Official Title: An Open-label, 8-Week Study of Safety and Efficacy of Pimavanserin Treatment in Adults with Parkinson's Disease and Depression

NCT Number: NCT03482882

STATISTICAL ANALYSIS PLAN

Protocol No.:	ACP-103-048
Protocol Title:	An Open-label, 8-Week Study of Safety and Efficacy of Pimavanserin Treatment in Adults with Parkinson's Disease and Depression
Drug:	Pimavanserin
Sponsor:	ACADIA Pharmaceuticals Inc. [redacted]
Version No. and Date	Version 2.0, 31 January 2019

SIGNATURE/APPROVAL PAGE

AUTHOR

[redacted]

01 MAR 2019
Date

Associate Director, Biostatistics
ACADIA Pharmaceuticals Inc.

APPROVERS

[redacted]

01 MAR 2019
Date

Vice President, Biometrics
ACADIA Pharmaceuticals Inc.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ADS	analysis dataset specification
AE	adverse event
ANCOVA	analysis of covariance
ATC	Anatomical Therapeutic Chemical
BMI	body mass index
CGI-I	Clinical Global Impression – Improvement scale
CGI-S	Clinical Global Impression – Severity scale
C-SSRS	Columbia-Suicide Severity Rating Scale
DPD	depression in Parkinson's disease
DS	daytime sleepiness
ECG	Electrocardiogram
eCRF	electronic case report form
EOS	end-of-study
ET	early termination
HAMD-17	Hamilton Depression Rating Scale-17 items
LOCF	last observation carried forward
MGH ATRQ	Massachusetts General Hospital Antidepressant Treatment Questionnaire
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	mixed model for repeated measures
MSEC	Milliseconds
NS	nighttime Sleep
OC	observed cases
PCI	potentially clinically important
PT	preferred term
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
SAE	serious adverse event
SAP	statistical analysis plan
SCOPA	Scale for Outcomes in Parkinson's Disease
SD	standard deviation
SE	standard error
SI	Système International
SOC	system organ class
SNRI	selective norepinephrine inhibitor
SSRI	serotonin reuptake inhibitor
TEAE	treatment-emergent adverse event
VAS	visual analog scale

1 INTRODUCTION

This statistical analysis plan (SAP) provides a technical and detailed elaboration of the statistical analyses of efficacy and safety data as described in Amendment 2 of the final study protocol dated 08 January 2019. Specifications for tables, figures, and listings are contained in separate documents.

2 OBJECTIVES

2.1 Primary Objective

The primary objective of this study is to assess the efficacy of pimavanserin for the treatment of depression in adults with Parkinson's disease (PD).

2.2 Secondary Objectives

The secondary objectives of this study are to explore the efficacy of pimavanserin in the following domains:

- Clinician's global assessment of treatment benefits
- Nighttime sleep and daytime sleepiness
- Quality of life (QoL)

2.3 Safety Objectives

The safety objectives of the study are to assess the safety of pimavanserin for the treatment of depression in adults with PD.

3 STUDY DESIGN

3.1 General Study Design

This study will be conducted as a Phase 2, 8-week, open-label, single arm, multi-center, outpatient study in subjects with PD and depression. Subjects can receive treatment as either a monotherapy (pimavanserin 34 mg alone), or adjunctively (pimavanserin 34 mg plus an antidepressant) if depression is inadequately controlled with current therapy.

Approximately 20 sites in the U.S will participate in this study. Subjects will participate in the study for up to 13 weeks, consisting of a screening period of up to 3 weeks, followed by a treatment period of approximately 8 weeks, and a 2-week safety follow-up (telephone call).

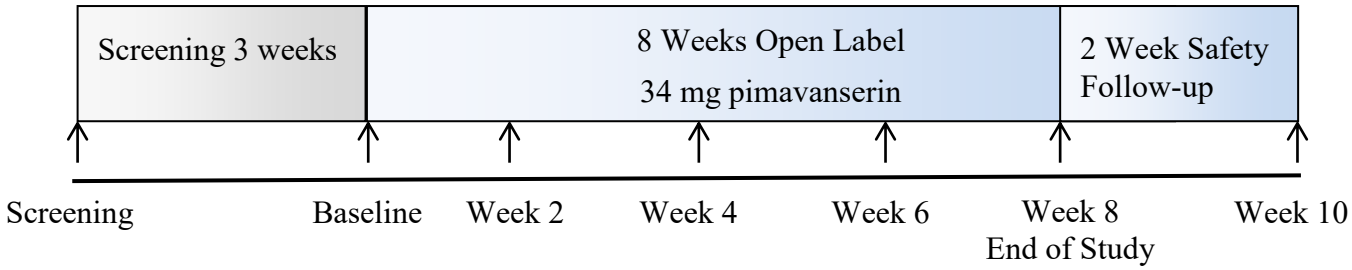
On the first day of the treatment phase (Baseline), eligible subjects will be enrolled and receive the first dose of adjunctive pimavanserin 34 mg per day. The remaining study

medication will be provided to the subjects to take home with instruction to take the medication at approximately the same time each day.

Clinic visits occurring after Baseline will be conducted at Weeks 2, 4, 6, and 8 (End-of-Study [EOS]/Early Termination [ET] visit).

Figure 1 illustrates the study design.

Figure 1 Schematic of Study Design



3.2 Schedule of Assessments

The schedule of events and assessments for the study is presented in Table 1.

Table 1 Schedule of Events and Assessments

	Screening	Baseline	Week 2	Week 4	Week 6	EOT Week 8[g]	Safety Follow-up Week 10
	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	Visit 7
Study Day(s)	-21 to -1	1	12 to 18	26 to 32	40 to 46	54 to 60	68 to 74
Informed consent	X						
Inclusion/exclusion assessment	X	X					
MGH ATRQ[a]	X						
Medical history and demographics	X						
Weight, height, BMI	X	X[f]				X[f]	
Physical and neurological examination	X					X	
12-lead ECG[b]	X	X	X			X	
Vital signs[c]	X	X	X	X	X	X	
Clinical laboratory tests[d]	X	X				X	
Screen for drugs of abuse	X	X					
Pregnancy test[e]	X	X				X	
HAMD-17	X	X	X	X	X	X	
C-SSRS	X	X	X	X	X	X	
MMSE	X	X		X		X	
CGI-I			X	X	X	X	
CGI-S		X	X	X	X	X	
SCOPA		X		X		X	
EQ-5D-5L		X		X		X	
UPDRS Part III		X		X		X	
Dispense study drug		X		X			

Study drug accountability			X	X	X	X	
Phone Follow-Up							X
Concomitant medication	X	X	X	X	X	X	X
AE assessment	X	X	X	X	X	X	X

Abbreviations: AE=adverse event; BMI=body mass index; C-SSRS=Columbia Suicide Severity Rating Scale; CGI-I=Clinical Global Impression–Improvement; CGI-S=Clinical Global Impression–Severity; ECG=electrocardiogram; EOT-end of treatment; EQ-5D-5L-5L=EuroQol 5 dimensions-5 levels; HAMD-17=Hamilton Depression Scale–17 Item; MGH ATRQ=Massachusetts General Hospital Antidepressant Treatment Questionnaire; MMSE=Mini-Mental State Exam; SCOPA=Sleep Scales for Outcomes in Parkinson's Disease-Sleep ; UPDRS Part III=Unified Parkinson's Disease Rating Scale Part III.

Notes:

[a] Applicable only to subjects who are currently taking an antidepressant.

[b] 12-lead ECG at Screening is to be completed in triplicate within a 3 minute period. The ECG may be repeated once at Screening in consultation with the Medical Monitor. Single ECG recordings are to be collected at subsequent visits. ECGs can be performed any time before blood sampling or at least 30 minutes after blood sampling during clinic visits

[c] Vital signs (sitting or supine [at least 3 minutes] blood pressure, pulse rate, oral temperature, and respiratory rate) will be performed at Screening and each study visit.

[d] Thyroid-stimulating hormone (TSH) will only be assessed at Screening. TSH will not be analyzed as part of the chemistry serum testing at other study visits.

[e] Applicable only to women of childbearing potential. A serum pregnancy test is performed at Screening and a urine pregnancy test at Baseline and Week 8.

[f] Only weight will be measured after Screening.

[g] Subject terminating early should at a minimum complete assessments of safety (vital signs, ECG, AEs, clinical laboratories, C-SSRS) and, if possible, complete all Week 8 assessments.

3.3 Randomization

Not applicable.

3.4 Blinding

This is an open-label study. Subjects, site staff, and the sponsor will be aware that the subject is being treated with a 34 mg dose of pimavanserin.

3.5 Determination of Sample Size

The initial sample size calculation was based on a standard deviation of 8.0 and a dropout rate of 10%. An interim review of the statistical assumptions was conducted after 9 subjects completed the Week 8 visit, and the sample size was recalculated. The standard deviation in HAMD-17 total score change from Baseline to Week 8 was observed to be lower, and the dropout rate higher, than the initial assumptions. Therefore, the sample size has been recalculated as follows.

Assuming the standard deviation for the change in HAMD-17 total score from Baseline to Week 8 is 6.0 points, 34 evaluable subjects will provide 80% power to detect a minimum mean reduction of 3 points from Baseline to Week 8 at a significance level of 0.05 using a 2-sided paired t-test.

Adjusting for a potential non-evaluable rate of up to 15%, approximately 40 subjects will be enrolled.

4 ANALYSIS SETS

4.1 Safety Analysis Set

The Safety Analysis Set will consist of a subset of enrolled subjects who received at least one dose of study drug (pimavanserin).

4.2 Full Analysis Set

The Full Analysis Set will consist of a subset of subjects in the Safety Analysis Set who have both a Baseline value and at least one post-Baseline value for the HAMD-17 total score.

5 DATA HANDLING CONVENTIONS

All data collected in the study will be listed.

5.1 General Data Reporting Conventions

5.1.1 Continuous Variables

For continuous variables, the following summary statistics will be provided: number of subjects, mean, standard error of the mean (SE), standard deviation (SD), minimum, maximum, and median. Unless specified otherwise, means, medians, and confidence intervals (CIs) will be presented to one more decimal place than the raw data, and the standard deviations and standard errors will be presented to 2 more decimal places than the raw data.

5.1.2 Categorical Variables

For categorical variables, summaries will include the number and percentage of subjects in each category, using the number of subjects with non-missing values as the denominator for the percentages (unless otherwise specified). Categories with zero counts will not have zero percentages displayed. Percentages will be presented to 1 decimal place.

5.1.3 Statistical Tests

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 (CIs) will be 2-sided 95% CIs. P-values will generally be presented to 4 decimal places; p-values less than 0.0001 will be presented as <0.0001.

5.1.4 Conversion of Days to Months

When converting number of days to months, it will be calculated as the number of days divided by 365.25 and then multiplied by 12. When converting number of days to years, it will be calculated as the number of days divided by 365.25.

5.2 Derived Variables

In general, the total scores and subscores for assessment scales will be derived within the analysis datasets. In the event that total scores and/or subscores are also collected on the electronic case report form (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 Depression Rating Scale 17-Items (HAMD-17)

The HAMD-17 is assessed at Screening, Baseline, Weeks 2, 4, 6, and Week 8/ET visits.

The HAMD-17 is a 17-item clinician-administered questionnaire used to provide an indication of depression, and as a guide to evaluate recovery (Hamilton, 1960). The

questionnaire assesses symptoms over the past week (7 days). The questionnaire is designed for adults to rate the severity of, and change in, depressive symptoms by probing mood, feelings of guilt, suicide ideation, insomnia, agitation or retardation, anxiety, weight loss, and somatic symptoms. Each item on the questionnaire is scored on a 3- or 5-point scale, depending on the item, and the total score is compared to the corresponding descriptor. The HAMD-17 total score can range from a minimum of 0 to a maximum of 52. A score of 0-7 is considered normal. Scores of 14 or higher indicate moderate, severe, or very severe depression.

If more than 3 items are missing, the total score will be set to missing. If 3 or fewer items are missing, then missing items will be imputed as follows:

- Items using a 3-point scale (4, 5, 6, 12, 13, 14, 16, and 17): missing values will be imputed using the arithmetic mean of the non-missing items that are also scored on a 3-point scale, then rounded to the nearest integer
- Items using a 5-point scale (1, 2, 3, 7, 8, 9, 10, 11, and 15): missing values will be imputed using the arithmetic mean of the non-missing items that are also scored on a 5-point scale, then rounded to the nearest integer

5.2.2 Clinical Global Impression – Severity (CGI-S) Scale

The CGI-S is assessed at Screening, Baseline, Weeks 2, 4, 6, and 8/ET visits.

The CGI-S is a clinician-rated, 7-point scale that is designed to rate the severity of the subject's depression at the time of assessment using Investigator’s judgment and past experience with subjects who have the same disorder (i.e., depression in Parkinson’s disease). The 7-point scores are: 1=normal, not ill; 2=minimally ill; 3=mildly ill; 4=moderately ill; 5=markedly ill; 6=severely ill; 7=among the most severely ill.

Missing CGI-S scores will not be imputed.

5.2.3 Clinical Global Impression– Improvement (CGI-I) Scale

The CGI-I is assessed at Weeks 2, 4, 6, and 8/ET visits.

The CGI-I is a clinician-rated, 7-point scale that is designed to rate the improvement in the subject’s depression at the time of assessment, relative to the symptoms at Baseline. The 7-point scores are: 1=very much improved; 2=much improved; 3=minimally improved; 4=no change; 5=minimally worse; 6=much worse; 7=very much worse.

Missing CGI-I scores will not be imputed.

5.2.4 Scale for Outcomes in Parkinson's Disease (SCOPA) – Sleep Scale

The SCOPA-Sleep scale is assessed at Baseline, Week 4, and Week 8/ET visits.

The SCOPA-Sleep was developed for research in Parkinson's disease to evaluate both nighttime sleep (NS) and daytime sleepiness (DS). The SCOPA-Sleep scale consists of 3 subscales, SCOPA-NS, SCOPA-DS, and one additional question that evaluates overall sleep quality. For each subscale, the time frame is the preceding month.

The SCOPA-NS subscale addresses problems in nighttime sleep and consists of 5 items: sleep initiation, sleep fragmentation, sleep efficiency, sleep duration, and early wakening. Each item has 4 response options (ranging from 0=not at all, to 3=a lot) such that the minimum score is 0 and the maximum score is 15, with higher scores indicating more severe NS problems.

The SCOPA-DS subscale addresses problems in daytime sleepiness and consists of 6 items: how often the subject 1) fell asleep unexpectedly, 2) fell asleep peacefully, 3) fell asleep watching TV/reading, 4) fell asleep while talking to someone, 5) had difficulty staying awake, and 6) whether falling asleep in the daytime was considered a problem. Each item has 4 response options (ranging from 0=never, to 3=often). Therefore, possible SCOPA-DS subscale scores can range from 0 to 18, with higher scores indicating more severe DS problems.

If only one SCOPA-NS item is missing, then the subscale score will be imputed by replacing the missing item with the mean (rounded to the nearest integer) of the non-missing items for a given subject and timepoint. If more than 1 item is missing, then the subscale total score will remain missing. The same imputation algorithm will be applied to the SCOPA-DS subscale.

Global sleep quality is assessed using a single-item 7-point scale (1=slept very well to 7=slept very badly). This item will not be included in the SCOPA-DS or SCOPA-NS total scores. Missing values for the global sleep quality items will not be imputed.

5.2.5 EQ-5D-5L Proxy Version 1

The EQ-5D-5L questionnaire is assessed at Baseline, Week 4, and Week 8/ET.

The EQ-5D-5L is a standardized measure of health status that provides a simple, generic measure of health for clinical and economic appraisal. In this study, Proxy Version 1 will be administered, in which the caregiver (the proxy) is asked to rate the subject's health-related quality of life in his/her (the proxy's) opinion. This questionnaire consists of 2 components – the EQ-5D-5L descriptive system and the EQ-5D-5L Visual Analogue scale (EQ-5D-5L VAS).

The descriptive system consists of 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has 5 levels (responses): no problem (1), slight problems (2), moderate problems (3), severe problems (4), and extreme problems (5). The digits for the 5 dimensions are combined into a 5-digit code that describes the subject's health state.

The EQ-5D-5L VAS records the subject's health on a vertical visual analogue scale, where the upper endpoint is labeled "The best health you can imagine" and is numbered 100, while the lower endpoint is labeled "The worst health you can imagine" and is numbered 0. The EQ-5D-5L VAS will be treated as a continuous endpoint.

Missing values will not be imputed for either the descriptive system or EQ-5D-5L VAS.

5.2.6 Columbia-Suicide Severity Rating Scale (C-SSRS)

The C-SSRS is assessed at Screening, Baseline, Weeks 2, 4, 6, and 8/ET visits.

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 Baseline, Week 2, Week 4, Week 6, and Week 8/ET.

The C-SSRS monitors changes in suicidal thinking and behavior over time in order to determine risk. Four constructs are measured: the severity of ideation, the intensity of ideation, behavior, and lethality.

There are 5 questions about 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, this subject will be counted as having suicidal ideation.

There are 5 questions about suicidal behavior: suicide, actual attempt; interrupted attempt; aborted attempt; preparatory acts or behavior. If a subject answers "yes" to any of these 4 questions, this subject will be counted as having suicidal behavior.

Missing values will not be imputed.

5.2.7 Movement Disorders Society-Unified Parkinson's Disease Rating Scale Part III

The Movement Disorders Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) Part III (motor examination) will be assessed at Baseline, Week 4, and Week 8/ET.

The MDS-UPDRS Part III is used to assess motor signs of PD and consists of 33 individual items based on 18 domains, due to left, right, and other body distributions. Each item has

5 possible responses (0, 1, 2, 3, and 4) with 0 being normal and 4 severe. The total score will be derived as the sum of the 33 item scores; thus, the total score has a potential range from 0 to 132, with higher scores indicating more severe extrapyramidal symptoms.

If 6 or fewer items are missing, then the total score for a given subject and timepoint will be imputed as the mean of the non-missing items multiplied by 33 and rounded to the nearest integer. If more than 6 items are missing, the total score will not be derived.

In addition, the MDS-UPDRS Part III assesses motor signs with the Hoehn and Yahr Stage on a 6-point scale, where asymptomatic is scored 0 and wheelchair bound or bedridden is scored 5. Missing values for Hoehn and Yahr Stage will not be imputed.

5.2.8 Mini-Mental State Examination (MMSE)

The MMSE is assessed at Screening, Baseline, Week 4, and Week 8/ET.

The MMSE is a 30-item questionnaire that includes simple questions and problems in the following areas: time and place of testing, repeating lists of words, arithmetic, language use and comprehension, and copying or drawing.

Each of the 30 items has 2 possible values, 0 (incorrect) or 1 (correct). The MMSE total score will be derived as the sum of the 30 item scores, thus the total score has a potential range of 0 to 30. Lower scores indicate more severe cognitive impairment.

If 6 or fewer items are missing, then the total score for a given subject and timepoint will be imputed as the mean of the non-missing items multiplied by 30 and rounded to the nearest integer. If more than 6 items are missing, the total score will not be derived.

5.3 Analysis Visit Windows

Baseline will be defined as the last non-missing result, including results from repeated and unscheduled measurements, before dosing.

Efficacy and safety assessments will be summarized by analysis visit as shown in Table 2 below.

Table 2 Analysis Visit Windows

Analysis Visit Name	Study Visit	Target Study Day[1]	Study Day Interval
Baseline	Visit 2 (Baseline)	1	≤1
Week 2	Visit 3 (Week 2)	15	2 to 21
Week 4	Visit 4 (Week 4)	29	22 to 35
Week 6	Visit 5 (Week 6)	43	36 to 49
Week 8	Visit 6 (Week 8/ET)	57	50 to 63
Follow-up	Visit 7 (Week 10/Safety Follow-up)	71	≥64

[1] 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).

5.3.1 Unscheduled Assessments

Both scheduled and unscheduled assessments, including the assessments at ET visits, will be considered for planned timepoint analyses. All assessments will be presented in data listings.

5.3.2 Multiple Measurements within Visit Windows

If more than one assessment falls within a given window, then the assessment closest to the target study day will be selected for the by-visit summaries. If two assessments are equidistant from the target 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 analysis where the most 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 will be considered, regardless of whether the value is selected for the by-visit summaries. All results will be presented in data listings.

When the 12-lead electrocardiogram (ECG) is collected in triplicate, the average of the triplicate parameter values will be considered as one assessment for the summaries.

5.4 Data Handling Conventions

If the last dose date of study drug is missing for a subject who completed or discontinued from the study, then the last study drug return date will be used in the calculation of treatment duration. If the study drug return dates are all missing (e.g., the subject never returned any

study drug), then the ET/EOS date (excluding follow-up visit) will be used in the calculation of treatment duration. For the incomplete last dose date of the study drug, the imputation algorithms will be detailed in the analysis dataset specification (ADS) document. The missing or incomplete dates will be displayed in the data listings as reported on the eCRF rather than the imputed dates.

For any data summarization that occurs before final database lock, if a subject is still ongoing, then that subject's last dose date will be imputed using the database extract date.

5.4.1 Missing or Incomplete Dates for Prior or Concomitant Medications

Missing or incomplete medication start or end dates will be imputed for the purpose of determining whether the medications are taken concomitantly (see Section 10 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 ADS document. The missing or incomplete dates will be displayed in the data listings as reported on the eCRF rather than the imputed dates.

5.4.2 Missing or Incomplete Dates 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 (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 ADS document. The missing or incomplete dates will be displayed in the data listings as reported on the eCRF rather than the imputed dates.

5.4.3 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, and the actual values will be used in data listings.

5.4.4 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, and the actual values will be presented in data listings.

5.4.5 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 ADS 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 enrolled in the study (i.e., 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 reason), 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 that screened at least 1 subject, number of sites that enrolled at least 1 subject, number of subjects screened, and number of unique subjects screened will be summarized by site and overall. In addition, the number of subjects enrolled at each site will also be tabulated by analysis set.

For enrolled subjects, the number and percentage of subjects in Safety Analysis Set and Full Analysis Set will be summarized. A listing will be provided displaying all subjects excluded from the Safety and Full Analysis Sets, and will include reason(s) for exclusion.

Within each analysis set, the number and percentage of subjects who completed the study or discontinued (all discontinued and by discontinuation reason) will also be summarized.

7 PROTOCOL DEVIATIONS

Protocol deviations will be reviewed periodically over the course of the study. 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.

A summary of the number and percentage of subjects with major protocol deviations for each deviation category will be presented for the Safety Analysis Set. A listing of protocol deviations by site and subject will also be provided.

8 DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS

8.1 Demographics

Demographics will be summarized for the Safety Analysis Set and the Full Analysis Set using descriptive statistics. Variables include age, sex, primary race (subjects of multi-racial background can only identify/select one primary race on eCRF, or choose "other" and specify), ethnicity, height (cm), weight (kg), body mass index (BMI), Baseline living situation category, and Baseline caregiver category.

The reported age will be calculated relative to the subject's age at the informed consent date. Age and BMI will be presented as both continuous and categorical variables. Age categories will be presented as <70 and ≥70 years old. BMI categories will be presented as <25, ≥25 to <30, and ≥30.

A listing of subject living situation and caregiver information will be provided.

8.2 Disease Characteristics at Baseline

Disease characteristics at Baseline will be summarized for the Safety Analysis Set and the Full Analysis Set using descriptive statistics. Variables include Baseline HAMD-17 total score, Baseline CGI-S, Baseline MMSE total score, Baseline SCOPA-NS score, Baseline SCOPA-DS score, Baseline SCOPA global sleep quality, and Baseline MDS-UPDRS Part III total score.

Baseline HAMD-17 total score, Baseline CGI-S, Baseline SCOPA global sleep quality, and Baseline MMSE total score will be presented as continuous and categorical variables. The categories will be displayed as follows:

- Baseline HAMD-17 total score: <24 and ≥24
- Baseline CGI-S: normal (1) to among the most extremely ill (7)
- Baseline SCOPA global sleep quality: very well (1) to very badly (7)
- Baseline MMSE total score: <25 and ≥25

PD and DPD disease history will include the following variables:

- Age (years) at onset of PD
- Duration (years) of PD
- Age (years) at onset of DPD
- Duration (years) of DPD

- Time (years) since first antidepressant treatment
- Current background antidepressant medication
- Duration (months) of current background antidepressant medication
- Ever had suicidal ideation or behavior (yes or no)
- Had suicidal ideation or behavior within past 6 months (yes or no)

Informed consent date will be used as the reference date for calculating the durations listed above.

The Massachusetts General Hospital (MGH) Antidepressant Treatment Response Questionnaire (ATRQ) is assessed at the Screening visit for subjects who are taking an SSRI/SNRI (i.e. adjunctive treatment). It is a clinician-assisted questionnaire used to determine response to antidepressant therapy. Drug type, generic name, drug taken during this current episode of depression (Yes/No), dose, and took at least this dose for at least 4 weeks (Yes/No), and the percentage of improvement in depression will be listed.

In addition, for all subjects lifetime antidepressant use and response will be collected, including start and stop dates, dose, units, and response (≤25%, 26% to <50%, 50% to <75%, and 75% to 100%). These data will only be listed.

9 MEDICAL HISTORY

Medical and surgical history data will be coded using 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 (PT) for the Safety Analysis Set and Full Analysis Set. A subject will be counted only once per SOC or per PT for the summary.

A listing of the SOC, PT, body system, verbatim term 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 PRIOR, CONCOMITANT AND POST-TREATMENT MEDICATION

Summaries of prior, concomitant, and post-treatment medications will be provided for the Safety Analysis Set.

For a subject, a prior medication is defined as any medication with the start and stop dates prior to the date of the first dose of study drug. A Concomitant medication is defined as any medication with a start date prior to the date of the first dose of study drug and continuing after 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. Prior, concomitant, and post-treatment medications will be summarized separately. Medications will be coded using the World Health Organization Drug Dictionary, version March 2017 or newer. The number and percentage of subjects taking each drug class (Anatomical Therapeutic Chemical Level 3) and medication PT will be tabulated. Multiple medication usage by a subject in the same category will be counted only once.

Background antidepressant therapies will be tabulated by medication PT and dose level.

11 EXTENT OF EXPOSURE AND TREATMENT COMPLIANCE

Summaries of exposure and compliance to study drug will be provided for the Safety Analysis Set and Full Analysis Set.

11.1 Exposure to Study drug

For each subject, the duration of exposure to study drug is calculated as follows:

$$Duration\ of\ exposure = last\ dose\ date - first\ dose\ date + 1$$

Duration of exposure will be summarized descriptively as both continuous and categorical variables. For the categorical presentation, the number and percentage of subjects in each of the following categories will be displayed: <1week (1 to 6 days), 1 to <2 weeks (7 to 13 days), 2 to <4 weeks (14 to 27 days), 4 to <6 weeks (28 to 41 days), and ≥6 weeks (42 days or longer). A Kaplan-Meier curve of duration of exposure will also be presented.

The pimavanserin dose levels are expressed as free base.

11.2 Measurement of Study Drug Compliance

Study drug compliance (in percentage) for a given subject is defined as follows:

$$Compliance = \left[\frac{tablets\ dispensed - tablets\ returned}{2 \times duration\ of\ exposure}\right] \times 100\%$$

Compliance will be summarized as both continuous and categorical variables. For categorical presentation, the number and percentage of subjects in each of the following categories will be presented: <80%, ≥80 to <120%, and ≥120%.

12 EFFICACY ANALYSES

All efficacy analyses will be performed on the Full Analysis Set.

12.1 Efficacy Variables

12.1.1 Primary Efficacy Endpoint

The primary efficacy endpoint is the change from Baseline to Week 8 in the HAMD-17 total score.

12.1.2 Secondary Efficacy Endpoints

Secondary efficacy endpoints are the following:

- Change from Baseline in HAMD-17 total score at Weeks 2, 4, and 6
- Proportion of responders (defined as ≥50% reduction from Baseline in HAMD-17 total score)
- CGI-I score
- Change from Baseline in CGI-S
- Change from Baseline in SCOPA-NS
- Change from Baseline in SCOPA-DS
- Change from Baseline in EQ-5D-5L

12.2 Adjustment for Covariates

For continuous variables (except CGI-I) analyzed using the mixed model for repeated measures (MMRM) , the Baseline value of the endpoint being analyzed will be included as a covariate as described in Section 13. For CGI-I, the Baseline CGI-S score will be included as a covariate in the MMRM analysis.

12.3 Handling of Missing Data

Total scores that are missing, after any imputation of individual missing items as described in Section 5.2.1, will not be imputed. Analysis of covariance (ANCOVA) and EQ-5D categorical analyses will be performed using the last observation carried forward (LOCF) method of imputation for missing data. The LOCF method will carry forward the Baseline value if necessary.

If the last dose date of study drug is missing for a subject who completed or early terminated from the study, then the date of the end-of-study/early termination visit will be used in the calculation of treatment duration. For the incomplete last dose date of the study drug, the imputation algorithms will be detailed in the ADS document. The missing or incomplete

dates will be displayed in the data listings as reported on the eCRF rather than the imputed dates.

For any data summarization that occurs before final database lock, if a subject is still ongoing, then that subject's last dose date will be imputed using the database extract date.

12.4 Multiple Comparisons / Multiplicity

No adjustments of p-values for multiple comparisons will be performed.

12.5 Examination of Subgroups

The MMRM analysis described in Section 13.1.1 will be performed separately for the primary efficacy variable (HAMD-17 total score), and for CGI-S and CGI-I as described in Section 13.1.6 and Section 13.1.7, respectively, for each of the following subgroups:

- treatment regimen (pimavanserin monotherapy or pimavanserin adjunctive therapy)
- age group (<70 or ≥70 years old)
- sex (male or female)
- primary race (white or non-white)
- Baseline symptom severity measured by HAMD-17 total score (<24 or ≥24)

The LS mean changes with corresponding 95% CIs from the subgroups will also be graphically presented in forest plots for the above efficacy variables.

13 METHODS OF EFFICACY ANALYSES

13.1 Primary Efficacy Analysis

The primary endpoint is the change from Baseline to Week 8 in the HAMD-17 total score. The primary analysis will be based on the Full Analysis Set.

The null and alternative hypotheses for the primary endpoint are as follows:

Let Δ be the mean change from Baseline to Week 8 in the HAMD-17 total score:

The null hypothesis for the primary efficacy endpoint is: $\Delta = 0$

The alternative hypothesis for the primary efficacy endpoint is: $\Delta \neq 0$

13.1.1 Primary Analysis using MMRM in the Full Analysis Set

The HAMD-17 total score will be analyzed using mixed model repeated measures (MMRM) in the Full Analysis Set. The dependent variable will be the change from Baseline in the

HAMD-17 total score. The independent variables in the model will include the following: visit (Week 2, Week 4, Week 6, and Week 8), Baseline-by-visit interaction, and the Baseline HAMD-17 total score as a continuous covariate. An unstructured covariance matrix will be used and the Kenward-Roger approximation will be used to adjust the denominator degrees of freedom. Least squares (LS) adjusted means will be estimated using observed cases (OC).

[redacted]

In the event that the model fails to converge using the unstructured covariance matrix, then the following covariance structures will be modeled in the order given: 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.

Summary statistics for the HAMD-17 total score (observed and change from Baseline) will be presented for all visits from Baseline through Week 8. For change from Baseline values at each post-Baseline visit, LS means and standard errors (SE), the corresponding 95% confidence interval and p-value will also be presented. In addition, LS mean ± SE over time for the change from Baseline values will also be presented in line plots.

The hypothesis testing will be performed based on the LS mean change at Week 8 and will be tested at an alpha level of 0.05 (2-sided).

13.1.2 LOCF Paired T-test for HAMD-17

A paired t-test using LOCF will be performed on the HAMD-17 total score change from Baseline at each post-baseline visit using the Full Analysis Set. Summary statistics for the HAMD-17 total score (observed and change from Baseline) will be presented for all visits from Baseline through Week 8. For change from Baseline values at each post-Baseline visit, 95% confidence intervals and p-values will be presented.

13.1.3 Secondary Efficacy Analyses

The Full Analysis Set will be used for all secondary efficacy analyses.

13.1.4 Change from Baseline to Weeks 2, 4, and 6 in HAMD-17 Total Score

See Section 13.1.1 for details of this analysis.

13.1.5 HAMD-17 Responder Analysis

For the HAMD-17 total score, a treatment response is defined as a ≥50% reduction from Baseline.

At each timepoint the proportion of responders will be descriptively summarized including 95% confidence intervals. Both OC, in which subjects with missing values at a given visit are excluded, as well as missing values imputed as non-responders will be presented.

13.1.6 CGI-S

The CGI-S score change from Baseline at each post-Baseline timepoint will be analyzed using an MMRM model similar to that used for the primary efficacy endpoint. The dependent variable will be the CGI-S score change from Baseline, and the independent variables in the model will include the following: visit (Weeks 2, 4, 6, and 8), Baseline CGI-S score (continuous covariate), and Baseline CGI-S score-by-visit interaction.

13.1.7 CGI-I

The CGI-I score at each post-Baseline timepoint will be analyzed using an MMRM model similar to that used for the primary efficacy endpoint. The dependent variable will be the CGI-I score, and the independent variables in the model will include the following: visit (Weeks 2, 4, 6, and 8), Baseline CGI-S score (continuous covariate), and Baseline CGI-S score-by-visit interaction.

The CGI-I scores will also be dichotomized by combining scores of 1 or 2 (very much improved or much improved) into an “Improved” category, and all remaining scores into a “Not Improved” category. At each timepoint, the proportion of responders (very much improved or much improved) will be descriptively summarized including 95% confidence intervals. Both OC, in which subjects with missing values at a given visit are excluded, as well as missing values imputed as not improved, will be presented.

The observed (un-dichotomized) scores, as well as dichotomized scores, will be descriptively summarized by visit.

13.1.8 SCOPA

13.1.8.1 SCOPA-NS

The SCOPA-NS score at each post-Baseline timepoint will be analyzed using an MMRM model similar to that used for the primary efficacy endpoint. The dependent variable will be

the SCOPA-NS score change from baseline, and the independent variables in the model will include the following: visit (Weeks 4 and 8), Baseline SCOPA-NS score (continuous covariate), and Baseline SCOPA-NS score-by-visit interaction.

13.1.8.2 SCOPA-DS

The SCOPA-DS score at each post-Baseline timepoint will be analyzed using an MMRM model similar to that used for the primary efficacy endpoint. The dependent variable will be the SCOPA-DS score change from Baseline, and the independent variables in the model will include the following: visit (Weeks 4 and 8), Baseline SCOPA-DS score (continuous covariate), and Baseline SCOPA-DS score-by-visit interaction.

13.1.8.3 SCOPA Global Sleep Quality

The SCOPA global sleep quality score at each post-Baseline timepoint will be analyzed using an MMRM model similar to that used for the primary efficacy endpoint. The dependent variable will be the global sleep quality score change from Baseline, and the independent variables in the model will include the following: visit (Weeks 4 and 8), Baseline global sleep quality score (continuous covariate), and Baseline global sleep quality score-by-visit interaction.

13.1.9 EQ-5D-5L

13.1.9.1 EQ-5D-5L Descriptive System

For each EQ-5D-5L dimension the proportion of subjects reporting no, slight, moderate, severe, and extreme/unable to perform activity will be summarized descriptively at each timepoint (Baseline, Week 4, and Week 8), using both OC and LOCF analysis methods.

Change from Baseline in EQ-5D-5L health state will be assessed with the Pareto classification of health change (PCHC) method. By using this methodology, at each post-Baseline visit (Week 4 and Week 8), each EQ-5D-5L health state will be classified into one of four categories, relative to the Baseline health state:

- Improved = improved on at least one dimension and not worsened on any other dimension
- Mixed = improved on at least one dimension and worsened on at least one other dimension
- No change = no changes in any dimension
- Worsened = deterioration on at least one dimension and no improvement on any other dimension

The proportion of subjects in each PCHC category will be summarized descriptively at Week 4 and Week 8, using both OC and LOCF analysis methods.

13.1.9.2 EQ-5D-5L VAS

EQ-5D-5L VAS score change from Baseline at each post-Baseline timepoint will be analyzed by using an MMRM model similar to that used for the primary efficacy endpoint. The dependent variable will be the EQ-5D-5L VAS change from Baseline, and the independent variables in the model will include the following: visit (Weeks 4 and 8), Baseline EQ-5D-5L VAS score (continuous covariate), and Baseline EQ-5D-5L VAS score-by-visit interaction.

13.1.9.3 EQ-5D-5L Index Scores

Since a United States value set to map EQ-5D-5L health states to preference weights (i.e., index values) is not yet available, the EQ-5D-5Lindex scores will be derived from EQ-5D-3L values. This procedure entails using a crosswalk dataset to map health states from the 5L system to the 3L system, and then applying EQ-5D-5L-3L weights from the US general population to obtain the index scores. The SAS software program to assign each health state to its corresponding index score was obtained from the EuroQol Research Foundation.

The EQ-5D-5L index score change from Baseline at each post-Baseline timepoint will be analyzed using an MMRM model similar to that used for the EQ-5D-5L VAS.

13.1.10 LOCF ANCOVA for Continuous Efficacy Variables

For each of the continuous secondary efficacy endpoints (HAMD-17, CGI-S, CGI-I, SCOPA-NS, SCOPA-DS, SCOPA global sleep quality, EQ-5D-5L VAS, EQ-5D-5L index scores), ANCOVA using LOCF will be performed at each post-baseline visit, similar to that described in Section 13.1.2.

To assess whether there is a linear correlation between the HAMD-17 total score change from Baseline and CGI-I rating, the HAMD-17 total score change from Baseline will also be summarized by CGI-I rating for all post-baseline visits. The mean $\pm$ SE HAMD-17 total score change from Baseline by CGI-I rating will also be plotted for Week 8OC.

14 SAFETY ANALYSES

The safety analysis will be performed based on the Safety Analysis Set using actual treatment. Safety variables include adverse events (AEs), clinical laboratory variables, vital signs, body weight, BMI, physical examinations, electrocardiogram (ECG), C-SSRS, MMSE, and MDS-UPDRS Part III variables.

14.1 Adverse Events

All Adverse events (AEs) will be coded using MedDRA version 19.0 or newer.

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

The event counts and the number and percentage of subjects reporting TEAEs will be tabulated by SOC and PT; and by SOC, PT, and maximum severity. If more than 1 AE occurs with the same PT for the same subject, the subject will be counted only once for that PT using the most severe occurrence for the summarization by severity to study drug. In addition, the event counts and the number and percentage of subjects with TEAEs classified by the Investigator as related to the study drug, with most frequently reported TEAEs (PTs reported by ≥ 5% of subjects), with treatment-emergent serious AEs (TESAEs), with fatal AEs (i.e., events that cause death), and with TEAEs leading to discontinuation of study drug will be summarized by SOC and PT. These tables will be sorted alphabetically by SOC and then by descending subject frequency for the PTs within each SOC.

The event counts and the number and percentage of subjects with any TEAEs will also be tabulated by PT without SOCs and will be sorted by descending subject frequency.

An AE listing by subject will display all events, including those which are not treatment-emergent, and will include the verbatim term in addition to the MedDRA SOC and PT. This listing will also include all relevant eCRF data associated with the event: e.g., date of onset, date resolved, date of first dose, date of last dose, study dose level at AE onset, severity, frequency, outcome, relationship to study drug, action taken with study drug, and required therapy. When a date is presented, the study day associated with the date will also be displayed. Separate listings will be presented for subjects with SAEs, subjects with AEs leading to discontinuation, and subjects who died (if any). In these listings, an indicator for treatment-emergent events will also be included.

14.2 Clinical Laboratory Variables

Laboratory evaluations will be completed according to the schedule presented in Table 3 below.

Table 3 Clinical Laboratory Assessments

Visit	Tests[a,b,c]
Screening	Clinical chemistry serum (including HbA1c, TSH, free T4 reflex), complete blood count, urinalysis, urine toxicity screen, and serum pregnancy test
Baseline	Clinical chemistry serum, complete blood count, urinalysis, urine toxicity screen, and urine pregnancy test
Week 8/ET	Clinical chemistry serum (including HbA1c), complete blood count, urinalysis, and urine pregnancy test

[a] Pregnancy tests are only required for women of child-bearing potential.

[b] An HbA1c test is only required at Screening and Week 8/ET.

[c] TSH and free T4 reflex are only required at Screening.

Hematology tests include the following:

- Complete blood count (CBC) including:
  - White blood cell (WBC) count
  - Complete differential (relative and absolute)
  - Hematocrit (Hct), hemoglobin (Hgb), red blood cells (RBC), platelets
  - Reticulocyte count

Serum chemistry 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)
- Glucose
- Albumin (ALB), total protein
- Prolactin
- HbA1c
- Lipid panel
  - Total cholesterol, high-density lipoprotein (HDL)-cholesterol, triglycerides, low-density lipoprotein(LDL)-cholesterol
- Thyroid stimulating hormone (TSH) and reflex free thyroxine (T4) (TSH and T4 are collected only at Screening)

Urinalysis tests include the following:

- Color, clarity, blood, RBCs, WBCs, protein, glucose, ketones, specific gravity, pH, leukocyte esterase, nitrite, microscopic analysis

Pregnancy tests include the following:

- A serum pregnancy test will be performed at Screening for women of child-bearing potential

- A urine pregnancy test will be performed at Baseline and Week 8/ET for women of child-bearing potential

All laboratory test results (including urine drug screen) are from a central laboratory and will be listed. The listings will include date and study day of collection. All units will be displayed in Système International [SI] units. Out-of-range values will be flagged in the data listings (e.g. 'L' or 'H').

Clinical laboratory values for hematology, chemistry and urinalysis (specific gravity and pH) will be summarized by descriptive statistics at Baseline and Week 8 visits. The change from Baseline values will also be summarized at the Week 8 visit. 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 (e.g., blood, protein, glucose, and ketones), the number and percentage of subjects will be tabulated by category at Baseline and Week 8, and the denominator is the number of subjects with non-missing values for the given parameter and visit.

Laboratory values (except for those that were only assessed at Screening or Baseline) will also be summarized in shift tables to determine the number and percentage of subjects with values classified as below (low), within (normal), and above (high) normal ranges at Week 8, overall post-Baseline minimum, and overall post-Baseline maximum. For the by-visit shift summary, the denominator is the total number of subjects with non-missing values at Baseline and the given visit for the given parameter. 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. The denominator is the number of subjects with non-missing Baseline and at least one post-Baseline value for the given parameter. For hemoglobin, hematocrit, and uric acid, the shift summaries will be presented for each gender and for both genders combined.

The number and percentage of subjects with potentially clinically important laboratory values (PCI) at Week 8 and overall post-Baseline will be summarized for selected parameters by Baseline status (All subjects and Within Normal Range). PCI criteria are listed in Table 4 and Table 5. 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 by-visit summary, the numerator for the percentage is the total number of subjects with a post-Baseline PCI value for the given parameter and visit, and the denominator is the total number of subjects with non-missing values for the given parameter and visit. For the overall post-Baseline summary, the numerator for the percentage is the total number of subjects with

at least 1 post-Baseline PCI laboratory value for the given parameter, and the denominator is the number of subjects with at least 1 post-Baseline laboratory value for the given parameter. Subjects with any 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	%	<30	>50 (F) >55 (M)	L/L	<0.3	>0.5 (F) >0.55 (M)
Leukocyte (White Blood Cell Count)	x $10^3/\mu L$	≤2.8	≥15	x $10^9/L$	≤2.8	≥15
Neutrophils	x $10^3/\mu L$	≤1.5	No upper limit	x $10^9/L$	≤1.5	No upper limit
Platelet Count	x $10^3/\mu L$	≤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
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
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	>300	mmol/L	Not	>3.39

		Applicable			Applicable	
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	≥+2
Protein	Not Applicable	≥+2
Glucose	Not Applicable	≥+2

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

14.3 Vital Signs

Vital signs are assessed at Screening, Baseline, and Weeks 2, 4, 6 and 8/ET visits.

Vital signs including weight, height (only at Screening), and derived BMI will be summarized at Baseline and all post-Baseline visits using descriptive statistics. The change from Baseline values will also be summarized at each post-Baseline visit.

Vital sign values will be considered PCI if they meet both the observed value criteria and the change from baseline criteria listed in Table 6. The number and percentage of subjects with post-Baseline vital signs that are PCI will be summarized 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 total number of subjects with a post-Baseline PCI vital sign for the given parameter and visit, and the denominator is the total number of subjects with non-missing values for the given parameter and visit. For the overall post-Baseline summary, the numerator for the percentage is the total number of subjects with at least 1 post-Baseline PCI vital sign for the given parameter, and the denominator is the number of subjects with at least 1 post-Baseline vital sign for the given parameter. A listing of all subjects with any PCI values will be provided.

Table 6 Criteria for Potentially Clinically Important Vital Signs

Vital Sign Parameter	Unit	Criteria[a]		
		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%

[a] A post-Baseline value is considered a PCI value if it meets both criteria for observed value and change from baseline.

14.4 Electrocardiogram

12-lead ECG assessments are performed at Screening, Baseline, Week 2 and Week 8/ET visits. ECGs will be completed in triplicate at Screening, and as a single tracing at all other timepoints. When ECG is collected in triplicate, the average of the triplicate will be considered as one assessment for the analyses. All tracings will be evaluated by a central reading laboratory. ECG data summaries will be performed using the centrally evaluated data, including the cardiologist's interpretation.

Observed values of ECG variables (heart rate, PR interval, RR interval, QRS interval, QT interval, and QTc intervals) and the changes from baseline at each assessment timepoint will be summarized by visit. QTc intervals include QTcB (Bazett's formula) and QTcF (Fridericia's formula).

QTcF (msec) observed and change from Baseline values will also be classified into the following categories, and the number and percentage of subjects in each category will be summarized at each visit and for the overall post-baseline maximum:

- Observed
    - ≤450, 451 to 480, 481 to 500, and >500
    - >450
    - >480
- Change from Baseline
    - ≤10, 11 to 30, 31 to 60, and >60
    - >30

Electrocardiogram variable values will be considered PCI if they meet or exceed the upper limit values listed in Table 7. The number and percentage of subjects with post-baseline PCI values will be summarized 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 and visit, and the denominator is the number of subjects with non-missing values for the given parameter and visit. 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 the denominator is the number of subjects with at least 1 post-Baseline ECG for the given parameter. A listing of all subjects with any PCI values 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	msec	>60

For cardiologist's interpretations, the number and percentage of subjects with ECG results that are determined as normal or abnormal will be summarized at scheduled visits. The overall post-baseline worst interpretation will also be summarized (i.e. if a subject has one or more ECG results that is/are considered as abnormal, this subject will be counted in the abnormal category). Cardiologist's interpretations will also be summarized in a shift table to determine the number and percentage of subjects with ECG results classified as normal or abnormal at scheduled post-Baseline visits relative to the same classification at the Baseline visit. The shifts from Baseline to overall post-Baseline worst interpretation will also be presented. For the by-visit shift summary, the denominator is the number of subjects with non-missing cardiologist's interpretation at Baseline and the given visit. For the summaries of shift from Baseline to the overall post-Baseline worst interpretation, the denominator is the number of subjects with non-missing Baseline and at least 1 post-Baseline cardiologist's interpretation.

14.5 Physical Examination and Neurological Examination

A general physical examination, in addition to a neurological examination (cranial nerves, motor, sensory, reflexes, gait, and coordination), are performed at Screening and Week 8/ET visits.

Physical and neurological examination results (normal, abnormal, and not done) at Screening and Week 8 will be summarized descriptively by body/neurological system and visit.

14.6 Other Safety Variables

14.6.1 MMSE

The MMSE total score will be summarized by timepoint (Baseline, Week 4, and Week 8) using descriptive statistics. The change from Baseline scores will also be summarized descriptively at Week 4 and Week 8. In addition, the MMSE total score change from Baseline be analyzed using an MMRM similar to the primary efficacy endpoint. The dependent variable will be the MMSE total score change from baseline, and the independent variables in the model will include the following: visit (Weeks 4 and 8), Baseline MMSE total score (continuous covariate), and Baseline MMSE total score-by-visit interaction.

The individual item scores will be listed but not summarized.

14.6.2 MDS-UPDRS Part III

The MDS-UPDRS Part III total score will be summarized by timepoint (Baseline, Week 4, and Week 8) using descriptive statistics. The MDS-UPDRS Part III total score change from Baseline will be summarized descriptively at Week 4 and Week 8, and will also be analyzed using an MMRM similar to the primary efficacy endpoint. The dependent variable will be the MDS-UPDRS Part III total score change from baseline, and the independent variables in the model will include the following: visit (Weeks 4 and 8), Baseline MDS-UPDRS Part III total score (continuous covariate), and Baseline MDS-UPDRS Part III total score-by-visit interaction.

The Hoehn and Yahr Stage will be summarized descriptively by visit as a categorical variable.

The individual item scores and all remaining questions will be listed but not summarized.

14.6.3 Suicidality

The C-SSRS is assessed at Screening, Baseline, Weeks 2, 4, 6, and Week 8/ET visits.

The event counts and the number and percentage of subjects in each suicide category at any post-Baseline timepoint (including unscheduled and out of window visits), will be summarized descriptively. Suicide categories will be defined as follows.

- Suicidality: Subjects who report at least one post-Baseline occurrence of suicidal behavior or suicidal ideation

- Suicidal ideation: Subjects who report at least one post-Baseline occurrence of the following
    - 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
- Suicidal behavior: Subjects who report at least one post-Baseline occurrence of the following
    - Preparatory acts or behavior
    - Aborted attempt
    - Interrupted attempt
    - Actual attempt
    - Suicide

15 CLINICAL PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES

Not applicable

16 DATA MONITORING/REVIEW COMMITTEE

Not applicable.

17 COMPUTER METHODS

Statistical analyses will be performed using Version 9.4 (or newer) of SAS software (SAS Institute Inc., Cary, NC, USA) 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.

18 CHANGES TO ANALYSES SPECIFIED IN PROTOCOL

Not applicable.

19 APPENDICES

19.1 Summary of Version Changes

Version No:	Document History Description of Update	Author(s)	Version Date
1.0	Original version	[redacted]	12 September 2018
2.0	Revised Section 3.5 due to sample size reduction per Protocol Amendment 2; and made update to Section 13.1.2 to use paired t-test (instead of ANCOVA) as per protocol. Updated primary objective to agree with Protocol Amendment 2 (adjunctive treatment no longer restricted to SSRI/SNRI). Updated Section 12.1.2 to agree with Protocol Amendment 2 (ANCOVA changed to t-test).	[redacted]	31 January 2019