9. DOCUMENTATION OF STATISTICAL METHODS

Statistical Analysis Plan

SAGE THERAPEUTICS INCORPORATED

Statistical Analysis Plan

Methods

Protocol Number 547-PPD-202B / NCT02942004

A Multicenter, Randomized, Double-Blind, Parallel-Group, Placebo-Controlled Study Evaluating the Efficacy, Safety, and Pharmacokinetics of SAGE-547 Injection in the Treatment of Adult Female Subjects with Severe Postpartum Depression and Adult Female Subjects with Moderate Postpartum Depression

Author of SAP: [redacted]

[redacted]

Version: Version 1.0

Date of SAP: 27 OCTOBER 2017

Sponsor Representative: [redacted] **PhD**

[redacted]

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

CONFIDENTIAL
This document contains confidential information. Any use, distribution, or disclosure without the prior written consent of Sage Therapeutics Incorporated is strictly prohibited except to the extent required under applicable laws or regulations. Persons to whom the information is disclosed must be informed that the information is confidential and may not be further disclosed by them.

Authorization Signature Page

A Multicenter, Randomized, Double-Blind, Parallel-Group, Placebo-Controlled Study Evaluating the Efficacy, Safety, and Pharmacokinetics of SAGE-547 Injection in the Treatment of Adult Female Subjects with Severe Postpartum Depression and Adult Female Subjects with Moderate Postpartum Depression

Author:

[redacted]

Name: [redacted] 30 Oct 2017
Date

Position: [redacted]
Company: [redacted]

Approved by:

[redacted], PhD, [redacted] 10/30/2017
Sage Therapeutics, Inc. Date

[redacted] 30Oct2017
Helen Colquhoun, MD, Vice President of Medical Science Date
Sage Therapeutics, Inc.

[redacted] 30 Oct 2017
[redacted], PhD Date
[redacted]
Sage Therapeutics, Inc.

1 TABLE OF CONTENTS

2 LIST OF ABBREVIATIONS

Abbreviation or Specialist Term	Explanation
AE	adverse event
ANCOVA	analysis of covariance
ATC	anatomical therapeutic chemical
AUC	area under the concentration-time curve
AUC_{∞}	area under the concentration-time curve from time zero to infinity
BIMF	Barkin Index of Maternal Functioning
BMI	body mass index
bpm	beats per minute
CGI-I	Clinical Global Impression - Improvement
CGI-S	Clinical Global Impression - Severity
C_{max}	maximum (peak) plasma concentration
CS	clinically significant
C_{ss}	steady-state drug concentration in the plasma during oral intake
C_{avg}	steady-state drug concentration in the plasma
C-SSRS	Columbia-Suicide Severity Rating Scale
ECG	electrocardiogram
eCRF	electronic case report form
EPDS	Edinburgh Postnatal Depression Scale
ET	early termination
FCS	fully conditional specification
GAD-7	Generalized Anxiety Disorder 7-Item Scale
GEE	Generalized Estimating Equation
HAM-D	Hamilton Rating Scale for Depression, 17-item
HCRU	healthcare resource utilization
HIV	human immunodeficiency virus
IV	intravenous
Kg	kilogram
m	meter
MADRS	Montgomery-Åsberg Depression Rating Scale
Max	maximum
MedDRA	Medical Dictionary for Regulatory Activities
Min	minimum
mmHg	millimeter of mercury
MMRM	mixed effects model for reported measures
msec	millisecond
n	number
NCS	not clinically significant
PHQ-9	Patient Health Questionnaire
PK	pharmacokinetic(s)
PPD	postpartum depression
PRO	patient reported outcome

PT	preferred term
QTcB	QT corrected with Bazett's formula
QTcF	QT corrected with Fridericia's formula
SAE	serious adverse event
SAP	statistical analysis plan
SBECD	betadex sulfobutyl ether sodium
SD	standard deviation
SF-36	Short Form-36
SI	International System of Units
SOC	system organ class
ss	steady state
TEAE	treatment-emergent adverse event
Tmax	time at maximum (peak) plasma concentration
WHO	World Health Organization
WHO-DDE	World Health Organization-Drug Dictionary Enhanced

3 INTRODUCTION

This Statistical Analysis Plan (SAP) is based on the approved clinical study protocol amendment 4, version 5.0 dated 16th March 2017, case report form (CRF) dated 02nd March 2017 and CRF dated 20th March 2017.

Protocol 547-PPD-202 is an umbrella protocol that describes methods for three parts, referred to as Parts A, B, and C (hereafter referred to as Study 547-PPD-202A, Study 547-PPD-202B, and Study 547-PPD-202C, respectively). These three parts were conducted as three separate studies and will be analyzed and reported as such: Study 547-PPD-202A, Study 547-PPD-202B, and Study 547-PPD-202C.

The purpose of the SAP is to describe in detail the statistical methodology and the statistical analyses to be conducted for data presentation for Study 547-PPD-202B, a randomized, multicenter, parallel-group, placebo-controlled study evaluating doses up to 60 or 90 µg/kg/h of SAGE-547 Injection in patients with severe postpartum depression.

All analyses and summary outputs will be generated using SAS® version 9.3 (or higher).

4 STUDY OBJECTIVES

4.1 Primary Objective

The primary objective of Study 547-PPD-202B is to determine if SAGE-547 Injection infused intravenously for 60 hours at up to 90 µg/kg/h reduces depressive symptoms in subjects with severe postpartum depression (PPD) compared to placebo injection as assessed by the change from baseline in Hamilton Rating Scale for Depression (HAM-D) total score.

4.2 Secondary Objectives

The secondary objectives of Study 547-PPD-202B are:

- To determine if SAGE-547 infusion at up to 60 µg/kg/h for 60 hours reduces depressive symptoms in subjects with severe PPD compared to placebo injection as assessed by the change from baseline in HAM-D total score.
- To determine if SAGE-547 Injection infused intravenously for 60 hours reduces depressive symptoms compared to placebo injection as assessed by HAM-D response, HAM-D remission, change from baseline in Montgomery-Asberg Depression Rating Scale (MADRS) total score, Clinical Global Impression - Improvement (CGI-I) response, and changes from baseline in HAM-D subscale and individual item scores.
- To determine if SAGE-547 Injection infused intravenously for 60 hours reduces other mood symptoms compared to placebo injection as assessed by changes from baseline in the Generalized Anxiety Disorder 7-Item Scale (GAD-7) total score.
- To evaluate the safety and tolerability of SAGE-547 Injection compared with placebo as assessed by the incidence of adverse events (AEs), vital sign measurement, clinical

laboratory evaluations, electrocardiogram (ECG) parameters, and the Columbia Suicide Severity Rating Scale (C-SSRS).

4.3 Other Objectives

The other objectives of Study 547-PPD-202B are:

- To determine if SAGE-547 Injection infused intravenously for 60 hours reduces subject-reported depressive symptoms compared to placebo injection as assessed by the change from baseline in Edinburgh Postnatal Depression Scale (EPDS) total score and the change from baseline in Patient Health Questionnaire (PHQ-9) total score.
- To determine if SAGE-547 Injection infused intravenously for 60 hours improves maternal behaviors compared to placebo injection as assessed by the change from baseline in Barkin Index of Maternal Functioning (BIMF) total and sub-scale scores.
- To determine if SAGE-547 Injection infused intravenously for 60 hours improves the general health status compared to placebo as assessed by the change from baseline in the Short Form-36 (SF-36) total score at Day 7 and Day 30.

4.4 Pharmacokinetic Objectives

The Pharmacokinetic (PK) objective of Study 547-PPD-202B is to assess the pharmacokinetic (PK) profile of SAGE-547, metabolites of SAGE-547, and betadex sulfobutyl ether sodium (SBECD).

5 STUDY ENDPOINTS

5.1 Primary Endpoint

- Change from baseline in HAM-D total score at the end of infusion (Day 3, 60 hours)

5.2 Secondary Endpoints

5.2.1 Key Secondary Efficacy Endpoints

- Change from baseline in HAM-D total score at the end of follow up (Day 30)

5.2.2 Other Secondary Efficacy Endpoints

- Change from baseline in HAM-D total score at each time point
- HAM-D response (defined as having a 50% or greater reduction from baseline in HAM-D total score) at each time point
- HAM-D remission (defined as having a HAM-D total score of ≤7) at each time point
- Change from baseline in MADRS total score at each time point
- CGI-I response (defined as a CGI-I score of “very much improved” or “much improved”) at each time point
- Change from baseline in HAM-D subscale and individual item scores at each time point

- Change from baseline in GAD-7 total score at each time point
- Time to first start or increase in the dose of any anti-depressant medication
- Time to first decrease in the dose of any anti-depressant medication

5.2.3 Safety Endpoints

- Frequency of AEs, severity and seriousness of AEs
- Change from baseline in clinical laboratory measures: hematology, serum chemistry, coagulation and urinalysis parameters (including observed values and change from baseline)
- Vital signs parameters (including observed values and change from baseline)
- ECG parameters (including observed values and change from baseline)
- Concomitant medication usage
- Suicidal ideation using the C-SSRS

5.3 Other Endpoints

- Change from baseline in EPDS total score at each time point
- Change from baseline in PHQ-9 total score at each time point
- Change from baseline in BIMF total and subscale scores at each time point
- Change from baseline in SF-36 subscale scores at each time point
- Healthcare resource utilization (HCRU)

5.4 Pharmacokinetic Endpoints

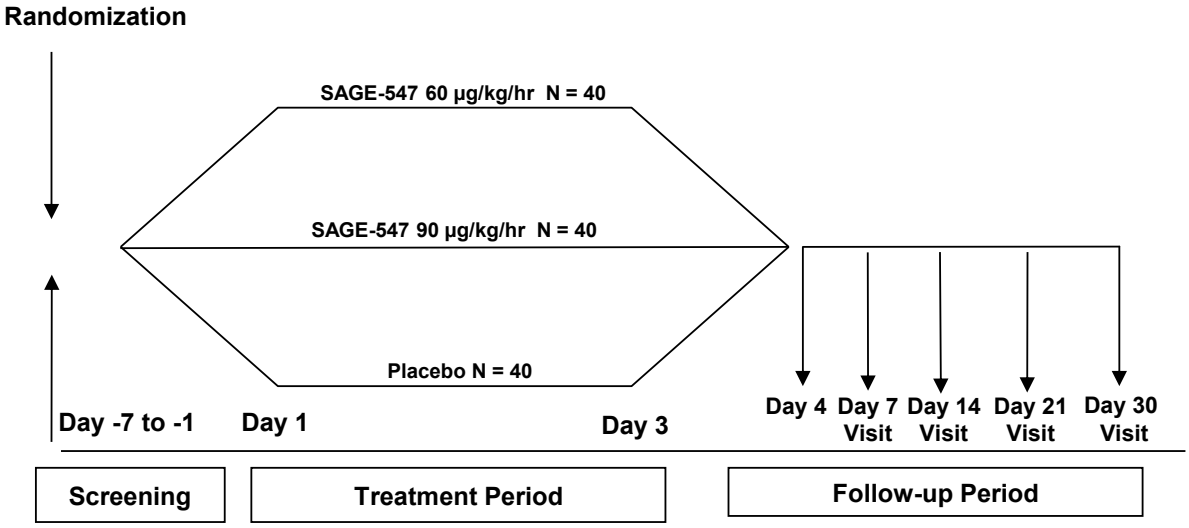
- Plasma concentrations of SAGE-547 and calculated PK parameters
- Plasma concentrations of metabolites of SAGE-547
- Plasma concentrations of SBECD

6 STUDY DESIGN

6.1 Overall Design

Study 547-PPD-202B will consist of a Screening Period (up to 7-days [Day -7 to -1]), a 3-day (60 hours of treatment and an additional 12 hours for completion of 72-hour assessments) Treatment Period, and a 30-day Follow-up Period. Subjects must remain as inpatients during the study Treatment Period, which is approximately 72 hours/3 days in duration. The Screening Period assessments may be conducted on an inpatient or an outpatient basis. The Follow-up Period assessments are conducted on an outpatient basis. The study design for 547-PPD-202B is presented in Figure 1.

Figure 1: Study Design - 547-PPD-202B



SAGE-547 Injection or placebo will be administered at the study center. Subjects will be monitored for safety during the Treatment and Follow-up Periods (through Study Day 30 [±3 days]) including monitoring for AEs/SAEs, routine clinical laboratory assessments, physical examination, vital signs, and ECG.

The end of the Screening Period coincides with the beginning of the Treatment Period. The Treatment Period is the period of Day 1 of study drug intravenous (IV) infusion through completion of the infusion on Day 2 and up to Day 3. Subjects will be confined to the study center from Day 1 until after the 72-hour assessments have been conducted on Day 3.

In Study 547-PPD-202B, once subjects are confirmed as eligible for the study, they will be randomized to one of three treatment groups (SAGE-547 60 µg/kg/hour, SAGE-547 90 µg/kg/hour, or placebo) on a 1:1:1 basis. For the 60 µg/kg/hour group, subjects will receive 30 µg/kg/hour (0-4 hours), then 60 µg/kg/hour (4-56 hours), followed by 30 µg/kg/hour (56-60 hours). For the 90 µg/kg/hour group, subjects will begin a 4-hour dose titration period of 30 µg/kg/hour (0-4 hours), then 60 µg/kg/hour (4-24 hours), then 90 µg/kg/hour (24-52 hours); followed by a decrease to 60 µg/kg/hour (52-56 hours), and 30 µg/kg/hour (56-60 hours) as shown in Table 1 and Figure 2. Subjects in the placebo group will receive infusion rates equivalent to either the 60 µg/kg/hour or 90 µg/kg/hour group.

Table 1: Infusion Rates for 547-PPD-202B

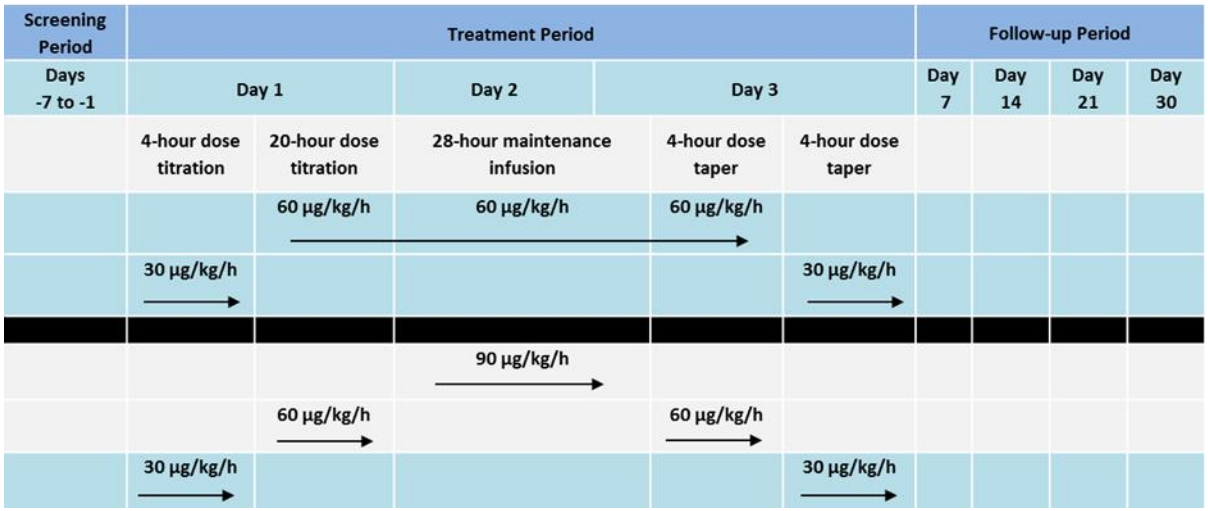
	Infusion Rate (µg/kg/hour)				
SAGE-547 Dose	**Day 1 0-4 hours**	**Day 1 4-24 hours**	**Day 2-3 24-52 hours**	**Day 3 52-56 hours**	**Day 3 56-60 hours**
60 µg	30	60	60	60	30
90 µg	30	60	90	60	30

Total SAGE-547 Injection or placebo dosing will occur over 60 hours.

The end of the Treatment Period coincides with the beginning of the Follow-up Period. Subjects will attend the clinic for safety follow-up assessment at 1 week (7±1d), 2 weeks (14±2d), 3 weeks (21±1d), and 1 month (30±3d) after the initiation of the study drug infusion.

The study design and timing of infusion is shown in Figure 2.

Figure 2: Study Design and Timeline for Dosing – Study 547-PPD-202B



6.2 Sample Size and Power

Using a two-sided test at an alpha level of 0.05, a sample size of 40 evaluable subjects per group (120 total) for 547-PPD-202B would provide 90% power to detect an effect size of 0.75 between the SAGE-547 and placebo groups with regard to the primary outcome variable of change from baseline in HAM-D total score. An effect size of 0.75 corresponds to a placebo-adjusted difference of 9.0 points in the change from baseline in HAM-D total score at 60 hours with an assumed common standard deviation of 12 points.

6.3 Randomization

Up to 120 subjects will be randomized. Subjects will be randomized to one of three treatment groups (SAGE-547 60 µg/kg/hour, SAGE-547 90 µg/kg/hour, or placebo) on a 1:1:1 basis.

For subjects screened to Version 3.0 or lower of the protocol, randomization will not be stratified. For subjects screened to Version 4.0 or greater of the protocol, randomization will be stratified by antidepressant use at baseline (yes/no). Subjects will be randomized within stratum to receive SAGE-547 or placebo according to a computer-generated randomization schedules.

6.4 Blinding and Unblinding

This is a double-blind study. Subjects will be randomized to receive SAGE-547 or matching placebo; subjects, clinicians, and the clinical site study team will be blinded to treatment allocation until the study is unblinded at final database lock. The pharmacist, who will

prepare the infusion bags according to the randomization schedule, and an unblinded Monitor, who will perform drug accountability during the study, will be unblinded.

In the event of a medical emergency, the Principal Investigator will discuss with the Medical Monitor if unblinding is warranted. If there is agreement to unblind treatment assignment, the unblinding procedure described in the Safety Management Plan for the study will be followed. In all cases where the study drug allocation for a subject is unblinded, pertinent information (including the reason for unblinding) must be documented in the subject's records and on the electronic case report form (eCRF).

7 MODIFICATIONS

7.1 Modifications from the Approved Clinical Study Protocol

Protocol Text	SAP Text
1. No protocol text.	1. Section 9.3.6: For efficacy analysis purposes, centers with fewer than 15 subjects per center for Part B will be pooled within geographic regions (North America). Within a region, first centers with fewer than 15 subjects will be ranked by size. Starting with the largest site, centers will be pooled until 15 or more subjects are reached to create a unique regional center. Continue this process of creating unique regional centers until there are less than 15 subjects remaining. The remaining centers will then be added to the last unique regional center created. Unless the combined number of subjects from all centers with fewer than 15 subjects in a region is fewer than 15 subjects, all pooled centers should have at least 15 subjects.
2. Section 13.2/page 62: A sensitivity analysis may be carried out to investigate the impact of missing data if more than 5% of subjects are missing primary endpoint assessments.	2. Section 9.1.3: Sensitivity analyses may be carried out to investigate the impact of missing data if more than 10% of subjects are missing HAM-D or MADRS total scores. The sensitivity analyses will only be performed on the endpoint(s) with sufficient missing data. Multiple imputation and Last Observation Carried Forward techniques will be considered.

7.2 Modifications from the Approved Statistical Analysis Plan

Not applicable.

7.3 Modifications from the Approved DMC Charter

Not applicable, as there is no DMC for this study.

8 ANALYSIS SETS

8.1 All Enrolled Set

The All Enrolled Set will include all subjects who have given written informed consent. This analysis set will be used for subject disposition.

8.2 All Randomized Set

The All Randomized Set will include the subset of subjects from the All Enrolled Set who have been randomized. Subjects will be classified according to randomized treatment. This analysis set will be used for subject disposition, demographic characteristics, and baseline characteristic summaries.

8.3 Safety Set

The Safety Set will include all randomized subjects who start the infusion of study drug. Subjects will be classified according to actual treatment received. This analysis set will be used for all safety analyses.

8.4 Efficacy Set

The Efficacy Set will include the subset of the Safety Set who have a valid baseline HAM-D assessment and at least one post-baseline HAM-D assessment. Subjects will be classified according to randomized treatment. This analysis set will be used for all efficacy analyses, as well as for demographic and baseline characteristics summaries.

8.5 Per Protocol Set

The Per Protocol Set will include the subset of the Efficacy Set who complete the 60-hour infusion without significant protocol violations or deviations. Completing the 60-hour infusion is defined as having total infusion time from all segments between 54 hours and 66 hours. Subjects will be classified according to randomized treatment. This analysis set will be used for supportive analyses of the primary and selected secondary endpoints.

8.6 Pharmacokinetics Set

The PK Set will include the subset of the Safety Set who has at least one evaluable PK sample. Subjects will be classified according to actual treatment received. This analysis set will be used for all PK analyses.

9 STATISTICAL ANALYSIS

9.1 General Considerations

Unless otherwise specified, continuous endpoints will be summarized with n, mean, standard deviation (SD), median, minimum (min) and maximum (max). The minimum and maximum will be reported with the same degree of precision (ie, the same number of decimal places) as the raw data. Measures of location (mean and median) will be reported to one degree of precision more than the raw data and measures of spread (standard deviation) will be reported to two degrees of precision more than the raw data. In addition, change from baseline values 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 one decimal place except for 0% which will be presented as “0” and 100% which will be presented as “100”.

All analyses and summary outputs will be generated using SAS® version 9.3 (or higher).

All subject data, including those derived, will be presented in the subject data listings; listings will display all subjects who were enrolled. In general, the subject data listings will be sorted by randomized treatment group, subject number and assessment visit and date (and time, if applicable). The summary tables will be presented descriptively by treatment group.

9.1.1 Study Day Definition

Study day is defined as follows:

- The day when the study drug infusion is started 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".

9.1.2 Baseline Definition

For the purpose of all safety and efficacy analyses where applicable, baseline is defined as the last non-missing measurement prior to the start of study drug administration.

9.1.3 Missing Data

Every attempt will be made to avoid missing data. All subjects will be used in the analyses, as per the analysis sets using all non-missing data available. SF-36 subscales will be calculated and missing responses will be handled using the built-in scoring tool. For all the other efficacy endpoints, the pro-rating approach will be considered when calculating the total scores. If no more than 20% of item responses are missing for a given subject on the same scale, we replace the missing responses with the mean score on all other non-missing responses, or the maximum possible values for the missing responses, whichever is small; otherwise, if more than 20% of item responses are missing for a given subject on the same scale, the total score will not be calculated and will be left as missing. See Section 9.3.1, 9.3.2 and 9.3.3 for the details of this approach on each endpoint.

A sensitivity analysis may be carried out to investigate the impact of missing data if more than 10% of subjects are missing primary endpoint assessment (ie, HAM-D total score). Two techniques will be considered for the sensitivity analysis: 1) Dropout reason based multiple imputation; 2) Stratified Wilcoxon rank-sum test. See Section 9.3.7.1.4 for details.

Safety data will not be subject to any imputation and will be summarized on an observed case basis.

The following conventions will be used for missing adverse event dates and prior and concomitant medication dates.

9.1.3.1 Adverse Events Onset Date

If the AE onset date is missing in which the day, month, and year are all unknown, the AE onset date is set to the date of initiation of the blinded treatment.

For the partial AE onset date and time,

- If the year is present and the month and day are missing:

- If year of AE onset = year of initiation of the blinded treatment, then set month and day to month and day of initiation of the blinded treatment
- If year of AE onset < year of initiation of the blinded treatment, then set month and day to December 31st.
- If year of AE onset > year of initiation of the blinded treatment, then set month and day to January 1st.

- If the year and day are present and the month is missing:
  - If year of AE onset = year of initiation of the blinded treatment, then set month to month of initiation of the blinded treatment
  - If year of AE onset < year of initiation of the blinded treatment, then set month to December.
  - If year > year of initiation of the blinded treatment, then set month to January.
- If the month and year are present and the day is missing:
  - If year of AE onset = year of initiation of the blinded treatment and:
    - if month = month of initiation of the blinded treatment then set day to day of initiation of the blinded treatment
    - if month < month of initiation of the blinded treatment then set day to last day of month
    - if month > month of initiation of the blinded treatment then set day to 1st day of month
  - If year of AE onset < year of initiation of the blinded treatment then set day to last day of month
  - If year of AE onset > year of initiation of the blinded treatment then set day to 1st day of month

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

9.1.3.2 Prior and Concomitant Medications Date

If the start date (or end date) of medication is completely missing in which the day, month, and year are all unknown or only the day is known, then the start date (or end date) will not be imputed.

For the partial start date of medication,

- If the year is present and the month and day are missing, set month and day to January 1.
- If the year and day are present and the month is missing, set month to January.
- If the year and month are present and the day is missing, set day to 1st day of month.
- If the imputed start date of medication is after the non-imputed end date 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, set month and day to December 31.
- If the year and day are present and the month is missing, set month to December.
- If the year and month are present and the day is missing, set day to last day of the month.
- If the imputed end date of medication is before the non-imputed start date of medication, then the end date will be set to the start date of medication.

9.2 Background Characteristics

9.2.1 Subject Disposition

The summaries of subject disposition will include the number and percentage of subjects who were enrolled, failed screening, randomized, received SAGE-547 Injection or placebo, prematurely discontinued, and completed the study. The number and percentage of subjects will also be summarized for each reason for premature discontinuation. In addition, the number of subjects whose data should be used for the planned analyses will be identified for each respective analysis set (ie, Safety, Efficacy, Per Protocol, and PK Set).

For screen failure subjects, reasons for screen failure will be listed. A screened subject is defined as any subject who signed the study specific informed consent. A screen failure subject is defined as any subject who is screened but failed to meet study requirements (inclusion/exclusion criterion) during screening. Subjects were screened for both Part B and Part C. Therefore a common, pooled set of screen failed subjects is displayed in disposition tables for both study parts.

Study completion and discontinuation details, eligibility, treatment allocation, population assignment, and protocol deviations will be presented in listings.

Subject disposition analysis will be performed for both All Enrolled Set and All Randomized Set.

9.2.2 Demographics and Baseline Characteristics

Demographic data, such as age, gender, race and ethnicity, and baseline characteristics, such as height, weight, body mass index (BMI, calculated as weight (kg)/ [height $(m)^2$]), and baseline antidepressants use, will be listed and summarized using the All Randomized Set and Efficacy Set. For this summary as well as for the subsequence analyses, baseline antidepressant use will be defined as taking any medications belonging to the anatomical therapeutic checmical (ATC) 3 term N06A or N05A, or with indications containing terms depression, postpartum depression, major depression, PPD, MDD, or mood on Day 1.

Age is calculated using the following SAS formula:

age = INT(INTCK('MONTH', date of birth, date of informed consent)/12);

IF MONTH(date of birth) = MONTH(date of informed consent)

THEN age = age -(DAY(date of birth)>DAY(date of informed consent));

Hepatitis, human immunodeficiency virus (HIV), drug and alcohol, and pregnancy screening results will be collected and listed but not summarized, as they are considered part of the inclusion/exclusion criteria.

9.2.3 Medical/family History

Medical history collected at screening will be coded using the Medical Dictionary for Regulatory Activities (MedDRA), Version 19.1. Medical history data will be summarized by system organ class (SOC) and preferred term (PT) and listed by subject using the Safety Set.

Type of psychiatric disorders, family history of psychiatric and perinatal conditions will be summarized using the Safety Set. The subject history of psychiatric disorders including all episodes of PPD will be listed. Family history of psychiatric and perinatal conditions will be listed.

9.2.4 Study Drug Exposure and Compliance

Drug administration information collected on the eCRF will be listed by subject using the Safety Set. Details will include type of infusion (loading/maintenance/taper), dose, infusion rate and unit, start date/time of dose, and end date/time of dose, infusion stopped after this dose. The dose categories and infusion rate for given dose are listed in Table 2.

Table 2: Dose Categories and Infusion Rate for Given Dose for 547-PPD-202B

Dose Category	**Infusion Rate for Given Dose (µg/kg/hour)**		
	60 µg	90 µg	
Titration	30 (Day 1 0-4 hours)	30 (Day 1 0-4 hours)	60 (Day 1 4-24 hours)
Maintenance	60 (Day 1 4 hours - Day 3 56 hours)	90 (Day 2 24 hours - Day 3 52 hours)	
Taper	30 (Day 3 56-60 hours)	60 (Day 3 52-56 hours)	30 (Day 3 56-60 hours)

Overall exposure to study drug is defined as total number of hours treated with study drug during the study, which is calculated as (end time of infusion – start time of infusion [in seconds])/3600.

Exposure to study drug in each planned infusion rate segment (loading dose, maintenance dose, and tapering dose) is defined as total number of hours treated with study drug during the corresponding segment, which is calculated as (end time of infusion in the corresponding segment-start time of infusion in the corresponding segment [in seconds])/3600.

Study drug compliance (%) is calculated as [Actual total dose (in µg/kg)/Planned total dose (in µg/kg)] x 100.

Overall exposure to study drug, exposure to study drug in each infusion rate segment and study drug compliance will be summarized using the Safety Set.

9.3 Efficacy Analysis

All efficacy analyses will be performed for the Efficacy Set unless otherwise specified.

9.3.1 Definition of Primary Efficacy Variable(s)

The primary efficacy variable is the Hamilton Rating Scale for Depression (HAM-D).

HAM-D consists of 17 items that will be used to rate the severity of depression in subjects who are already diagnosed as having a Major Depressive Episode, no earlier than the third trimester and no later than the first 4 weeks following delivery using the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) Axis I Disorders (SCID-I).

The 17-item HAM-D comprises 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 three individual items are missing, the HAM-D total score will not be calculated and will be left as missing. If less than or equal to three individual item scores are missing, the missing item scores will be imputed by the maximum possible value of the item or the mean of all other available item scores, whichever is smaller to calculate the HAM-D total score. Imputed individual scores will be rounded to the nearest integer.

9.3.2 Definition of Secondary Efficacy Variable(s)

The secondary efficacy variables are HAM-D Response, HAM-D Remission, Montgomery-Asberg Depression Rating Scale (MADRS), Clinical Global Impression-Improvement (CGI-I), Clinical Global Impression-Severity (CGI-S), HAM-D Subscales and Individual Items, the Generalized Anxiety Disorder 7-Item scale (GAD-7) and post-baseline antidepressant use. These secondary efficacy variables are described in the following sections.

9.3.2.1 HAM-D Response and HAM-D Remission

HAM-D response will be defined as having a 50% or greater reduction from baseline in HAM-D total score. Any subject who meets this criterion will be defined as a HAM-D Responder.

HAM-D remission will be defined as having a HAM-D total score of ≤7. Any subject who meets this criterion will be defined as a subject in HAM-D Remitter.

9.3.2.2 Montgomery-Åsberg Depression Rating Scale (MADRS)

The MADRS is a 10-item diagnostic questionnaire that psychiatrists use to measure the severity of depressive episodes in subjects with mood disorders. It was designed as an adjunct to the HAM-D that would be more sensitive to the changes brought on by antidepressants and other forms of treatment than the Hamilton Scale.

Higher MADRS scores indicate more severe depression, and each item yields a score of 0 to 6. 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 two individual items are missing, the MADRS total score will not be calculated and will be left as missing. If less than or equal to two 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. Imputed individual scores will be rounded to the nearest integer.

9.3.2.3 Clinical Global Impression – Improvement (CGI-I) and Clinical Global Impression – Severity (CGI-S)

The CGI-I employs a 7-point Likert scale to measure the overall improvement in the subject's condition posttreatment. The Investigator will rate the subject's total improvement whether or not it is due entirely to drug treatment. Response choices include: 0=not assessed, 1=very much improved, 2=much improved, 3=minimally improved, 4=no change, 5=minimally worse, 6=much worse, and 7=very much worse.

CGI-I Response will be defined as having a CGI-I score of "very much improved" or "much improved".

The Clinical Global Impression - Severity (CGI-S) uses a 7-point Likert scale to rate the severity of the subject's illness at the time of assessment, relative to the clinician's past experience with subjects who have the same diagnosis. Considering total clinical experience, a subject 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.

9.3.2.4 HAM-D subscales and individual item scores

The HAM-D can be summarized using four subscales: Core, Anxiety, Bech-6 and Meier. Table 3 lists the items used to calculate each subscale.

Table 3: HAM-D Subscales and Items for Calculating Each Subscale

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 calculate HAM-D total score to calculate the subscale.
Anxiety	Anxiety psychic Anxiety somatic Somatic symptoms gastrointestinal Somatic symptoms general Hypochondriasis Loss weight	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.
Meier	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.

9.3.2.5 Generalized Anxiety Disorder 7-item Scale (GAD-7)

The GAD-7 is a patient-rated generalized anxiety symptom severity scale. Scoring for GAD-7 is calculated by assigning scores of 0, 1, 2, and 3 to the response categories, respectively, of "not at all sure," "several days," "over half the days," and "nearly every day."

GAD-7 total score will be calculated as the sum of the 7 individual item scores, which ranges from 0 to 21. If more than one individual items are missing, the GAD-7 total score will not be calculated and will be left as missing. If only one individual item score is missing, the missing item score will be imputed by the mean of all other available item scores to calculate the GAD-7 total score. As a measure of severity of anxiety, GAD-7 total score can be categorized as: 0 to 4 = minimal anxiety, 5 to 9 = mild anxiety, 10 to 14 = moderate anxiety, and 15 to 21 = severe anxiety. As a measure of severity of anxiety, GAD-7 total score can be categorized as: 0 to 4 = minimal anxiety, 5 to 9 = mild anxiety, 10 to 14 = moderate anxiety, and 15 to 21 = severe anxiety.

The GAD-7 response is defined as having a GAD-7 total score of 0 to 4.

9.3.2.6 Post-baseline Antidepressant Use

The doses of all antidepressant medications will be recorded throughout the study. No changes and/or additions to antidepressant or anxiolytic medicine will be allowed from 14 days prior to the initiation of the blinded study drug infusion to completion of the H72 assessments. Rescue antidepressant medications will be identified as all medications belonging to the ATC 3 term N06A or N05A, or with indication containing terms, depression, postpartum depression, major depression, PPD, MDD, or mood, starting or with increase in dose on or after 4 days from the initiation of infusion. But if the dosing frequency is "as needed", the medication will not be classified as rescue antidepressant medications.

For each rescue antidepressant medication, the time to start of this medication is calculated as the date of start minus the start date of the study drug infusion. For each antidepressant medication, the time to increase in dose of this medication is calculated as the date of increase minus the start date of the study drug infusion. For subjects who do not have any start/increase in any antidepressant dose, the date of start/increase is censored using the study completion date if available, otherwise the last date on study. For each subject, the time to first start in rescue antidepressant medications or first increase in dose of any antidepressant medication is the minimum time across all antidepressant medications.

For each antidepressant medication, the time to stop or decrease in dose of this medication will be calculated as the date of stop/decrease in dose minus the start date of the infusion. For subjects who do not have any stop/decrease in antidepressant dose, the date of stop/decrease will be censored using the study completion date if available, otherwise the last date on study. For each subject, the time to first stop/decrease in dose of any antidepressant medication will be calculated as the minimum time across all antidepressant medications.

9.3.3 Other Efficacy Endpoints

9.3.3.1 Edinburgh Postnatal Depression Scale (EPDS)

The EPDS is a patient-rated depressive symptom severity scale specific to the perinatal period. 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. For all items, 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 two individual items are missing, the EPDS total score will not be calculated and will be left as missing. If one or two individual item scores are missing, the missing item scores will be imputed by the mean of all other available item scores to calculate the EPDS total score.

9.3.3.2 Patient Health Questionnaire (PHQ-9)

The PHQ-9 is a patient-rated depressive symptom severity scale. This scale consists of 9 items with scoring based on item responses as follows: "not at all" = 0; "several days" = 1; "more than half the days" = 2; and "nearly every day" = 3.

The PHQ-9 total score will be calculated as the sum of the 9 individual item scores. If more than one individual items are missing, the PHQ-9 total score will not be calculated and will

be left as missing. If only one individual item score is missing, the missing item scores will be imputed by the mean of all other available item scores to calculate the PHQ-9 total score. In addition, the PHQ-9 total score will be categorized as follows: 1-4 = minimal depression, 5-9 = mild depression, 10-14 = moderate depression, 15-19 = moderately severe depression; and 20-27 = severe depression.

The PHQ-9 response is defined as having a PHQ-9 total score of 0 to 4.

9.3.3.3 Barkin Index of Maternal Functioning (BIMF)

The BIMF is a patient-reported outcome scale. BIMF covers a broad range of functional areas (self-care, infant care, mother-child interaction, psychological well-being of mother, social support, management, adjustment). Each item is rated on a scale of 0 (strongly disagree) to 6 (strongly agree). Except for items 16 and 18, higher items scores indicate higher functioning.

The BIMF total score will be calculated as the sum of the 20 item scores (after the reverse coding of items 16 and 18). If more than four individual item scores are missing, the BIMF total score will not be calculated and will be left as missing. If less than or equal to four individual item scores are missing, the missing item scores will be imputed by the mean of all other available item scores to calculate the BIMF total score. The Mom's Competency subscale score is calculated as the sum of items 1, 3, 4, 5, 10, 12, 14, 15, 17, 19, and 20; if there are more than two missing values, the Mom's Competency subscale will be left as missing, otherwise the missing values will be imputed with the mean of non-missing values to calculate the sum. The Mom's Needs subscale score is calculated as the sum of items 2, 6, 7, 8, 9, 11, and 13; if there is more than one missing value, the Mom's Needs subscale will be set as missing, otherwise the missing values will be imputed with the mean of non-missing values to calculate the sum.

9.3.3.4 Short Form-36 (SF-36)

The SF-36 covers eight health dimensions including four physical health status domains (physical functioning, role participation with physical health problems [role-physical], bodily pain, and general health) and four mental health status domains (vitality, social functioning, role participation with emotional health problems [role-emotional], and mental health). In addition, two summary scores, physical component summary (PCS) and mental component summary (MCS), are produced by taking a weighted linear combination of the eight individual domains. All of the eight health dimension scale scores and two summary scores will be provided by the scoring tool and score calculations will not be provided in this SAP.

9.3.3.5 Healthcare Resource Utilization (HCRU)

Subject-reported healthcare resource utilization data, including baseline diagnosis history, baseline antidepressant treatment history, and healthcare visits, inpatient visits, and medication use, will be collected at screening and on Day 30 of follow-up (or at early termination).

9.3.4 Multiplicity adjustment

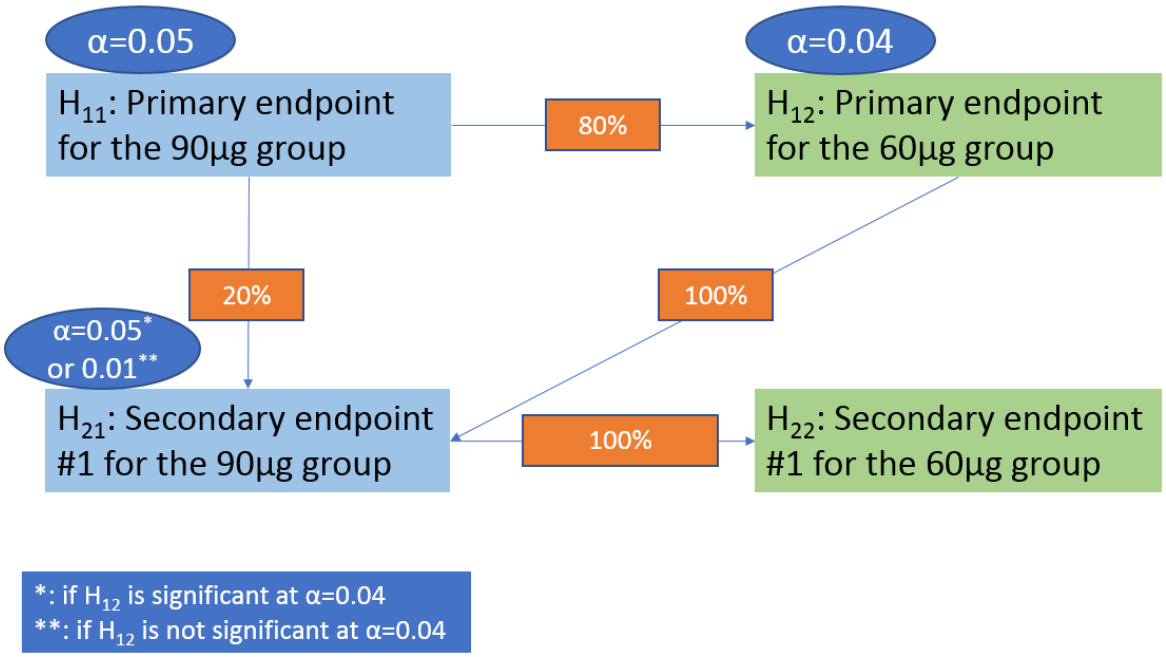
To control the type 1 error rate for conducting multiple comparisons, a stepwise method will be considered to adjust for multiplicity. For the analyses of primary and secondary endpoints and multiple comparisons (90 µg vs placebo and 60 µg vs placebo), the primary comparison will be on change from baseline in HAM-D total score between 90 µg SAGE-547 and placebo at the 60-hour assessment timepoint; if this comparison is significant at the 0.05 level, then the 60 µg group will be compared to placebo at the 0.04 level of significance.

The testing of the key secondary endpoints, which is change from baseline in HAM-D total score at the Day 30 assessment will also be adjusted for multiplicity.

- If the comparison of the HAM-D total score at the 60-hour assessment is significant for both the 90 µg SAGE-547 and the 60 µg SAGE-547 groups vs. placebo as described above, then the comparison of the HAM-D total score at the Day 30 assessment between 90 µg SAGE-547 and placebo will be tested at the 0.05 level.
- If the comparison of the HAM-D total score is significant for the 90 µg SAGE-547 vs. placebo but not for the 60 µg SAGE-547 groups vs. placebo, then the comparison of the HAM-D total score at the Day 30 assessment between 90 µg SAGE-547 and placebo will be tested at the 0.01 level.

A schematic diagram of the multiple testing method is presented in Figure 3:

Figure 3: Multiple Testing Method for Study 547-PPD-202B



9.3.5 Visit Windows

The unscheduled or early termination (ET) visit will be mapped to a scheduled visit for analysis using the date/time of collection/assessment as a basis to determine study day/hour and then study day/hour will be mapped to the intended visit. Once analysis visit windows get assigned, all visits, including scheduled visits, unscheduled visits, and ET visits will be eligible for being flagged as the "analyzed record" within the analysis window. A subject'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:

- If there is a scheduled visit/day for the analysis visit window, then the scheduled visit/day data will be used.
- If there is no scheduled visit/day for the analysis visit window, the data closest to the scheduled study day/hour will be used.
- If there is no scheduled visit/day for the analysis visit window and there are visits equally close to the day/hour of scheduled visit, the latest data will be used.

The data not flagged as the "analyzed record" will also be listed in subject listings.

Table 4 displays the visit windows for analysis of HAM-D and CGI-I; Table 5 displays the visit windows for analysis of MADRS; Table 6 displays the visit windows for analysis of EPDS, GAD-7 andPHQ-9; Table 7 displays the visit windows for analysis of BIMF and SF-36.

Table 4: Visit Windows for Analysis of HAM-D and CGI-I

Scheduled Visit	Study Day/Hour of Expected Visit	Study Day/Hour Window for Visit
Screening [a]	D -1	D -7 to D -1
Baseline [a]	D1H0	D1H0
D1H2	D1H2	>D1H0, <=D1H3
D1H4	D1H4	>D1H3, <=D1H6
D1H8 [a]	D1H8	>D1H6, <=D1H10
D1H12	D1H12	>D1H10, <=D1H18
D1H24	D1H24	>D1H18, <=D2H30
D2H36	D2H36	>D2H30, <=D2H42
D2H48	D2H48	>D2H42, <=D3H54
D3H60	D3H60	>D3H54, <=D3H66
D3H72	D3H72	>D3H66, <=D3H72
D7/ET	D7	>D4H0, <=D10
D14	D14	>D10, <=D17
D21	D21	>D17, <=D25
D30	D30	>D25, <=D33

Note: D=Day, H=Hour, ET=Early Termination.
a This scheduled visit only applies to HAM-D.

Table 5: Visit Windows for Analysis of MADRS

Scheduled Visit	Study Day/Hour of Expected Visit	Study Day/Hour Window for Visit
Screening	D -1	D -7 to D -1
Baseline	D1H0	D1H0
D1H24	D1H24	>D1H0, <=D2H36
D2H48	D2H48	>D2H36, <=D3H54
D3H60	D3H60	>D3H54, <=D3H66
D3H72	D3H72	>D3H66, <=D3H72
D7/ET	D7	>D4H0, <=D10
D14	D14	>D10, <=D17
D21	D21	>D17, <=D25
D30	D30	>D25, <=D33

Note: D=Day, H=Hour, ET=Early Termination.

Table 6: Visit Windows for Analysis of EPDS, GAD-7 and PHQ-9

Scheduled Visit	Study Day/Hour of Expected Visit	Study Day/Hour Window for Visit
Baseline	D1H0	D1H0
D3H60	D3H60	>D1H0, <=D3H72
D7/ET	D7	>D4H0, <=D10
D14	D14	>D10, <=D17
D21	D21	>D17, <=D25
D30	D30	>D25, <=D33

Note: D=Day, H=Hour, ET=Early Termination.

Table 7: Visit Windows for Analysis of BIMF and SF-36

Scheduled Visit	Study Day/Hour of Expected Visit	Study Day/Hour Window for Visit
Baseline	D1H0	D1H0
D7/ET	D7	>D1H0, <=D10
D14	D14	>D10, <=D17
D21	D21	>D17, <=D25
D30	D30	>D25, <=D33

Note: D=Day, H=Hour, ET=Early Termination.

9.3.6 Pooling of Centers

For efficacy analysis purposes, centers with fewer than 15 subjects per center will be pooled within geographic regions (North America). Within a region, first centers with fewer than 15 subjects will be ranked by size. Starting with the largest site, centers will be pooled until 15 or more subjects are reached to create a uniqe regional center. Continuing this process of creating unique regional centers until there are less than 15 subjects remaining. The remaining centers will then be added to the last unique regional center created.

Unless the combined number of subjects from all centers with fewer than 15 subjects in a region is fewer than 15 subjects, all pooled centers should have at least 15 subjects.

9.3.7 Analysis of Efficacy Variable(s)

Efficacy data will be summarized using appropriate descriptive statistics and other data presentation methods where applicable; subject listings will be provided for all efficacy data. Subjects will be analyzed according to randomized treatment.

Descriptive statistics including n, mean, SD, median, minimum, and maximum of actual values and change from baseline, and percentage change from baseline values will be presented by treatment group and assessment time point for the following continuous efficacy variables:

- HAM-D total score;
- MADRS total score;
- HAM-D subscale and individual item scores (Note: percentage change from baseline values will not be presented for HAM-D individual item scores);
- CGI-I and CGI-S scale scores (Note: change and percentage change from baseline are not applicable);
- GAD-7 total score;
- PHQ-9 total score;
- EPDS total score;
- BIMF and SF-36 scale scores

Raw values in HAM-D, MADRS, GAD-7, EPDS total score and CGI-I scale scores over time will be presented graphically by treatment group.

Descriptive statistics including counts and percentages will be summarized by treatment group and assessment time point for the following categorical efficacy variables:

- HAM-D response;
- HAM-D remission;
- CGI-I scale scores;
- CGI-I response

Bar plot will be presented for the percentage of HAM-D responders and HAM-D remitters.

9.3.7.1 Analysis of Primary Efficacy Endpoints

9.3.7.1.1 Hamilton Rating Scale for Depression (HAM-D)

Mixed Effects Model for Repeated Measures (MMRM):

Change from baseline to each assessment timepoint in HAM-D total score will be analyzed using a mixed effects model for repeated measures (MMRM).

The model will include the change from baseline at each visit time point as the dependent variable, center (pooled), treatment, baseline antidepressant use, baseline HAM-D total score,

visit time point, and visit time point-by-treatment interaction terms as explanatory variables. All explanatory variables including pooled center will be treated as a fixed effect in the primary analysis.

The primary comparison will be between 90 µg SAGE-547 and placebo at the 60-hour assessment; if this comparison is significant at the 0.05 level, then the 60 µg group will be compared to placebo at the 0.04 level of significance. Comparisons at other time points, including the one for a key secondary endpoint at the Day 30 time point, will be conducted to support the findings for the primary comparison. Model-based point estimates (ie, LS means), 95% confidence intervals, and p-values will be reported for each assessment. An unstructured covariance structure will be used to model the within-subject errors. If there is a convergence issue with the unstructured covariance model, Toeplitz compound symmetry or Autoregressive (1) [AR(1)] covariance structure will be used, follwing this sequence until convergence is achieved. If the model still does not converge with AR(1) structure, no results will be reported See sample SAS code for MMRM in section 15.2.

In addition, the LS means (+/- SE) estimates of change from baseline in HAM-D total score over time will be presented graphically by treatment group.

9.3.7.1.2 Supportive Analysis of the Primary Endpoint

The following supportive analyses will be performed for the primary endpoints:

- A MMRM method similar to those described in section 9.3.7.1.1 will be used for the analysis of the HAM-D total score. Center (pooled) will be treated as random effect. See sample SAS code for MMRM in section 15.2.
- The Per Protocol Set will also be used for supportive analysis.

9.3.7.1.3 Subgroup Analysis of the Primary Endpoint

Summaries using descriptive analyses and MMRM model (if there are at least 5 subjects in each treatment arm for a subgroup) as described in section 9.3.7.1.1 will be performed base on the following subgroups.Due to smaller smaple sizes, center (pooled) and baseline antidepressant use will be removed as covariates in the MMRM model.

- Race category: White, Black or African American, and Other
- Age category: 18-24, 25-45
- Baseline antidepressant use: Yes, No
- BMI category: <=18.4, 18.5-24.9, 25-29.9, >=30 kg/m^2
- Onset of PPD: 3rd trimester and within 4 weeks of delivery
- Family history of PPD: Yes, No

9.3.7.1.4 Sensitivity Analysis of the Primary Endpoint

1) Sensitivity analyses will be carried out to investigate the impact of missing data if more than 10% of subjects are missing primary endpoint assessment (ie, HAM-D total score).

A dropout reason based imputation will be used. An analysis of covariance (ANCOVA) model, including center (pooled), treatment, baseline antidepressant use, and baseline HAM-D total score, will be used to assess the treatment difference in change from baseline in HAM-D total score at Hour 60 (MI imputed data). The missing change from baseline in HAM-D total score at Hour 60 will be imputed using multiple imputation.

Imputation distribution:
The imputation distribution for the missing change from baseline in HAM-D total score at Hour 60 will be a normal distribution. All randomized subjects will be classified as non-missing category, missing category 1, or missing category 2, based on the following rules:

- Non-missing category: Subject with non-missing change from baseline in HAM-D total score at Hour 60.
- Missing category 1: Subject discontinued due to adverse events, physician decision, protocol violation or other, and is missing change from baseline in HAM-D total score at Hour 60.
- Missing category 2: Subject discontinued due to pregnancy, study terminated by sponsor, or withdrawal by subject, and is missing change from baseline in HAM-D total score at Hour 60, or subject completed study but is missing change from baseline in HAM-D total score at Hour 60.

Imputation algorithm:
Missing values of change from baseline in HAM-D total score at Hour 60 will be imputed separately within each treatment group using the following missing reason based algorithm:

- **Missing category 1:** randomly draw a sample from the normal distribution $N(\mu_{25}, \sigma^2)$, where μ_{25} is the 75th percentile of the non-missing change from baseline in HAM-D total score and σ^2 is the sample variance estimated using the non-missing change at Hour 60. This respresents a conservative approach since higher values of change from baseline represents worse outcome.
- **Missing category 2:** randomly draw a sample from the normal distribution $N(\mu, \sigma^2)$, where μ is the mean of the non-missing change from baseline in HAM-D total score at Hour 60 and σ^2 is the sample variance estimated using the non-missing change at Hour 60.

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 Efficacy Set subjects will have non-missing change from baseline in HAM-D total score at Hour 60.
- Repeat the process K (K=5) times, using the procedure described above to form K imputed complete datasets.
- Fit the ANCOVA model including center (pooled), treatment, baseline antidepressant use, and baseline HAM-D total score, to each imputed dataset, to estimate the treatment effect and its variance at Hour 60.
- Combine the results from the K imputed datasets using the SAS procedure MIANALYZE, to derive the MI estimator.

We fit the analysis model (ANCOVA 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_V distribution where $V = (K-1)\left(1 + \frac{W}{B}\right)^2$.

2) The following sensitivity analysis will be performed to account for the possible scenario that the data is not normally distributed.

Stratified Wilcoxon rank-sum test on change from baseline in the HAM-D total score at the 60 hour timepoint, stratified by antidepressant use, will be used as a sensitivity analysis. If necessary (e.g., serious violation of the primary analysis model assumptions on normality, as assessed by using QQ plot on the residual values from the primary analysis for HAM-D), this will be considered the primary analysis.

See sample SAS code for stratified Wilcoxon rank-sum test using PROC FREQ in section 15.2.

9.3.7.2 Analysis of Secondary Efficacy Endpoints

9.3.7.2.1 Montgomery-Åsberg Depression Rating Scale (MADRS)

A MMRM method similar to those described in section 9.3.7.1.1will be used for the analysis of the change from baseline in MADRS total score. The comparison of interest will be between each SAGE-547 dose and placebo at the 60-hour assessment. Model based point estimates (ie, LS means), 95% confidence intervals, and p-values will be reported.

In addition, the LS means (+/- SE) estimates of change from baseline in MADRS total score over time will be presented graphically by treatment group.

9.3.7.2.2 Generalized Anxiety Disorder 7-item Scale (GAD-7)

A MMRM method similar to those described in section 9.3.7.1.1 will be used for the analysis of the change from baseline in GAD-7 total score. The comparison of interest will be

between each SAGE-547 dose and placebo at the 60-hour assessment. Model based point estimates (ie, LS means), 95% confidence intervals, and p-values will be reported.

In addition, the LS means (+/- SE) estimates of change from baseline in GAD-7 total score over time will be presented graphically by treatment group.

Number and percent of subjects GAD-7 response will be tabulated by treatment and timepoint. Odds ratios and the 95% confidence intervals and the p-values for each SAGE-547 treatment group vs. placebo will be estimated using Fisher's exact test.

See sample SAS code for Fisher's exact test using PROC FREQ in section 15.2.

9.3.7.2.3 HAM-D subscales and individual item scores

A MMRM method similar to those described in section 9.3.7.1.1 will be used for the analysis of the change from baseline in HAM-D subscales and individual item scores. The comparison of interest will be between each SAGE-547 dose and placebo at the 60-hour assessment. Model based point estimates (ie, LS means), 95% confidence intervals, and p-values will be reported.

9.3.7.2.4 HAM-D Response and HAM-D Remission

A Generalized Estimating Equation (GEE) method will be used for the analysis of HAM-D response and HAM-D remission. GEE models will include the response variables above at each visit time point as the dependent variable, center (pooled), treatment, baseline antidepressant use, baseline HAM-D total score, visit time point, and visit time point-by-treatment interaction terms as explanatory variables. An unstructured working correlation structure for the binary outcome variables will be assumed for the model parameters estimation. Exchangable, independence or compound symmetry working correlation structure will be used if there is a convergence issue with the unstructured working correlation model. Only data from the Hour 24 visit and later will be included in the model to avoid non-convergence arising from sparse data.

The comparison of interest will be the difference between each SAGE-547 dose and placebo at the 60-hour assessment. Model based point estimates (ie, odds ratios), 95% confidence intervals, and p-values will be reported. See sample SAS code for GEE in section 15.2.

9.3.7.2.5 Clinical Global Impression – Improvement (CGI-I)

A GEE method similar to those described in section 9.3.7.2.4 will be used for the analysis CGI-I response. For the CGI-I response analysis, baseline CGI-S score will be included as the baseline value in the model.

The comparison of interest will be the difference between each SAGE-547 dose and placebo at the 60-hour assessment. Model based point estimates (ie, odds ratios), 95% confidence intervals, and p-values will be reported.

In addition, the CGI-I score will be analyzed using a MMRM method similar to those described in section 9.3.7.1.1. The LS means (+/- SE) estimates of CGI-I score over time will be presented graphically by treatment group.

9.3.7.2.6 Post-baseline Antidepressant Use

The time to first start/increase and the time to first stop/decrease in dose of any rescue antidepressant medication will be summarized descriptively using Kaplan-Meier methods. Descriptive statistics including n, median, 1st quartile, 3rd quartile, minimum, maximumand censored values will be presented.

Kaplan-Meier plots for the analysis of antidepressant use will also be presented.

9.3.7.2.7 Supportive Analysis of the Secondary Endpoints

The Per Protocol Set will also be used for supportive analysis for change from baseline in MADRS total score.

9.3.7.2.8 Subgroup Analysis of the Secondary Endpoints

Summaries using descriptive analyses and MMRM model will be performed for MADRS, in a similar way as described in Section 9.3.7.1.3. Only subgroup by baseline antidepressant use will be considered.

Summary analysis using descriptive statistics and GEE method as described in 9.3.7.2.4 will be performed for HAM-D response and HAM-D remission based on subgroup by baseline antidepressant use as described in 9.3.7.1.3.

Summary analysis using descriptive statistics and GEE method as described in 9.3.7.2.5 will be performed for CGI-I response based on subgroup by baseline antidepressant use as described in 9.3.7.1.3.

9.3.7.2.9 Sensitivity Analyses of the Secondary Endpoints

- Sensitivity analyses similar to those described in 9.3.7.1.4 will be performed on the change from baseline in HAM-D total score at each timepoint and change from baseline in MADRS total score at the 60 hour assessment.
- Logistic regression methods will be used for the analysis of the following binary variables as a sensitivity analyses for the sencondary endpoints: HAM-D response, HAM-D remission and CGI-I response. Logistic regression models will include the response variables above at each visit time point as the dependent variable, center (pooled), treatment, baseline antidepressant use, and baseline score as explanatory variables. The Firth penalized likelihood method will be used to reduce bias in the parameter estimates and to avoid non-convergence due to quasi separation. For the CGI-I response analysis, baseline CGI-S score will be included in the model as the baseline score. The comparison of interest will be the difference between each SAGE-547 dose and placebo. Model-based point estimates (i.e., odds ratios), 95% confidence intervals, and p-values will be reported. See sample SAS code for logistic regression in section 15.2.
- As a sensitivity analysis to assess the impact of rescue antidepressant use, subject will be set as HAM-D non-responders and not achieving HAM-D remission following the use of rescue antidepressant. Logistic regression methods as described above will be

used to compare the treatment difference in HAM-D response and HAM-D remission at these follow-up visits: Day 7, Day 14, Day 21, and Day 30.

9.3.7.3 Analysis of Other Efficacy Endpoints

9.3.7.3.1 Edinburgh Postnatal Depression Scale (EPDS)

A MMRM method similar to those described in section 9.3.7.1.1 will be used for the analysis of the change from baseline in EPDS total score. The comparison of interest will be between each SAGE-547 dose and placebo at the 60 hours and the follow-up visits. Model based point estimates (ie, LS means), 95% confidence intervals, and p-values will be reported.

9.3.7.3.2 Patient Health Questionnaire (PHQ-9)

A MMRM method similar to those described in section 9.3.7.1.1 will be used for the analysis of the change from baseline in PHQ-9 total score. The comparison of interest will be between each SAGE-547 dose and placebo at Day 3, 60 hours and the follow-up visits. Model based point estimates (ie, LS means), 95% confidence intervals, and p-values will be reported.

Number and percent of subjects with PHQ-9 response will be tabulated by treatment and timepoint. Odds ratios and the 95% confidence intervals and the p-values for each SAGE-547 treatment group vs. placebo will be estimated using Fisher's exact test.

9.3.7.3.3 Barkin Index of Maternal Functioning (BIMF)

A MMRM method similar to those described in section 9.3.7.1.1 will be used for the analysis of the change from baseline in BIMF total score. The comparison of interest will be between each SAGE-547 dose and placebo at the follow-up visits. Model based point estimates (ie, LS means), 95% confidence intervals, and p-values will be reported.

The same analysis will be done for the Mom's competency subscale score and Mom's needs subscale score.

9.3.7.3.4 Short Form-36 (SF-36)

A MMRM method similar to those described in section 9.3.7.1.1 will be used for the analysis of the change from baseline in each of the 8 health dimension scale scores and 2 summary scores. The comparison of interest will be between each SAGE-547 dose and placebo at the Day 3, 60 hours and the follow-up visits. Model based point estimates (ie, LS means), 95% confidence intervals, and p-values will be reported.

9.3.7.3.5 Healthcare Resource Utilization (HCRU)

Data on each type of HCRU will be reported by reasons for utilization (depression related, pregnancy/labor/delivery related, other) with appropriate descriptive statistics at screening and on the Day 30 (end of study) in patients comprising the Efficacy Set. Table 8 summarized the type of HCRU, reasons for utilization and descriptive statistics to be used to analyze these data.

Table 8: Descriptive Statistics to be Used to Analyze HCRU Data

Utilization Type	Reason for Utilization	Descriptive Statistics
Emergency room visit, use of an ambulance, outpatient primary care physician visit, outpatient specialist visit (e.g. OB/GYN, surgeon), outpatient counselling visit	Depression related Pregnancy/labor/delivery related Other	Number of patients reporting use; Percentage of patients reporting use out of the total number of patients; Number of visits, mean, median, minimum and maximum of those reporting use of that utilization type and for that reason; Number of visits mean and median for the total number of patients
Hospital admission	Depression related Pregnancy/labor/delivery related Other	Number of patients reporting use; Percentage of patients reporting use out of the total number of patients; Number of visits, mean, median, minimum/maximum of those reporting hospital admission; Length of stay mean, median, minimum and maximum of those reporting hospital admission; Number of visits mean and median for the total number of patients; Length of stay mean and median for the total number of patients

9.4 Safety Analysis

The secondary objectives include the safety and tolerability of SAGE-547 Injection, as evaluated by AEs, clinical laboratory tests, vital signs, ECGs and concomitant medication usage. Suicidality will be monitored by the C-SSRS.

Safety data will be listed by subject and summarized descriptively by treatment group. All safety analyses will be performed on the Safety Set.

The safety endpoints and variables considered in the summary tables for this study are summarized in Table 9.

Table 9: Safety Endpoints and Variables in the Summary Tables and Listings

Safety Evaluation	Incidence	Actual Value	Change from Baseline	Abnormality/Clinical Significance (CS)
AEs	X	*		
Con Meds	X	*		
Labs		X	X	*
ECG		X	X	*
Vital Signs		X	X	*
Physical Examination		*		*
C-SSRS	X	*		

X = Safety Assessment will be summarized in tables

* = Safety Assessment will be presented in individual subject data listings only

9.4.1 Visit Windows

The mapping rule of unscheduled or early termination (ET) visit and the rule of flagging "analyzed record" are the same as described in section 9.3.5.

Table 10 displays the visit windows for analysis of clinical laboratory parameters (including hematology, serum chemistry, coagulation) parameters; Table 11 displays the visit windows for analysis of urinalysis parameters; Table 12 displays the visit windows for analysis of 12-lead ECG; Table 13 displays the visit windows for analysis of vital signs parameters; Table 14 displays the visit windows for analysis of C-SSRS parameters.

Table 10: Visit Windows for Analysis of Clinical Laboratory Parameters

Scheduled Visit	Study Day/Hour of Expected Visit	Study Day/Hour Window for Visit
Baseline	D -1	D -7 to D1H0
D3H72	D3H72	>D1H0, <=D3H72
D7/ET	D7	>D4H0, <=D33

Note: D=Day, H=Hour, ET=Early Termination.

Table 11: Visit Windows for Analysis of Urinalysis Parameters

Scheduled Visit	Study Day/Hour of Expected Visit	Study Day/Hour Window for Visit
Baseline	D -1	D -7 to D1H0
D7/ET	D7	>D1H0, <=D33

Note: D=Day, H=Hour, ET=Early Termination.

Table 12: Visit Windows for Analysis of 12-lead ECG Parameters

Scheduled Visit	Study Day/Hour of Expected Visit	Study Day/Hour Window for Visit
Baseline	D -1	D -7 to D1H0
D2H48	D2H48	>D1H0, <=D3H72
D7/ET	D7	>D4H0, <=D33

Note: D=Day, H=Hour, ET=Early Termination.

Table 13: Visit Windows for Analysis of Vital Signs Parameters

Scheduled Visit	Study Day/Hour of Expected Visit	Study Day/Hour Window for Visit
Screening	D -1	D -7 to D -1
Baseline	D1H0	D1H0
D1H2	D1H2	>D1H0, <=D1H3
D1H4	D1H4	>D1H3, <=D1H6
D1H8	D1H8	>D1H6, <=D1H10
D1H12	D1H12	>D1H10, <=D1H15
D1H18	D1H18	>D1H15, <=D1H21
D1H24	D1H24	>D1H21, <=D2H27
D2H30	D2H30	>D2H27, <=D2H33
D2H36	D2H36	>D2H33, <=D2H39
D2H42	D2H42	>D2H39, <=D2H45
D2H48	D2H48	>D2H45, <=D3H51
D3H54	D2H54	>D3H51, <=D3H57
D3H60	D3H60	>D3H57, <=D3H63
D3H66	D3H66	>D3H63, <=D3H69
D3H72	D3H72	>D3H69, <=D3H72
D7/ET	D7	>D4H0, <=D33

Note: D=Day, H=Hour, ET=Early Termination.

Table 14: Visit Windows for Analysis of C-SSRS Parameters

Scheduled Visit	Study Day/Hour of Expected Visit	Study Day/Hour Window for Visit
Baseline	D1H0	D -7 to D1H0
D1H24	D1H24	>D1H0, <=D2H42
D3H60	D3H60	>D2H42, <=D3H66
D3H72	D3H72	>D3H66, <=D3H72
D7/ET	D7	>D4H0, <=D10
D14	D14	>D10, <=D17
D21	D21	>D17, <=D25
D30	D30	>D25, <=D33

Note: D=Day, H=Hour, ET=Early Termination.

9.4.2 Adverse Events

The analysis of AEs will be based on the concept of treatment-emergent AEs (TEAEs). A TEAE is defined as an AE with onset on or after the start of study drug infusion, or any

worsening of a pre-existing medical condition/AE with onset on or after the start of study drug infusion.

All adverse events will be coded using MedDRA version 19.1 and summarized by SOC and PT. Multiple occurrences of an AE are counted only once per subject per SOC and PT for summary tables. Incidences will be presented in order of decreasing frequency of the SAGE-547 90 µg/kg/hour group by SOC and PT. Incidences will also be presented in order of decreasing frequency of the SAGE-547 90 µg/kg/hour group by PT only. In addition, summaries will be provided by severity (mild, moderate, severe) and by causality (related, not related) to study drug (see protocol section 14.2.2.1 for definitions). TEAEs leading to discontinuation and Serious AEs (SAEs) (see protocol section 14.1.4 for definition) with onset after the start of randomized infusion will also be summarized.

Summary tables of TEAEs will be presented by treatment and will summarize the number and percentage of subjects for the following:

- Any TEAE
- TEAEs by relatedness to study drug (not related, related)
- TEAEs by severity to study drug (mild, moderate, severe)
- Treatment-emergent serious AEs (SAEs)
- TEAEs leading to discontinuation of study drug
- Overall summary of the number and percentage of subjects reporting TEAEs, drug-related TEAEs, severe TEAEs, SAEs and TEAEs leading to study drug discontinuation.

Subjects will be counted only once within each SOC and PT at the maximum severity in the following order: severe, moderate, and mild. An AE with missing severity will be considered as a severe AE. Subjects will be counted only once within each SOC and PT at the strongest relationship to study drug in the following order: related, not related to study drug. If the relationship between the adverse event and the study drug is determined to be "possible" or "probable", the event will be considered as related to the study drug. An AE with missing relationship to study drug will be considered as related to study drug.

All AEs, SAEs (including those with onset or worsening before the start of randomized infusion), AEs Leading to study drug discontinuation and AEs leading to death through the Day 30 Follow-up Visit (±3 days) will be listed.

During the study, the categories of outcome of AEs were changed slightly. Originally, the follow categories were used:

- Recovered/resolved
- Recovered/resolved with sequelae
- Recovering/resolving
- Not recovered/not resolved

- Fatal
- Unknown

However, starting from Version 4.0 of the protocol, the following categories were used:

- Ongoing
- Resolved
- Resolved with sequelae
- Unknown
- Death

In the data listings, the following mapping will be applied:

From	To
Resolved	Recovered/resolved
Resolved with sequelae	Recovered/resolved with sequelae
	Recovering/resolving
Ongoing	Not recovered/not resolved
Death	Fatal
Unknown	Unknown

9.4.3 Prior and Concomitant Medications

Concomitant medications will be coded using World Health Organization-Drug dictionary Enhanced (WHO-DDE) September 1, 2016.

Medications will be presented according to whether they are being taken prior to and/or during the study (concomitant):

- Prior medications are defined as all medications taken prior to the infusion of study treatment.
- Concomitant medications are defined as all medications taken on or after the infusion of study treatment starts through the Day 30 (±3 days) visit. If medication dates are incomplete and it is not clear whether the medication was concomitant, it will be assumed to be concomitant.

Note that medication that started prior to the infusion of study drug and continued after the start of infusion will be summarized as prior medication and separately as concomitant medication. Details of prior and concomitant medications will be listed by subject, start date, and verbatim term.

Medication summaries will be performed by ATC level 2 term and PT based on the Safety Set.

Antidepressant medications will be summarized by ATC level 2 term and PT based on the Safety Set.

9.4.4 Clinical Laboratory

Blood samples will be collected for hematology, serum chemistry, coagulation, specific hormone parameters, and exploratory biochemistry; pregnancy testing; and genetic analysis. Urine samples for urinalysis and selected drugs of abuse will also be collected.

Summary tables will include descriptive statistics for the actual values and changes from baseline by scheduled time point in hematology, serum chemistry, coagulation and quantitative urinalysis test results by treatment group. For qualitative urinalysis parameters, test results will be categorized as normal and abnormal. Frequency counts and percentages will be presented over time for these categorical data by treatment. Genetic sample results will be listed.

Shift analysis pre- and post-treatment will be presented for the following laboratory categories (low, normal and high).

Number of percent of subjects with liver injury as defined by Hy's law will be presented. Hy's law is defined as ALT or AST > 3 times of upper limit of normal; and total bilirubin > 2 times of upper limit of normal, and alkaline phosphatase < 2 time of upper limit of normal.

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

All clinical laboratory results will be listed by subject and timing of collection. Out-of-range values will be flagged as low, high, or abnormal, where applicable.

The number and percentage of subjects with potentially clinically significant (PCS) values at any time after study drug infusion will be summarized by treatment group for the hematology and chemistry parameters defined in Table 15 and Table 16. PCS laboratory values will also be listed by subject.

Table 15: Hematology PCS Values

Hematology parameter	Potentially clinically significant values
Hemoglobin	<110 g/L or >165 g/L
Hematocrit (HCT)	<0.359 or >0.446
Platelet count	<125 GI/L or >600 GI/L
Leukocytes	<2.5 GI/L or >15 GI/L
Basophils	>0.5 GI/L
Eosinophils	>1.5 GI/L
Lympocytes	<0.5 GI/L or >6.0 GI/L
Monocytes	>1.4 GI/L
Neutrophils	<1.5 GI/L

GI/L = giga/L = 10^9/L

Table 16: Chemistry PCS Values

Chemistry parameter	Potentially clinically significant values
Alanine aminotransferase (ALT)	>3 x ULN
Albumin	<28 g/L or >70 g/L
Aspartate aminotransferase (AST)	>3 x ULN
Bicarbonate	<18 mmol/L or >30 mmol/L
Bilirubin	>2 x ULN
Blood urea nitrogen (BUN)	>10.71 mmol/L
Calcium	<2.0 mmol/L or >2.75 mmol/L
Chloride	<90 mmol/L or >120 mmol/L
Creatine kinase	>3 x ULN
Creatinine	>140 μmol/L
Potassium	<3.5 mmol/L or >5.2 mmol/L
Sodium	<132 mmol/L or >145 mmol/L
Protein	<45 g/L
Glucose	<2.8 mmol/L or >13.9 mmol/L
Gamma Glutamyl Transferase (GGT)	>3 x ULN

ULN = upper limit of normal.

A summary of subjects with abnormal liver enzymes and liver function tests post-baseline will be presented, including abnormalities that worsened for subjects with baseline abnormal levels and ALT and AST shifts from baseline.

9.4.4.1 Hematology, Serum Chemistry, Coagulation

Blood samples will be collected for analysis of the following:

- Hematology: complete blood count including white blood cell count with differential (basophils, eosinophils, lymphocytes, monocytes, and neutrophils) platelet count, red blood cell count, hemoglobin and hematocrit, mean corpuscular volume, and mean corpuscular hemoglobin
- Serum chemistry: albumin, alkaline phosphatase, alanine aminotransferase, aspartate aminotransferase, bilirubin (total), blood urea nitrogen, calcium, chloride, creatinine, gamma glutamyl transferase, glucose, phosphate, potassium, sodium, total protein, and triglycerides (Screening only)
- Coagulation: activated partial thromboplastin time, prothrombin time, international normalized ratio

9.4.4.2 Hormones and Exploratory Biochemistry

Blood samples will be collected and may be analyzed for thyroid stimulating hormone, estrogen, progesterone, progesterone metabolites, oxytocin, tryptophan, kynurenine, and markers of inflammation.

9.4.4.3 Pregnancy Test

All subjects will be tested for pregnancy by serum at screening and urine human chorionic gonadotropin on Day 1 prior to administration of study drug and Day 30. Subjects with a positive pregnancy test at screening or Day 1 will be ineligible for study participation. Preganancy test results will be listed by subject.

9.4.4.4 Urinalysis

Urinalysis will include assessment of bilirubin, glucose, ketones, leukocytes, nitrite, pH, protein, and specific gravity.

9.4.5 Electrocardiogram

A baseline 12-lead ECG will be performed during screening. The following ECG parameters will be recorded: heart rate (bpm), PR (msec), QRS (msec), QT (msec), and QT corrected with Fridericia's formula (QTcF) (msec).

QTcF (msec) is calculated as: QT (msec) / $RR^{1/3}$, where RR = 60 / heart rate (bpm).

All ECG results will be interpreted by the Investigator as Normal, Abnormal; not clinically significant (NCS), or Abnormal; clinically significant (CS). If Abnormal, details will be provided.

The actual value at each time point and change from baseline at each post-baseline time point will be summarized by treatment group. The number and percentage of subjects with 'normal', 'abnormal, not clinically significant' and 'abnormal, clinically significant' ECG results will be summarized at baseline and each post-baseline time point.

All ECG data will be listed by subject and time of measurement. Mean values of ECG parameters over time will be presented graphically by treatment group.

Additionally, the number and percentage of subjects in each treatment group who meet any of the following criteria for QT or QTcF will be tabulated.

- Maximum value >450 to 480 msec
- Maximum value >480 to 500 msec
- Maximum value >500 msec
- Maximum increase from baseline >30 to 60 msec
- Maximum increase from baseline >60 msec

9.4.6 Vital Signs

Descriptive summaries of actual values and changes from baseline will be calculated for vital signs, which include oral temperature (°C), respiratory rate (breaths/min), heart rate (bpm), and blood pressure (mmHg, supine and standing). A full set of vital signs will be obtained at all specified time points (±30 minutes), unless the subject is asleep between the hours of 23.00h and 06.00h each day.

To evaluate potential orthostatic hypotension, difference between supine and standing systolic blood pressure, and difference between supine and standing diastolic blood pressure (supine – standing), will be calculated. Orthostatic hypotension is defined as decrease in

systolic blood pressure of at least 20 mmHg or decrease in diastolic blood pressure of at least 10 mmHg.

All vital sign data will be listed by subject and time of measurement. Mean values of vital signs parameters, including the difference between supping and standing blood pressure values over time will be presented graphically by treatment group. Number and percent of subjects who developed orthostatic hypotension will be presented in a table by visit and treatment group.

9.4.7 Physical Examination

Body weight and height will be measured at screening. Body mass index (BMI) will be programmatically calculated in the eCRF. The date of assessment and study day of physical examinations will be listed.

Note: Any condition present at the post-treatment physical examination that was not present at or worsened since the baseline examination is to be documented as an AE and will be presented in AE listings and tables.

9.4.8 Columbia Suicide Severity Rating Scale (C-SSRS)

Suicidality will be monitored during the study using the C-SSRS. This scale consists of a pre-dose evaluation that assesses the lifetime/ past 3 months of the subject with suicidal ideation and the lifetime/ past 1 year of the subject with suicidal behavior, and a post-baseline evaluation that focuses on suicidality since the last study visit. The C-SSRS includes "yes" or "no" responses for assessment of suicidal ideation and behavior as well as numeric ratings for severity of ideation, if present (from 1 to 5, with 5 being the most severe). The "Baseline/Screening" C-SSRS form will be completed on Day 1 prior to dosing. The "Since Last Visit" C-SSRS form will be completed for all subsequent assessments.

Suicidality data collected on the C-SSRS at baseline and by visit during the Treatment Period will be listed for all subjects. Listings will include behavior type and/or category for Suicidal Ideation and Suicidal Behavior of the C-SSRS. In addition, the number and percentage of subjects with a response of 'Yes' to any C-SSRS Suicidal Ideation or Suicidal Behavior item will be presented.

9.5 Pharmacokinetic Analysis

Plasma will be collected to assay for concentrations of SAGE-547, SAGE-547 metabolites, and SBECD. PK analyses will be performed using the PK Population.

9.5.1 Collection schedule

Blood samples for PK analysis Blood samples for PK analysis will be collected at pre-infusion and at 4 (before change in infusion rate, if applicable), 8, 12, 24 (before change in infusion rate, if applicable), 30, 36, 48, 60 (before end of infusion), and 72 hours after the start of the infusion. PK blood draws after the start of infusion will have a window of ±10 minutes. In the event of an unplanned dose adjustment, an unscheduled PK sample should be collected just prior to the infusion rate change (see section 15.1).

9.5.2 Derived PK parameters

The following PK parameters will be derived from the plasma concentrations of SAGE-547 (where evaluable):

Table 17: PK parameters and definitions

$AUC_{0\text{-}60}$ (or AUC)	Area under the plasma concentration time curve from time 0 to 60 hours
AUC_{∞}	AUC from time 0 to infinity
C_{max}	Maximum (peak) plasma concentration
T_{max}	Time at maximum (peak) plasma concentration
C_{ss}	Steady-state drug concentration in the plasma during constant-rate infusion
C_{avg}	Average drug concentration in the plasma at steady state during a dosing interval

9.5.3 Handling of dropouts or missing data

PK parameters will be provided by a third party. Missing data will not be addressed.

9.5.4 Statistical analysis

The plasma concentrations along with time point deviation from scheduled time will be listed by subject.

Pharmacokinetic parameters will be summarized using appropriate descriptive statistics. All PK parameters will be summarized using n, geometric mean, geometric coefficient of variation, coefficient of variation, median, minimum, and maximum and listed by subject.

9.5.5 Data presentation

The descriptive statistics will be generated as discussed above in Section 9.5.4.

The following figures will be produced:

- Mean ± SD plasma concentration-time profiles through Day 3, 60 hours will be plotted on linear and semi-log scales
- Individual subject concentration-time profiles on linear and semi-logarithmic concentration scales
- Spaghetti plots for each treatment through Day 3, 60 hours (linear and semi-logarithmic scale)

10 SUMMARY OF INTERIM AND DMC ANALYSES

Not applicable

11 INDEX OF TABLES

Table 14.2.1.6.1	Hamilton Rating Scale for Depression (HAM-D) Total Score - Descriptive Statistics Subgroup Analysis: Baseline Antidepressant Use (Yes vs No)	Efficacy Set
Table 14.2.1.6.2	Change from Baseline in Hamilton Rating Scale for Depression (HAM-D) Total Score - MMRM Analysis (Pooled Center as Fixed Effect) Subgroup Analysis: Baseline Antidepressant Use (Yes vs No)	Efficacy Set
Table 14.2.1.7.1	Hamilton Rating Scale for Depression (HAM-D) Total Score - Descriptive Statistics Subgroup Analysis: BMI Category (<=18.4, 18.5-24.9, 25-29.9, >=30 kg/m^2)	Efficacy Set
Table 14.2.1.7.2	Change from Baseline in Hamilton Rating Scale for Depression (HAM-D) Total Score - MMRM Analysis (Pooled Center as Fixed Effect) Subgroup Analysis: BMI Category (<=18.4, 18.5-24.9, 25-29.9, >=30 kg/m^2)	Efficacy Set
Table 14.2.1.8.1	Hamilton Rating Scale for Depression (HAM-D) Total Score - Descriptive Statistics Subgroup Analysis: Onset of PPD (3rd Trimester and Within 4 Weeks of Delivery)	Efficacy Set
Table 14.2.1.8.2	Change from Baseline in Hamilton Rating Scale for Depression (HAM-D) Total Score - MMRM Analysis (Pooled Center as Fixed Effect) Subgroup Analysis: Onset of PPD (3rd Trimester and Within 4 Weeks of Delivery)	Efficacy Set
Table 14.2.1.9.1	Hamilton Rating Scale for Depression (HAM-D) Total Score - Descriptive Statistics Subgroup Analysis: Family history of PPD (Yes vs No)	Efficacy Set
Table 14.2.1.9.2	Change from Baseline in Hamilton Rating Scale for Depression (HAM-D) Total Score - MMRM Analysis (Pooled Center as Fixed Effect) Subgroup Analysis: Family history of PPD (Yes vs No)	Efficacy Set
Table 14.2.1.10.1	Sensitivity Analysis - Change from Baseline in Hamilton Rating Scale for Depression (HAM-D) Total Score - Dropout Reason Based Imputation and ANCOVA Model	Efficacy Set
Table 14.2.1.10.2	Sensitivity Analysis - Change from Baseline in Hamilton Rating Scale for Depression (HAM-D) Total Score - Stratified Wilcoxon Rank-Sum Test	Efficacy Set
Table 14.2.1.11.1	Hamilton Rating Scale for Depression (HAM-D) Subscale and Individual Item Scores - Descriptive Statistics	Efficacy Set
Table 14.2.1.11.2	Change from Baseline in Hamilton Rating Scale for Depression (HAM-D) Subscale and Individual Item Scores - MMRM Analysis (Pooled Center as Fixed Effect)	Efficacy Set
Table 14.2.1.12	Hamilton Rating Scale for Depression (HAM-D) Response - Descriptive Statistics and GEE Analysis	Efficacy Set
Table 14.2.1.13.1	Hamilton Rating Scale for Depression (HAM-D) Response - Descriptive Statistics and Logistic Regression Analysis	Efficacy Set
Table 14.2.1.13.2	Hamilton Rating Scale for Depression (HAM-D) Response - Descriptive Statistics and GEE Analysis Subgroup Analysis: Baseline Antidepressant Use (Yes vs No)	Efficacy Set

Table 14.2.1.14	Hamilton Rating Scale for Depression (HAM-D) Remission - Descriptive Statistics and GEE Analysis	Efficacy Set
Table 14.2.1.15.1	Hamilton Rating Scale for Depression (HAM-D) Remission - Descriptive Statistics and Logistic Regression Analysis	Efficacy Set
Table 14.2.1.15.2	Hamilton Rating Scale for Depression (HAM-D) Remission - Descriptive Statistics and GEE Analysis Subgroup Analysis: Baseline Antidepressant Use (Yes vs No)	Efficacy Set
Table 14.2.1.16.1	Hamilton Rating Scale for Depression (HAM-D) Non-Response - Descriptive Statistics and Logistic Regression Analysis	Efficacy Set
Table 14.2.1.16.2	Hamilton Rating Scale for Depression (HAM-D) Non-Remission - Descriptive Statistics and Logistic Regression Analysis	Efficacy Set
Table 14.2.2.1	Montgomery-Åsberg Depression Rating Scale (MADRS) Total Score - Descriptive Statistics	Efficacy Set
Table 14.2.2.2	Change from Baseline in Montgomery-Åsberg Depression Rating Scale (MADRS) Total Score - MMRM Analysis (Pooled Center as Fixed Effect)	Efficacy Set
Table 14.2.2.3	Change from Baseline in Montgomery-Åsberg Depression Rating Scale (MADRS) Total Score - MMRM Analysis (Pooled Center as Fixed Effect) - Per Protocol Set	Per Protocol Set
Table 14.2.2.4.1	Montgomery-Åsberg Depression Rating Scale (MADRS) Total Score - Descriptive Statistics Subgroup Analysis: Baseline Antidepressant Use (Yes vs No)	Efficacy Set
Table 14.2.2.4.2	Change from Baseline in Montgomery-Åsberg Depression Rating Scale (MADRS) Total Score - MMRM Analysis (Pooled Center as Fixed Effect) Subgroup Analysis: Baseline Antidepressant Use (Yes vs No)	Efficacy Set
Table 14.2.2.5.1	Sensitivity Analysis - Change from Baseline in Montgomery-Åsberg Depression Rating Scale (MADRS) Total Score - Dropout Reason Based Imputation and ANCOVA Model	Efficacy Set
Table 14.2.2.5.2	Sensitivity Analysis - Change from Baseline in Montgomery-Åsberg Depression Rating Scale (MADRS) Total Score - Stratified Wilcoxon Rank-Sum Test	Efficacy Set
Table 14.2.3.1	Generalized Anxiety Disorder 7-item Scale (GAD-7) Total Score - Descriptive Statistics	Efficacy Set
Table 14.2.3.2	Change from Baseline in Generalized Anxiety Disorder 7-item Scale (GAD-7) Total Score - MMRM Analysis (Pooled Center as Fixed Effect)	Efficacy Set
Table 14.2.3.3	Generalized Anxiety Disorder 7-item Scale (GAD-7) Response - Descriptive Analysis and Fisher's Exact Test	Efficacy Set
Table 14.2.4.1	Clinical Global Impression-Severity (CGI-S) Score (Continuous) - Descriptive Statistics	Efficacy Set
Table 14.2.4.2.1	Clinical Global Impression-Improvement (CGI-I) Score (Continuous) - Descriptive Statistics	Efficacy Set
Table 14.2.4.2.2	Clinical Global Impression-Improvement (CGI-I) Score - MMRM Analysis (Pooled Center as Fixed Effect)	Efficacy Set

Table 14.2.4.3	Clinical Global Impression-Improvement (CGI-I) Score (Categorical) - Descriptive Statistics	Efficacy Set
Table 14.2.4.4	Clinical Global Impression-Improvement (CGI-I) Response - Descriptive Analysis and GEE Analysis	Efficacy Set
Table 14.2.4.5.1	Clinical Global Impression-Improvement (CGI-I) Response - Descriptive Statistics and Logistic Regression Analysis	Efficacy Set
Table 14.2.4.5.2	Clinical Global Impression-Improvement (CGI-I) Response - Descriptive Analysis and GEE Analysis Subgroup Analysis: Baseline Antidepressant Use (Yes vs No)	Efficacy Set
Table 14.2.5	Changes in Post-baseline Antidepressant Medications - Kaplan Meier Analysis	Efficacy Set
Table 14.2.6.1	Edinburgh Postnatal Depression Scale (EPDS) Total Score - Descriptive Statistics	Efficacy Set
Table 14.2.6.2	Change from Baseline in Edinburgh Postnatal Depression Scale (EPDS) Total Score - MMRM Analysis (Pooled Center as Fixed Effect)	Efficacy Set
Table 14.2.7.1	Patient Health Questionnaire (PHQ-9) Total Score - Descriptive Statistics	Efficacy Set
Table 14.2.7.2	Change from Baseline in Patient Health Questionnaire (PHQ-9) Total Score - MMRM Analysis (Pooled Center as Fixed Effect)	Efficacy Set
Table 14.2.7.3	Patient Health Questionnaire (PHQ-9) Response - Descriptive Analysis and Fisher's Exact Test	Efficacy Set
Table 14.2.8.1	Barkin Index of Maternal Functioning (BIMF) Total Score and Subscales - Descriptive Statistics	Efficacy Set
Table 14.2.8.2	Change from Baseline in Barkin Index of Maternal Functioning (BIMF) Total Score and Subscales - MMRM Analysis (Pooled Center as Fixed Effect)	Efficacy Set
Table 14.2.9.1	Short Form-36 (SF-36) - Descriptive Statistics	Efficacy Set
Table 14.2.9.2	Change from Baseline in Short Form-36 (SF-36) Scales - MMRM Analysis (Pooled Center as Fixed Effect)	Efficacy Set
Table 14.2.10	Health Care Resource Utilization (HCRU) - Descriptive Statistics	Efficacy Set
Table 14.3.1.1	Summary of Overall Treatment-Emergent Adverse Events (TEAEs)	Safety Set
Table 14.3.1.2	Treatment Emergent Adverse Events (TEAEs) by System Organ Class and Preferred Term	Safety Set
Table 14.3.1.3	Treatment Emergent Adverse Events (TEAEs) by Preferred Term	Safety Set
Table 14.3.1.4	Treatment Emergent Serious Adverse Events (SAEs) by System Organ Class and Preferred Term	Safety Set
Table 14.3.1.5	Treatment Emergent Adverse Events (TEAEs) by System Organ Class, Preferred Term and Maximum Severity	Safety Set
Table 14.3.1.6	Treatment Emergent Adverse Events (TEAEs) by System Organ Class, Preferred Term and Strongest Relatedness to SAGE-547	Safety Set

Table 14.3.1.7	Treatment Emergent Adverse Events (TEAEs) Leading to Study Drug Discontinuation by System Organ Class and Preferred Term	Safety Set
Table 14.3.4.1.1	Hematology - Descriptive Statistics	Safety Set
Table 14.3.4.1.2	Shift from Baseline in Hematology Parameters	Safety Set
Table 14.3.4.1.3	Number and Percentage of Subjects with Potentially Clinically Significant (PCS) values in Hematology Parameters	Safety Set
Table 14.3.4.2.1	Serum Chemistry - Descriptive Statistics	Safety Set
Table 14.3.4.2.2	Shift from Baseline in Serum Chemistry Parameters	Safety Set
Table 14.3.4.2.3	Number and Percentage of Subjects with Potentially Clinically Significant (PCS) values in Serum Chemistry Parameters	Safety Set
Table 14.3.4.2.4	Number and Percentage of Subjects with Liver Function Abnormalities Meeting Hy's Law	Safety Set
Table 14.3.4.2.5	Number and Percentage of Subjects with Worsened Abnormal Liver Enzymes and Liver Function Tests Results	Safety Set
Table 14.3.4.3.1	Coagulation - Descriptive Statistics	Safety Set
Table 14.3.4.3.2	Shift from Baseline in Coagulation Parameters	Safety Set
Table 14.3.4.4.1	12-lead ECG - Descriptive Statistics	Safety Set
Table 14.3.4.4.2	Investigator Interpretation of 12-lead ECG Result - Descriptive Statistics	Safety Set
Table 14.3.4.4.3	Number and Percentage of Subjects Meeting ECG Threshold Criterion	Safety Set
Table 14.3.4.5.1	Vital Signs - Descriptive Statistics	Safety Set
Table 14.3.4.5.2	Number and Percentage of Subjects with Orthostatic Hypotension	Safety Set
Table 14.3.4.6.1	Number and Percentage of Subjects with a Response of Yes to Any Columbia Suicide Severity Rating Scale (C-SSRS) Suicidal Ideation Item	Safety Set
Table 14.3.4.6.2	Number and Percentage of Subjects with a Response of Yes to Any Columbia Suicide Severity Rating Scale (C-SSRS) Suicidal Behavior Item	Safety Set
Table 14.4.1.1	Plasma Pharmacokinetic Parameters of SAGE-547 - Descriptive Statistics	PK Set

12 INDEX OF FIGURES

13 INDEX OF LISTINGS

Number	Title	Analysis Set
Listing 16.2.1.1	Subject Disposition	All Subjects
Listing 16.2.1.2	Subjects not Meeting Study Eligibility Criteria	All Subjects
Listing 16.2.2	Protocol Deviations	Protocol Deviations
Listing 16.2.3	Analysis Sets	All Enrolled Set
Listing 16.2.4.1	Demographics and Baseline Characteristics	All Enrolled Set
Listing 16.2.4.2	Drug and Alcohol Screening Results	All Enrolled Set
Listing 16.2.4.3	Hepatitis and HIV Screening Results	All Enrolled Set
Listing 16.2.4.4	Genetic Testing	All Enrolled Set
Listing 16.2.4.5	Medical History	All Enrolled Set
Listing 16.2.4.6	Subject History of Psychiatric Disorders	All Enrolled Set
Listing 16.2.4.7	Post-Partum Depression Episode Current and Previous	All Enrolled Set
Listing 16.2.4.8	Psychiatric Family History	All Enrolled Set
Listing 16.2.4.9	Perinatal Family History	All Enrolled Set
Listing 16.2.4.10.1	Prior and Concomitant Medications	All Enrolled Set
Listing 16.2.4.10.2	Antidepressant Use	All Enrolled Set
Listing 16.2.5.1	Study Drug Administration	All Enrolled Set
Listing 16.2.5.2	Plasma Pharmacokinetic Sampling and SAGE-547 Concentrations Over Time	All Enrolled Set
Listing 16.2.5.3	Individual Pharmacokinetic Parameters of SAGE-547	All Enrolled Set
Listing 16.2.5.4	Plasma Pharmacokinetic Sampling and SAGE-547 Metabolites Concentrations Over Time	All Enrolled Set
Listing 16.2.5.5	Plasma Pharmacokinetic Sampling and SBECD Concentrations Over Time	All Enrolled Set
Listing 16.2.6.1.0	Hamilton Rating Scale for Depression (HAM-D) - Preface	
Listing 16.2.6.1.1	Hamilton Rating Scale for Depression (HAM-D) - Individual Item Scores	All Enrolled Set
Listing 16.2.6.1.2	Hamilton Rating Scale for Depression (HAM-D) - Subscale and Total Scores	All Enrolled Set
Listing 16.2.6.2.0	Montgomery-Åsberg Depression Rating Scale (MADRS) - Preface	
Listing 16.2.6.2.1	Montgomery-Åsberg Depression Rating Scale (MADRS)	All Enrolled Set
Listing 16.2.6.3.0	Generalized Anxiety Disorder 7-item Scale (GAD-7) - Preface	
Listing 16.2.6.3.1	Generalized Anxiety Disorder 7-item Scale (GAD-7)	All Enrolled Set

Listing 16.2.6.4	Clinical Global Impression(CGI)	All Enrolled Set
Listing 16.2.6.5.0	Edinburgh Postnatal Depression Scale (EPDS) - Preface	
Listing 16.2.6.5.1	Edinburgh Postnatal Depression Scale (EPDS)	All Enrolled Set
Listing 16.2.6.6.0	Patient Health Questionnaire (PHQ-9) - Preface	
Listing 16.2.6.6.1	Patient Health Questionnaire (PHQ-9)	All Enrolled Set
Listing 16.2.6.7.0	Barkin Index of Maternal Functioning (BIMF) - Preface	
Listing 16.2.6.7.1	Barkin Index of Maternal Functioning (BIMF)	All Enrolled Set
Listing 16.2.6.8.0	Short Form-36 (SF-36) - Preface	
Listing 16.2.6.8.1	Short Form-36 (SF-36) - Individual Item Scores	All Enrolled Set
Listing 16.2.6.8.2	Short Form-36 (SF-36) - Subscale and Summary Scores	All Enrolled Set
Listing 16.2.6.9	Healthcare Resource Utilization (HCRU) Questionnaire	All Enrolled Set
Listing 16.2.7.1	Adverse Events	All Enrolled Set
Listing 16.2.7.2	Serious Adverse Events	All Enrolled Set
Listing 16.2.7.3	Adverse Events Leading to Study Drug Discontinuation	All Enrolled Set
Listing 16.2.7.4	Adverse Events Leading to Death	All Enrolled Set
Listing 16.2.8.1	Serum Chemistry	All Enrolled Set
Listing 16.2.8.2	Hematology	All Enrolled Set
Listing 16.2.8.3	Coagulation	All Enrolled Set
Listing 16.2.8.4	Urinalysis	All Enrolled Set
Listing 16.2.8.5	Hormones	All Enrolled Set
Listing 16.2.8.6	Exploratory Biochemistry	All Enrolled Set
Listing 16.2.8.7	Serum and Urine Pregnancy Test	All Enrolled Set
Listing 16.2.8.8	Vital Signs	All Enrolled Set
Listing 16.2.8.9	12-lead ECG	All Enrolled Set
Listing 16.2.8.10	Physical Examination	All Enrolled Set
Listing 16.2.8.11	Columbia Suicide Severity Rating Scale (C-SSRS)	All Enrolled Set

14 REFERENCES

Clinical Study Protocol: amendment 4, version 5.0 (16th March, 2017), Company: Sage Therapeutics Inc.

Implementation of Pattern-Mixture Models Using Standard SAS/STAT Procedures. Bohdana Ratitch, Quintiles, Montreal, Quebec, Canada and Michael O'Kelly, Quintiles, Dublin, Ireland.

15 LIST OF APPENDICES

15.1 Appendix A: Schedule of Assessments

Table 15:Schedule of Events - Part B

	Screening Period	Treatment Period Clinic Period (Day 1 to Day 3)															Follow-up Period			
Visit Days	**Screening D-7 to -1**	**D1 H0***	**D1 H2**	**D1 H4**	**D1 H8**	**D1 H12**	**D1 H18**	**D1 H24**	**D2 H30**	**D2 H36**	**D2 H42**	**D2 H48**	**D3 H54**	**D3 H60**	**D3 H66**	**D3 H72**	**D7/ET** (±1d)	**D14 (+2d)**	**D21 (+3d)**	**D30** (±3d)
Study Procedure																				
Informed Consent	X																			
Inclusion/Exclusion Criteria	X	X																		
Randomization		X																		
Demographics	X																			
Medical/Family History	X																			
Physical Examination	X															X	X			
Body Weight/Height	X																			
Clinical Lab Assessments [a]	X															X	X			
Urinalysis [a]	X																X			
Drug & Alcohol Screen[b]	X	X																		
Pregnancy Test [c]	X	X																		X
Genetic Sample [d]	O																			
Vital Signs [e]	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X			
12-Lead ECG [f]	X											X					X			
C-SSRS [g]		X						X						X		X	X	X	X	X
Confinement		X							X											
CGI-I [h]			X	X		X		X		X		X		X		X	X	X	X	X
CGI-S	X	X																		
SCID-I	X																			
HAM-D [h]	X	X	X	X	X	X		X		X		X		X		X	X	X	X	X

Visit Days	Screening D-7 to -1	D1 H0*	D1 H2	D1 H4	D1 H8	D1 H12	D1 H18	D1 H24	D2 H30	D2 H36	D2 H42	D2 H48	D3 H54	D3 H60	D3 H66	D3 H72	D7/ET (±1d)	D14 (+2d)	D21 (+3d)	D30 (±3d)
Study Procedure																				
MADRS [h]	X	X						X				X		X		X	X	X	X	X
BIMF [h]		X															X	X	X	X
EPDS [h]		X												X			X	X	X	X
GAD-7 [h]		X												X			X	X	X	X
PHQ-9 [h]		X												X			X	X	X	X
SF-36 (acute version)		X															X	X	X	X
HCRU	X	X																		
Plasma PK[i]		X		X	X	X		X	X	X		X		X		X				
Instructions for Lactating Subjects[j]		X							X				X							
Study Drug Infusion		X																		
Adverse Events	X																			
Prior/Concomitant Medications[k]	X																			

O = optional; * = All H0 procedures to be completed prior to dosing

a Safety laboratory tests will include hematology, serum chemistry, coagulation, and select hormone parameters. The urine test will include a urinalysis. Lab assessments are to be completed within ±30 minutes of the scheduled time point.

b Urine for selected drugs of abuse and alcohol (serum or breath)

c Serum at screening and urine for all other time points; lactation status (ie, subject is breastfeeding, subject is lactating but not breastfeeding, or subject is not lactating) will be recorded at screening

d A blood sample for genetic testing, where consent is given

e Vital signs include oral temperature (°C), respiratory rate, heart rate, and blood pressure (supine and standing). Vital signs will be obtained within ±30 minutes of the scheduled time point, unless the subject is asleep between the hours of 23.00h and 06.00h.

f Performed within ±30 minutes of the scheduled time point on Day 2.

g The "Baseline/Screening" C-SSRS form will be completed on Day 1. The "Since Last Visit" C-SSRS form will be completed at all subsequent time points.

h To be completed within ±30 minutes of the scheduled time point during the Treatment Period.

i Blood samples for PK analysis will be collected at pre-infusion and at 4 (before change in infusion rate, if applicable), 8, 12, 24 (before change in infusion rate, if applicable), 30, 36, 48, 60 (before end of infusion), and 72 hours after the start of the infusion. PK blood draws after the start of infusion will have a window of ±10 minutes. In the event of an unplanned dose adjustment, an unscheduled PK sample should be collected just prior to the infusion rate change.

j Breast milk will be pumped and discarded by subjects who are lactating. On Day 3, subjects who are lactating will be reminded that they must continue to pump and discard breast milk through Day 12 of the study.

k To include those taken within 60 days prior to signing the informed consent through the Day 30 visit.
BIMF = Barkin Index of Maternal Functioning; CGI-I = Clinical Global Impression of Improvement; C-SSRS = Columbia Suicide Severity Rating Scale; ECG = electrocardiogram; EPDS = Edinburgh Postnatal Depression Scale; GAD-7 = Generalized Anxiety Disorder 7-Item Scale; HAM-D = Hamilton Rating Scale for Depression, 17-item; HCRU = Health Care Resource Utilization; MADRS = Montgomery-Asberg Depression Rating Scale; PHQ-9 = Patient Health Questionnaire; PK = pharmacokinetic; SF-36 = Short Form-36; SCID-I = Structured Clinical Interview for DSM-IV Axis I Disorders.

15.2 Appendix B: Details of Statistical Methodology

Sample SAS code for MMRM (center as fixed effect):

```
PROC MIXED DATA = xxx;
  CLASS center trt avisit usubjid strata;
  MODEL change = base center trt avisit avisit*trt strata / ddfm=kr;
  REPEATED avisit/subject = usubjid type= un;
* if type= un does not converge, use type= TOEP;TOEP;
* if type=TOEP does not converge, use type= ar(1);
  LSMEANS avisit*trt / diff=all cl alpha=0.05;
RUN;
```

Sample SAS code for MMRM (center as random effect):

```
PROC MIXED DATA = xxx;
  CLASS center trt avisit usubjid strata;
  MODEL change = base center trt avisit avisit*trt strata / DDFM=KR;
  REPEATED subject = usubjid /type= un;
  RANDOM center /type= un;
  RANDOM intercept / subject=usubjid type = un;
* if type= un does not converge, use type= toep;
* if type=toep does not converge, use type ar(1);
  LSMEANS avisit*trt / diff=all cl alpha=0.05;
RUN;
```

Sample SAS code for GEE:

```
PROC GENMOD DATA = xxx;
  CLASS center trt avisit usubjid strata;
  MODEL resp = base center trt avisit avisit*trt strata/ DIST= bin link=logit;
  REPEATED subject=usubjid / type = un;
  LSMEANS avisit*trt / diff exp cl;
RUN;
```

Sample SAS code for Stratified Wilcoxon Rank Sum Test:

```
PROC FREQ DATA = xxx;
  WHERE trtan in (1, 2);
  TABLES strata*trtan*chg/cmh2 scores=modridit noprint;
RUN;
```

Sample SAS code for Fisher's Exact Test:

```
PROC FREQ DATA = xxx ORDER=DATA;
  WHERE trta in ('PLACEBO','SAGE-547 90ug');
  BY avisitn;
  TABLE trta*crit2fl / NOCOL NOPERCENT;
  EXACT OR Fisher;
  ODS OUTPUT CrossTabFreqs=freq FishersExact=fisher OddsRatioCL=cl;
RUN;
```

Sample SAS code for Logistic Regression:

```
PROC LOGISTIC DATA = xxx;
 BY avisit;
  CLASS center trt usubjid strata /PARAM=GLM;
  MODEL resp (event='1') = base center trt strata; Note: for CGI-S, add cgisbase independent
variable as well
  LSMEANS trtan / diff oddsratio cl exp;
RUN;
```

Sample SAS code for MI with FCS statement:

```
PROC MI DATA = xxx SEED = 1213 NIMPUTE = 4 ROUND = 1 OUT = fcs_reg;
  CLASS usubjid trt strata avisit center;
  FCS DISCRIM (usubjid trt avisit center/ details classeffects=include);
  FCS NBITER = 20 LOGISTIC (strata = usubjid/details);
  FCS NBITER = 20 REG(change = usubjid base center trt avisit avisit*trt strata/details);
  VAR change strata usubjid trt avisit center base;
RUN;
```

Sample SAS code for ANOVA:

```
PROC ANOVA DATA = xxx;
  CLASS center trt strata;
  MODEL aval = center trt strata;
RUN;
```

Sample SAS code for ANCOVA:

```
PROC MIXED DATA = xxx;
  CLASS center trt strata;
  MODEL change = base center trt strata;
RUN;
```