Official Title: A 16-Week Open-Label Study of the Effects of Treatment With Pimavanserin on Activities of Daily Living in Subjects With Parkinson's Disease Psychosis

NCT Numbers: NCT04292223

Document Date: 28 Mar 2022

STATISTICAL ANALYSIS PLAN

Protocol No.:	ACP-103-063
Protocol Title:	A 16-Week Open-Label Study of the Effects of Treatment With Pimavanserin on Activities of Daily Living in Subjects With Parkinson's Disease Psychosis
Drug:	Pimavanserin
Sponsor:	Acadia Pharmaceuticals Inc. [redacted]
Version No. and Date	Version 1.0, 28 March 2022

SIGNATURE/APPROVAL PAGE

AUTHOR

See appended electronic signature page

[redacted] Date
Director, Biostatistics
Acadia Pharmaceuticals Inc.

APPROVERS

See appended electronic signature page

Date

[redacted]
Executive Director, Biostatistics
Acadia Pharmaceuticals Inc.

See appended electronic signature page

[redacted] Date
Vice President, Medical Affairs
Acadia Pharmaceuticals Inc.

Signature Page for ACP-103-063 Statistical Analysis Plan

Approve	[redacted]
Approve	[redacted]
Approve	[redacted]

Signature Page for ACP-103-063 Statistical Analysis Plan

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ADL	Activities of Daily Living
AE(s)	adverse event(s)
ATC	Anatomical Therapeutic Chemical
BMI	body mass index
CGI-I	Clinical Global Impression – Improvement
CGI-S	Clinical Global Impression – Severity
CI	confidence interval
COVID-19	Coronavirus disease 2019
CSI	Caregiver Strain Index
C-SSRS	Columbia-Suicide Severity Rating Scale
DMP	Data Management Plan
ECG	electrocardiogram
eCRF	electronic case report form
EOT	End-of-Treatment
ET	Early Termination
GDS	Geriatric Depression Scale
MDS-UPDRS	Movement Disorders Society-Unified Parkinson's Disease Rating Scale
MedDRA	Medical Dictionary for Regulatory Activities
mFSQ	modified Functional Status Questionnaire
MMRM	mixed-effect model repeated measures
MMSE	Mini-Mental State Examination
msec	milliseconds
NPI	Neuropsychiatric Inventory
NPI-Q	Neuropsychiatric Inventory-Questionnaire
PCI	potentially clinically important
PDP	Parkinson's disease psychosis
PGI-I	Patient Global Impression - Improvement
QTcB	QT interval corrected for heart rate using Bazett's formula
QTcF	QT interval corrected for heart rate using Fridericia's formula
SAE(s)	serious adverse event(s)
SAP	statistical analysis plan
SD	standard deviation
SE	standard error
SOC	system organ class

TEAE(s)	treatment-emergent adverse event(s)
VRFCAT	Virtual Reality Functional Capacity Assessment Tool

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 the protocol Amendment 1.0 dated 20 May 2021 and the electronic case report form (eCRF). This SAP will be approved prior to database lock. Specifications for tables, figures, and listings are contained in separate documents (ACP-103-063 Table Shells, ACP-103-063 Figure Shells, and ACP-103-063 Listing Shells, respectively).

2 OBJECTIVES AND ENDPOINTS

2.1 Primary Objective

The primary objective of this study is to assess the effect of pimavanserin on the activities of daily living (ADLs) in subjects with Parkinson's disease psychosis (PDP) measuring the subject's physical, psychological, social and role functions.

2.1.1 Primary Endpoint

Change from baseline (Week 0) to Week 16 on the modified Functional Status Questionnaire (mFSQ) total score.

2.2 Secondary Objectives

- To evaluate the efficacy and benefits of pimavanserin in subjects with PDP on activities of daily living
- To assess the clinician's global impression of severity and improvement of hallucination/delusions

2.2.1 Secondary Endpoints

- Change from baseline to Week 16 on the Schwab and England ADL Scale (Caregiver and Patient Version)
- Change from baseline to Week 16 on the Movement Disorders Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) Parts I and II (Caregiver and Patient Version)
- Week 16 Clinical Global Impression – Improvement (CGI-I) score for hallucinations and delusions
- Change from baseline to Week 16 on the Clinical Global Impression – Severity of Illness (CGI-S) score for hallucinations and delusions

- Week 16 Patient Global Impression of Improvement (PGI-I) score for hallucinations and delusions

2.3 Exploratory Objectives

- To assess the extent to which caregivers and families experience additional demands, responsibilities, and difficulties
- To explore the effects of pimavanserin in subjects with PDP on:
    - Functional capacity
    - Cognition
    - Depressive symptoms
    - Nighttime behavior
    - Caregiver burden

2.3.1 Exploratory Endpoints

- Change from baseline to Week 16 on the Virtual Reality Functional Capacity Assessment Tool (VRFCAT) including total time to complete, total number of errors, and total number of forced progressions
- Change from baseline to Week 16 on the Caregiver Strain Index (CSI) total score
- Change from baseline to Week 16 on the Neuropsychiatric Inventory-Questionnaire (NPI-Q) Nighttime Behavior Domain Severity and Caregiver Distress
- Change from baseline to Week 16 on the Geriatric Depression Scale (GDS) (Short Form)
- Change from baseline to Week 16 on the mFSQ score by subscale

2.4 Safety Objectives

- To assess the safety and tolerability of pimavanserin in adults with PDP

2.4.1 Safety Endpoints

The safety and tolerability of pimavanserin will be assessed using the following:

- Treatment-emergent adverse events (TEAEs)
- Columbia-Suicide Severity Rating Scales (C-SSRS) score
- Potentially clinically important (PCI) findings on:
    - Physical examinations

- Clinical laboratory tests
- Vital signs
- Electrocardiogram parameters

3 STUDY DESIGN

3.1 General Study Design

This study will be conducted as a 16-week, multi-center, single-arm, open-label study. Pimavanserin (ACP-103) will be administered at a dose of 34 mg to approximately 53 subjects with PDP. Approximately 42 subjects are expected to complete the study, assuming a 20% drop-out rate. Approximately 25 sites in the US will participate in this study.

The duration of participation for individual study subjects will be up to 26 weeks, consisting of the following 3 periods:

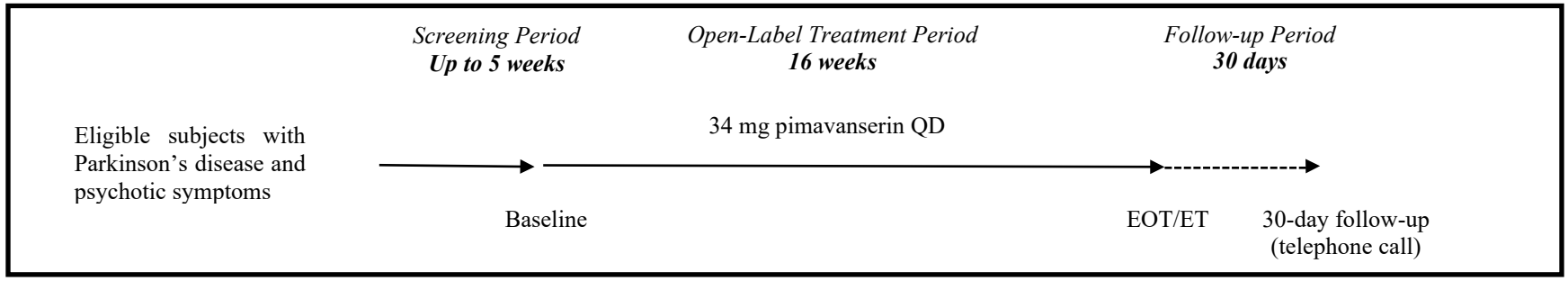
- Screening period (3-35 days)
- Treatment period (16 weeks)
- Safety follow-up period (30-34 days)

The study will be conducted on an outpatient basis with visits performed at Screening, Baseline (Week 0), Week 2, Week 4, Week 8, Week 12, Week 16 (end of Treatment Period) and Follow-up (at least 30 days and not to exceed 34 days after last dose of study drug). The first dose of study drug (pimavanserin) should be taken on the morning of the day after the Baseline visit.

The study start date is defined as the date the first subject is enrolled, which is the baseline visit date for the first subject. The primary completion date is the last date that subject data was collected for the primary outcome measure. The study completion date is defined as the last date that subject data was collected, which includes the safety follow-up telephone call visit.

Figure 1 illustrates the study design.

Figure 1 Schematic of Study Design for ACP-103-063



Abbreviations: EOT: end of trial; ET: early termination; QD: once daily

3.2 Schedule of Assessments

The schedule of events and assessments for the study is presented in the Appendix (Schedule of Events and Assessments).

3.3 Randomization

This is a single-arm study; therefore, randomization is not applicable.

3.4 Blinding

This is an open-label study; therefore, blinding is not applicable.

3.5 Determination of Sample Size

This is an exploratory study and is not powered for statistical significance. Formal sample size calculations were not used to determine sample size.

3.6 Coronavirus Disease 2019

In March 2020, the emerging coronavirus disease 2019 (COVID-19) pandemic resulted in the implementation of urgent safety measures designed to ensure subject safety (FDA 2020). 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 Appendix B of the Data Management Plan [DMP]). The impact of COVID-19 on the statistical analysis is discussed in each of the relevant sections of this SAP.

Because of possible restrictions due to COVID-19, all references to performing assessments, procedures, and evaluations at the clinic should be understood as including the options of performing these at the subject's home in the presence of a member of the clinical site staff and the subject's caregiver.

4 ANALYSIS SETS

4.1 Safety Analysis Set

The Safety Analysis Set will consist of all subjects who have taken at least 1 dose of study drug. Safety analyses will be based on the Safety Analysis Set.

4.2 Full Analysis Set

The Full Analysis Set will consist of all subjects who receive at least 1 dose of study drug, and have both baseline and at least one post-baseline value for the mFSQ total score. Efficacy analyses will be based on the Full Analysis Set.

5 DATA HANDLING CONVENTIONS

All data collected in the study will be listed.

5.1 General Data Reporting Conventions

For continuous variables, the following summary statistics will be provided: number of subjects, mean, standard error (SE) of the mean, 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 SDs and SEs will be presented to 2 more decimal places than the raw data. In general, the maximum number of decimal places is 4 and values will be truncated to 4 decimal places in situations where there are more than 4 decimal places. Wherever possible data will be decimal aligned.

For categorical variables, summaries will include the number and percentage of subjects in each category; the number of subjects and the percentage of subjects with missing data will be summarized for demographic and baseline characteristics (if applicable). Categories with zero counts will not have zero percentages displayed. Percentages will be presented to 1 decimal place. For demographic (except for Childbearing Potential) and baseline summaries (except for NPI-Q distress), percentages will be calculated by using the total number of subjects as the denominator.

Unless specified otherwise, all statistical tests will be 2-sided hypothesis tests performed at the significance level of 0.05 for main effects. All CIs will be 2-sided, 95% CIs, p-values will generally be presented to 4 decimal places, and p-values less than 0.0001 will be presented as <0.0001

Values that are collected with “<” or “>” signs will generally be analyzed as the numerical value without the sign in tables and figures. In listings, these data will be reported as collected with the sign.

Duration in months will be calculated as (number of days/365.25)*12 and will be rounded to 1 digit after the decimal point (tenths) for presentation purposes.

Duration in years will be calculated as number of days/365.25 and will be rounded to 1 digit after the decimal point (tenths) for presentation purposes.

Baseline is defined as the last non-missing value prior to the first dose of study drug. All change from baseline endpoints are calculated as the value of the corresponding post-baseline visit minus the value at baseline.

5.2 Derived Efficacy Variables

In general, all assessment scale total scores and subscores will be derived within the analysis datasets (with the exception of the VRFCAT variables, which are derived by the VRFCAT software). In the event that total scores and/or subscores are also collected on the eCRF, the derived values will be used for all analyses. Both the raw and derived scores will be presented in data listings.

5.2.1 Mini-Mental State Examination

The Mini-Mental State Examination (MMSE) is assessed only at the Screening visit. The MMSE is being used in this study to screen for cognitive impairment.

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 is scored as 0 for an incorrect response, and 1 for a correct response. 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.

The total score will be set to missing if there are one or more missing items.

5.2.2 NPI Domains A and B

At Screening (Visit 1) and Baseline (Visit 2), the degree of the patient's neuropsychiatric symptoms will be evaluated using two of the domains of the Neuropsychiatric Inventory (NPI), Domain A (Delusions) and Domain B (Hallucinations). For each domain, three separate ratings are collected, Frequency, Severity and Distress, and are scored as follows:

- Frequency of the symptom: 0=no symptom, 1=occasionally, 2=often, 3=frequently, 4=very frequently
- Severity of the symptom: 0=no symptom, 1=mild, 2=moderate, 3=severe
- Distress (how it affects the caregiver): 0=no distress, 1=minimal, 2=mild, 3=moderate, 4=severe, 5=extreme or very severe

The individual domain scores are calculated as the frequency score × severity score. The NPI psychosis score is derived as the sum of the domain A and domain B scores. If either A or B domain scores are missing then the psychosis total score will be missing.

Missing NPI scores (frequency, severity, and distress) will not be imputed. However, each of the individual domains contains a screening question that is used to elicit whether the

symptom is present or absent. If the answer to the screening question is "No" then the derived domain score will be set to 0.

5.2.3 NPI-Q Nighttime Behavior Domain

The NPI-Q is a self-administered questionnaire completed by informants about patients for whom they care. In this study, the informant is typically the caregiver. There are 12 domains in the full questionnaire; however, only the Nighttime Behavior domain will be completed by informants in this study. Two separate ratings are collected, Severity and Distress, and are scored as follows:

- Severity of the symptom (how it affects the subject): 0=no symptom, 1=mild, 2=moderate, 3=severe
- Distress (how it affects the informant): 0=not distressing at all, 1=minimal, 2=mild, 3=moderate, 4=severe, 5=extreme or very severe

If the informant reports that nighttime behaviors are not present, the Severity score will be set to 0, and the Distress score will remain a missing value. Otherwise, missing Severity and Distress scores will not be imputed.

5.2.4 Modified Functional Status Questionnaire

The Functional Status Questionnaire (FSQ) is a self-administered questionnaire that provides an assessment in ambulatory patients of physical, psychological, social, and role function (Jette et al. 1986). It comprises 34 core items that produce 6 summary scale scores: basic activities of daily living (ADL), intermediate activities of daily living, psychological function and mental health, work performance, social activity, and quality of interaction. In addition, it also includes 6 single-item scores: work situation; days per month in bed due to illness or injury; days per month when illness injury reduced activities normally performed for half a day or more; satisfaction with sexual relationships; satisfaction with own health; and frequency of social interaction.

Basic ADL consists of the following items:

1. Taking care of yourself, that is, eating, dressing or bathing
2. Moving in or out of a bed or chair
3. Walking indoors, such as around your home

Intermediate ADL consists of the following items:

1. Walking several blocks

2. Walking one block or climbing one flight of stairs
3. Doing work around the house, such as cleaning, light yard work, home maintenance
4. Doing errands such as grocery shopping
5. Driving a car or using public transportation
6. Doing vigorous activities such as running, lifting heavy objects or participating in strenuous sports

Social Activity consists of the following items:

1. Had difficulty visiting with relatives or friends
2. Had difficulty participating in community activities, such as religious services, social activities, or volunteer work
3. Had difficulty taking care of other people such as family members

Each of the Basic ADL, Intermediate ADL, and Social Activity items are scored as follows: 0=usually did not do for other reasons; 1=usually did not do because of health; 2=much difficulty; 3=some difficulty; 4=usually did with no difficulty.

Psychological Function consists of the following items:

1. Have you been a very nervous person
2. Have you felt calm and peaceful
3. Have you felt downhearted and blue
4. Were you a happy person
5. Did you feel so down in the dumps that nothing could cheer you up

Quality of Interaction consists of the following items:

1. Isolated yourself from people around you
2. Acted affectionate toward others
3. Acted irritable toward those around you
4. Made unreasonable demands on your family and friends
5. Gotten along well with other people

The Psychological Function and Quality of Interaction items are scored as follows: 1=all of the time; 2=most of the time; 3=a good bit of the time; 4=some of the time; 5=a little of the time; 6=none of the time.

Work Performance (for those employed during the preceding month) consists of the following items:

1. Done as much as others in similar jobs
2. Worked for short periods of time or taken frequent rests because of your health
3. Worked your regular number of hours
4. Done your job as carefully and accurately as others with similar jobs
5. Worked at your usual job but with some changes because of your health
6. Feared losing your job because of your health

The Work Performance items are scored as follows: 1=all of the time; 2=most of the time; 3=some of the time; 4=none of the time.

The modified FSQ (mFSQ) score will be calculated as the unweighted mean of the following subscale scores: Basic ADL, Intermediate ADL, Psychological Function, Social Activity, and Quality of Social Interaction. The Work performance subscale is excluded due to the age and non-employment status of the intended patient population. Prior to deriving the mFSQ total score, the following items scores are reversed: items 2 and 4 of the Psychological Function, and items 2 and 5 of the Quality of Interaction.

Subscale scores (SS, or transformed scale scores) are calculated according to this formula [Erratum 1986]:

$$SS = (100/k) \times (\text{Sum of individual item scores} - n)/n$$

Where,

n = number of questions in the subscale for which non-missing, valid information is available

k = maximum minus minimum valid response score

Activities that are not attempted because of non-health-related reasons are scored as 0 and will be excluded from average calculations [Tedesco et al. 1990]. Therefore, the maximum minus minimum valid response score for the Basic ADL, Intermediate ADL, and Social Activity will have a value of 3.

The mFSQ score will be calculated as (Basic ADL SS + Intermediate ADL SS + Psychological Function SS + Social Activity SS + Quality of Social Interaction SS)/5.

If any of the 5 subscale scores are missing then neither the mFSQ nor the subscale will be imputed.

5.2.5 Schwab and England ADL Scale (Patient and Caregiver Versions)

The Schwab and England ADL scale is an assessment of an individual's ability in activities of daily living, rated in 10% increments where 100%=completely independent, and 0%=vegetative. The scale will be completed independently by both the subject and the caregiver. Missing scores will not be imputed.

5.2.6 MDS-UPDRS Parts I and II (Patient and Caregiver Version)

The MDS-UPDRS is a comprehensive battery of motor and behavioral indices. The MDS-UPDRS Parts I and II will be used to assess non-motor aspects of experiences of daily living (Part I) and motor aspects of experiences of daily living (Part II).

5.2.6.1 MDS-UPDRS Part I

Part I of the MDS-UPDRS assesses the non-motor aspects of experiences of daily living. Part I consists of 13 items: cognitive impairment, hallucinations and psychosis, depressed mood, anxious mood, apathy, features of dopamine dysregulation syndrome, sleep problems, daytime sleepiness, pain and other sensations, urinary problems, constipation problems, light headedness on standing, and fatigue. The first 6 items are assessed by the investigator (with input from patients or caregivers), and the last 7 items are completed by the patient with or without the aid of a caregiver.

Each item is scored on a 5-point Likert scale: 0=normal, 1=slight, 2=mild, 3=moderate, 4=severe.

The total score will be derived as the sum of the 13 individual items from Part I. If 1 or 2 items are missing, the total score will be imputed as the mean of the non-missing items multiplied by 13 and rounded to the nearest integer. If more than 2 items are missing, the total score will be set to missing.

5.2.6.2 MDS-UPDRS Part II

Part II of the MDS-UPDRS concerns the motor aspects of experiences of daily living. Part II consists of 13 items: speech, saliva and drooling, chewing and swallowing, eating tasks, dressing, hygiene, handwriting, doing hobbies and other activities, turning in bed, tremor, getting out of bed (or a car or a deep chair), walking and balance, and freezing. All 13 items are completed by the patient with or without the aid of a caregiver.

Each item is scored on a 5-point Likert scale: 0=normal, 1=slight, 2=mild, 3=moderate, 4=severe.

The total score will be derived as the sum of the 13 individual items from Part II. If 1 or 2 items are missing, the total score will be imputed as the mean of the non-missing items multiplied by 13 and rounded to the nearest integer. If more than 2 items are missing, the total score will be set to missing.

5.2.7 CGI-S for Hallucinations and Delusions

The CGI-S is a single-item, clinician-rated, 7-point scale that is designed to rate the severity of the subject's hallucinations and delusions at the time of assessment using Investigator's judgment and past experience with subjects who have the same disorder. The 7-point scores are: 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. Higher scores denote more severe symptoms of hallucinations and delusions. Missing CGI-S scores will not be imputed.

5.2.8 CGI-I for Hallucinations and Delusions

The CGI-I is a clinician-rated, 7-point scale that is designed to rate the total improvement in the subject's hallucinations and delusions at the time of the 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. Higher scores denote more severe symptoms of hallucinations and delusions. Missing CGI-I scores will not be imputed.

5.2.9 PGI-I for Hallucinations and Delusions

The PGI-I is a single-item, global index used to rate the patient's impression of response of a condition to a therapy. The PGI-I asks the patient to rate their symptoms (of hallucinations and delusions) now, as compared with how it was at Baseline before beginning treatment, according to a 7-point scale. 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 PGI-I scores will not be imputed.

5.2.10 Virtual Reality Functional Capacity Assessment Tool

The Virtual Reality Functional Capacity Assessment Tool (VRFCAT) is a tablet-based assessment of functional capacity that simulates key instrumental activities of daily living in a realistic and interactive virtual environment. It will be used to explore the effects of pimavanserin on functional capacity and cognition. VRFCAT tablets will be provided by

VeraSci. The specific VRFCAT software utilized in this study are Pathway version 1.16.15 and software build 20200206.1

Study participants are asked to complete 12 objectives associated with multiple components of a shopping trip, including searching the pantry, making a shopping list, taking the correct bus to the store, shopping, paying for purchases, and getting home (see Table 1). Two different versions of the 4 mini scenarios are utilized in this study. Each version has a unique recipe, unique bus fares, and unique amounts of money in the wallet.

Table 1 VRFCAT Mini Scenarios and Objectives

Mini Scenario	Objective	Description
Apartment	1	Pick-up the recipe on the counter
	2	Search for ingredients
	3	Cross off ingredients and pick up bus schedule
	4	Pick up wallet
	5	Exit the apartment
Bus to Store	6	Board the bus to grocery store
Store	7	Pay for the bus
	8	Select food aisle
	9	Shop for groceries
	10	Pay for groceries
Bus to Apartment	11	Board the bus for home
	12	Pay for the bus

For each objective, participants who are unable to complete the objective within a pre-specified time period or who make too many errors are automatically progressed (i.e., forced progression) to the next objective.

For each objective the VRFCAT derives the following quantities:

1) Adjusted time (msec): Time to complete an objective, corrected for time of forced inactivity when an error occurs (as well as other factors). For example, if a subject makes an error, an error message is given during which the subject is not able to

proceed. The time required for the error message to display is removed from the adjusted time.

2) Forced progression: Progressions are scored using a binary value, such that 1 indicates the subject was progressed to the next objective at least once in a given testing period because the current objective was not completed in a timely manner.

3) Error count: total number of errors for a given objective.

The total adjusted time, total forced progressions, and total errors are then calculated as the sum of the adjusted time, forced progression, and error count, respectively, across all 12 objectives. Adjusted times will be converted from msec to seconds for purposes of analysis.

The VRFCAT software then takes the total adjusted time, total errors, and total forced progression scores and transforms them into T-scores according to the following algorithm:

1) calculate Z-score using age and sex corrected mean and standard deviation values from Verasci's proprietary normative data, i.e.

$$\text{Z-score} = -1 \times (\text{raw value} - \text{normative\_mean}) \div \text{normative standard deviation}$$

2) calculate T-score as (Z-score×10) + 50

Missing scores will not be imputed for any of the VRFCAT variables.

5.2.11 Caregiver Strain Index

The Caregiver Strain Index (CSI) is a 13-item questionnaire that measures strain related to care provision, and can be used to quickly identify families with potential caregiving concerns. There is at least one item for each of the following major domains: employment, financial, physical, social and time. Positive responses to seven or more items indicate a greater level of strain. Each item is scored zero (answered no) or one (answered yes). The total score is the sum score of the 13 items, resulting in a minimum possible score of 0 (no strain) and a maximum possible score of 13 (maximum possible strain). A score of 7 or higher indicates a high level of stress.

If 1 or 2 items are missing, the total score will be imputed as the mean of the non-missing items multiplied by 13 and rounded to the nearest integer. If more than 2 items are missing, the total score will be set to missing.

5.2.12 Geriatric Depression Scale (Short Form)

The Geriatric Depression Scale (GDS) (Short Form) is a brief 15-item questionnaire in which participants are asked to respond by answering yes or no in reference to how they felt over

the past week. For items 1, 5, 7, 11, and 13 a “no” response is scored as 1, and a “yes” response as 0. For items 2, 3, 4, 6, 8, 9, 10, 12, 14, and 15 a “yes” response is scored as 1, and a “no” response is scored as 0.

The GDS total score will be calculated as the sum of the 15 item scores. A total score of ≥5 suggests depression.

If 1 or 2 items are missing, the total score will be imputed as the mean of the non-missing items multiplied by 15 and rounded to the nearest integer. If more than 2 items are missing, the total score will be set to missing.

5.3 Analysis Visit Windows

Baseline is defined as the last non-missing value, including results from repeated and unscheduled measurements, prior to the first dosing of study drug. The first dose date is defined as the date of the first dose of study drug. Efficacy and safety assessments will be summarized by analysis visit as presented in Table 2 below, excluding Follow-up.

Table 2 Analysis Visit Windows

Analysis Visit Name	Target Study Day[1]	Study Day Interval
Baseline	1	≤1
Week 2	15	2 - 21
Week 4	29	22 – 43
Week 8	57	44 – 71
Week 12	85	72 - 99
Week 16	113	100 – 127
Follow-up	141	128 to maximum

[1] If the assessment date ≥ first dose date, study day = assessment date - first dose date + 1, otherwise study day = assessment date – first dose date. Study day 1 is the day of first dose.

5.3.1 Unscheduled Assessments

Both scheduled and unscheduled assessments, including early termination visits, will be considered for planned timepoint summaries, unless noted otherwise. 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 for summary. Exceptions may be made for incomplete assessments, in which case, more complete assessments may be given priority. If multiple measurements occur on the same day

at the same time (or the time is not available), the average of the measurements will be taken. Details are provided in a separate programming specifications document.

For safety summaries where the most extreme post-baseline values should be selected, e.g. overall post-baseline minimum and overall post-baseline maximum, all non-missing post-Baseline values should be considered, regardless of whether the value is selected for the by-visit summaries.

When replicate ECGs are collect at a given timepoint (typically, in triplicate), the average of the replicate values (i.e., PR interval, QRS duration, QT interval, QTcB interval, QTcF interval, and RR interval) will be considered as one assessment for the analyses.

5.3.3 Data Listings

All results will be presented in data listings, and the values selected for analysis will be marked.

5.4 Data Handling Conventions

5.4.1 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 the non-missing drug return date of the last dispensed drug kits, the last drug dispensed date plus the number of days that the dosing would continue per protocol or the end-of-treatment/early termination visit (EOT/ET) date, whichever occurs earlier. For the incomplete last dose date of the study drug, the imputation will be compared against the last expected dosing date defined as above. Detailed algorithms will be documented in a separate programming specifications document. The missing or incomplete dates will be displayed in the data listings as reported on the eCRF rather than the imputed dates.

For data summarizations before final database lock, if a subject is still ongoing, then the subject's last dose date will be imputed using the database cutoff date.

5.4.2 Missing or Incomplete Dates for Concomitant or Post-Treatment 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 programming specifications

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

5.4.4 Missing Severity Assessment for Adverse Events

If the severity is missing for a treatment-emergent AE, 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 presented in data listings.

5.4.5 Missing Relationship to Study Drug for Adverse Events

If the relationship to study drug is missing for a treatment-emergent AE, 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.

5.4.6 Character Values of Clinical Laboratory Variables

If the reported value of a clinical laboratory variable cannot be used in a summary due to, for example, a character string reported for a numeric variable, an appropriately determined coded value will be used in the summary. 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

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. In addition, the number of subjects enrolled at each site will also be tabulated.

For enrolled subjects, the number and percentage of subjects in the analysis sets will be summarized. A listing will be provided displaying all subjects excluded from the analysis sets (if any), and will include reason(s) for exclusion.

For subjects who participate in the screening phase and are screen failures, their demographics information (including age, sex, ethnicity, and primary race), screen failure reasons (the specific inclusion/exclusion criterion [or criteria] not met or other reason), informed consent information, and protocol version will be provided in a listing. 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 a 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.

Early terminations of subjects due to COVID-19 related reasons will be captured in the eCRFs and tabulated. The number and percentage of subjects who completed the study or discontinued (all discontinued and by discontinuation reason) will also be summarized for the Safety and Full Analysis sets.

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 will be detailed in the study Protocol Deviation Management Plan. Protocol deviations will also be assessed with respect to relationship to COVID-19.

For Safety subjects, a summary of the number and percentage of subjects with major protocol deviations for each deviation category will be presented by relationship to COVID-19 (overall, COVID-19 related, and non-COVID-19 related). A listing of all protocol deviations will also be provided by relationship to COVID-19 (overall, COVID-19 related, and non-COVID-19 related).

8 DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS

Demographics and baseline characteristics will be summarized for the Safety and Full Analysis sets 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, weight, body mass index (BMI), childbearing potential, subject/caregiver living situation, and caregiver relationship.

The reported age reflects a 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

<65, 65 to 74, 75 to 84, and ≥85 years old. BMI categories (kg/m^2) will be presented as <25, 25 to 30, and >30. Race will be categorized as White vs. Non-White.

8.1 Baseline Disease Characteristics

Disease characteristics at Baseline will be summarized for the Safety and Full Analysis sets using descriptive statistics. MMSE total score, CGI-S, NPI-Q nighttime behavior (severity and distress), MDS-UPDRS Part I total score, and MDS-UPDRS Part II total score, will be presented as continuous variables. CGI-S and NPI-Q will also be presented as categorical variables using discrete categories. The number and percentage of subjects reporting suicidal ideation (within the past 6 months) and suicidal behavior (within the past 12 months) and lifetime, and suicidal ideation or behavior since the Screening visit, will be tabulated.

8.2 Neuropsychiatric Inventory

Baseline NPI will summarized for the Safety and Full Analysis sets using descriptive statistics. The delusions total score, hallucinations total score, and psychosis total score will be presented as continuous variables. In addition, for each of the delusions and hallucinations domains the frequency, severity, and distress scores will be presented as categorical variables.

9 MEDICAL HISTORY

Medical and surgical history data will be coded using Medical Dictionary for Regulatory Activities (MedDRA) version 23.0 or newer. The subject incidence will be summarized for each system organ class (SOC) and preferred term for the Safety Analysis Set and Full Analysis Set. 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 term for the medical history condition/event, start and stop dates (when available), and an indicator for whether or not the condition is ongoing at Screening will be provided.

10 PRIOR, CONCOMITANT, AND POST-TREATMENT MEDICATION

Medications will be coded using WHO Drug Dictionary (WHODRUG-Global-B3) March 2019 or later. Relationship to COVID-19 will be assessed by the investigator as specified in the DMP, Appendix B.

For medications, the number and percentage of subjects taking each drug class (ATC Level 3) and medication preferred term will be tabulated. A subject will be counted only once per drug class or per medication preferred term for the summary. Prior, concomitant, and

post-treatment medications will be summarized separately using the Safety Analysis Set. Concomitant medications will also be summarized by relationship to COVID-19.

Listings of the prior, concomitant, and post-treatment medications will also be provided.

Prior medication is defined as any medication with a start and stop date 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 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.

Post-treatment medication is defined as any medication with a start date after the last dose of study drug.

11 EXTENT OF EXPOSURE AND TREATMENT COMPLIANCE

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

11.1 Exposure to Treatment

Pimavanserin dose levels are expressed as free base.

Duration of pimavanserin exposure (days) for each subject is calculated as follows:

Duration of exposure = last dose date – first dose date + 1

The number and percentage of subjects in each of the following categories will be displayed: ≤2 weeks (1 to 14 days), ≤4 weeks (1 to 28 days), ≤8 weeks (1 to 56 days), ≤12 weeks (1 to 84 days), and ≤16 weeks (1 to 112 days). An additional category may be added depending on the length of maximum exposure.

11.2 Measurement of Treatment Compliance

Treatment compliance is defined as the total number of tablets actually taken by a subject divided by the number of tablets expected to be taken while on treatment, 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). If one or more kits are not returned, the tablets expected to be taken will be imputed; the details will be provided in the analysis dataset specification document.

Treatment compliance 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 presented: <80%, 80 to 120%, and >120%.

12 EFFICACY ANALYSES

All efficacy analyses will be performed using the Full Analysis Set. Descriptive summaries of all efficacy endpoints will be provided. No hypothesis testing is planned.

12.1 Primary Efficacy Endpoint

The primary efficacy endpoint is the change from Baseline to Week 16 in the mFSQ total score.

12.2 Secondary Efficacy Endpoints

The secondary efficacy endpoints are the following:

- Change from Baseline to Week 16 on the Schwab and England ADL Scale (Caregiver and Patient version)
- Change from Baseline to Week 16 on the MDS-UPDRS Parts I total score and MDS-UPDRS Part II total score (Caregiver and Patient version)
- Change from Baseline to Week 16 on the CGI-S score for hallucinations and delusions
- Week 16 CGI-I score for hallucinations and delusions
- Week 16 PGI-I score for hallucinations and delusions

12.3 Exploratory Efficacy Endpoints

The exploratory efficacy endpoints are the following:

- Change from Baseline to Week 16 on the VRFCAT including adjusted total time T-score, total number of errors T-score, and total number of forced progressions T-score
- Change from baseline to Week 16 on the CSI total score
- Change from Baseline to Week 16 on the NPI-Q Nighttime Behavior Domain Severity and Caregiver Distress scores
- Change from Baseline to Week 16 on the GDS (Short Form)
- Change from Baseline to Week 16 mFSQ score by subscale

- Change from Baseline to Week 16 for each item in MDS-UPDRS Part II

12.4 Handling of Missing Data

Scores that are missing, after any imputation of individual missing items as described in Section 5.2, will not be imputed. MMRM analyses will be used as the statistical approach for handling of missing data. Model parameters are simultaneously estimated using restricted maximum likelihood estimation. The MMRM method is unbiased in the estimation of treatment effect under the missing at random assumption and can be thought of as attempting to estimate the treatment effect that could have been observed if all subjects had continued on study for the full study duration.

12.5 Multiple Comparisons/Multiplicity

Multiplicity adjustments will not be made as this is an exploratory study and no hypothesis testing is planned.

12.6 Examination of Subgroups

Selected subgroup analyses may be performed.

12.7 Analysis of Efficacy Endpoints

Descriptive statistics for all exploratory efficacy endpoints listed in Section 12.3 will be tabulated at scheduled timepoints (excluding Screening). In addition, descriptive statistics may be provided for various subscales and individual scale items.

12.8 Mixed Model Repeated Measures (MMRM)

The MMRM method with restricted maximum likelihood estimation (REML) will be used for selected efficacy endpoints. The dependent variable will be the change from Baseline. The independent variables in the model will include effects for visit, Baseline score, and Baseline score-by-visit-interaction. An unstructured covariance matrix will be used to model the within patient errors, and the Kenward-Roger approximation will be used to adjust the denominator degrees of freedom.

In the event that the model fails to converge using the unstructured covariance matrix, then only descriptive statistics will be presented.

Summary statistics (observed and change from Baseline, LS means with corresponding two-sided 95% confidence intervals (CIs), and p-values will be presented at each post-Baseline visit for the Full Analysis Set. The primary comparison is based on the LS mean change from Baseline to Week 16.

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

The above MMRM model will be used for each of the primary, secondary, and exploratory endpoints. CGI-I score will be analyzed using a similar MMRM model with effects for visit, Baseline CGI-S score, and Baseline CGI-S score-by-visit interaction. PGI-I score will be analyzed using an MMRM model with an effect for visit.

NPI-Q Caregiver and Distress scores will be summarized in shift tables to determine the number and percentage of subjects with Severity scores classified as none, mild, moderate, or severe relative to the same classification at the Baseline visit; and the number and percentage of subjects with Distress scores classified as not distressing at all, minimal, mild, moderate, severe, and extreme or very severe relative to the same classification at the Baseline visit. For the by-visit shift summary, the denominator is the number of subjects with non-missing values at Baseline and the given visit for the given parameter.

For each of the continuous efficacy endpoints (except for CGI-I and PGI-I) the null hypothesis is that the change from baseline to a certain time-point is equal to 0. The corresponding alternative hypothesis is that the change from baseline to a certain endpoint is not equal to 0.

For the CGI-I and PGI-I endpoints the null hypothesis is that the score at a certain time-point is equal to 4 (no change); the corresponding alternative hypothesis is that the score at a certain time-point is not equal to 4.

For each of the continuous efficacy endpoints (except for CGI-I and PGI-I), in addition to MMRM, a paired t-test using a last observation carried forward imputation method will be performed at each post-baseline visit. For CGI-I and PGI-I, a one sample t-test (against a value of 4, indicating no change) will be performed instead of the paired t-test.

Additional exploratory analyses may be conducted on VRFCAT data to adjust for deficits in cognition.

13 SAFETY ANALYSES

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

13.1 Adverse Events

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

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

For the following summaries, the event counts and the number and percentage of subjects reporting TEAEs will be tabulated:

- TEAEs by SOC and preferred term (PT)
- Most frequently reported TEAEs by SOC and PT
- TEAEs by PT
- TEAEs by maximum severity, SOC, and PT
  - 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
- TEAEs related to study drug by SOC and PT
- Treatment-emergent serious adverse events (TESAEs) by SOC and PT
- TEAEs leading to study drug discontinuation or study termination by SOC and PT
- AEs with fatal outcome by SOC and PT
- TEAEs with fatal outcome (that occurred within 30 days of last dose) by SOC and PT

The relationship of selected AEs to COVID-19 will be assessed as detailed in the DMP, Appendix B. COVID-19 related TEAEs will be tabulated separately.

For tabulations that include SOC and PT the display will be sorted alphabetically by SOC and then by descending subject frequency for the PTs within each SOC. SOCs will not be included in the TEAEs by PT tabulation. This display 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 the first and last dose of study drug, 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 TEAEs leading to study drug discontinuation or study termination, subjects who died (if any), and COVID-19 related AEs. In these listings, an indicator for TEAEs will also be included.

13.2 Other Safety Analyses

13.2.1 Clinical Laboratory Variables

Clinical laboratory tests are performed at Screening, Baseline, Week 4, and Week 16/ET visits. All laboratory test results are from a central laboratory. All units will be displayed in Système International [SI] units.

13.2.2 Chemistry

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), aspartate aminotransferase (AST), gamma-glutamyl transpeptidase (GGT), alkaline phosphatase (ALP), total bilirubin (TBIL), lactate dehydrogenase (LDH)
- Glucose
- Albumin (ALB), total protein
- Creatine kinase (CK)/creatine phosphokinase (CPK)
- Lipid panel
    - total cholesterol, high-density lipoprotein (HDL)-cholesterol, triglycerides, LDL-cholesterol, cholesterol/HDL ratio, ratio, non-HDL-cholesterol
- Magnesium (Mg), Vitamin B12, glycosalated hemoglobin (HbA_{1c}), thyroid stimulating hormone (TSH) and free T4 (if TSH is out of range)
    - these tests should only be performed at Screening
- Serum pregnancy test
    - performed only at Screening for women of childbearing potential

13.2.3 Hematology

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

13.2.4 Urinalysis

Urinalysis tests include the following:

- Blood, RBCs, WBCs, leukocyte esterase, protein, glucose, ketones, specific gravity, pH
- A urine pregnancy test will be performed for all women of childbearing potential at Baseline and at each scheduled post-baseline visit (except Week 2)
- A urine toxicity screen will test for controlled substances at screening (and Baseline for those subjects enrolled under the original Protocol), and may include the following: amphetamine, barbiturates, benzodiazepines, cocaine, methadone, morphine/opiates, methamphetamine, marijuana (THC), phencyclidine (PCP), ecstasy (3,4-methylenedioxy- methamphetamine [MDMA]). Negative drug screens are required for study eligibility.

13.2.5 Methods of Analysis for Clinical Laboratory Variables

13.2.5.1 Observed Values and Change from Baseline

Clinical laboratory values reported as continuous values for hematology, chemistry and urinalysis will be summarized using descriptive statistics at Baseline and post-Baseline visits. The change from Baseline values will also be summarized at post-Baseline visits.

The overall post-Baseline minimum, maximum, and last post-Baseline observed and change from Baseline values will also be summarized (for continuous values only).

For hemoglobin, hematocrit and uric acid, by-visit as well as overall post-Baseline minimum, maximum, and last post-Baseline 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 each scheduled post-Baseline visit. For the categorical urinalysis by-visit summary, the denominator is the number of subjects with non-missing values for the given parameter and visit.

13.2.5.2 Shift Tables

Laboratory values will be summarized in shift tables to determine the number and percentage of subjects with values classified as below (low), within (normal), or above (high) normal ranges at each scheduled post-Baseline visit relative to the same classification at the Baseline visit. For the by-visit shift summary, the denominator is the number of subjects with non-missing values at Baseline and the given visit for the given parameter.

The shifts from Baseline to overall post-Baseline minimum and maximum will also be

presented. The denominator is the number of subjects with non-missing Baseline and at least 1 post-Baseline value for the given parameter.

13.2.5.3 Potentially Clinically Important (PCI) Laboratory Values

The number and percentage of subjects with PCI laboratory values at scheduled post-Baseline visits Week 4 and Week 8 and overall post-Baseline will be summarized by Baseline status (all or within normal range) for selected parameters. PCI criteria are listed in Table 3 and Table 4. Subjects with multiple PCI values for a given parameter will be counted only once for that parameter. All post-Baseline values will be considered, including unscheduled and out-of-window values, for the overall post-Baseline visit.

For the by-visit summary, the numerator for the percentage is the number of subjects with a post-Baseline PCI value 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 value for the given parameter, and the denominator is the number of subjects with at least 1 post-Baseline value for the given parameter. 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.

Table 3 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
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 4 Criteria for Potentially Clinically Important Laboratory Values - Urinalysis

Urinalysis (Occult Blood)	Low PCI Criteria	High PCI Criteria
Blood (occult blood)	Not Applicable	≥+2
Protein	Not Applicable	≥+2
Glucose	Not Applicable	≥+2

13.2.5.4 Data Listings

All laboratory test results will be listed. The listings will include date and study day of collection. Out of range values will be flagged in the data listings (e.g., as 'L' or 'H').

A listing of all PCI values will be provided. This listing will include all observations for those subjects and parameters for which at least one PCI value was observed.

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

13.2.6 Vital Signs and Height, Weight, and BMI

13.2.6.1 Vital Signs Variables

Vital signs will be collected throughout the study, and include body temperature, resting respiration rate, sitting or supine systolic and diastolic blood pressure, and pulse rate. Blood pressure should be measured after the subject has been sitting or supine for at least 3 minutes.

13.2.6.2 Height, Weight, and BMI

Height and weight will be measured at Screening, Baseline, and Week 16/EOT.

BMI will be derived as: Weight (kg)/ [height (m)]^2.

13.2.6.3 Vital Signs Methods of Analysis

Vital signs will be summarized using descriptive statistics at Baseline and all scheduled post-Baseline visits. The change from Baseline values will also be summarized at the scheduled post-Baseline visits. The overall post-Baseline minimum, maximum, and last post-Baseline observed and change from Baseline values will also be summarized.

Vital signs and weight values will be considered PCI if they meet the criteria listed in Table 5. The number and percentage of subjects with post-Baseline vital signs that are PCI will be summarized at scheduled post-Baseline visits 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 value 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 value for the given parameter, and the denominator is the number of subjects with at least 1 post-Baseline value for the given parameter.

A listing of all PCI values will be provided. This listing will include all observations for those subjects and parameters for which at least one PCI value was observed.

Table 5 Criteria for Potentially Clinically Important 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%

13.2.7 Electrocardiogram (ECG)

12-lead ECGs are collected at Screening, Baseline, Week 4, and Week 16.

All 12-lead ECGs will be complete, standardized recordings. All ECGs will be centrally read; the interpretation by the central cardiologist is considered the official interpretation. Each ECG will be recorded continuously for 5 minutes. For subjects enrolled under the original Protocol, a 12-lead ECG was completed in triplicate at Screening, and as a single ECG at other timepoints. All tracings will be evaluated by a central reading laboratory. At the Baseline visit the machine-read results will also be recorded. ECG data summaries, including the cardiologist's interpretation, will be analyzed using the centrally evaluated data. All data, including the machine-read Baseline results, will be listed.

13.2.7.1 ECG Variables

ECG variables include heart rate, PR interval, RR interval, QRS interval, QT interval, and QTc intervals. QTc intervals include QTcB (Bazett's formula) and QTcF (Fridericia's formula). QTcF will also be categorized into the following categories (msec):

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

13.2.7.2 ECG Methods of Analysis

13.2.7.2.1 Observed Values and Change from Baseline

Observed values and the changes from Baseline of ECG variables will be summarized using descriptive statistics at Baseline and all scheduled post-Baseline visits. The overall post-Baseline minimum, maximum, and last post-Baseline observed and change from Baseline values will also be summarized.

For the QTcF categorical analysis, the number and percentage of subject in each category will be summarized at each scheduled visit as well as overall post-Baseline maximum and last post-Baseline visit.

13.2.7.2.2 PCI Values

Electrocardiogram values will be considered PCI if they meet or exceed the criteria listed in Table 6. The number and percentage of subjects with PCI values will be summarized at each scheduled post-Baseline visits 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 value 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 value for the given parameter, and the denominator is the number of subjects with at least 1 post-Baseline value for the given parameter; all post-Baseline values will be considered, including unscheduled and out-of-window values.

A listing of all subjects with any PCI value will be provided for those parameters for which at least 1 PCI value was observed.

Table 6 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

13.2.7.2.3 Cardiologist Interpretations

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 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.

13.2.8 Physical Examinations

A physical examination, including neurological examination (cranial nerves, motor, sensory, reflexes, gait, and coordination), will be performed at Screening and Week 16/ET visits.

Physical and neurological examination results (normal, abnormal, and not done) at Screening and Week 16/ET will be summarized in a frequency table by body system and visit. For purposes of the PCI analysis, any abnormal result will be considered as PCI.

13.2.9 C-SSRS

The C-SSRS will be used to assess suicidal ideations and behaviors. 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, 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, that subject will be counted as having suicidal ideation.

The C-SSRS will be administered at Screening, Baseline, Weeks 2, 4, 8, 12 and 16/ET. The C-SSRS Baseline/Screening version will be used at the Screening visit, and the version assessing suicidality since the last visit will be used at all following visits (including the Baseline visit).

The event counts and the number and percentage of subjects reporting any post-Baseline 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), suicidal behavior (preparatory acts or behavior; aborted attempt; interrupted attempt; actual attempt; suicide), or suicidality (any suicidal ideation or behavior) will be tabulated.

The event counts and the number and percentage of subjects reporting any post-Baseline non-suicidal self-injurious behavior will also be tabulated.

For calculating the percentages, the denominator will be the number of subjects with at least one post-Baseline C-SSRS.

14 CLINICAL PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES

No pharmacokinetic samples are collected in this study.

15 INTERIM ANALYSIS

An efficacy analysis may be performed when 50% of subjects have completed the study or terminated early. Additional analyses may be performed as necessary throughout the course of the study.

16 DATA MONITORING/REVIEW COMMITTEE

There is no data monitoring/review committee for this open-label study.

17 COMPUTER METHODS

Statistical analyses will be performed using Version 9.4 (or newer) of SAS® software 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

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

19 REFERENCES

Erratum. *Journal of General Internal Medicine*. 1986; 1: 427.

FDA Guidance on Conduct of Clinical Trials of Medical Products during COVID-19 Pandemic. March 2020; updated May 14, 2020.

Jette AM, Davies AR, Cleary PD, Calkins DR, Rubenstein LV, Fink A, Kosecoff J, Young RT, Brook RH, Delbanco TL. The Functional Status Questionnaire: Reliability and Validity When Used in Primary Care. *J Gen Intern Med*. 1986; 1(3):143-149.

Tedesco C, Manning S, Lindsay R, Alexander C, Owen R, Smucker ML. Functional assessment of elderly patients after percutaneous aortic balloon valvuloplasty: New York Heart Association classification versus functional status questionnaire. *Heart Lung*. 1990; 19:118-125.

20 APPENDICES

20.1 Summary of Version Changes

Version No:	Document History Description of Update	Author(s)	Version Date
1.0	Original version	[redacted]	01 March 2022

20.2 Schedule of Events and Assessments for ACP-103-063

Period	Screening	Treatment						Safety Follow-up
		Baseline						
Visit Week	**-5 to 0**	**0**	**2**	**4**	**8**	**12**	**16/EOT**	**20[a]**
Visit Number	**1**	**2**	**3**	**4**	**5**	**6**	**7/ET[b]**	**8**
Visit window (days)			**±3**	**±3**	**±3**	**±3**	**±3**	**+4**
Type of Visit	clinic	clinic	clinic	clinic	clinic	clinic	clinic	telephone
Informed consent of subject	X							
Informed consent of caregiver or study partner	X							
Inclusion/exclusion criteria	X	X						
Medical history and demographics	X							
Weight, height, BMI	X	X					X	
Physical examination	X						X	
12-lead ECG[c]	X	X		X			X	
Vital signs[d]	X	X	X	X	X	X	X	
Clinical laboratory tests	X	X		X			X	
Urine drug screen	X							
Pregnancy test[e]	X	X		X	X	X	X	
MMSE	X							
C-SSRS	X	X	X	X	X	X	X	
FSQ	X	X		X	X	X	X	
Schwab & England ADL Scale (Patient)	X	X	X	X	X	X	X	
Schwab & England ADL Scale (Caregiver)	X	X	X	X	X	X	X	
NPI domains A and B	X	X						
NPI-Q Nighttime Behavior Domain only		X	X	X	X	X	X	

20.2 Schedule of Events and Assessments for ACP-103-063 (continued)

Period	Screening	Treatment						Safety Follow-up
		Baseline						
Visit Week	**-5 to 0**	**0**	**2**	**4**	**8**	**12**	**16/EOT**	**20[a]**
Visit Number	**1**	**2**	**3**	**4**	**5**	**6**	**7/ET[b]**	**8**
Visit window (days)			**±3**	**±3**	**±3**	**±3**	**±3**	**+4**
Type of Visit	clinic	clinic	clinic	clinic	clinic	clinic	clinic	telephone
MDS-UPDRS Parts I and II (Patient)	X	X	X	X	X	X	X	
MDS-UPDRS Parts I and II (Caregiver)	X	X	X	X	X	X	X	
CGI-S specific to H&D	X	X	X	X	X	X	X	
CGI-I specific to H&D			X	X	X	X	X	
PGI-I specific to H&D			X	X	X	X	X	
VRFCAT	X	X		X			X	
Caregiver Strain Index	X	X	X	X	X	X	X	
Geriatric Depression Scale (Short Form)	X	X	X	X	X	X	X	
Concomitant medication	X	X	X	X	X	X	X	X
Assessment of adverse events		X	X	X	X	X	X	X
Dispense study drug		X	X	X	X	X		
Study drug return			X	X	X	X	X	
Study drug accountability			X	X	X	X	X	

Abbreviations: ADL=activities of daily living; BMI=body mass index; CGI-I=clinical global impression – improvement; CGI-S=clinical global impression – severity; C-SSRS=Columbia-Suicide Severity Rating Scale; ECG=electrocardiogram; EOT=end of trial; ET=early termination; FSQ= functional status questionnaire; H&D=hallucinations and delusions; MMSE=mini-mental state examination; NPI=neuropsychiatric inventory; NPI-Q=neuropsychiatric inventory – questionnaire; PGI-I=patient global impression – improvement; MDS-UPDRS=Movement Disorders Society – Unified Parkinson's Disease Rating Scale; VRFCAT=virtual reality functional capacity assessment tool

[a] The safety follow-up telephone call visit is to occur 30 (+4) days after the last dose of study drug, nominally at Week 20 for completers and earlier for ET subjects.

[b] A 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 16 assessments.

[c] The ECG may be repeated once at Screening in consultation with the Medical Monitor. ECGs can be performed any time before blood sampling or at least 30 minutes after blood sampling during clinic visits.

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

[e] Applicable only to women of childbearing potential. A serum pregnancy test is performed at Screening and a urine pregnancy test at Baseline and all subsequent visits (except Visit 3 at which neither serum nor urine pregnancy test is performed).