set
Figure 14.2.1.5.3	Bar Chart of Hamilton Rating Scale for Depression (HAM-D) Remission over Time by Antidepressant Use at Baseline	Full Analysis Set
Figure 14.2.1.5.4	Sensitivity Analysis: Bar Chart of Hamilton Rating Scale for Depression (HAM-D) Remission over Time by Antidepressant Use at Baseline	Full Analysis Set
Figure 14.2.1.6	Histogram of Hamilton Rating Scale for Depression (HAM-D) Total Score in Categories over Time	Per Protocol Set
Figure 14.2.1.7	Forest Plot of LS Mean (95% CI) Change from Baseline in Hamilton Rating Scale for Depression (HAM-D) Subscales on Day 15 by Treatment Group	Full Analysis Set

Figure Number	Title	Analysis Set
Figure 14.2.1.8	Forest Plot of LS Mean (95% CI) Change from Baseline at Day 15 in Hamilton Rating Scale for Depression (HAM-D) for Individual Items by Treatment Group	Full Analysis Set
Figure 14.2.2	Bar Chart of CGI-I Response over Time by Treatment Group	Full Analysis Set
Figure 14.2.3	Line Plot of LS Mean (±SE) Change from Baseline in CGI-S over Time by Treatment Group	Full Analysis Set
Figure 14.2.4	Line Plot of LS Mean (±SE) Change from Baseline in Hamilton Rating Scale for Anxiety (HAM-A) Total Score over Time by Treatment Group	Full Analysis Set
Figure 14.2.5	Line Plot of LS Mean (±SE) Change from Baseline in Montgomery-Åsberg Depression Rating Scale (MADRS) Total Score over Time by Treatment Group	Full Analysis Set
Figure 14.2.5.1.1	Bar Chart of Montgomery-Åsberg Depression Rating Scale (MADRS) Response over Time	Full Analysis Set
Figure 14.2.5.1.2	Sensitivity Analysis: Bar Chart of Montgomery-Åsberg Depression Rating Scale (MADRS) Response over Time	Full Analysis Set
Figure 14.2.5.2.1	Bar Chart of Montgomery-Åsberg Depression Rating Scale (MADRS) Remission over Time	Full Analysis Set
Figure 14.2.5.2.2	Sensitivity Analysis: Bar Chart of Montgomery-Åsberg Depression Rating Scale (MADRS) Remission over Time	Full Analysis Set
Figure 14.2.6	Line Plot of Hamilton Rating Scale for Depression (HAM-D) Percent Retention (%) of Day 15 Change from Baseline over Time	Full Analysis Set – Participants who Had HAM-D Improvement at Day 15
Figure 14.2.7	Line Plot of Hamilton Rating Scale for Depression (HAM-D) Percent Retention (%) of Day 15 Reduction from Baseline over Time, for HAM-D Responders at Day 15	Full Analysis Set-Day 15 HAM-D Responders Only
Figure 14.2.8	Bar Chart of Hamilton Rating Scale for Depression (HAM-D) Response at Post-Day 15 Visits, for HAM-D Responders at Day 15	Full Analysis Set-Day 15 HAM-D Responders Only
Figure 14.2.9	Bar Chart of Hamilton Rating Scale for Depression (HAM-D) Remission at Post-Day 15 Visits, for HAM-D Remitters at Day 15	Full Analysis Set-Day 15 HAM-D Remitters Only

Figure Number	Title	Analysis Set
Figure 14.2.10	Bar Chart of Clinical Global Impression-Improvement (CGI-I) Response at Post-Day 15 Visits, for CGI-I Responders at Day 15	Full Analysis Set-Day 15 CGI-I Responders Only
Figure 14.2.11	Line Plot of LS Mean (±SE) Change from Baseline in Edinburgh Postpartum Depression Scale (EPDS)	Full Analysis Set
[redacted]		

Listings

Listing Number	Title	Analysis Set
Listing 16.1.7	Participant Randomization	Randomized Set
Listing 16.2.1.1	Participant Disposition	Randomized Set
Listing 16.2.1.2	Premature Discontinuation from Investigational Product and/or Withdrawal from the Study	Safety Set
Listing 16.2.2.1.1	Protocol Deviations	Full Analysis Set
Listing 16.2.2.1.2	A Listing of All Participants Affected by the COVID-19 Related Study Disruptions	Full Analysis Set
Listing 16.2.2.2	Inappropriate Investigational Product Consumption	Safety Set
Listing 16.2.3.1	Inclusion/Exclusion Criteria Violations	Randomized Set
Listing 16.2.4.1	Demographics	Safety Set
Listing 16.2.4.2	Baseline Characteristics	Safety Set
Listing 16.2.4.2.1	Incorrect Stratification related to Antidepressant Use at Baseline	Safety Set
Listing 16.2.4.2.2	Diagnostic Laboratory Results	Safety Set
Listing 16.2.4.3.1	Disease History	Safety Set
Listing 16.2.4.3.2	Medical and Surgical History	Safety Set
Listing 16.2.4.3.3	Participant History of Psychiatric Disorder	Safety Set
Listing 16.2.4.3.4	Family History of Postpartum Depression	Safety Set
Listing 16.2.4.4.1	Prior and Concomitant Psychotropic Medications	Safety Set
Listing 16.2.4.4.2	Prior and Concomitant Non-Psychotropic Medications	Safety Set
Listing 16.2.4.4.3	Listing of Use of Antidepressant during the Study	Safety Set

Listing Number	Title	Analysis Set
Listing 16.2.4.4.4	Concomitant Procedures	Safety Set
Listing 16.2.5.1	Investigational Product Administration	Safety Set
Listing 16.2.5.2	Investigational Product Exposure	Safety Set
Listing 16.2.5.3	Investigational Product Adherence	Full Analysis Set
[redacted]		
Listing 16.2.6.1	Hamilton Rating Scale for Depression (HAM-D)	Full Analysis Set
Listing 16.2.6.2	Clinical Global Impression (CGI) – Improvement	Full Analysis Set
Listing 16.2.6.3	Clinical Global Impression (CGI) – Severity	Full Analysis Set
Listing 16.2.6.4	Hamilton Anxiety Rating Scale (HAM-A)	Full Analysis Set
Listing 16.2.6.5	Montgomery-Asberg Depression Rating Scale (MADRS)	Full Analysis Set
Listing 16.2.6.6	Summary of Edinburgh Postpartum Depression Scale (EPDS) Individual Item Scores	Full Analysis Set
Listing 16.2.6.7	Patient Health Questionnaire (PHQ-9)	Full Analysis Set
Listing 16.2.6.8	Hamilton Rating Scale for Depression (HAM-D) with Response Retention (%), Response, Remission, Relapse and Rebound	Full Analysis Set, SAGE-217 Only
Listing 16.2.7.1	Treatment-Emergent Adverse Events	Safety Set
Listing 16.2.7.2	Non-fatal Serious Adverse Events	Safety Set
Listing 16.2.7.3	Adverse Events Leading to Death	Safety Set
Listing 16.2.7.4	Adverse Events with Onset Prior to First Dose of Investigational Product	Safety Set
Listing 16.2.7.5	Adverse Events leading to Premature Discontinuation of Investigational Product	Safety Set
Listing 16.2.7.6	Adverse Events leading to Premature Withdrawal from the Study	Safety Set
Listing 16.2.7.7	Listing of Adverse Events Experienced by Participants with Investigational Product Dose Reduction due to Adverse Events	Safety Set
Listing 16.2.7.8	Listing of Participants with COVID-19 Related Adverse Events	Safety Set
[redacted]		

Listing Number	Title	Analysis Set
[redacted]		
Listing 16.2.8.1.6	Pregnancy Test	Safety Set
[redacted]		
Listing 16.2.8.4	Physical Examination	Safety Set
[redacted]		