Official Title: A Pilot Randomized, Double-blind, Placebo-controlled Study to Evaluate Safety and Daytime Sedation in Subjects With Parkinson's Disease With Neuropsychiatric Symptoms Treated With Pimavanserin or Low-Dose Quetiapine

NCT Numbers: NCT04164758

Document Date: 19 Oct 2020

STATISTICAL ANALYSIS PLAN

Protocol No.:	ACP-103-056
Protocol Title:	A Pilot Randomized, Double-blind, Placebo-controlled Study to Evaluate Safety and Daytime Sedation in Subjects With Parkinson's Disease With Neuropsychiatric Symptoms Treated With Pimavanserin or Low-Dose Quetiapine
Drug:	Pimavanserin
Sponsor:	ACADIA Pharmaceuticals Inc.
Version No. and Date	Version 1.0, 19 October 2020

SIGNATURE/APPROVAL PAGE

AUTHORS

[redacted] 19-Oct-2020

[redacted] Date
Associate Director, Biostatistics
ACADIA Pharmaceuticals Inc.

APPROVERS

[redacted] 20-Oct-2020

[redacted] Date
Temporary Senior Medical Director, Medical Affairs
ACADIA Pharmaceuticals Inc.

[redacted] 19-Oct-2020

[redacted] Date
Executive Director, Biostatistics
ACADIA Pharmaceuticals Inc.

TABLE OF CONTENTS

LIST OF TABLES

ABBREVIATIONS

AE	Adverse event
ANCOVA	analysis of covariance
CGI-I	Clinical Global Impressions – Improvement
CGI-S	Clinical Global Impressions – Severity
COVID-19	Coronavirus disease 2019
CRO	Contract Research Organization
C-SSRS	Columbia Suicide Severity Rating Scale
ECG	electrocardiogram
eCRF	electronic case report form
EDC	electronic data capture
EMA	Ecological Momentary Assessment
GSD	Guidance for Site Documentation and Data Management Querying of Data Impacted by COVID-19
IRT	Interactive Response Technology
KSS	Karolinska Sleep Scale
MDS-UPDRS	Movement Disorders Society-Unified Parkinson's Disease Rating Scale
MedDRA	Medical Dictionary for Regulatory Activities
MMSE	Mini-Mental State Examination
OC	observed cases
PCI	potentially clinically important
PD	Pharmacodynamic(s)
PD NMS	Parkinson's disease non-motor symptoms
PDP	Parkinson's disease psychosis
PGI-I	Patient Global Impression - Improvement
PROMIS	Patient-Reported Outcomes Measurement Information System
SAE	serious adverse event
SAP	statistical analysis plan
SD	Standard Deviation
SE	Standard Error
SOC	system organ class
TEAE	treatment-emergent adverse event

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 Protocol Amendment 3, dated 31 January 2020. Specifications for tables, figures, and listings are contained in a separate document. Actigraphy and Ecological Momentary Analysis (EMA) data collected on the Handheld Device will be analyzed as described in a separate, stand-alone SAP.

2. OBJECTIVES

2.1 Primary Objective

The primary objective of this study is to assess the safety and tolerability of pimavanserin and low-dose quetiapine compared to placebo in subjects with Parkinson's disease.

2.2 Exploratory Objective

The exploratory objectives of this study are:

- to explore the sensitivity of assessments/instruments that are novel for evaluating the experience of subjects with Parkinson's disease with respect to sedation/sleep, cognition, and other non-motor symptoms.
- to explore the effects of pimavanserin and low-dose quetiapine compared to placebo in subjects with Parkinson's disease on:
    - Daytime sleepiness
    - Nighttime sleep
    - Cognition
    - Other non-motor symptoms
    - Treatment satisfaction
    - Safety and tolerability

3. STUDY DESIGN

3.1 General Study Design

This study is a multicenter, randomized, double-blind, placebo and active-control, parallel-group pilot study in subjects with Parkinson's disease. The study will have 3 periods:

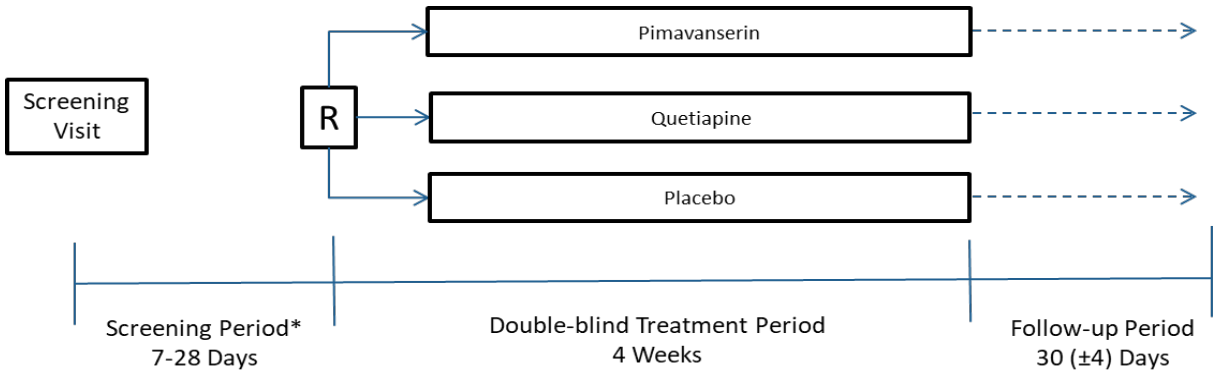
- Screening period (7-28 days)
- Double-blind treatment period (4 weeks)
- Safety follow-up period (30 [±4] days)

The study completion date is defined as the date the final subject, across all sites, completes their final protocol-defined assessment (this includes the safety follow-up phone call). If the study is terminated for any reason, subjects remaining in the study will return to standard of care.

Up to 60 subjects will be randomized in a 1:1:1 ratio to pimavanserin, quetiapine, or placebo. Subjects for whom the Investigator does not increase the dose of blinded study drug consistent with 50 mg quetiapine at Week 1 (Visit 3) will be withdrawn from the study and replaced. A maximum of 12 subjects across treatment groups (20%) will be replaced for this and only this reason.

The study design is summarized in Figure 1.

Figure 1 Schematic of Study Design

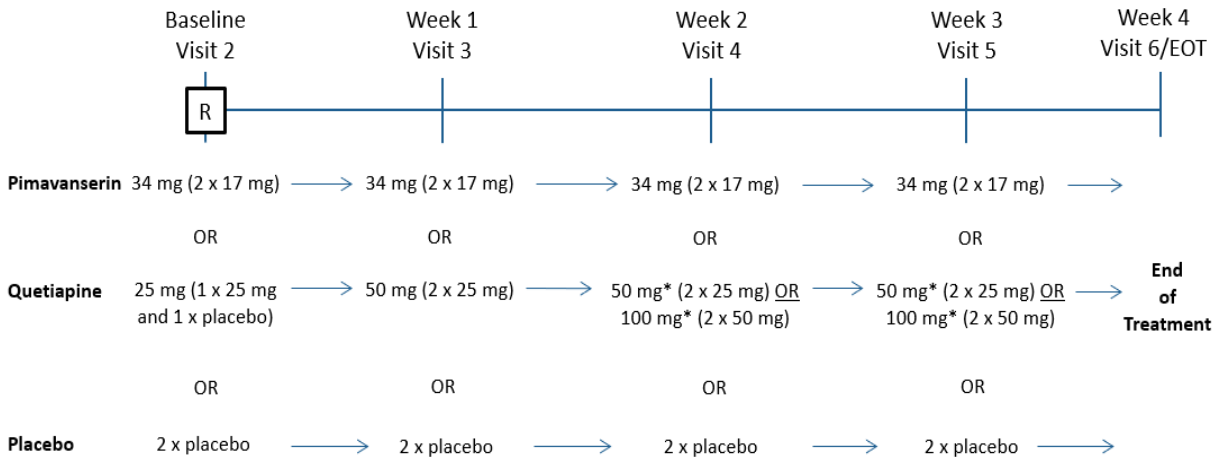


R= randomization

* Pre-drug actigraphy data and handheld device assessments are conducted prior to randomization (Baseline)

Note: A schematic for study drug dispensing and dose adjustments is provided in Figure 2.

Figure 2 Schematic for Study Drug Dispensing and Dose Adjustments



* Week 2 (Visit 4) & Week 3 (Visit 5): Subjects may remain on blinded study drug dose consistent with 50 mg quetiapine or be optionally increased to 100 mg quetiapine.

Abbreviations: EOT=end of treatment; R=randomization

Note: Dose adjustments and study drug dispensing are only allowed at scheduled visits (i.e., not at unscheduled visits). Subjects should not take study drug on the mornings of clinic visit days until instructed to do so by the site staff. At Baseline and at each subsequent clinic visit, subjects will be dosed and will remain in the clinic for observation for a minimum of 60 minutes after dosing. At Week 1 (Visit 3), Week 2 (Visit 4), and Week 3 (Visit 5), the Investigator will be encouraged to increase the dose of blinded study drug, consistent with 50-100 mg quetiapine, based on the Investigator's assessment of clinical response. Dose adjustments will be implemented through the IRT. In actuality, only the dose of quetiapine would be increased (to 50 mg at Week 1 [Visit 3] and then optionally to 100 mg at Week 2 [Visit 4] or Week 3 [Visit 5]); the pimavanserin dose will remain fixed at 34 mg per day throughout the study. Subjects for whom the Investigator does not increase the dose of blinded study drug consistent with 50 mg quetiapine at Week 1 (Visit 3) will be withdrawn from the study and replaced. The Investigator will be blinded to study drug and thus will not know whether their decision to increase the dose of blinded study drug results in an actual increase in study drug dose. The dose of study drug (pimavanserin, quetiapine or placebo) may not be decreased. If a subject cannot tolerate their dose, they must be withdrawn from the study.

3.2 Schedule of Assessments

The schedule of events and assessments, the schedule of assessments for subjects on the handheld device, and the schedule of assessments for informants on the handheld device are presented in Tables 6, 7 and 8 in the appendix.

3.3 Randomization

Eligible subjects will be randomized into one of three treatment groups (pimavanserin, quetiapine, or matching placebo) in a 1:1:1 ratio using an Interactive Response Technology (IRT) system. The assignments will be based on a pre-generated permuted-block randomization schedule.

3.4 Blinding

This is a double-blind study. Blinding is assured by restricting access of Investigators and Sponsor personnel and/or designee to the treatment codes, and providing identical encapsulated tablets and packaging for the pimavanserin, quetiapine, and placebo treatments.

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

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

3.5 Determination of Sample Size

This is a pilot study and is not powered for statistical significance. Up to 60 subjects will be randomized into the study.

3.6 Coronavirus Disease 2019

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

Due to COVID-19, Screening and enrollment into this study were temporarily halted in March 2020, and the decision was made to permanently stop the study as of 24 September 2020. The impact of COVID-19 on the statistical analysis is discussed in each of the relevant sections of this SAP.

4. ANALYSIS SETS

The Randomized Analysis Set will consist of all subjects who were randomized. Subjects will be classified according to the randomized treatment assignment.

The Safety Analysis Set will consist of all subjects who have taken at least 1 dose of study drug. Subjects will be classified according to the actual treatment received. Safety analyses will be based on the Safety Analysis Set.

The Exploratory Analysis Set will consist of randomized subjects in the Safety Analysis Set excluding subjects who were replaced due to no increase of the dose of blinded study drug consistent with 50 mg quetiapine at Week 1 (Visit 3). Subjects will be classified according to the randomized treatment assignment. Exploratory efficacy analyses will be based on the Exploratory Analysis Set.

5. DATA HANDLING CONVENTIONS

All data collected in the study will be listed, excluding actigraphy data or data collected on the Handheld Device.

5.1 General Data Reporting Conventions

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

Categorical and count variables will be summarized by the number of subjects and the percent of subjects in each category. Categories with zero counts will not have zero percentages displayed. Percentages will be presented with one decimal place.

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

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

5.2 Derived Efficacy Variables

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

5.2.1 Mini-Mental State Examination (MMSE)

MMSE is assessed at Screening. The MMSE is an 11-area, 30 items questionnaire that is used to measure cognitive impairment, with lower scores indicating more severe cognitive impairment. The total score (0-30) is calculated as the sum of the 30 item scores. The total score will be missing if there are at least one missing items. The MMSE is being used in this study to screen for cognitive impairment.

5.2.2 Neuropsychiatric Inventory (NPI)

For subjects enrolled under the Original Protocol or Protocol Amendments 1 or 2, the NPI is assessed at Screening. At least one of the following domains must have a Frequency×Severity greater than or equal to 3: Domain A Delusions, Domain B Hallucinations, Domain D

Depression/Dysphoria, Domain G Apathy/Indifference, Domain H Disinhibition, or Domain I Irritability/Lability. The NPI was removed from the list of assessments in Protocol Amendment 3.

Domain scores will not be derived. Scores, as reported in RAVE, will be listed.

5.2.3 Patient-Reported Outcomes Measurement Information System (PROMIS®)

Data will be collected by the Handheld Device and the analysis method will not be covered by this SAP.

5.2.4 Sleep-Related Impairment Scale

Data will be collected by the Handheld Device and the analysis method will not be covered by this SAP.

5.2.5 Nighttime Sleep Scale

Data will be collected by the Handheld Device and the analysis method will not be covered by this SAP.

5.2.6 Trail Making Test (TMT) Parts A and B

The TMT is a neuropsychological test of visual attention and task switching. The TMT has two parts, Parts A and B. Part A measures rote memory and consists of 25 circles on a piece of paper with the numbers 1-25 written randomly in the circles. The subject is to connect the circles in numerical order. Part B measures executive functioning and consists of 24 circles on a piece of paper, half of the circles have the numbers 1-12 in them and the other half (12) contain the letters A-L. The subject must alternate connecting circles with numbers and circles with letters. Time to completion or maximum time (seconds) for Parts A and B are recorded. The maximum completion time for Part A is 150 seconds, and 300 seconds for Part B. Higher completion or maximum time denote slower visual attention and task switching. Missing values will not be imputed.

5.2.7 Hopkins Verbal Learning Test-Revised (HVLT-R)

The HVLT is a brief verbal learning and memory test with six alternate forms. The HVLT is ideal in situations calling for repeated neuropsychological examinations, but it lacks a delayed recall trial which is essential for the assessment of abnormal forgetting and therefore the revised version of the HVLT (HVLT-R) was developed. The HVLT-R test consists of three trials of free-recall of a 12-item, semantically categorized list, followed by yes/no recognition. Approximately 20-25 min later, a delayed recall trial (trial 4) and a delayed recognition trial are completed.

Total recall will be derived as the sum of total correct responses for Trials 1, 2, and 3. Total recall will be set to missing if all 3 trials are not completed.

Retention (%) will be derived as (Trial 4 score ÷ Higher score of Trials 2 and 3) × 100.

The Delayed Recognition Trial will be used to derive the recognition discrimination index, which will be calculated as the total number of true-positive responses minus the total number of false-positive errors (ranging from -12 to 12). Higher indexes denote better verbal learning and memory test results. Missing total numbers of true-positive responses and/or false-positive errors will not be imputed. The index will be missing if total number of true-positive responses and/or total number of false-positive errors are missing.

5.2.8 Parkinson's Disease Non-motor Symptoms (PD NMS) Questionnaire

The Non-motor Symptoms questionnaire is a validated, clinician rated questionnaire that assesses non-motor issues that may be present in Parkinson's disease. The PD NMS is comprised of 30 items grouped into nine domains: cardiovascular including falls, sleep/fatigue, mood/cognition, perceptual problems/hallucinations, attention/memory, gastrointestinal tract, urinary, sexual function, and miscellaneous.

Each item is scored with respect to frequency (1=rarely, 2=often, 3=frequent, 4=very frequent) and severity (0=none, 1=mild, 2=moderate, 3=severe). Individual item scores are calculated as frequency × severity scores. The total score, ranging from 0 to 360, is calculated as the sum of the 30 item scores. Higher total scores denote worse non-motor symptoms. If 6 or fewer item scores are missing, then the total score will be imputed as the mean of the non-missing item scores multiplied by 30 and rounded to the nearest integer. The total score will be missing if there are at least 7 missing item scores. Missing item scores will not be imputed.

5.2.9 Patient Global Impression – Improvement (PGI-I)

The PGI-I is a global index used to rate the response of a condition to a therapy. The PGI-I asks the patient to rate their symptoms now, as compared with how it was at Baseline before beginning treatment. Severity ratings ranging from 1 to 7 are based on the behavioral domain(s) of clinical concern:

1 = Very much better	4 = No change
2 = Much better	5 = A little worse
3 = A little better	6 = Much worse
	7 = Very much worse

Higher PGI-I scores denote worse symptoms compared with Baseline before beginning treatment. Missing PGI-I scores will not be imputed.

Beginning with Protocol Amendment 3, rather than being assessed on a single global scale as described above, subjects were assessed on the following symptom-specific PGI-I scales:

delusions, hallucinations, depression/dysphoria, apathy/indifference, disinhibition, irritability/lability. For those subjects assessed with symptom-specific PGI-I scales, a single, global score will be derived by taking the arithmetic mean of the non-missing symptom-specific scores at the given visit, rounded to the nearest integer.

5.2.10 Clinical Global Impression – Severity (CGI-S)

The CGI-S scale is a clinician-rated, 7-point scale that is designed to rate the severity of the subject's neuropsychiatric symptoms at the time of assessment using the Investigator's judgment and past experience with subjects who have the same disorder. The CGI-S will be assessed at Screening and at Baseline. Severity ratings ranging from 0 to 7 are based on the behavioral domain(s) of clinical concern used for eligibility:

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

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

Beginning with Protocol Amendment 3, rather than being assessed on a single global scale as described above, subjects were assessed on the following symptom-specific CGI-S scales: delusions, hallucinations, depression/dysphoria, apathy/indifference, disinhibition, irritability/lability. For those subjects assessed with symptom-specific CGI-S scales, a single, global score will be derived by taking the arithmetic mean of the non-missing symptom-specific scores at the given visit, rounded to the nearest integer.

5.2.11 Clinical Global Impression – Improvement (CGI-I)

The CGI-I is a clinician-rated, 7-point scale that is designed to rate the improvement in the subject's symptoms at the time of assessment, relative to the symptoms at Baseline. Severity ratings ranging from 0 to 7 are based on the behavioral domain(s) of clinical concern used for eligibility:

0 = Not assessed	4 = No change
1 = Very much improved	5 = Minimally worse
2 = Much improved	6 = Much Worse
3 = Minimally improved	7 = Very much worse

Higher CGI-I scores denote less improvement in illness. Missing CGI-I scores will not be imputed.

Beginning with Protocol Amendment 3, rather than being assessed on a single global scale as described above, subjects were assessed on the following symptom-specific CGI-I scales: delusions, hallucinations, depression/dysphoria, apathy/indifference, disinhibition, irritability/lability. For those subjects assessed with symptom-specific CGI-I scales, a single, global score will be derived by taking the arithmetic mean of the non-missing symptom-specific scores at the given visit, rounded to the nearest integer.

5.2.12 Patient and Informant Satisfaction Question

Data will be collected by the Handheld Device and the analysis method will not be covered by this SAP.

5.3 Analysis Visit Windows

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

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

Table 1 Analysis Visit Windows

Analysis Visit	Study Visit	Target Study Day @	Study Day Interval
Baseline (Day 1)	2 (Week 0)	1	≤ 1
Week 1	3 (Week 1)	8	2 – 11
Week 2	4 (Week 2)	15	12 – 18
Week 3	5 (Week 3)	22	19 – 25
Week 4	6 (Week 4)	29	26 – 32
Safety Follow-up	7 (Week 8)	59	≥ 33

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

5.3.1 Unscheduled Assessments

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

5.3.2 Multiple Measurements within Visit Windows

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

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

5.4 Missing or Incomplete Date for Last Dose of Study Drug

In the Safety Analysis Set, if the last dose date of study drug is missing for a subject who completed or early terminated from the study, then it will be imputed using the last expected dosing date (LED), defined as the earlier of the following three dates: the date of the end-of-treatment (EOT)/early termination (ET) visit, the last expected dosing date per protocol, and the kit return date of the last dispensed drug kit. For the incomplete last dose date of study drug, the imputation algorithms will be detailed in the analysis dataset specification document. The missing or incomplete dates will be displayed in the data listings as reported on the eCRF rather than the imputed dates.

5.5 Missing or Incomplete Dates for Prior or Concomitant Medications

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

5.6 Missing or Incomplete Date for Adverse Events

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

5.7 Missing Severity Assessment for Adverse Events

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

5.8 Missing Relationship to Study Drug for Adverse Events

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

5.9 Character Values of Clinical Laboratory Variables

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

6. SUBJECT DISPOSITION

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

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

For randomized subjects, the number and percentage of subjects in the Safety Analysis Set and Exploratory Analysis Set will be summarized by treatment group and overall. A listing will be provided displaying all subjects excluded from the Safety or Exploratory 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 reasons including the reason due to COVID-19) will also be summarized by treatment group and overall.

7. PROTOCOL DEVIATIONS

Protocol deviations will be reviewed periodically over the course of the 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. Protocol deviations will also be assessed with respect to relationship to COVID-19.

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

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

8. DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS

Demographics and baseline characteristics will be summarized by treatment group and overall for the Randomized and Safety analysis sets, using descriptive statistics. Variables include age, sex, race, ethnicity, height, weight, and BMI.

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

Parkinson's disease history will be summarized by treatment and overall for the Randomized and Safety analysis sets using descriptive statistics. Variables include Age at Parkinson's disease diagnosis, prior antipsychotic use (yes/no), and presence of insomnia (yes/no).

9. MEDICAL HISTORY

Medical history reported terms will be coded with 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 by treatment group and overall for the Safety 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 for the medical history condition/event, start and stop dates (when available), and an indicator for whether or not the condition is ongoing will be provided.

10. EXTENT OF EXPOSURE AND TREATMENT COMPLIANCE

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

10.1 Exposure to Study drug

Duration of exposure to study drug will be calculated for each subject as (last dose date – first dose date + 1). The number and percentage of subjects within each of the following exposure levels in terms of duration of exposure will also be tabulated: <1 week (1 to 6 days), 1 to <2 weeks (7 to 13 days), 2 to <3 weeks (14 to 20 days), 3 to <4 weeks (21 to 27 days), and >4 weeks (28 days or longer).

10.2 Measurement of Treatment Compliance

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

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

11. PRIOR, CONCOMITANT, AND POST-TREATMENT MEDICATION

Prior medication is defined as any medication with the start and stop dates prior to the date of the first dose of study drug. Concomitant medication is defined as any medication with a start date prior to the date of the first dose of study drug and continuing past the first dose of study drug or with a start date between the dates of the first and last doses of study drug, inclusive. Any medication with a start date after the date of the last dose of study drug will be considered as post-treatment medication. Medications will be coded using WHO Drug Dictionary March 2018 or newer version. The number and percentage of subjects taking each drug class (ATC Level 3) and medication preferred term will be tabulated by treatment group and overall for the Safety Analysis Set. Multiple medication usage by a subject in the same category will be counted only once.

Prior and Concomitant Medications

Prior and concomitant medications will be summarized for the Safety Analysis Set by ATC Level 3, preferred term, and the study drug (placebo, quetiapine, and pimavanserin).

Post-Treatment Medications

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

COVID-19 Related Medications

Relationship to COVID-19 will be assessed for selected medications as detailed in the GSD. COVID-19 related concomitant medications will be listed separately.

12. EFFICACY ANALYSES

All efficacy analyses will be performed using the planned treatment assignments based on the randomization schedule for the Exploratory Analysis Set.

12.1 Efficacy Variables

Exploratory Efficacy Endpoints Measured at Study Visits

- Change from Baseline to Week 4 in
  - Time to completion or maximum time of TMT Parts A and B,
  - HVLT-R recognition discrimination index, and
  - PD NMS total score
- CGI-I score at Week 4
- PGI-I score at Week 4

12.2 Adjustment for Covariates

The corresponding baseline value will be included as a covariate for the analysis of time to completion or maximum time of TMT Parts A and B, HVLT-R recognition discrimination index, and PD NMS total score. Baseline CGI-S score will be used as a covariate for the CGI-I analysis.

12.3 Handling of Missing Data

Missing data will not be imputed in the analysis of exploratory efficacy endpoints.

12.4 Multiple Comparisons / Multiplicity

Multiplicity will not be adjusted in the analysis of exploratory efficacy endpoints due to the exploratory nature of this trial.

12.5 Examination of Subgroups

No subgroup analyses are planned.

13. METHODS OF EFFICACY ANALYSES

The change from Baseline in the exploratory efficacy endpoints measured at scheduled visits including time to completion or maximum time of TMT Parts A and B, HVLT-R recognition discrimination index, and PD NMS total score will be analyzed using an analysis of covariance model (ANCOVA) using observed scores. The model will include treatment group as a factor and the corresponding Baseline measure as a covariate.

Summary statistics (observed and change from Baseline), LS means, the between-group differences (Pimavanserin – Placebo and Quetiapine – Placebo) in LS means with the corresponding 95% confidence intervals, p-values, and the effect sizes (Cohen's d) will be presented at each post-Baseline visit for the Exploratory Analysis Set.

CGI-I score will be analyzed using a similar ANCOVA model with treatment group as a factor and Baseline CGI-S score as a covariate.

PGI-I score will be analyzed using an analysis of variance model with treatment group as a factor.

14. SAFETY ANALYSES

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

14.1 Adverse Events

The primary endpoint for this study is treatment-emergent adverse events (TEAEs). Adverse events will be coded using MedDRA dictionary, Version 23.0.

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

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

The relationship of selected AEs to COVID-19 will be assessed as detailed in the GSD. COVID-19 related adverse events will be listed separately.

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

The incidence of most frequently reported (preferred terms reported by 2 or more subjects) TEAEs, treatment emergent SAEs, and TEAEs leading to discontinuation of study drug will be summarized by SOC, preferred term, and treatment group. The tables will be sorted alphabetically by SOC and then by descending frequency within each SOC. In addition, the incidence of fatal treatment-emergent AEs (i.e., events that cause death) will be summarized separately by preferred term and treatment group.

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

14.2 Clinical Laboratory Variables

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

Clinical laboratory assessments are performed at Screening Visit 1, Baseline (Week 0), Visit 4 (Week 2), and Visit 6 (Week 4/EOT/ET).

- Clinical chemistry serum tests include the following:
    - Sodium (Na), potassium (K), chloride (Cl), phosphorus (P), calcium (Ca), magnesium (Mg), carbon dioxide (CO_2), blood urea nitrogen (BUN), creatinine (CR), uric acid
        - Mg is only performed at Visit 1 (Screening)
    - Alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma-glutamyl transpeptidase (GGT), alkaline phosphatase (ALP), total bilirubin (TBIL), lactate dehydrogenase (LDH)
    - Vitamin B12 (Screening only)
    - HbA1c (Screening only)
    - Glucose
    - Albumin (ALB), total protein
    - Prolactin (Week 4/EOT will be blinded, also performed at Baseline)
    - Thyroid stimulating hormone (TSH) and reflex free T4, if TSH is out of range
        - TSH/free T4 is only performed at Visit 1 (Screening)
    - Creatine kinase (CK)/creatine phosphokinase (CPK)
    - Insulin (at Baseline and Week 4 only)
    - Lipid panel
        - Total cholesterol, HDL-cholesterol, triglycerides, LDL-cholesterol, Cholesterol/HDL ratio, Non-HDL cholesterol, very low density lipoprotein cholesterol
- Hematology tests include the following:
    - Complete blood count (CBC) including:

- White blood cell (WBC) count
- Absolute neutrophil count
- Complete differential (relative and absolute)
- Hematocrit (Hct), hemoglobin, red blood cells (RBC), platelets
- Reticulocyte count

- Urinalysis tests include the following:
  - Blood, RBCs, WBCs, protein, glucose, ketones, specific gravity, pH

Clinical laboratory values (in Système International [SI] units) and the change from Baseline values will be summarized by treatment group at Weeks 2 and 4(EOT/ET) using descriptive statistics. The overall minimum, maximum and last post-Baseline observed and change from Baseline values will also be summarized. For hemoglobin, hematocrit and uric acid, the above summaries will be presented for each gender as well as for both genders combined. For urinalysis with categorical results, the number and percentage of subjects will be tabulated by category at Baseline, Weeks 2 and 4 (EOT/ET); and the denominator is the number of subjects with non-missing values for the given parameter, visit and treatment group.

Laboratory values will also be summarized in shift tables by treatment group, to determine the number and percentage of subjects with values classified as below, within, and above normal ranges at Weeks 2 and 4 (EOT/ET) 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 and treatment group. For the shift to the overall post-Baseline minimum or maximum, all post-Baseline values will be considered, including unscheduled and out of window values and the denominator is the number of subjects with non-missing Baseline value and at least 1 post-Baseline value for the given parameter and treatment group. For hemoglobin, hematocrit and uric acid, the shift summaries will be presented for each gender as well as for both genders combined.

Clinical laboratory values are potentially clinically important (PCI) if they meet either the low or high PCI criteria listed in Tables 2 and 3. The number and percentage of subjects with post-Baseline PCI values for each of the categories in Table 7 and 8 will be summarized by treatment group for selected parameters. For the overall post-Baseline summaries of PCI values, all post-Baseline values will be considered, including unscheduled and out of window values. Subjects with multiple PCI values for a given parameter will be counted only once for that parameter. For the overall post-Baseline summary, the numerator of the percentage is the number of subjects with at least 1 post-Baseline PCI laboratory value for the given parameter and treatment group,

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

Table 2 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

Table 2 Criteria for Potentially Clinically Important Laboratory Values – Hematology and Chemistry (Continued)

Analyte	Conventional Unit	Low PCI Criteria	High PCI Criteria	SI Unit	Low PCI Criteria	High PCI Criteria
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 3 Criteria for Potentially Clinically Important Laboratory Values – Urinalysis

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

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

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

14.3 Vital Signs and Weight

Vital signs will be collected throughout the study. Weight and BMI will be observed and/or derived at Screening Visit 1, Baseline (Week 0) and Visit 6 (Week 4/EOT/ET). Observed values and the changes from Baseline at each post-Baseline visit will be summarized by treatment group using descriptive statistics.

Vital sign and weight values will be considered PCI if they meet the criteria listed in Table 4. The number and percentage of subjects with post-Baseline vital signs and weight that are PCI will be

summarized by treatment group at each post-Baseline visit and for overall post-Baseline. For the overall post-Baseline summaries, all post-Baseline values will be considered, including unscheduled and out of window values. Subjects with multiple PCI values for a given parameter will be counted only once for that parameter. For the by-visit summary, the numerator for the percentage is the number of subjects with a post-Baseline PCI value for the given parameter, visit and treatment group, and the denominator is the number of subjects with non-missing values for the given parameter, visit and treatment group. For the overall post-Baseline summary, the numerator for the percentage is the number of subjects with at least 1 post-Baseline PCI value for the given parameter and treatment group, and the denominator is the number of subjects with at least 1 post-Baseline value for the given parameter and treatment group. A listing of subjects with any PCI value will be provided.

Table 4 Criteria for Potentially Clinically Important (PCI) Vital Signs and Weight

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

14.4 Electrocardiogram (ECG)

12-lead ECGs are collected throughout the study. Observed values of ECG variables (e.g., heart rate, PR interval, QRS interval, QT interval, and QTc interval) and the changes from Baseline at each assessment time point will be summarized by treatment group.

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

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

Electrocardiogram variable values will be considered PCI if they meet the criteria listed in Table 5. The number and percentage of subjects with post-baseline PCI values will be summarized by treatment group at each post-Baseline visit and for overall post-Baseline. For the by-visit summary, the numerator for the percentage is the number of subjects with a post-Baseline PCI ECG for the given parameter, visit and treatment group, and the denominator is the number of subjects with non-missing values for the given parameter, visit and treatment group. For the overall post-Baseline summary, the numerator for the percentage is the number of subjects with at least 1 post-Baseline PCI ECG for the given parameter and treatment group, and the denominator is the number of subjects with at least 1 post-Baseline ECG value for the given parameter and treatment group. A listing of all subjects with any PCI value will be provided.

Table 5 Criteria for Potentially Clinically Important ECG Values

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

Cardiologist interpretations (normal and abnormal) will be summarized in a frequency table by treatment group and visit and for overall post-Baseline.

14.5 Physical Examination

Physical examinations are performed at Screening Visit 1, Baseline (Week 0), and Week 4 (EOT/ET). Physical examination results (normal, abnormal, and not done) will be summarized in a frequency table by treatment group, body system and visit.

14.6 Other Safety Endpoints

Columbia-Suicide Severity Rating Scale (C-SSRS)

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

There are 5 questions about suicidal ideation, representing 5 types of suicidal ideation: wish to be dead; non-specific active suicidal thoughts; active suicidal ideation with any methods (not plan) without intent to act; active suicidal ideation with some intent to act, without specific plan; active suicidal ideation with specific plan and intent. If a subject answers “yes” to any of these

5 questions at any post-Baseline visit including unscheduled and out of window visits, this subject will be counted as having suicidal ideation in this study.

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

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

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

Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) Part III (motor examination)

The MDS-UPDRS is a comprehensive battery of motor and behavioral indices derived from the Columbia Scale. The MDS-UPDRS Part III will be used to assess motor function and is a motor examination consisting of 18 rater based items plus Hoehn and Yahr scale (describing how the symptoms of Parkinson's disease progress). Each item rating ranges from 0 to 4:

0: Normal; 1: Slight; 2: Mild; 3: Moderate; 4: Severe

Some of the 18 items have subitems. In these cases, the item scores are defined as the sum of the subitem scores. If any of the subitem scores are missing then the corresponding item score will be missing.

The following item scores will be derived as the sum of the Right and Left subitem scores: finger tapping, hand movements, pronation-supination movement of hands, toe tapping, leg agility, postural tremor of the hands, and kinetic tremor of the hands.

The rigidity item score is derived as the sum of the neck, right upper extremity, left upper extremity, right lower extremity, and left lower extremity subitem scores.

The rest tremor amplitude item score is derived as the sum of the lip/jaw, right upper extremity, left upper extremity, right lower extremity, and left lower extremity subitem scores.

The Hoehn and Yahr scale is a single item question that is used to assess the stage of Parkinson's disease; the score ranges from 0 to 5:

0: Asymptomatic

1: Unilateral involvement only

2: Bilateral involvement without impairment of balance

3: Mild to moderate involvement; some postural instability but physically independent; needs assistance to recover from pull test

4: Severe disability; still able to walk or stand unassisted

5: Wheelchair bound or bedridden unless aided

The assessment is to be completed at the clinic visits at Baseline (Week 0) and Week 4 (EOT/ET).

The observed and change from Baseline values will be summarized. The change from Baseline to Week 4 in the score for each of the 18 items as well as Hoehn and Yahr Stage score will be analyzed using an ANCOVA model with treatment group as a factor and the Baseline value as a covariate. Missing data will not be imputed.

15. CLINICAL PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES

No pharmacokinetic samples are collected in this study.

16. INTERIM ANALYSIS

No interim analysis is planned in this study.

17. DATA MONITORING/REVIEW COMMITTEE

There is no data monitoring/review committee in this study.

18. COMPUTER METHODS

Statistical analyses will be performed using Version 9.4 (or newer) of SAS® on a suitably qualified and validated environment.

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

19. CHANGES TO ANALYSES SPECIFIED IN PROTOCOL

Because this trial was stopped early after randomizing only 11 subjects, all of the mixed-model repeated measures analyses were changed to ANCOVA in order to minimize possible non-convergence issues.

20. APPENDICES

20.1 Summary of Version Changes

Version No:	Document History Description of Update	Author(s)	Version Date
1.0	Original version	[redacted]	19 October 2020

20.2 Supplementary Tables

Table 6 Schedule of Events and Assessments

	Screening	Baseline	Double-blind Treatment Period				Safety Follow-up
Visit Week (Study Day)	**-4**	**0**	**1 (Day 7)**	**2 (Day 14)**	**3 (Day 21)**	**4/EOT (Day 28)**	**8 [l] (Day 58)**
Visit Number	**1**	**2**	**3**	**4**	**5**	**6/ET**	**7**
Visit window (days)	**0**	**0**	**±3**	**±3**	**±3**	**±3**	**±4**
Type of Visit	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Telephone
Informed consent	X						
Inclusion/exclusion criteria	X	X					
Medical history and demographics	X						
Psychiatric and neurological history	X						
Physical examination (including neurologic examination)	X	X				X	
Training on the use of the handheld device and the actigraph[a]	X	X					
Orthostatic vital signs	X	X	X	X	X	X	
Weight and BMI[b]	X	X				X	
Height	X						
12-lead ECG[c]	X	X	X	X	X	X	
Clinical laboratory tests	X	X		X		X	
Pregnancy test[d]	X	X				X	
Urine drug screen	X	X[k]				X	
Capsule swallowing test[e]	X						
MMSE	X						
Dispense study drug[g]		X	X	X	X		
CGI-S[f]	X	X					
Trail Making Test[f]		X	X	X		X	
HVLT-R[f]		X	X	X		X	
Handheld device questionnaires[h]		X---------X					
Actigraphy data collection[i]		X---------X					
PD NMS Questionnaire[f]		X		X	X	X	
PGI-I[f]			X	X	X	X	
CGI-I[f]			X	X	X	X	
MDS UPDRS[f]		X				X	
Subject and Informant satisfaction question			X	X	X	X	
Concomitant medications	X	X	X	X	X	X	X
C-SSRS[f,j]	X	X	X	X	X	X	

	Screening	Baseline	Double-blind Treatment Period				Safety Follow-up
Visit Week (Study Day)	**-4**	**0**	**1 (Day 7)**	**2 (Day 14)**	**3 (Day 21)**	**4/EOT (Day 28)**	**8 [l] (Day 58)**
Visit Number	**1**	**2**	**3**	**4**	**5**	**6/ET**	**7**
Visit window (days)	**0**	**0**	**±3**	**±3**	**±3**	**±3**	**±4**
Type of Visit	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Telephone
Assessment of adverse events	X	X	X	X	X	X	X
Study drug accountability			X	X	X	X	

Abbreviations: CGI-I=Clinical Global Impression-Improvement; CGI-S=Clinical Global Impression-Severity; C-SSRS=Columbia-Suicide Severity Rating Scale; ECG=electrocardiogram; EOT=end of treatment; ET=early termination; HVLT-R=Hopkins Verbal Learning Test-Revised; MDS UPDRS=Movement Disorders Society Unified Parkinson's Disease Rating Scale; MMSE=Mini-Mental State Examination; PD NMS=Parkinson's Disease Non-motor Symptoms; PGI-I=Patient Global Impression-Improvement

[a] Training on the handheld device and actigraph may be done on the first day of Screening but must take place at least 7 days before the Baseline visit.

[b] Measurement of weight and BMI are required at Screening and Baseline; measurement of weight (not BMI) is required at Week 4/EOT.

[c] The ECG will be completed in triplicate at Visit 1 (Screening) and collected within an approximately 5-minute period. A single ECG tracing should be completed at all other designated visits.

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

[e] Subjects will be assessed for their ability to swallow a test capsule (i.e., placebo).

[f] This clinician-administered assessment should be conducted at least one hour after dosing when conducted at Weeks 1-4. The Baseline clinician-administered assessments should be conducted prior to dosing.

[g] The site staff must observe the subject for a minimum of 60 minutes in the clinic after dosing. Subjects should not take study drug on the mornings of clinic visit days until instructed to do so by the site staff.

[h] Baseline handheld device data will be collected during the 7-day period before the Baseline visit. Assessments completed on the handheld device should not be done on the days of clinic visits. See Table 2 and Table 3 for further details on the assessments for the handheld device.

[i] Baseline actigraphy data will be collected during the 7-day period before the Baseline visit. The handheld device is to be brought with the subject for the clinic visits at Visit 2 (Baseline), Visit 3 (Week 1), Visit 4 (Week 2), Visit 5 (Week 3), and Visit 6 (Week 4). Site staff are to review completion of assessments on the handheld device and use of actigraph at Visit 3 (Week 1), Visit 4 (Week 2), Visit 5 (Week 3), and Visit 6 (Week 4).

[j] The Baseline/Screening version of the C-SSRS will be administered at Screening, and the "Since Last Visit" version of the C-SSRS will be administered at all other designated visits.

[k] Urine drug screen at Visit 2 (Baseline) can be performed at the clinical site for rapid results to qualify the subject for study drug administration or exclude the subject without dosing.

[l] The safety follow-up visit is to occur 30 (±4) days after the last dose of study drug.

Table 7 Schedule of Assessments for Subjects on the Handheld Device (EMAs)

	Three Times Each Week				Weekly
	Sleep Assessments			PROMIS Cognitive Function	Satisfaction Question
	Nighttime Sleep[b]	Sleep-Related Impairment	Karolinska Sleepiness Scale (KSS)		
9 am ± 2 hours[a]	X	X	X	X	X
2 pm ± 2 hours		X	X	X	
7 pm ± 2 hours		X	X	X	

Abbreviations: PROMIS=Patient-Reported Outcomes Measurement Information System

Note: Assessments completed on the handheld device should not be done on the days of clinic visits. **It is recommended that assessments on the handheld device be done on the first, third and fifth days following clinic visits** (except Visit 6, the last visit); however, flexibility is permitted as long as assessments are completed 3 times each week.

[a] Testing should be completed at least 1 hour after the daily administration of study drug. If subject does not complete morning assessment, then no assessments can be completed that day.

[b] Questions related to nighttime sleep will be asked only at the morning assessment.

Table 8 Schedule of Assessments for Informants on the Handheld Device (EMAs)

	Three Times Each Week			Weekly
	Sleep Assessments		PROMIS Cognitive Function	Satisfaction Question
	Nighttime Sleep[b]	Sleep-Related Impairment		
9 am ± 2 hours[a]	X	X		X
7 pm ± 2 hours		X	X	

Abbreviations: PROMIS=Patient-Reported Outcomes Measurement Information System

Note: Assessments completed on the handheld device should not be done on the days of clinic visits. **It is recommended that assessments on the handheld device be done on the first, third and fifth days following clinic visits** (except Visit 6, the last visit); however, flexibility is permitted as long as assessments are completed 3 times each week.

[a] Testing should be completed at least 1 hour after the subject's daily administration of study drug.

[b] Questions related to nighttime sleep will be asked only at the morning assessment.