Official Title: A Phase 3, Multicenter, Randomized, Double-blind, Placebo-controlled Study to Evaluate the Efficacy and Safety of Adjunctive Pimavanserin in Subjects With Major Depressive Disorder and Inadequate Response to Antidepressant Treatment (Sponsor Protocol Numbers: ACP-103-054, ACP-103-059)

NCT Numbers: NCT03999918, NCT03968159

Document Date: 22 Jun 2020

STATISTICAL ANALYSIS PLAN

Protocol No.:	ACP-103-059 and ACP-103-054
Protocol Title:	A Phase 3, Multicenter, Randomized, Double-blind, Placebo-controlled Study to Evaluate the Efficacy and Safety of Adjunctive Pimavanserin in Subjects With Major Depressive Disorder and Inadequate Response to Antidepressant Treatment
Drug:	Pimavanserin
Sponsor:	ACADIA Pharmaceuticals Inc.
Version No. and Date	Version 1.0, 22 June 2020

SIGNATURE/APPROVAL PAGE

AUTHORS

[redacted]

22JUN2020

Date

Senior Director, Biostatistics
ACADIA Pharmaceuticals Inc.

APPROVERS

[redacted]

22 June 2020

Date

Executive Director, Clinical Research
ACADIA Pharmaceuticals Inc.

[redacted]

22 JUN 2020

Date

Executive Director, Biostatistics
ACADIA Pharmaceuticals Inc.

TABLE OF CONTENTS

LIST OF TABLES

ABBREVIATIONS

AE	Adverse event
ANCOVA	analysis of covariance
ATC	Anatomical/Therapeutic/Chemical
BIS-11	Barratt Impulsiveness Scale
BMI	body mass index
CI	Confidence Interval
CGI-I	Clinical Global Impressions – Global Improvement
CGI-S	Clinical Global Impressions – Severity of Illness
CMH	Cochran-Mantel-Haenszel
COVID-19	coronavirus disease 2019
CRO	Contract Research Organization
CSFQ-14	Changes in Sexual Functioning Questionnaire Short Form
C-SSRS	Columbia Suicide Severity Rating Scale
eCRF	electronic case report form
ECG	electrocardiogram
EDC	electronic data capture
ESRS-A	Extrapyramidal Symptom Rating Scale–Abbreviated
FAS	Full Analysis Set
FDA	Food and Drug Administration
GSD	Guidance for Site Documentation and Data Management Querying of Data Impacted by COVID-19
HAMD-17	Hamilton Depression Scale (17 Items)
IRT	Interactive Response Technology
ITT	Intention-to-treat
KSS	Karolinska Sleepiness Scale
LOCF	last observation carried forward
MADRS	Montgomery Asberg Depression Scale
MDD	Major Depressive Disorder
MDE	Major Depressive Episode
MedDRA	Medical Dictionary for Regulatory Activities
MGH-ATRQ	Massachusetts General Hospital Antidepressant Treatment Questionnaire
MGH-SFI	Massachusetts General Hospital Sexual Functioning Index
MINI	Mini-International Neuropsychiatric Interview

MMRM	mixed model for repeated measures
MNAR	missing not at random
OC	observed cases
PCI	potentially clinically important
PD	Pharmacodynamic(s)
PK	Pharmacokinetic(s)
PMM	Pattern-Mixture Models
QoL	quality of life
QTcB	QT Interval Corrected for Heart Rate using Bazett's Formula
QTcF	QT Interval Corrected for Heart Rate using Fridericia's Formula
QTcNi	QT Interval Corrected for Heart Rate using individual corrections
SAE	serious adverse event
SAFER	State versus trait, Assessability, Face validity, Ecological validity, and Rule of three Ps (pervasive, persistent, and pathological)
SAP	statistical analysis plan
SD	Standard Deviation
SDS	Sheehan Disability Scale
SE	Standard Error
SIS	Sheehan Irritability Scale
SNRI	serotonin-norepinephrine reuptake inhibitors
SOC	system organ class
SSRI	selective serotonin reuptake inhibitors
TEAE	treatment-emergent adverse event

1. INTRODUCTION

This statistical analysis plan (SAP) provides a technical and detailed elaboration of the statistical analyses of combined efficacy and safety data as described in both ACP-103-059 Protocol Amendment 3 dated 21 June 2020 and ACP-103-054 Protocol Amendment 4 dated 21 June 2020.

The two studies have exactly the same design with the exception of where they were conducted. Study ACP-103-059 was conducted in the United States (US) and Study ACP-103-054 was conducted outside of the US. Unless specified otherwise, discussions about study design apply to both of the 2 studies.

2. OBJECTIVES

2.1 Primary Objective

The primary objective of the two studies is to evaluate the efficacy of adjunctive pimavanserin compared to placebo in subjects with major depressive disorder who have an inadequate response to antidepressant therapy.

2.2 Secondary Objective

Secondary objective of the two studies is to evaluate the efficacy and benefits of adjunctive pimavanserin compared to placebo in subjects with major depressive disorder who have an inadequate response to antidepressant therapy on the following:

- Functional impairment
- Clinician's global impression of severity and improvement of depressive symptoms
- Sexual functioning
- Sleepiness
- Treatment response and remission
- Anxiety
- Impulsiveness
- Early response to treatment

2.3 Safety Objective

The safety objective of the two studies is to assess the safety and tolerability of pimavanserin compared to placebo in subjects with major depressive disorder who have an inadequate response to antidepressant therapy.

3. STUDY DESIGN

3.1 General Study Design

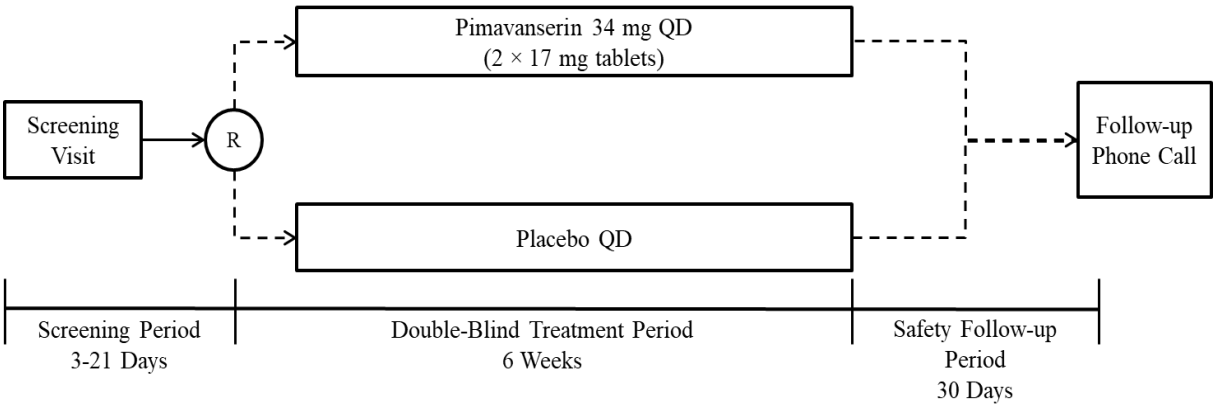
The two studies are multicenter, randomized, double-blind, placebo-controlled, parallel-group studies in subjects with major depressive disorder (MDD) and inadequate response to a selective serotonin reuptake inhibitor (SSRI) or serotonin-norepinephrine reuptake inhibitor (SNRI).

The duration of participation for individual study subjects will be up to 14 weeks, consisting of a screening period of up to 4 weeks (i.e., 3 weeks, with possible extension of up to 7 days), a 6-week double-blind treatment period, and a safety follow-up period of at least 30 days (for those subjects who discontinue prematurely or who do not enroll in the open-label extension study). The study completion date is defined as the date the final subject, across all sites, completes their final protocol-defined assessment, which includes the safety follow-up visit/contact. If the studies are terminated for any reason, subjects remaining in the studies will return to standard of care.

Approximately 40 sites in the United States and approximately 40 sites outside of the United States will participate in ACP-103-059 and ACP-103-054 respectively. Approximately 280 subjects will be randomized in a 1:1 ratio to either the pimavanserin or the placebo treatment groups in each of the two individual studies.

The study design is summarized in Figure 1.

Figure 1 Schematic of Study Design



Abbreviations: QD=once daily; R=randomization
Note: Subjects who enroll in an open-label extension will not enter the Safety Follow-up Period.

3.2 Schedule of Assessments

Schedule of assessments can be found in Appendix A.

3.3 Randomization

Eligible subjects will be randomized at the Baseline visit (Visit 2) in a 1:1 ratio to receive in addition to their SSRI/SNRI antidepressant: Pimavanserin (34 mg) or placebo. The assignments will be based on a pre-generated permuted-block randomization schedule. Randomization will be electronically performed using an Interactive Response Technology (IRT) based randomization system.

3.4 Blinding

Blinding is achieved by restricting access of Investigators and Sponsor personnel and/or designee to the treatment codes, and providing identical tablets and packaging for the pimavanserin and placebo treatments.

Unblinding of individual treatment assignment during the study is discouraged. The Investigator may break the blind in the event of a medical emergency if it is considered necessary for the care of the subject. The Investigator should attempt whenever possible to contact the Medical Monitor before unblinding a subject's treatment to discuss the event.

For the final analysis, the treatment codes for all subjects will be released to ACADIA after all subjects have completed the study and the clinical database is locked.

3.5 Determination of Sample Size

A total sample size of 266 evaluable subjects was estimated to provide at least 90% power at a two-sided significance level of 0.05 when assuming a treatment effect size of 0.4 (an estimated common standard deviation of 8.0 based on data from the Phase 2 study ACP-103-042) for pimavanserin compared to placebo on the change from Baseline to Week 5 in the HAMD-17 total score. Adjusting for a potential non-evaluable rate of up to 5%, approximately 280 subjects will be randomized.

3.6 Coronavirus Disease 2019

In March, 2020, the emerging coronavirus disease 2019 (COVID-19) pandemic resulted in implementation of urgent safety measures designed to ensure subject safety. Mechanisms to record information on the potential impact of the COVID-19 pandemic on data itself, as well as data collection and integrity, were implemented (as detailed in the "Guidance for Site Documentation and Data Management Querying of Data Impacted by COVID-19" [GSD] of the Data Management Plan).

Screening and enrollment into studies ACP-103-054 and ACP-103-059 were temporarily halted. The impact of COVID-19 on the statistical analysis is discussed in each of the relevant sections of this SAP.

4. ANALYSIS SETS

Randomized Analysis Set

The Randomized Analysis Set includes all subjects who were randomized in the 2 studies. Subjects will be classified according to the randomized treatment assignment.

Safety Analysis Set

The Safety Analysis Set includes all subjects who received at least one dose of study drug in the 2 studies. Subjects will be classified according to the actual treatment received.

Full Analysis Set

The Full Analysis Set includes all subjects who were randomized, received at least one dose of study drug, and have both a Baseline (Week 0) value and at least one post-Baseline value for the HAMD-17 total score in the 2 studies. Subjects will be classified according to the randomized treatment assignment.

Per-protocol Analysis Set

The Per-protocol Analysis Set includes a subset of subjects in the Full Analysis Set without any protocol deviations that could have a significant effect on the study conclusions. The Per-protocol Analysis Set will be determined prior to unblinding the studies for the final analysis. Subjects will be classified according to the randomized treatment assignment.

5. DATA HANDLING CONVENTIONS

All data collected in the study will be listed.

5.1 General Data Reporting Conventions

Continuous variables will be summarized using the following descriptive statistics: number of subjects, mean, median, standard deviation, standard error, minimum, and maximum. Unless specified otherwise, means, medians, and confidence intervals will be presented to one more decimal place than the raw data, and the standard deviations and standard errors will be presented to two more decimal places than the raw data.

Categorical and count variables will be summarized by the number of subjects and the percent of subjects in each category. Unless specified otherwise, the denominator of the percentage is the number of subjects with non-missing data. Categories with zero counts will not have zero percentages displayed. Percentages will be presented with one decimal place.

Duration in months will be calculated as ([the number of days / 365.25] *12).

Unless specified otherwise, all statistical tests will be 2-sided hypothesis tests performed at the significance level of 5% for main effects and all confidence intervals will be 2-sided 95% confidence intervals. P-values will generally be presented to 4 decimal places; values less than 0.0001 will be presented as <0.0001.

5.2 Derived Efficacy Variables

In general, assessment total scores and subscores will be derived within the analysis datasets. In the event that total scores and/or subscores are also collected on the electronic case report from (eCRF), the derived values will be used for all analyses. Both the raw and derived scores will be presented in listings.

5.2.1 Hamilton Rating Scale for Depression – 17 Items

The HAMD-17 is completed with the Structured Interview Guide for the Hamilton Depression Rating Scale (SIGH-D) by the Investigator or designee based on an assessment of a patient's symptoms. The time frame for this scale is the past 7 days.

The HAMD-17 consists of 8 items with a score on a 3 point scale and 9 items with a score on a 5 point scale. The total score ranging from 0 to 52 will be calculated as the sum of the scores for all 17 items. Higher total scores denote more severe depression. Missing item scores will be imputed and rounded to the nearest integer as follows:

- Missing scores for items with a score on a 3 point scale will be imputed using the arithmetic mean of the non-missing scores for items with a score on a 3 point scale
- Missing scores for items with a score on a 5 point scale will be imputed using the arithmetic mean of the non-missing scores for items with a score on a 5 point scale

The total score will be considered as missing if there are missing scores for 4 or more items.

The Anxiety/Somatization factor of the HAMD-17 includes 6 items: psychic anxiety, somatic anxiety, gastrointestinal somatic symptoms, general somatic symptoms, hypochondriasis, and insight. The HAMD-17 Anxiety/Somatization factor score ranging from 0 to 18 will be calculated as the sum of the scores for the 6 items. Higher factor scores denote more severe anxiety/somatization condition. Missing item scores will be imputed and rounded to the nearest integer as follows:

- Missing scores for items (e.g. gastrointestinal somatic symptoms, general somatic symptoms, and insight) with a score on a 3 point scale will be imputed using the arithmetic mean of the non-missing scores for items with a score on a 3 point scale among the 6 items

- Missing scores for items (e.g. psychic anxiety, somatic anxiety, and hypochondriasis) with a score on a 5 point scale will be imputed using the arithmetic mean of the non-missing scores for items with a score on a 5 point scale among the 6 items

The HAMD-17 Anxiety/Somatization factor score will be considered as missing if there are missing scores for 2 or more items.

5.2.2 Sheehan Disability Scale (SDS)

The SDS is a 3-item subject-facing questionnaire used to evaluate impairments in the domains of work, social life/leisure, and family life/home responsibility. Subjects rate each item using an 11-point scale ranging from 0 (not at all) to 10 (extremely).

The primary analysis will be performed on the mean score ranging from 0 to 10 calculated as the arithmetic mean of the scores for all 3 items and rounded to the nearest 2 decimal places. Higher mean scores denote greater disability. If a subject indicates that no work/school occurred, the mean score will be calculated as the arithmetic mean of the other two item responses (Social Life and Family Life/Home Responsibilities) and rounded to the nearest 2 decimal places. If either of the other two item responses is missing, or if the work/school response is missing and the subject does not indicate that no work/school occurred, the mean score will be missing.

The following supplementary analysis will be performed:

Total Score without Imputation

The SDS total score ranging from 0 to 30 will be calculated as the sum of the scores for all 3 items. Higher total score denote greater disability. The total score will be missing if at least one of the 3 item responses is missing.

The mean score imputed based on only two item responses (Social Life and Family Life/Home Responsibilities) will be flagged in data listings. These results will be considered observed values for data summaries and analyses.

5.2.3 Clinical Global Impressions – Severity (CGI-S) and Clinical Global Impressions – Improvement (CGI-I) Scales

The CGI-S and CGI-I are scales used by the Investigator or designee to rate the severity of the disorder and the global improvement since beginning of the study. The CGI-S rates the severity of a subject's depression over the past 7 days and the CGI-S score ranges from 1 to 7:

1 = Normal, not at all ill
2 = Borderline mentally ill
3 = Mildly ill
4 = Moderately ill
5 = Markedly ill
6 = Severely ill
7 = Among the most extremely ill patients

The CGI-I rates the change in a subject's depression over the past 7 days relative to the subject's symptoms at Baseline and the CGI-I score ranges from 1 to 7:

1 = Very much improved	5 = Minimally worse
2 = Much improved	6 = Much worse
3 = Minimally improved	7 = Very much worse
4 = No change	

Higher CGI-S and CGI-I scores denote more severe depression and less improvement in depression respectively. Missing CGI-S and CGI-I scores will not be imputed.

5.2.4 Changes in Sexual Functioning Questionnaire Short Form (CSFQ-14)

The CSFQ-14 is a 14-item version of the CSFQ. This is a patient-facing questionnaire, with a male version and a female version. The total score ranging from 14 to 70 will be calculated as the sum of the scores for all 14 items. Higher total scores denote better sexual functioning. Missing item scores will be imputed using the arithmetic mean of the non-missing item scores and rounded to the nearest integer. The total score will be considered as missing if there are missing scores for 3 or more items.

5.2.5 Karolinska Sleepiness Scale (KSS)

The KSS is a scale that measures the subject's drowsiness and is frequently used in studies measuring subjective sleepiness. Scoring is based on a 9-point verbally anchored scale going from "1 = extremely alert" to "9 = very sleepy, great effort to keep awake, fighting sleep". Higher scores denote more drowsiness. Missing scores will not be imputed.

5.2.6 Barratt Impulsiveness Scale (BIS-11)

The BIS-11 is a questionnaire designed to assess the personality/behavioral construct of impulsiveness. It is composed of 30 items describing common impulsive or non-impulsive (reverse scored items: 1, 7, 8, 9, 10, 12, 13, 15, 20, 29, and 30) behaviors and preferences. Items are scored on the following 4-point scale: Rarely/Never = 1; Occasionally = 2; Often = 3; Almost Always/Always = 4. For reverse scored items, a response of 1 is recoded to 4; 2 is recoded to 3; 3 is recoded to 2; and 4 is recoded to 1. The BIS-11 score ranging from 30 to 120 will be calculated as the sum of the scores for all 30 items. Higher scores denote more impulsiveness. Missing item scores will be imputed with the arithmetic mean of the non-missing item scores and rounded to the nearest integer. The BIS-11 score will be considered as missing if there are missing scores for 7 or more items.

5.2.7 Montgomery Asberg Depression Rating Scale (MADRS)

The MADRS is a 10-item, clinician-rated instrument measuring depression severity. It is scored on a fixed scale of 7 points (0-6) following a structured clinician interview. The total score ranging from 0 to 60 will be calculated as the sum of the scores for all 10 items. Higher scores reflect more severe symptomatology. The total score will be considered missing if there are at least one missing item scores.

5.2.8 Extrapyramidal Symptom Rating Scale–Abbreviated (ESRS-A)

The ESRS was developed to assess drug induced movement disorders such as parkinsonism, akathisia, dystonia and tardive dyskinesia with established reliability, validity, and sensitivity. The ESRS-A, an accepted modified form of the original ESRS, is used during the study to monitor for any worsening in extrapyramidal symptoms or signs at scheduled and unscheduled visits. The CGI-S for each movement subtype (i.e., parkinsonism, dystonia, dyskinesia, and akathisia) rates the severity of a subject's movement disorder and the CGI-S score for each movement subtype ranges from 0 to 5:

0 = Absent	1 = Minimal	2 = Mild
3 = Moderate	4 = Severe	5 = Extreme

Higher CGI-S scores denote more severe movement disorder. Missing scores will not be imputed.

5.3 Analysis Visit Windows

Baseline of the study is defined as the last non-missing result, including results from repeated and unscheduled measurements, before first dosing.

Efficacy, safety, and PK assessments will be summarized by analysis visit as presented in Table 1 below.

Table 1 Analysis Visit Windows

Analysis Visit	Study Visit	Target Study Day @	Study Day Interval
Baseline (Day 1)	2 (Week 0)	1	≤ 1
Week 1	3 (Week 1)	8	2 – 11
Week 2	4 (Week 2)	15	12 – 18
Week 3	5 (Week 3)	22	19 – 25
Week 4	6 (Week 4)	29	26 – 32
Week 5	7 (Week 5)	36	33 – 39
Week 6	8 (Week 6)	43	40 – 46
Safety Follow-up	9 (Week 10)	73	≥47

@ Derivation of study day: study day = assessment date - first dose date + 1 if the assessment date ≥ first dose date, otherwise study day = assessment date – first dose date. Study day 1 is the day of first administration of study drug (pimavanserin or placebo).

5.3.1 Unscheduled Assessments

Both Scheduled and Unscheduled assessments, including the assessments at early termination visits, will be used for planned timepoint analyses. All assessments will be presented in data listings.

5.3.2 Multiple Measurements within Visit Windows

In the event that more than one assessment falls within a given window the assessment closest to the target study day will be selected for the by-visit analysis. If two assessments are equidistant from the target study day then the chronologically last assessment will be used. Exceptions may be made for incomplete assessments, in which case, more complete assessments may be given priority. Details are provided in a separate programming conventions document.

For safety analyses where the extreme values should be selected (e.g., overall post-Baseline minimum, overall post-Baseline maximum, and potentially clinically important values), all non-missing post-Baseline values should be considered, regardless of whether the value is selected for the by-visit summaries. All assessments will be presented in data listings.

5.4 Missing or Incomplete Date for Last Dose of Study Drug

In the Safety Analysis Set, if the last dose date of study drug is missing for a subject who completed or early terminated from the study, then the missing last dose date of study drug will be imputed using the last expected dosing date which is defined as the minimum of

(last drug dispense date + 7) or EOT/ET date, whichever occurs earlier. For the incomplete last dose date of study drug, the imputation algorithms will be detailed in the analysis dataset specification document. The missing or incomplete dates will be displayed in the data listings as reported on the eCRF rather than the imputed dates.

5.5 Missing or Incomplete Dates for Prior or Concomitant Medications

Missing or incomplete medication start or stop dates will be imputed for the purpose of determining whether the medication is taken concomitantly or not (see Section 11 for definition). When the chronological order of medication use relative to the study drug treatment period is unclear due to missing or incomplete date(s), the medication will be considered as concomitant. The imputation algorithms will be detailed in the analysis dataset specification document. The missing or incomplete dates as captured on the eCRF will be displayed in the data listings.

5.6 Missing or incomplete Date for Adverse Events

Missing or incomplete adverse event (AE) start dates will be imputed for the purpose of determining whether the AEs are treatment-emergent or not (see Section 14.1 for definition). When the chronological order of an AE onset relative to the study drug treatment period is unclear due to missing or incomplete date(s), the AE will be considered as treatment-emergent. The imputation algorithms will be detailed in the analysis dataset specification document. The missing or incomplete dates captured on the eCRF will be displayed in the data listings.

5.7 Missing Severity Assessment for Adverse Events

If the severity is missing for an AE starting on or after the date of the first dose of study drug, then a severity of "Severe" will be assigned. The imputed values for severity assessment will be used for incidence summaries, while the actual values will be used in data listings.

5.8 Missing Relationship to Study Drug for Adverse Events

If the relationship to study drug is missing for an AE starting on or after the date of the first dose of study drug, a causality of "Related" will be assigned. The imputed values for relationship to study drug will be used for incidence summaries, while the actual values will be presented in data listings.

5.9 Character Values of Clinical Laboratory Variables

If the reported value of a clinical laboratory variable cannot be used in a statistical analysis due to, for example, a character string reported for a numeric variable, an appropriately determined coded value may be used in the statistical analysis. The coding algorithms will be detailed in the analysis dataset specification document. The actual values as reported in the database will be presented in data listings.

6. SUBJECT DISPOSITION

For subjects who participate in the screening phase but are not randomized (screen failures), their demographics information (including age, sex, and primary race), screen failure reasons (the specific inclusion/exclusion criterion (or criteria) not met or other reasons including the reason due to coronavirus disease 2019 [COVID-19]) and protocol version will be listed. If a subject is re-screened, then the re-screening subject ID and the final enrollment status (whether eventually enrolled) will also be displayed in this listing. In addition, the frequency that the screen failure reasons are reported will also be summarized. Note that one subject may be deemed ineligible for multiple inclusion/exclusion criteria and may be allowed to rescreen with the permission of the Medical Monitor, provided the screen failure was due to a temporary condition that subsequently resolved.

The number of sites (and the countries the sites belongs to) that screened at least 1 subject, number of sites that randomized at least 1 subject, number of subjects screened, and number of unique subjects screened will be tabulated. In addition, the number of subjects enrolled at each site (and the country the site belongs to) will also be tabulated by analysis set and by treatment group and overall.

For randomized subjects, number and percentage of subjects in the Safety Analysis Set, Full Analysis Set, and Per-protocol Analysis Set will be summarized by treatment group and overall. A listing will be provided displaying all subjects excluded from the Safety, Full or Per-protocol Analysis Sets, and will include reason(s) for exclusion. The number and percentage of subjects who are excluded from the Per-protocol Analysis Set will be presented in a summary table by reason, and by treatment group and overall.

Within each analysis set, the number and percentage of subjects who completed the studies or discontinued (all discontinued and by discontinuation reasons including the reason due to COVID-19) will also be summarized by treatment group and overall.

7. PROTOCOL DEVIATIONS

Protocol deviations will be reviewed periodically over the course of the studies. The review process, definition of the deviation categories, and the classification of a deviation as major or minor are detailed in the Protocol Deviation Management Plan. Protocol deviations will also be assessed with respect to relationship to COVID-19.

A summary of the number and percentage of subjects with major protocol deviations for each deviation category will be presented by treatment group for the Randomized Analysis Set in three ways: all protocol deviations, COVID-19 related protocol deviations, and non COVID-19 related protocol deviations.

Two listings of major protocol deviations will be provided: all deviations and non COVID-19 related protocol deviations. A listing of all COVID-19 related protocol deviations including the major and the minor will be provided.

8. DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS

Demographics and baseline characteristics will be summarized by treatment group and overall for the Randomized, Safety, Full, and Per-Protocol Analysis Sets using descriptive statistics. In addition, demographics and baseline characteristics will also be summarized by treatment group and overall for both ACP-103-054 and ACP-103-059 separately using the Full Analysis Set. Variables include age, age group (18-40 years and > 40 years), sex, race, ethnicity, height, weight, BMI, Baseline Montgomery Asberg Depression Rating Scale (MADRS) total score, Baseline HAMD-17 total score, Baseline HAMD-17 total score < 24 and ≥ 24, Baseline SDS mean score, and Baseline Clinical Global Impressions-Severity Illness (CGI-S). Baseline CGI-S will be summarized as both a continuous variable and a categorical variable.

Race will also be categorized by White vs. Non-White. The reported age reflects a subject's age at the informed consent date.

Depression history will be summarized by treatment group for the Randomized, Safety, Full, and Per-protocol Analysis Sets using descriptive statistics. Variables include:

- Age at first onset of depression symptoms (years)
- Age at MDD diagnosis (years)
- Number of depression episodes during the subject's lifetime
    - Summary statistics for both the continuous variable and the categorical variable (1 to 5 and ≥6)
- Number of hospitalizations for depression during the subject's lifetime
    - Summary statistics for both the continuous variable and the categorical variable (0; 1 to 5; and ≥6)
- Duration of the current episode of depression (months relative to the informed consent date)
- Ever had suicidal ideation or behavior in the past 6 months or lifetime (yes or no)

Depression history as described above, date of last hospitalization (if number of hospitalizations is greater than 0), and onset date for the current episode of depression will also be presented in a data listing.

The Mini-International Neuropsychiatric Interview (MINI) is performed at Screening visit. The MINI data will be summarized separately.

The data captured in SAFER remote interview at Screening visit including the percentage improvement, MADRS score, and CGIS score will be summarized.

9. MEDICAL HISTORY

Medical history reported terms will be coded with Medical Dictionary for Regulatory Activities (MedDRA), version 19.0 or newer. The subject incidence will be summarized for each system organ class (SOC) and preferred term by treatment group and overall for the Safety, Full, and Per-protocol Analysis Sets. A subject will be counted only once per SOC or per preferred term for the summary.

A listing of the SOC, preferred term, body system, verbatim for the medical history condition/event, start and stop dates (when available), and an indicator for whether or not the condition is ongoing will be provided.

10. EXTENT OF EXPOSURE AND TREATMENT COMPLIANCE

Extent of exposure and treatment compliance will be summarized as continuous variables by treatment group for the Safety Analysis Set.

10.1 Exposure to Study drug

Duration of exposure to study drug will be calculated for each subject as (last dose date – first dose date + 1). The number and percentage of subjects within each of the following exposure levels in terms of duration of exposure will also be tabulated: <1 week (1 to 6 days), 1 to <2 weeks (7 to 13 days), 2 to <3 weeks (14 to 20 days), 3 to <5 weeks (21 to 34 days), and >5 weeks (35 days or longer). Kaplan-Meier curves of duration on study drug will also be presented for each treatment group.

10.2 Measurement of Treatment Compliance

Study drug dosing compliance is defined as the total number of tablets actually taken by a subject divided by the number of tablets expected to be taken and then multiplied by 100. The total number of tablets actually taken is calculated by the total number of tablets dispensed minus the number of tablets returned. The number of tablets expected to be taken is calculated as the duration of exposure multiplied by 2 (the planned number of tablets taken per day).

Treatment compliance will also be summarized as a categorical variable. The number and percentage of subjects within each of the following compliance levels will be tabulated: <80%, 80 to 120%, and >120%.

11. PRIOR, CONCOMITANT, AND POST-TREATMENT MEDICATION

Prior medication is defined as any medication with the start and stop dates prior to the date of the first dose of study drug. Concomitant medication is defined as any medication with a start date prior to the date of the first dose of study drug and continuing past the first dose of study drug or with a start date between the dates of the first and last doses of study drug, inclusive. Any medication with a start date after the date of the last dose of study drug will be

considered as post-treatment medication. Medications will be coded using WHO Drug Dictionary March 2020. The number and percentage of subjects taking each drug class (ATC Level 3) and medication preferred term will be tabulated by treatment group and overall for the Safety Analysis Set. Multiple medication usage by a subject in the same category will be counted only once.

Psychiatric Medication History

The number and percentage of subjects having psychiatric medication history collected on the Psychiatric Medication History CRF pages will be tabulated by Anatomical/Therapeutic/Chemical (ATC) Level 3, preferred term, treatment group, and overall for the Safety Analysis set. Multiple medication usage by a subject in the same category will be counted only once. Listing of all psychiatric medications will be provided.

Prior and Concomitant Medications

Prior and concomitant medications will be summarized separately in the same way as the psychiatric medication history.

Post-Treatment Medications

Post-treatment medications will be summarized for the Safety Analysis Set by ATC Level 3, preferred term, and the study drug (placebo vs. pimavanserin) of which subjects received their last doses.

COVID-19 Related Medications

Relationship to COVID-19 will be assessed for selected medications as detailed in the GSD. Concomitant and post-treatment medication analyses described above will also be summarized by relationship to COVID-19 (Not related to COVID-19 vs. Related to COVID-19).

12. EFFICACY ANALYSES

Unless otherwise specified, all efficacy analyses will be performed using the planned treatment assignments based on the randomization schedule for the Full Analysis Set.

12.1 Efficacy Variables

Primary Efficacy Endpoint

The primary efficacy endpoint is the change from Baseline to Week 5 in the HAMD-17 total score.

Secondary Efficacy Endpoints

The secondary efficacy endpoints are the following:

- Change from Baseline to Week 5 in the SDS mean score
- Change from Baseline to Week 5 in Clinical Global Impression–Severity (CGI-S) score for depressive symptoms
- Clinical Global Impression–Improvement (CGI-I) score for depressive symptoms at Week 5
- Change from Baseline to Week 5 in the Changes in Sexual Functioning Questionnaire Short Form (CSFQ-14)
- Change from Baseline to Week 5 in Karolinska Sleepiness Scale (KSS) score
- Treatment responder rates at Week 5. Treatment response is defined as a reduction from Baseline in HAMD-17 total score of 50% or more.
- Treatment remission rates at Week 5. Treatment remission is defined as a HAMD-17 total score ≤7.
- Change from Baseline to Week 5 in the Hamilton Depression (HAMD) Anxiety/Somatization factor score
- Change from Baseline to Week 5 in the Barratt Impulsiveness Scale (BIS-11)
- Change from Baseline to Week 1 in the HAMD-17 total score

12.2 Adjustment for Covariates

The corresponding baseline value will be included as a covariate and study (ACP-103-059 and ACP-103-054) will be included as a factor for the analysis of HAMD-17 total score, CGI-S score, SDS mean score, CSFQ-14 score, KSS score, and BIS-11 score using the mixed model for repeated measures (MMRM) as described in Section 13.

12.3 Handling of Missing Data

The primary analyses of

- Change from Baseline to Week 5 in the HAMD-17 total score
- Change from Baseline to Week 5 in CGI-S score for depressive symptoms
- Change from Baseline to Week 5 in the SDS mean score
- Change from Baseline to Week 5 in the CSFQ-14 score
- Change from Baseline to Week 5 in KSS score
- Change from Baseline to Week 1 in the HAMD-17 total score

will be performed assuming missing at random (MAR) using the direct likelihood-based MMRM. Scores that are missing, after any imputation of individual missing items as described in Section 5.2, will not be imputed. The MMRM method is unbiased in the estimation of treatment effect under the MAR assumption and can be thought of as aiming to estimate the treatment effect that could have been observed if all subjects had continued on treatment for the full study duration (EMA, 2009).

12.4 Multiple Comparisons / Multiplicity

Family-wise type I error will be controlled at the 2-sided significance level of 5% in hypothesis testing for 6 efficacy endpoints as a family using a sequential testing strategy. The 6 efficacy endpoints will be tested sequentially in the following order:

1. Change from Baseline to Week 5 in the HAMD-17 total score (primary efficacy endpoint)
2. Change from Baseline to Week 5 in CGI-S score for depressive symptoms (secondary efficacy endpoint)
3. Change from Baseline to Week 5 in the SDS mean score (secondary efficacy endpoint)
4. Change from Baseline to Week 5 in the CSFQ-14 score (secondary efficacy endpoint)
5. Change from Baseline to Week 5 in KSS score (secondary efficacy endpoint)
6. Change from Baseline to Week 1 in the HAMD-17 total score (secondary efficacy endpoint)

The hypothesis testing for the primary efficacy endpoint will be performed first. If the primary efficacy endpoint is not statistically significant at the 2-sided significance level of 5%, no further testing for the secondary efficacy endpoints will be performed. Otherwise, the hypothesis testing for the CGI-S score will be performed. In general, the hypothesis testing for an endpoint in the sequence will be performed only if the preceding hypothesis test is statistically significant at the 2-sided significance level of 5%. P-values for endpoints not formally tested will be interpreted as exploratory. Unadjusted p-values will be reported and the testing sequence will be used to determine statistical significance.

12.5 Examination of Subgroups

Treatment effect will be examined with respect to the primary and the selected secondary efficacy endpoints listed in Section 12.4 within the subgroups of study (ACP-103-059 and ACP-103-054), sex (male and female), race (white and non-white), age (18-40 years and > 40 years), the severity of depression at Baseline (Baseline HAMD-17 total score < 24 and ≥ 24), and the background antidepressant type (SSRI and SNRI).

13. METHODS OF EFFICACY ANALYSES

13.1 Primary Efficacy Analysis

13.1.1 Primary Analysis

Estimand

Treatment difference in the change from Baseline to Week 5 in the HAMD-17 total score if all subjects tolerated or adhered.

Population: Subjects with MDD and inadequate response to an SSRI/SNRI, as defined by the inclusion/exclusion criteria of the study.

Variable: Change from Baseline to Week 5 in the HAMD-17 total score.

Intercurrent events: Possible intercurrent events include, but are not limited to, treatment discontinuation due to various reasons, treatment non-compliance, and COVID-19. The treatment policy and hypothetical strategies will be used in the primary analysis. Observations on the primary efficacy endpoint will be used regardless of the occurrence of intercurrent events. Subjects who discontinued early with missing data will be assumed to follow the same clinical course as those in the same treatment group who complete the study treatment.

Summary measure: Treatment difference in the mean change from Baseline to Week 5 in the HAMD-17 total score.

The estimand addresses the hypothesis about the causal effects of the initially randomized study drug. It is assumed that all subjects take the study drug as instructed, when in fact there will always be some subjects that fail to adhere.

Hypotheses

Let Δ ($\boldsymbol{\mu}_{PIM} - \boldsymbol{\mu}_{PBO}$) be the difference in the mean change from Baseline to Week 5 in the HAMD-17 total score between the pimavanserin ($\boldsymbol{\mu}_{PIM}$) and placebo ($\boldsymbol{\mu}_{\mathrm{PBO}}$) groups.

The null hypothesis is: $\Delta = 0$.

The alternative hypothesis is: $\Delta \neq 0$.

Analysis

The hypothesis testing will be performed for the Full Analysis Set using the direct likelihood, MMRM method assuming missing data are missing at random (MAR). The response variable of the MMRM is the change from Baseline in the HAMD-17 total score and the effects include treatment group, visit, treatment-by-visit interaction, Baseline HAMD-17 total score, the Baseline HAMD-17 total score-by-visit interaction, and study. An unstructured covariance

matrix will be used and the Kenward-Roger approximation will be used to adjust the denominator degrees of freedom.

The following is sample SAS code for the MMRM:

In the event that the model fails to converge using the unstructured covariance matrix, the following covariance structures will be modeled in the order given (i.e., from least parsimonious to most parsimonious): heterogeneous Toeplitz, heterogeneous compound symmetry, heterogeneous autoregressive(1), Toeplitz, compound symmetry, autoregressive(1), variance components. The first covariance structure that allows for convergence will be selected for the final model and the asymptotically consistent sandwich estimator for the variance-covariance matrix will be used.

Summary statistics for the HAMD-17 total score (observed and change from Baseline), LS means, the between-group difference in LS mean with the corresponding 95% confidence interval, p-value, and the effect size (Cohen's d) will be presented at each post-Baseline visit for the Full Analysis Set.

The treatment effect size (Cohen's d) is calculated using the following formula:

Effect Size = LS mean difference / SD

SD is the model-based estimate, i.e., the estimated standard deviation from the unstructured covariance matrix. The sign (+ or -) of the effect size will be chosen so that a positive value favors pimavanserin.

LS mean ± SE from the MMRM model over time for the change from Baseline values by treatment group will be displayed in line plots for the Full Analysis Set.

The analysis will also be performed for the subgroups of study (ACP-103-059 and ACP-103-054), sex (male and female), race (white and non-white), age (18-40 years and > 40 years), the severity of depression at Baseline (Baseline HAMD-17 total score < 24 and ≥ 24), and the background antidepressant type (SSRI [Citalopram, Escitalopram, Paroxetine, Fluoxetine, or Sertraline] and SNRI [Duloxetine, Venlafaxine, Desvenlafaxine, or Venlafaxine XR]).

13.1.2 Sensitivity Analyses

The following 4 sensitivity analyses of the HAMD-17 total score are planned.

Pattern-Mixture Models (PMM) Assuming Missing Not At Random (MNAR) - Placebo-based Pattern Imputation

The sensitivity analysis is implemented for the Full Analysis Set using multiple imputations. The underlying assumption is that subjects with missing data due to early withdrawal evolve in the same way as placebo subjects that remain in the study.

The following steps are involved and the imputed values will be constrained to be within the limits of 0 - 52:

1 Non-monotone (intermediate) missing data at Baseline, Week 1, Week 2, …, and Week 6 will be multiply imputed using the Markov chain Monte Carlo (MCMC) method and a random seed of 103059 to create 100 monotone datasets. The imputation model will include the effect for treatment group.

2 The monotone missing data will be imputed using a parametric, sequential linear regression method in which the missing data are imputed only on data from the placebo arm. A single imputation will be performed sequentially at each visit for each of the 100 imputed datasets using the random seed of 103059. The predictors and their order in the PROC MI VAR statement for each visit are summarized in Table 2.

Table 2 Imputation Predictors – Placebo-based Pattern Imputation

Visit (Week)	Predictors
1	Study and Baseline HAMD-17 total score
2	Study, Baseline, and Week 1 HAMD-17 total scores
3	Study, Baseline, Week 1, and Week 2 HAMD-17 total scores
4	Study, Baseline, Week 1, Week 2, and Week 3 HAMD-17 total scores
5	Study, Baseline, Week 1, Week 2, …, and Week 4 HAMD-17 total scores

3 The change from Baseline to Week 5 values will then be calculated and analyzed for each of the 100 fully imputed datasets using the analysis of co-variance (ANCOVA) model. The ANCOVA model will have treatment group and study as factors and Baseline HAMD-17 total score as a covariate.

4 The results will be summarized using the SAS MIANALYZE procedure to yield a combined estimate for treatment effect with its associated 95% confidence interval and p-value.

PMM Assuming MNAR - Pattern Imputation with Delta Adjustment

The sensitivity analysis is implemented for the Full Analysis Set using multiple imputations. The underlying assumption is that subjects with missing data due to early withdrawal from the pimavanserin treatment would have, on average, their unobserved/imputed HAMD-17 total score worse by a delta amount (+1, +2, and +3 to be evaluated). Subjects with missing data due to early withdrawal from the placebo treatment will have their imputed values following the projection as if they continue in the study.

The following steps are involved and the imputed values will be constrained to be within the limits of 0 - 52:

1 Non-monotone (intermediate) missing data at Baseline, Week 1, Week 2, …, and Week 6 will be multiply imputed using the Markov chain Monte Carlo (MCMC) method and a random seed of 103059 to create 100 monotone datasets. The imputation model will include the effect for treatment group.

2 The monotone missing data will be imputed using a parametric, sequential linear regression method in which the missing data are imputed on data from their own treatment groups. A single imputation will be performed sequentially at each visit for each of the 100 imputed datasets using the random seed of 103059. The predictors and their order in the PROC MI VAR statement for each visit are summarized in Table 3. Penalties will then be applied to subjects on the pimavanserin treatment with missing data by shifting their multiply-imputed HAMD-17 total score at Week 5 by +1, +2, and +3 (with the limit of no more than 52).

Table 3 Imputation Predictors – Pattern Imputation with Delta Adjustment

Visit (Week)	Predictors
1	Study, Treatment, Baseline HAMD-17 total score
2	Study, Treatment, Baseline, and Week 1 HAMD-17 total scores
3	Study, Treatment, Baseline, Week 1, and Week 2 HAMD-17 total scores
4	Study, Treatment, Baseline, Week 1, Week 2, and Week 3 HAMD-17 total scores
5	Study, Treatment, Baseline, Week 1, Week 2, …, and Week 4 HAMD-17 total scores

3 The change from Baseline to Week 5 values will then be calculated and analyzed for each of the 100 fully imputed datasets using the ANCOVA model. The ANCOVA model will have treatment group and study as factors and Baseline HAMD-17 total score as a covariate.

4 The results will be summarized using the SAS MIANALYZE procedure to yield a combined estimate for treatment effect with its associated 95% CI and p-value.

COVID-19 Related Sensitivity Analysis 1

The sensitivity analysis will be performed for the Full Analysis Set. The underlying assumption is that the subjects with HAMD-17 completed at Week 5 clinic visit are different from those with HAMD-17 completed remotely due to COVID-19 with respect to the measurement of HAMD-17 total score at Week 5.

An indicator variable of remote will be created. The values of the indicator variable are:

"Yes" = subjects with Week 5 HAMD-17 completed remotely due to COVID-19 or
"No " = subjects with HAMD-17 completed at Week 5 clinic visit or
"N/A" = subjects with missing HAMD-17 total score at Week 5

The following MMRM that is similar to that in the primary analysis will be performed: The response variable of the MMRM is the change from Baseline in the HAMD-17 total score and the effects include treatment group, visit, treatment-by-visit interaction, Baseline HAMD-17 total score, the Baseline HAMD-17 total score-by-visit interaction, study, and remote. An unstructured covariance matrix will be used and the Kenward-Roger approximation will be used to adjust the denominator degrees of freedom.

COVID-19 Related Sensitivity Analysis 2

The sensitivity analysis will be performed using the data cut on March 20, 2020 when the official notification was sent to sites to stop screening/randomization in the studies due to COVID-19. The underlying assumption is that HAMD-17 completed after the date of notification may be drastically different such that the measure of HAMD-17 total score no longer measures the treatment effect correctly as on or before the date of notification.

The MMRM exactly the same as that in the primary analysis will be performed.

A supportive analysis similar to the primary analysis will be performed for the Per-protocol Analysis Set.

13.2 Secondary Efficacy Analyses

Analyses for the secondary efficacy endpoints in this section will be performed for the Full Analysis Set using the observed values. Missing data will not be imputed.

<u>SDS Mean Score, CGI-S Score, CSFQ-14 Total Score, KSS Score, HAMD Anxiety/Somatization Factor Score, and BIS-11 Score</u>

The change from Baseline will be analyzed using the MMRM method similar to the primary analysis for the primary efficacy endpoint. The MMRM model will include effects for treatment group, visit, the treatment-by-visit interaction, the corresponding Baseline (mean/total) score, the Baseline (mean/total) score-by-visit interaction, and study.

Summary statistics (observed and change from Baseline), LS means, the between-treatment difference in LS mean with the corresponding 95% confidence interval, p-value, and the effect size (Cohen's d) will be presented at each post-Baseline visit.

For the SDS mean score, the analysis will also be performed for the subgroups of the background antidepressant type (SSRI and SNRI).

For the CSFQ-14 total score, the analysis will also be performed for the subgroups of the background antidepressant type (SSRI and SNRI), sex (male and female), the background antidepressant type by sex (SSRI and male, SSRI and female, SNRI and male, and SNRI and female), and the sexual dysfunction status at Baseline (dysfunction and not dysfunction; Refer to Section 14.6 for the definition of sexual dysfunction).

CGI-I Score

The CGI-I score will be analyzed in a similar way as the primary analysis of the primary efficacy endpoint. The differences are (1) the response is CGI-I score (as opposed to the change from Baseline) and (2) the Baseline CGI-S score will be used as the covariate in the MMRM model.

Treatment Response Rate

Treatment response is defined as a reduction from Baseline in the HAMD-17 total score of 50% or more. Missing HAMD-17 total scores will be considered as non-responses and the subjects with missing HAMD-17 total scores will be included in the denominator of the response rate. The response rate will be compared between the pimavanserin and placebo groups using the Cochran-Mantel-Haenszel (CMH) test stratified by study. The treatment difference in the response rate and the corresponding Newcombe 95% confidence interval will be presented.

Treatment Remission Rate

Treatment remission is defined as a HAMD-17 total score $\leq$7. The analysis method for treatment remission rate is the same as that for the treatment response rate but with remission replacing response.

Change from Baseline to Week 1 in the HAMD-17 total score

The treatment comparison with respect to the mean change from Baseline to Week 1 in the HAMD-17 total score will be performed using the estimates from the primary analysis MMRM model for the primary efficacy endpoint described in Section 13.1.1.

13.3 Exploratory Efficacy Analyses

Exploratory analyses will be performed using the methods discussed in Sections 13.1 and 13.2 for the following:

- Change from Baseline to Weeks other than Weeks 1 and 5 in HAMD-17 total score
- Change from Baseline to Weeks other than Week 5 in SDS mean score, CGI-S score, CSFQ-14 score, KSS score, HAMD Anxiety/Somatization factor score, and BIS-11 score
- CGI-I score at Weeks other than Week 5
- Treatment responder rate and remission rate at Weeks other than Week 5

14. SAFETY ANALYSES

All safety analyses will be performed using the actual treatment for the Safety Analysis Set.

14.1 Adverse Events

Adverse events will be coded using MedDRA dictionary, Version 19.0 or newer.

An AE (classified by preferred term) will be considered a treatment-emergent AE (TEAE) if started after first study dose administration and no later than last study dose date + 30. AEs reported on Day 1 based on Baseline (pre-dose) findings (e.g., clinically significantly abnormal vital signs, laboratory test results, or electrocardiogram parameters) will not be considered as TEAEs.

The number and percentage of subjects reporting TEAEs in each treatment group will be tabulated by system organ class (SOC) and preferred term; by SOC, preferred term, and maximum severity; and by SOC, preferred term, and relationship to study drug. If more than 1 AE occurs with the same preferred term for the same subject, then the subject will be counted only once for that preferred term using the most severe and most related occurrence for the summarization by severity and by relationship to study drug.

The relationship of selected AEs to COVID-19 will be assessed as detailed in the GSD. TEAEs tabulated by SOC and preferred term will be presented with and without COVID-19 related TEAEs. In addition, COVID-19 related TEAEs will also be tabulated separately.

TEAEs will also be tabulated by treatment group without displaying the SOC terms; this table will be sorted in descending order of frequency of preferred term within the pimavanserin group.

The incidence of most frequently reported (preferred terms reported by ≥ 5% of subjects in any treatment group) TEAEs, treatment emergent SAEs, and TEAEs leading to discontinuation of study drug will be summarized by SOC, preferred term, and treatment

group. The tables will be sorted alphabetically by SOC and then by descending frequency within each SOC. In addition, the incidence of fatal treatment-emergent AEs (i.e., events that cause death) will be summarized separately by preferred term and treatment group. These summary tables except for the most frequently reported TEAEs table will also be presented for COVID-19 related events.

An AE listing by subject will display all events, including those which occur during screening, and will include the verbatim term in addition to the MedDRA SOC and preferred term. This listing will also include all relevant eCRF data associated with the event: date of onset, date resolved, date of last dose, severity, frequency, outcome, relationship to study drug, and action taken with study drug. Separate listings will be presented for subjects with treatment-emergent SAEs, subjects with TEAEs leading to discontinuation, subject who died (if any), and subjects with all COVID-19 related events.

14.2 Clinical Laboratory Variables

Due to COVID-19 disruptions it is possible that some test results may be collected from a local laboratory. Local laboratory results will not be included in any data analysis. Local laboratory results will, however, be included in data listings.

Clinical laboratory assessments are performed at Screening Visit 1, Baseline (Week 0), and Week 6 (EOT/ET).

- Clinical chemistry serum tests include the following:
    - Sodium (Na), potassium (K), chloride (Cl), phosphorus (P), calcium (Ca), carbon dioxide (CO_2), blood urea nitrogen (BUN), creatinine (CR), uric acid
    - Alanine aminotransferase (ALT/SGPT), aspartate aminotransferase (AST/SGOT), gamma-glutamyl transpeptidase (GGT), alkaline phosphatase (ALP), total bilirubin (TBIL), lactate dehydrogenase (LDH)
    - HbA1c (Screening only)
    - Glucose
    - Albumin (ALB), total protein
    - Prolactin
    - Creatine kinase (CK)/creatine phosphokinase (CPK)
    - Lipid panel
        - Total cholesterol, HDL-cholesterol, triglycerides, LDL-cholesterol, Cholesterol/HDL ratio, Non-HDL cholesterol

- Hematology tests include the following:
    - Complete blood count (CBC) including:
        - White blood cell (WBC) count
        - Complete differential (relative and absolute)
        - Hematocrit (Hct), hemoglobin, red blood cells (RBC), platelets
        - Reticulocyte count
- Urinalysis tests include the following:
    - Blood, RBCs, WBCs, protein, glucose, ketones, specific gravity, pH

Clinical laboratory values (in Système International [SI] units) and the change from Baseline values will be summarized by treatment group at Week 6 (EOT/ET) using descriptive statistics. The overall minimum and maximum post-Baseline observed and change from Baseline values will also be summarized. For hemoglobin, hematocrit and uric acid, the above summaries will be presented for each gender as well as for both genders combined. For urinalysis with categorical results, the number and percentage of subjects will be tabulated by category at Baseline and Week 6 (EOT/ET), and the denominator is the number of subjects with non-missing values for the given parameter, visit and treatment group. For glucose and lipid panel, the summary table will also be presented by fasting status (fasting vs. non-fasting).

Laboratory values will also be summarized in shift tables by treatment group, to determine the number and percentage of subjects with values classified as below, within, and above normal ranges at Week 6 (EOT/ET) visit relative to the same classification at the Baseline visit. For the by-visit shit summary, the denominator is the number of subjects with non-missing values at Baseline and the given visit for the given parameter and treatment group. For the shift to the overall post-Baseline minimum or maximum, all post-Baseline values will be considered, including unscheduled and out of window values and the denominator is the number of subjects with non-missing Baseline value and at least 1 post-Baseline value for the given parameter and treatment group. For hemoglobin, hematocrit and uric acid, the shift summaries will be presented for each gender as well as for both genders combined. For glucose and lipid panel, the shift table will be presented by fasting status (fasting vs. non-fasting).

Clinical laboratory values are potentially clinically important (PCI) if they meet either the low or high PCI criteria listed in Tables 4 and 5. The number and percentage of subjects with post-Baseline PCI values for each of the categories in Table 5 and 6 will be summarized by treatment group for selected parameters. For the overall post-Baseline summaries of PCI values, all post-Baseline values will be considered, including unscheduled and out of window values. Subjects with multiple PCI values for a given parameter will be counted only once for that parameter. For the overall post-Baseline summary, the numerator of the

percentage is the number of subjects with at least 1 post-Baseline PCI laboratory value for the given parameter and treatment group, and the denominator is the number of subjects with at least 1 post-Baseline laboratory value for the given parameter and treatment group. For hemoglobin, hematocrit and uric acid, the count and percentage of subjects with PCI values will be presented for each gender as well as for both genders combined. Subjects with PCI values will be presented in an additional listing.

Table 4 Criteria for Potentially Clinically Important Laboratory Values – Hematology and Chemistry

Analyte	Conventional Unit	Low PCI Criteria	High PCI Criteria	SI Unit	Low PCI Criteria	High PCI Criteria
Hematology (whole blood)						
Hemoglobin (male)	g/dL	<11	>18	g/L	<110	>180
Hemoglobin (female)	g/dL	<10	>17	g/L	<100	>170
Hematocrit (male)	%	<30	>55	L/L	<0.3	>0.55
Hematocrit (female)	%	<30	>50	L/L	<0.3	>0.5
Leukocyte (White Blood Cell Count)	x 10^3/uL	≤2.8	≥15	x 10^9/L	≤2.8	≥15
Neutrophils	x 10^3/uL	≤1.5	No upper limit	x 10^9/L	≤1.5	No upper limit
Platelet Count	x 10^3/uL	≤75	≥700	10^9/L	≤75	≥700
Chemistry (serum or plasma)						
ALT (SGPT)	U/L	No lower limit	≥3 X ULN	U/L	No lower limit	≥3 X ULN
AST (SGOT)	U/L	No lower limit	≥3 X ULN	U/L	No lower limit	≥3 X ULN
Total Bilirubin	mg/dL	No lower limit	≥1.5 ULN	umol/L	No lower limit	≥1.5 ULN
BUN	mg/dL	No lower limit	≥30.0	mmol/L	No lower limit	≥10.71
Creatine Kinase (CK)	U/L	No lower limit	≥3 ULN	U/L	No lower limit	≥3 ULN
Sodium	mEq/L	≤125	≥155	mmol/L	≤125	≥155
Potassium	mEq/L	≤3.0	≥5.5	mmol/L	≤3.0	≥5.5
Calcium, total	mg/dL	<8.0	>11.0	mmol/L	<2.0	>2.75
Lactate Dehydrogenase (LDH)	U/L	No lower limit	≥3 X ULN	U/L	No lower limit	≥3 X ULN
Alkaline Phosphatase	U/L	No lower limit	≥3 X ULN	U/L	No lower limit	≥3 X ULN
Uric acid (male)	mg/dL	No lower limit	≥10.5	umol/L	No lower limit	≥624.75
Uric acid (female)	mg/dL	No lower limit	≥8.5	umol/L	No lower limit	≥505.75
Albumin	g/dL	≤2.6	≥6.0	g/L	≤26	≥60
Total Protein	g/dL	≤5.0	≥10.0	g/L	≤50	≥100

Table 4 Criteria for Potentially Clinically Important Laboratory Values – Hematology and Chemistry (Continued)

Analyte	Conventional Unit	Low PCI Criteria	High PCI Criteria	SI Unit	Low PCI Criteria	High PCI Criteria
Chloride	mEq/L	≤85	≥120	mmol/L	≤85	≥120
Glucose (random)	mg/dL	≤45.1	≥200.0	mmol/L	≤2.48	≥11
Serum Creatinine	mg/dL	Not Applicable	>1.5 ULN	umol/L	Not Applicable	>1.5 ULN
Triglycerides	mg/dL	Not Applicable	>300	mmol/L	Not Applicable	>3.39
Gamma-Glutamyl Transferase (GGT)	U/L	Not Applicable	≥3 ULN	U/L	Not Applicable	≥3 ULN

Table 5 Criteria for Potentially Clinically Important Laboratory Values – Urinalysis

Urinalysis (qualitative dipstick)	Low PCI Criteria	High PCI Criteria
Blood (occult blood)	Not Applicable	≥ Moderate
Protein	Not Applicable	≥ 100 mg/dL
Glucose	Not Applicable	≥ 500 mg/dL

Clinical laboratory data will be displayed in data listings with date and study day of collection. All units will be displayed according to SI conventions for units. Out of range values will be flagged in the data listings (i.e., 'L' or 'H'). A separate listing will be provided for a subset of the chemistry, hematology, and urinalysis analytes with values classified as PCI.

The number and percentage of subjects with the change from Baseline to Week 6 meeting the following criteria in the blood lipid parameters and the glucose will be summarized by treatment group and fasting status (fasting vs. non-fasting):

- Total Cholesterol – (<200 mg/dL to ≥240 mg/dL)
- Triglycerides – (<150 mg/dL to ≥200 mg/dL)
- LDL Cholesterol – (<100 mg/dL to ≥160 mg/dL)
- HDL Cholesterol – (≥40 mg/dL to <40 mg/dL)
- Glucose fasting – (<100 mg/dL to ≥126 mg/dL)
- Glucose non-fasting – (<140 mg/dL to ≥140 mg/dL)
- Glucose fasting – (≥100 mg/dL and <126 mg/dL to ≥126 mg/dL)
- Glucose non-fasting - (≥140mg/dL and <200 mg/dL to ≥200 mg/dL)

The pregnancy results (positive or negative) for female subjects will be presented in a listing.

14.3 Vital Signs

Vital signs including weight and the derived BMI will be collected throughout the study at every visit. Observed vital signs and the changes from Baseline at each post-Baseline visit will be summarized by treatment group using descriptive statistics.

Vital sign values will be considered PCI if they meet the criteria listed in Table 6. The number and percentage of subjects with post-Baseline vital signs that are PCI will be summarized by treatment group at each post-Baseline visit and for overall post-Baseline. For the overall post-Baseline summaries, all post-Baseline values will be considered, including unscheduled and out of window values. Subjects with multiple PCI values for a given parameter will be counted only once for that parameter. For the by-visit summary, the numerator for the percentage is the number of subjects with a post-Baseline PCI vital sign for the given parameter, visit and treatment group, and the denominator is the number of subjects with non-missing values for the given parameter, visit and treatment group. For the overall post-Baseline summary, the numerator for the percentage is the number of subjects with at least 1 post-Baseline PCI vital sign for the given parameter and treatment group, and the denominator is the number of subjects with at least 1 post-Baseline vital sign for the given parameter and treatment group. A listing of subjects with any PCI value will be provided.

Table 6 Criteria for Potentially Clinically Important (PCI) Vital Signs

Vital Sign Parameter	Unit	Criteria		
		Observed Value	And/Or	Change Relative to Baseline
Systolic blood pressure (supine or sitting)	mmHg	≥180	And	Increase of ≥20
		≤90	And	Decrease of ≥20
Diastolic blood pressure (supine or sitting)	mmHg	≥105	And	Increase of ≥15
		≤50	And	Decrease of ≥15
Pulse (supine or sitting)	bpm	≥120	And	Increase of ≥15
		≤50	And	Decrease of ≥15
Weight	kg	Not Applicable		Increase of ≥7%
				Decrease of ≥7%

14.4 Electrocardiogram (ECG)

12-lead ECGs are collected at Screening Visit 1, Baseline (Week 0), Week 1, and Week 6 (EOT/ET). Observed values of ECG variables (e.g., heart rate, PR interval, QRS interval, QT interval, and QTc interval) and the changes from Baseline at each assessment time point will be summarized by treatment group.

QTcF will also be categorized into the following categories (msec) and the number and percentage of subjects in each category will be summarized by treatment group at each visit and for the overall post-Baseline maximum:

- Observed: ≤450, 451 - ≤480, 481 - ≤500, and >500; >450; >480.
- Change from Baseline: ≤10, 11 – 30, 31 – 60, and >60; >30.

The above summaries will also be presented for the following subgroups: subjects with background antidepressant of Citalopram, subjects with background antidepressant of Escitalopram, and subjects with background antidepressant of Venlafaxine.

Electrocardiogram variable values will be considered PCI if they meet the criteria listed in Table 7. The number and percentage of subjects with post-baseline PCI values will be summarized by treatment group at each post-Baseline visit and for overall post-Baseline. For the by-visit summary, the numerator for the percentage is the number of subjects with a post-Baseline PCI ECG for the given parameter, visit and treatment group, and the denominator is the number of subjects with non-missing values for the given parameter, visit and treatment group. For the overall post-Baseline summary, the numerator for the percentage is the number of subjects with at least 1 post-Baseline PCI ECG for the given parameter and treatment group, and the denominator is the number of subjects with at least 1 post-Baseline ECG value for the given parameter and treatment group. A listing of all subjects with any PCI value will be provided.

Table 7 Criteria for Potentially Clinically Important ECG Values

ECG Parameter	Unit	High PCI Criteria
QRS Interval	msec	≥120
PR Interval	msec	≥220
QTcB or QTcF	msec	>500
QTcB or QTcF: change from baseline	>60 msec	

14.5 Physical Examination

Physical examinations are performed at Screening Visit 1, Baseline (Week 0), and Week 6 (EOT/ET). Physical examination results (normal, abnormal, and not done) will be summarized in a frequency table by treatment group, body system and visit.

14.6 Other Safety Endpoints

Unless otherwise specified, scores derived based on sub-scores for the safety endpoints in this section will be considered missing if any corresponding sub-scores are missing. No imputations will be performed.

Columbia-Suicide Severity Rating Scale (C-SSRS)

The C-SSRS Baseline/Screening version will be completed at the Screening visit and the version assessing information since the last visit will be completed at all following visits (including the Baseline visit).

There are 5 questions about suicidal ideation, representing 5 types of suicidal ideation: wish to be dead; non-specific active suicidal thoughts; active suicidal ideation with any methods (not plan) without intent to act; active suicidal ideation with some intent to act, without specific plan; active suicidal ideation with specific plan and intent. If a subject answers "yes" to any of these 5 questions at any post-Baseline visit including unscheduled and out of window visits, this subject will be counted as having suicidal ideation in this study.

There are 5 questions about suicidal behavior, representing 5 types of suicidal behavior: actual attempt; interrupted attempt; aborted attempt; preparatory acts or behavior; suicide. If a subject answers "yes" to any of these 5 questions at any post-Baseline visit including unscheduled and out of window visits, this subject will be counted as having suicidal behavior in this study.

Suicidality is defined as a subject who reported at least 1 occurrence of suicidal ideation or at least 1 occurrence of suicidal behavior at any post-Baseline visit including unscheduled and out of window visits.

All data will be listed. The event counts and the number and percentage of subjects reporting any post-Baseline suicidal ideation, suicidal behavior, or suicidality will be summarized by treatment group.

Change from Baseline to Week 5 in Extrapyramidal Symptom Rating Scale-Abbreviated (ESRS-A) score

The ESRS-A will be completed at each visit throughout the study except for the Screening visit. Observed values and change from Baseline in the ESRS-A CGI-S score for each movement subtype (i.e., parkinsonism, dystonia, dyskinesia, and akathisia) will be summarized by treatment group and visit. The change from Baseline in the ESRS-A CGI-S score for each movement subtype will be analyzed using the MMRM method. The response variable of the MMRM is the change from Baseline in the ESRS-A CGI-S score and the effects include treatment group, visit, treatment-by-visit interaction, Baseline ESRS-A CGI-S score, the Baseline ESRS-A CGI-S score-by-visit interaction, and study. An unstructured covariance matrix will be used and the Kenward-Roger approximation will be used to adjust the denominator degrees of freedom.

Sexual Dysfunction Based on the CSFQ-14 Total Score and Summary for Subgroups

Sexual dysfunction is defined as a CSFQ-14 total score of $\leq$47 for men and $\leq$41 for women. The following analyses will be performed overall for the Safety Analysis set and for the subgroups of the background antidepressant type (SSRI and SNRI), sex (male and female),

and the background antidepressant type by sex (SSRI and male, SSRI and female, SNRI and male, and SNRI and female):

1. The number and percentage of subjects with sexual dysfunction will be summarized by treatment group at Baseline and each post-Baseline visit. The numerator for the percentage is the number of subjects with sexual dysfunction for the visit and treatment group, and the denominator is the number of subjects with non-missing values for the visit and treatment group.

2. The sexual dysfunction classification will also be summarized in shift tables by treatment group, to determine the number and percentage of subjects with and without sexual dysfunction at Week 5 visit and Week 6 (EOT/ET) visit relative to the same classification at the Baseline visit. The denominator is the number of subjects with non-missing values at Baseline and the given visit for the given treatment group.

15. CLINICAL PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES

Blood concentration for the concomitant SSRIs/SNRIs collected at the Screening visit will be listed.

16. INTERIM ANALYSIS

No formal interim analysis is planned.

17. DATA MONITORING/REVIEW COMMITTEE

There is no data monitoring committee in the studies.

18. COMPUTER METHODS

Statistical analyses will be performed using Version 9.4 (or newer) of SAS® (SAS® Institute, Inc., Cary, North Carolina) on a suitably qualified and validated environment.

Validation and quality control of the tables, listings and figures containing the results of the statistical analyses will follow appropriate standard operating procedures.

19. CHANGES TO ANALYSES SPECIFIED IN PROTOCOL

No changes are made to the analyses specified in the protocols.

20. REFERENCES

EMA (2009). *Guideline on Missing Data in Confirmatory Clinical Trials, European Medicines Agency*, London, UK.

FDA Guidance on Conduct of Clinical Trials of Medical Products during COVID-19 Pandemic | Guidance for Industry, Investigators, and Institutional Review Boards (March 2020).

21. APPENDICES

Appendix A Schedule of Assessments

Period	Screening	Baseline	Double-blind Treatment Period							Safety Follow-up[a]
Visit Number	1	2	3	4	5	6	7	(EOT/ET) 8		9
Visit Day/Week	Day -28 to -3	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Unscheduled Visit[j]	Week 10
Type of Visit	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Phone call
Visit window (# days)	+7		±3	±3	±3	±3	±3	±3		+7
Informed consent and if applicable, privacy forms	X									
Inclusion/exclusion criteria	X	X								
Medical history, medication history, and demographics	X									
Psychiatric history	X									
MGH ATRQ[b]	X									
SAFER remote interview[c]	X									
MINI	X									
Physical examination	X	X						X		
Vital signs and weight	X	X	X	X	X	X	X	X	X	
Height	X									
12-lead ECG[d]	X	X	X					X		
Clinical laboratory tests[e]	X	X						X		
Pregnancy test[f]	X	X						X		
Background antidepressant blood level[g]	X									
Urine toxicity (drug) screen	X	X						X		
MADRS	X	X								

Appendix A Schedule of Assessments (Continued)

Period	Screening	Baseline	Double-blind Treatment Period							Safety Follow-up[a]
Visit Number	1	2	3	4	5	6	7	(EOT/ET) 8	Unscheduled Visit[j]	9
Visit Day/Week	Day -28 to -3	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6		Week 10
Type of Visit	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Phone call
Visit window (# days)	+7		±3	±3	±3	±3	±3	±3		+7
HAMD-17[h]		X	X	X	X	X	X	X		
SDS		X	X	X	X	X	X	X		
CGI-S	X	X	X	X	X	X	X	X		
CGI-I			X	X	X	X	X	X		
CSFQ-14		X	X	X	X	X	X	X		
KSS		X	X	X	X	X	X	X		
BIS-11		X	X		X		X	X		
ESRS-A		X	X	X	X	X	X	X	X	
C-SSRS[i]	X	X	X	X	X	X	X	X	X	
Concomitant medications	X	X	X	X	X	X	X	X	X	X
Assessment of adverse events	X	X	X	X	X	X	X	X	X	X
Dispense study drug		X	X	X	X	X	X		X[k]	
Study drug accountability			X	X	X	X	X	X	X	
Review of background antidepressant adherence	X	X	X	X	X	X	X	X	X	X

Abbreviations: BIS-11=Barratt Impulsiveness Scale; CGI-I=Clinical Global Impression–Improvement; CGI-S=Clinical Global Impression–Severity; CSFQ-14=Changes in Sexual Functioning Questionnaire Short Form; C-SSRS=Columbia–Suicide Severity Rating Scale; ECG=electrocardiogram; eCRF=electronic case report form; EOT=end of treatment; ESRS-A=Extrapyramidal Symptom Rating Scale–Abbreviated; ET=early termination; HAMD-17=Hamilton Depression Scale (17 items); KSS=Karolinska Sleepiness Scale; MADRS=Montgomery-Asberg Depression Rating Scale; MGH ATRQ=Massachusetts General Hospital Antidepressant Treatment Response Questionnaire; MINI=Mini-International Neuropsychiatric Interview; SAFER=State versus trait, Assessability, Face validity, Ecological validity, and Rule of three Ps (pervasive, persistent, and pathological); SDS=Sheehan Disability Scale; SNRI=serotonin-norepinephrine reuptake inhibitor; SSRI=selective serotonin reuptake inhibitor

Table footnotes on next page.

[a] This visit is a safety follow-up telephone call visit for subjects who discontinue prematurely from the study or who do not participate in the open-label extension study. This visit will occur at least 30 days after the last dose of study drug.

[b] Inadequate response is defined as a response of <50% to a course of treatment of at least 8 weeks at the minimum effective dose (as listed on the ATRQ), which has been stable for the 4 weeks prior to the SAFER interview.

[c] The SAFER remote interview may be conducted via telephone off-site or on-site. If conducted on-site, the site staff should not be present during the interview.

[d] At Visit 1, ECG should be completed in triplicate and collected within a 3-minute period. At all other visits, a single 12-lead ECG should be completed. ECGs should be completed with the subject in a supine position after 5 minutes of rest. The ECG should occur before blood sampling or at least 30 minutes after blood sampling.

[e] To include hematology, serum chemistry, prolactin levels, and urinalysis.

[f] A pregnancy test is only required for women of childbearing potential. Serum pregnancy should only be performed at Visit 1; a urine pregnancy test should be performed at Baseline and Week 6 (EOT). If urine cannot be obtained in women of childbearing potential, a serum pregnancy test should be done in its place.

[g] At Screening, blood samples will be collected for the analysis of concomitant SSRI/SNRI concentrations. The presence of SSRI or SNRI must be confirmed in order to qualify the subject for further consideration in the study.

[h] The HAMD-17 is to be the first scale completed at each visit.

[i] Suicidal assessment is required. The Baseline/Screening version of the C-SSRS will be administered at Screening. The Since Last Visit version of the C-SSRS will be administered at all subsequent visits.

[j] At a minimum the safety assessments indicated should be completed at unscheduled visits. Other assessments may be completed at unscheduled visits at the discretion of the Investigator.

[k] Study drug may be dispensed to the subject at unscheduled visits if needed.

Appendix B Summary of Version Changes

Version No:	Document History Description of Update	Author(s)	Version Date
1.0		[redacted]	22 June 2020