Official Title: A Phase 3b, Multicenter, Randomized, Double-blind, Placebo-controlled, Safety Study of Pimavanserin Therapy in Adult and Elderly Subjects Experiencing Neuropsychiatric Symptoms Related to Neurodegenerative Disease

NCT Number: NCT03575052

Document Date: 23 May 2022

STATISTICAL ANALYSIS PLAN

Protocol No.:	ACP-103-046
Protocol Title:	A Phase 3b, Multicenter, Randomized, Double-blind, Placebo-controlled, Safety Study of Pimavanserin Therapy in Adult and Elderly Subjects With Neuropsychiatric Symptoms Related to Neurodegenerative Disease
Drug:	Pimavanserin
Sponsor:	Acadia Pharmaceuticals Inc. [redacted]
Version No. and Date	Version 4.0, 23 May 2022

SIGNATURE/APPROVAL PAGE

AUTHOR

See appended electronic signature page

[redacted] Date
Director, Biostatistics
Acadia Pharmaceuticals Inc.

APPROVERS

See appended electronic signature page

[redacted] Date
Director, Clinical Research
Acadia Pharmaceuticals Inc.

See appended electronic signature page

[redacted] Date
Executive Director of Biostatistics, Biometrics
Acadia Pharmaceuticals Inc.

Signature Page for ACP-103-046 Statistical Analysis Plan v4

Approve	[redacted] 26-May-2022 13:18:41 GMT+0000

Approve	[redacted] 26-May-2022 15:26:05 GMT+0000

Approve	[redacted] 27-May-2022 13:40:29 GMT+0000

Signature Page for ACP-103-046 Statistical Analysis Plan v4

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

AD	Alzheimer's disease
ADP	Alzheimer's disease psychosis
AE	adverse event
ANCOVA	analysis of covariance
ATC	Anatomical Therapeutic Chemical
BMI	body mass index
CGI-I	Clinical Global Impression – Improvement
CGI-S	Clinical Global Impression – Severity
CI	confidence interval
C-SSRS	Columbia-Suicide Severity Rating Scale
COVID-19	coronavirus disease 2019
DMP	Data Management Plan
DRP	dementia related psychosis
DSMB	Data and Safety Monitoring Board
ECG	Electrocardiogram
eCRF	electronic case report form
EOS	End-of-Study
EQ-5D-5L	5-level EuroQol-5D
ESRS-A	Extrapyramidal Symptom Rating Scale-Abbreviated
ET	Early Termination
GCAS	Global Clinical Assessment of Suicidality
IA	interim analysis
ISG	Independent Statistical Group
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	mixed-model repeated measures
msec	Milliseconds
PCHC	Pareto Classification of Health Change
PCI	potentially clinically important
QTcB	QT interval corrected for heart rate using Bazett's formula
QTcF	QT interval corrected for heart rate using Fridericia's formula
SAE	serious adverse event
SAP	statistical analysis plan
SD	standard deviation
SDI	Sleep Disorders Inventory
SE	standard error

SOC	system organ class
TEAE	treatment-emergent adverse event
VAS	visual analogue scale

1 INTRODUCTION

This statistical analysis plan (SAP) provides a technical and detailed elaboration of the statistical analyses of efficacy and safety data as described in the study protocol amendment 7 dated 05 March 2021. Specifications for tables, figures, and listings are contained in a separate document.

In order to support potential regulatory submissions, the analysis plan was modified from that as specified in the Protocol by including formal statistical comparisons between treatment groups for selected efficacy and safety endpoints, and defining various subgroups. In addition, a factor for Baseline anti-dementia medication use was added for CGI-S and CGI-I analysis models.

For the Czech Republic, a country-specific protocol amendment (Amendment 7-CZ dated 23 July 2019) specifies additional exclusion criteria for tachycardia and blood pressure readings, as well as other procedural details.

For Bulgaria, a country-specific protocol amendment (Amendment 6-BG dated 23 July 2019) specifies that the designated study partner/caregiver must be in daily contact with the subject, as well as other procedural details.

2 OBJECTIVES

2.1 Primary Objective

The primary objective of this study is to assess the safety and tolerability of pimavanserin compared to placebo in adult and elderly subjects with neuropsychiatric symptoms related to neurodegenerative disease.

2.1.1 Primary Measure

The primary measure for this study is treatment-emergent adverse events (TEAEs).

2.2 Secondary Objectives

The secondary objective of this study is to assess the safety and tolerability of pimavanserin compared to placebo in adult and elderly subjects with neuropsychiatric symptoms related to neurodegenerative disease, with respect to extrapyramidal symptoms and cognition.

2.2.1 Secondary Safety Endpoints

The secondary endpoints for this study are as follows:

- Change from Baseline to Week 8 in Extrapyramidal Symptom Rating Scale-Abbreviated (ESRS-A)
- Change from Baseline to Week 8 in Mini-Mental State Examination (MMSE)

2.3 Exploratory Objectives

The exploratory objectives of this study are:

- To assess the safety and tolerability of pimavanserin compared to placebo in adult and elderly subjects with neuropsychiatric symptoms related to neurodegenerative disease, as described by suicidality
- To evaluate the benefit of pimavanserin compared with placebo in adult and elderly subjects with neuropsychiatric symptoms related to neurodegenerative disease

2.3.1 Exploratory Endpoints

The exploratory safety endpoint for this study to assess the safety and tolerability of pimavanserin is:

- Columbia-Suicide Severity Rating Scale (C-SSRS) or Global Clinician Assessment of Suicidality (GCAS) scale (if the subject is not able to complete the C-SSRS in the Investigator's judgement)

The exploratory efficacy endpoints for this study to assess the benefit of pimavanserin are:

- Change from Baseline to Week 8 in Clinical Global Impression-Severity (CGI-S) score for neuropsychiatric symptoms
- Clinical Global Impression-Improvement (CGI-I) score at Week 8 for neuropsychiatric symptoms
- Change from Baseline to Week 8 in the 5-level EuroQol-5D (EQ-5D-5L) scores, including both the Visual Analogue Scale (VAS) and the 5 level descriptive system
- Change from Baseline to Week 8 in sleep disturbances (assessed by the Sleep Disorder Inventory [SDI])

3 STUDY DESIGN

3.1 General Study Design

This study will be conducted as a Phase 3b, 8-week, randomized, double-blind, placebo-controlled, parallel group study in adult and elderly subjects with neuropsychiatric symptoms related to neurodegenerative disease.

Subjects will receive once daily (QD) doses of pimavanserin 34 mg (provided as 2 x 17 mg tablets) or matching placebo. During the Treatment Period, clinic visits will be conducted at Baseline and Weeks 1, 2, 4, 6, and 8, or upon early termination (ET) from the study.

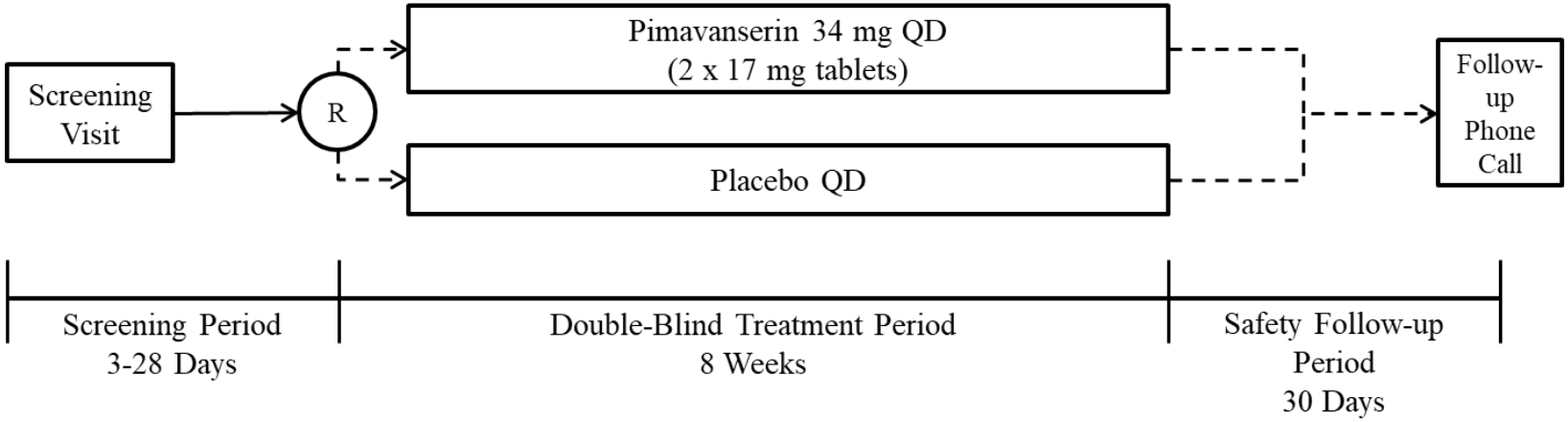
Study drug will be dispensed to the subject to take home at the Baseline visit and at each subsequent visit starting with Week 2. The subject and their study partner/caregiver will be

provided instructions for subject's first dose of study drug on the day after the Baseline visit. It is recommended that the subject take the study drug at approximately the same time each day as a single, oral dose.

A follow-up safety assessment will be conducted by telephone call at approximately 30 days after the last dose of study drug.

Figure 1 illustrates the study design.

Figure 1 Schematic of Study Design



Abbreviation: R=randomization
Note: Subjects who enroll in an open-label extension will not complete the Follow-up Safety Period.

3.2 Schedule of Assessments

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

Table 1 Schedule of Events and Assessments for ACP-103-046

Period	Screening	Baseline	Double-blind Treatment Period						Safety Follow-up[a]
Visit Day/Week	Day -28 to -3	Week 0	Week 1	Week 2	Week 4	Week 6	Week 8	Unscheduled Visit	Week 12
Visit Number	1	2	3	4	5	6	(EOT/ET) 7		8
Visit window (days)			±3	±3	±3	±3	±3		+4
Type of Visit	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Phone call
Informed consent	X								
Inclusion/exclusion criteria	X	X							
Medical history and demographics	X								
Psychiatric, dementia, and neurological history	X								
Physical examination	X	X			X		X		
Vital signs and weight	X	X	X	X	X	X	X	X	
Height	X								
12-lead ECG[b]	X	X		X			X		
Clinical laboratory tests[c]	X	X			X		X		
Pregnancy test[d]	X	X					X		
Urine drug screen	X				X		X		
MMSE	X	X	X	X	X	X	X		
NPI	X	X							
CGI-S	X	X	X	X	X	X	X	X	
CGI-I			X	X	X	X	X		
C-SSRS[e] or GCAS[e]	X	X	X	X	X	X	X	X	
ESRS-A		X	X	X	X	X	X	X	

Table 1 Schedule of Events and Assessments for ACP-103-046 (Continued)

Period	Screening	Baseline	Double-blind Treatment Period						Safety Follow-up[a]
Visit Day/Week	Day -28 to -3	Week 0	Week 1	Week 2	Week 4	Week 6	Week 8	Unscheduled Visit	Week 12
Visit Number	1	2	3	4	5	6	(EOT/ET) 7		8
Visit window (days)			±3	±3	±3	±3	±3		+4
Type of Visit	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Phone call
EQ-5D-5L		X					X		
SDI		X			X		X		
CT/MRI[f]	X								
Concomitant medications	X	X	X	X	X	X	X	X	X
Assessment of adverse events	X	X	X	X	X	X	X	X	X
Dispense study drug		X		X	X	X		X	
Study drug accountability				X	X	X	X	X	

Abbreviations: CGI-I=Clinical Global Impression-Improvement; CGI-S=Clinical Global Impression-Severity; CRF=case report form; C-SSRS=Columbia-Suicide Severity Rating Scale; CT=computed tomography; ECG=electrocardiogram; EOT=end of treatment; EQ-5D-5L=5-level EuroQol-5D; ESRS-A=Extrapyramidal Symptom Rating Scale-Abbreviated; ET=early termination; GCAS=Global Clinician Assessment of Suicidality; MMSE=Mini–Mental State Examination; MRI=magnetic resonance imaging; NPI=Neuropsychiatric Inventory; SDI=Sleep Disorders Inventory

[a] This visit is a safety follow-up telephone call visit for subjects who discontinue prematurely from the study or who do not participate in the long-term extension study. This visit will occur 30 (+4) days after the last dose of study drug.

[b] At Visit 1, ECG should be completed in triplicate. At all other visits, a single 12-lead ECG should be completed. These should occur before blood sampling or at least 30 minutes after blood sampling.

[c] To include hematology, serum chemistry, prolactin levels, and urinalysis. Urinalysis requirement is not applicable to subjects who are unable to provide urine sample (e.g., incontinent subjects). Where collection of a urine sample proves impractical or impossible, failure to collect a urine sample should be recorded in the subject's CRF, and will not be considered a protocol deviation.

[d] A pregnancy test is only required for women of childbearing potential. Serum pregnancy should only be performed at Visit 1; a urine pregnancy test should be performed at all subsequent visits. If urine cannot be obtained in women of childbearing potential, a serum pregnancy test should be done in its place.

[e] Suicidal assessment is required. For subjects who are able to complete the assessment, 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 subsequent visits. The C-SSRS is preferred. If the subject is not able to complete the C-SSRS in the Investigator's judgment, the GCAS should be administered and used thereafter in the study.

[f] A CT or MRI should only be completed if a CT or MRI scan has not been performed in the last 3 years prior to screening.

3.3 Randomization

At Baseline (Visit 2/Week 0), eligible subjects who meet all inclusion criteria and do not meet any exclusion criteria will be randomized in a 1:1 ratio to receive either pimavanserin 34 mg or matched placebo, using an interactive response technology system. Randomization will be stratified by geographic region (North America, Europe, or rest of world).

The treatment assignments will be based on a pre-generated permuted-block randomization schedule. Blinding will be assured by restricting assess of Investigators and Sponsor personnel and/or designee to the treatment codes, and providing identical tablets and packaging for the pimavanserin and placebo treatments.

3.4 Blinding

Treatment assignments will be blinded to all study subjects, study partners/caregivers, Investigators, raters, site personnel, and Sponsor personnel who oversee the study. See Section 16 regarding interim analysis (IA) unblinding in order to support safety and efficacy evaluations for regulatory submissions. Details regarding medical emergency unblinding procedures are provided in Section 9.8 of the ACP-103-046 protocol.

3.5 Determination of Sample Size

Approximately 1500 subjects will be screened to enroll approximately 750 subjects, allowing for a 50% screen failure rate. Approximately 375 subjects will be randomized to the pimavanserin group as well as the placebo group (i.e., 375 subjects in each treatment group). With an assumed dropout rate of 30% in the double-blind phase, 525 subjects in total are expected to complete the study. In the event that the dropout rate is higher than anticipated, more subjects may be enrolled to ensure that at least 500 subjects complete the study.

A sample size of 375 subjects in each treatment group ensures that within each treatment group, the probability of observing at least 1 AE with a true incidence of 1.0% is 0.977; the probability of observing at least 1 AE with a true incidence of 1.5% is 0.997; and the probability of observing at least 1 AE with a true incidence of 2.0% is 0.999.

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 Data Management Plan [DMP] Appendix B). Recruitment efforts into ACP-103-046, and roll-overs into the open-label extension study (ACP-103-047), were paused. The impact of COVID-19 on statistical analysis is discussed in each of the relevant sections of this SAP.

3.7 Geopolitical Conflict in Ukraine

Due to the ongoing geopolitical conflict in Ukraine, some sites in Ukraine may not be able to respond to all database queries, source data verification by clinical monitors may not be possible at specific sites, and Investigators may not be able to sign-off on subject eCRFs prior to database lock. However, all data will be included in the final analyses as specified in this SAP.

In order to assess the potential impact of any affected Ukraine data on the final analysis, all analyses will also be performed excluding all data for those subjects in Ukraine where the data has been impacted by any of the non-standard data scenarios noted above.

The study team will document all subject data impacted by the geopolitical conflict in Ukraine in a comprehensive note-to-file that will be included in the Trial Master File at study closure.

4 ANALYSIS SETS

4.1 Randomized Analysis Set

The Randomized Analysis Set will include all subjects who were randomized. Subjects will be classified according to their randomized treatment assignment.

4.2 Safety Analysis Set

The Safety Analysis Set will include all randomized subjects who took at least 1 dose of study drug. Subjects will be classified according to their actual treatment received.

4.3 Full Analysis Set

The Full Analysis Set will include all subjects in the Randomized Analysis Set who received at least one dose of study drug and have both a Baseline CGI-S and at least one post-Baseline value for either the CGI-I or CGI-S. Subjects will be classified according to the randomized treatment assignment.

5 DATA HANDLING CONVENTIONS

All data collected in the study will be listed.

5.1 General Data Reporting Conventions

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

standard deviations and standard errors will be presented to 2 more decimal places than the raw data. .

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

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

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

All safety endpoints will be summarized using the Safety Analysis Set. The Full Analysis Set will be used for summaries of all efficacy endpoints.

Descriptive summaries of all safety and efficacy endpoints will be provided.

5.2 Derived Variables

In general, all assessment scale 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 form (eCRF), the derived values will be used for all analyses. Both the raw and derived total scores will be presented in data listings.

5.2.1 Neuropsychiatric Inventory

The Neuropsychiatric Inventory (NPI) is assessed at Screening and Baseline.

The NPI was developed to assess psychopathology in dementia patients. The original NPI evaluated 10 neuropsychiatric disturbances common in dementia: delusions (domain A), hallucinations (domain B), agitation/aggression (domain C), depression/dysphoria (domain D), anxiety (domain E), elation/euphoria (domain F), apathy/indifference (domain G), disinhibition (domain H), irritability/lability (domain I), and aberrant motor behavior (domain J). Two additional domains, sleep (domain K) and appetite and eating disorders (domain L), were subsequently added to the NPI. At Screening, the degree of the patient's neuropsychiatric symptoms will be evaluated using several of the domains of the NPI (delusions, hallucinations, depression/dysphoria, apathy/indifference, disinhibition, irritability/lability, and sleep). Patients with at least one individual domain score (frequency ×

severity) greater than or equal to 4 at Screening and Baseline (the domains at Screening and Baseline can be different) will be eligible for the study.

Individual domain total scores are calculated as frequency × severity. Missing scores will not be imputed. Thus, if either the frequency or severity is missing for a given domain then the domain total score will be missing.

Each domain has an associated caregiver distress question, rated on a 5 point scale from 0=no distress to 5=very severe or extreme. The caregiver distress items will not be summarized but will be displayed in data listings.

5.2.2 Clinical Global Impression Scale – Severity

The CGI-S is assessed at Screening, Baseline, Weeks 1, 2, 4, 6, and Week 8/ET visits.

The CGI-S 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 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 ill; 3=mildly ill; 4=moderately ill; 5=markedly ill; 6=severely ill; 7=among the most extremely ill. Higher scores denote more severe neuropsychiatric symptoms.

Missing CGI-S scores will not be imputed.

5.2.3 Clinical Global Impression Scale – Improvement

The CGI-I is assessed at Weeks 1, 2, 4, 6, and Week 8/ET visits.

The CGI-I is a clinician-rated, 7-point scale that is designed to rate the improvement in the subject's neuropsychiatric symptoms 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 neuropsychiatric symptoms.

Missing CGI-I scores will not be imputed.

5.2.4 EQ-5D-5L Proxy Version 1

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

The EQ-5D-5L is a standardized instrument used as a measure of health outcome. The EQ-5D-5L Proxy version 1 will be used in this study. In this version, a caregiver (the proxy) is asked to rate the subject's health-related quality of life in their (the proxy's) opinion. It consists of 2 components: the descriptive system and the visual analogue scale (VAS).

The EQ-5D-5L descriptive system consists of 5 dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression) each of which has 5 potential responses. The

responses record 5 levels of severity within a particular EQ-5D-5L dimension: 1=no problems; 2=slight problems; 3=moderate problems; 4=severe problems; 5=extreme problems. The responses for the 5 dimensions are combined into a single 5-digit code that describes the subject's health state.

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

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

5.2.5 Sleep Disorders Inventory

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

The SDI is an expanded version of one item of the Neuropsychiatric Inventory (NPI). It consists of the 7 subquestions from the NPI sleep disturbance item. Each of the subquestions was made into a separate question with frequency, severity, and caregiver distress rated by the caregiver with respect to the patient-participant for the 2 weeks prior to the visit. Thus, in contrast to a single rating for frequency and severity for all sleep disturbance-related behaviors, which would be incorporated into an overall NPI score, the SDI score is derived after the caregiver rates the frequency and severity of each of the 7 separate sleep disturbance symptoms. Caregiver distress ratings are not part of the SDI total score, but distress is measured. The SDI total score is calculated as the average of seven frequency ratings × average of seven severity ratings; thus, the total score has a range of 0-12 (higher scores indicate more severe sleep disorder).

If one frequency rating and/or one severity rating is missing then the total score will be imputed as the average of the non-missing frequency ratings × the average of the non-missing severity ratings. Otherwise, the total score will not be derived.

5.2.6 Columbia-Suicide Severity Rating Scale

The C-SSRS will be used to assess suicidal ideations and behaviors in subjects who are able to complete the assessment (subjects without dementia, defined as an MMSE≥25, and subjects with dementia who the Investigator has determined can complete the C-SSRS). The C-SSRS is assessed (when applicable) at Screening, Baseline, Weeks 1, 2, 4, 6, and Week 8/ET visits .

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.

For subjects who are able to complete the assessment, the Baseline/Screening version will be administered at Screening, and the Since Last Visit version will be administered at subsequent visits. The C-SSRS results for each subject should be reviewed by the Investigator at each visit. If at any time the C-SSRS results for a given subject reveal potential suicidality, then the Investigator should assess the clinical significance of such results. If a clinically significant risk of suicidality is identified for a subject, then the Investigator should discontinue the subject and implement appropriate treatment.

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, this subject will be counted as having suicidal ideation.

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

Missing C-SSRS item scores will not be imputed.

5.2.7 Global Clinician Assessment of Suicidality

If the subject is not able to reliably complete the C-SSRS in the Investigator’s judgment, the GCAS will be administered and used thereafter in the study. The GCAS is assessed (when applicable) at Screening, Baseline, Weeks 1, 2, 4, 6, and Week 8/ET visits.

If the GCAS is administered at Visit 1 (Screening) lifetime suicidality and suicidality for the past 3 months will be assessed and at all other visits, suicidality since the previous visit will be assessed. As with the C-SSRS, if at any time the GCAS results for a given subject reveal potential suicidality, then the Investigator should assess the clinical significance of such results. If a clinically significant risk of suicidality is identified for a subject, then the Investigator should discontinue the subject and implement appropriate treatment.

The GCAS is a clinician-rated, 5-point scale that is designed to rate the subject’s suicidality based on the report of the subject, the report of the study partner/caregiver, and the clinician’s global assessment. Ratings can be 0 (“Absent”), 1 (“Feels life is not worth living”), 2 (“Wishes he/she were dead or any thoughts of possible death to self”), 3 (“Suicidal ideas or gesture”), or 4 (“Attempt at suicide”). The Investigator will record a subject rating, a study partner/caregiver rating, and a clinician rating.

Missing scores will not be imputed.

5.2.8 Extrapyramidal Symptom Rating Scale-Abbreviated

The ESRS-A is assessed at Baseline, Weeks 1, 2, 4, 6, and Week 8/ET.

The ESRS was developed to assess 4 types of drug-induced movement disorders (Parkinsonism, akathisia, dystonia, and tardive dyskinesia) with established reliability, validity, and sensitivity. The ESRS-A, an accepted modified form of the original ESRS, will be used during the study to monitor for any worsening in extrapyramidal symptoms or signs at scheduled and unscheduled visits.

The ESRS-A consists of 4 subscales and 4 clinical global impression movement severity (CGI-S) scales of Parkinsonism, dyskinesia, dystonia, and akathisia.

The Parkinsonism subscale consists of 10 items, the dyskinesia subscale consists of 6 items, the dystonia subscale consists of 6 items, and the akathisia subscale consists of 2 items. Each item is scored on a 6-point scale: 0=absent; 1=minimal; 2=mild; 3=moderate; 4=severe; 5=extreme. The ESRS-A total score is the sum of the 24 item scores with a possible range of 0 to 120. Higher scores denote more severe drug-induced movement disorders.

Each CGI-S scale of Parkinsonism, dyskinesia, dystonia, and akathisia is a single-item 6-point scale: 0=absent; 1=minimal; 2=mild; 3=moderate; 4=severe; 5=extreme, with higher scores indicating more severe movement disorders.

If 4 or fewer items are missing, then the total score for a given subject and timepoint will be imputed as the mean of the non-missing items multiplied by 24 and rounded to the nearest integer. If more than 4 items are missing, the total score will not be derived. Missing ESRS-A item scores and subscale total scores of the 4 subscales will not be imputed. Missing CGI-S scores will not be imputed.

5.2.9 Mini-Mental State Examination

The MMSE is assessed at Screening, Baseline, Weeks 1, 2, 4, 6, and Week 8/ET visits.

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

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

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

5.3 Analysis Visit Windows

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

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

Table 2 Analysis Visit Windows

Analysis Visit Name	Target Study Day[1]	Study Day Interval
Baseline	1	≤1
Week 1	8	2 to 11
Week 2	15	12 to 21
Week 4	29	22 to 35
Week 6	43	36 to 49
Week 8	57	50 to 63
Follow-up	85	64 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 date of first dose of study drug.

5.3.1 Unscheduled Assessments

Both scheduled and unscheduled assessments, including early termination visits, will be considered for planned timepoint summaries. 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. Details are provided in a separate programming conventions document.

For safety summaries where the most extreme 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. All results will be presented in data listings.

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 date of the end-of-study/early termination visit will be used in the calculation of treatment duration. For the incomplete last dose date of the study drug, the imputation algorithms will be detailed in the 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.

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

5.4.3 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 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 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 randomized 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. The number and percentage of subjects who completed the study or discontinued (all discontinued and by discontinuation reason) will also be summarized for the Randomized, Safety, and Full Analysis sets. A summary by region for the Safety Analysis set will also be presented.

For subjects who were not included in the first interim analysis (IA-1) (see Section 16.2), early terminations of subjects due to COVID-19 related reasons will be captured in the eCRFs and summarized.

For randomized 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, and primary race), screen failure reasons (the specific inclusion/exclusion criterion [or criteria] not met or other reason) and protocol version will be listed. If a subject is re-screened, then the re-screening subject ID and the final enrollment status (whether eventually enrolled) will also be displayed in this listing. In addition, the frequency that the screen failure reasons are reported will also be summarized. Note that 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.

For subjects who were not included in IA-1, screen failures due to COVID-19 will be captured in the eCRFs, and summarized.

7 PROTOCOL DEVIATIONS

Protocol deviations will be reviewed periodically over the course of the study. The review process, definition of the deviation categories, and the classification of a deviation as major

or minor will be detailed in the Protocol Deviation Management Plan. For subjects who were not included in IA-1, protocol deviations will also be assessed with respect to relationship to COVID-19.

For randomized subjects, a summary of the number and percentage of subjects with major protocol deviations for each deviation category will be presented in three ways: all protocol deviations, COVID-19 related protocol deviations, and non COVID-19 related protocol deviations.

Similarly, three data listings of major protocol deviations will be provided: All deviations, COVID-19 related protocol deviations, and non COVID-19 related protocol deviations. Additionally, a data listing of all minor COVID-19 related protocol deviations will be provided.

In order to assess the potential impact of COVID-19 on missed visits, the proportion of missed visits will be tabulated by treatment arm for the Safety analysis set, computed as the number of total number missed visits divided by the total number of expected visits. This will be tabulated separately for IA-1 subjects, and subjects enrolled after IA-1, as a proxy for pre- and post-COVID-19 analyses.

8 DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS

Demographics and Baseline characteristics will be summarized for the Randomized, Safety, and Full Analysis sets using descriptive statistics. Summaries by region using the Safety Analysis set will also be presented.

8.1 Demographics

Demographic 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), region, subject living situation, and partner/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.

8.2 Disease Characteristics

Disease characteristics at Baseline will be summarized for the Randomized, Safety, and Full Analysis sets using descriptive statistics. MMSE total score, CGI-S, EQ-5D-5L VAS, SDI total score, and ESRS-A total score will be presented as continuous variables. CGI-S will be presented as a categorical variable using categories of 1 to 7, and MMSE total score will be

presented as a categorical variable using categories of ≤10, 11 to 17, 18 to 24, and 25 to 30. The number and percentage of subjects reporting suicidal ideation and suicidal behavior within the past 6 months and lifetime will be tabulated (for those subjects assessed with C-SSRS), and number and percentage of subjects reporting suicidality within the past 3 months and lifetime will be tabulated (for those subjects assessed with the GCAS).

Neuropsychiatric history will be summarized descriptively for continuous and categorical variables. Informed consent date will be used as the reference date when calculating duration of disease variables.

For both the Disease Characteristics and Neuropsychiatric History, summaries by region will be presented using the Safety Analysis Set.

8.3 Neuropsychiatric Inventory (NPI)

NPI will be summarized at Baseline for the Randomized, Safety and Full Analysis sets. Each of the individual domain total scores will be presented as continuous variables using descriptive statistics.

9 MEDICAL HISTORY

Medical and surgical history data will be coded using Medical Dictionary for Regulatory Activities (MedDRA) version 20.0 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-DDE-B2) 2018 March or newer version. For concomitant 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.

10.1 Prior Medication

Prior medication is defined as any medication with a start and stop date prior to the date of the first dose of study drug.

10.2 Concomitant Medication

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.

10.3 Post-Treatment Medication

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

10.4 COVID-19 Related Medications

For those subjects not included in IA-1, relationship to COVID-19 will be assessed by the investigator as detailed in the DMP Appendix B. Concomitant and post-treatment medication analyses described in Section 10.2 and Section 10.3 above will also be summarized by relationship to COVID-19 (Not related to COVID-19 vs. Related to COVID-19).

11 EXTENT OF EXPOSURE AND TREATMENT COMPLIANCE

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

11.1 Exposure to Study drug

Pimavanserin dose levels are expressed as free base.

Duration of exposure is calculated as follows:

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

For the categorical presentation of duration of exposure the number and percentage of subjects in each of the following categories will be displayed: <1 week (1 to 6 days), 1 to <2 weeks (7 to 13 days), 2 to <4 weeks (14 to 27 days), 4 to <6 weeks (28 to 41 days), 6 to <8 weeks (42 to 55 days) and ≥8 weeks (≥56 days).

Kaplan-Meier curves of duration of exposure will also be provided for each treatment group.

11.2 Measurement of Study Drug Compliance

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

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

Study drug 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 and the planned treatment assignments based on the randomization schedule. Descriptive summaries of all efficacy endpoints will be provided.

12.1 Exploratory Efficacy Endpoints

All of the efficacy endpoints are considered exploratory, and include the following:

- Change from Baseline to Week 8 in CGI-S score for neuropsychiatric symptoms
- CGI-I score at Week 8 for neuropsychiatric symptoms
- Change from Baseline to Week 8 in SDI score
- Change from Baseline to Week 8 in EQ-5D-5L scores (VAS and 5 level descriptive system)

12.2 Adjustment for Covariates

For continuous variables (except CGI-I) analyzed using a mixed model for repeated measured (MMRM) or an analysis of covariance (ANCOVA), the Baseline value of the endpoint being analyzed and region will be included as covariates as described in Sections 13.2 and 13.3. For CGI-I the Baseline CGI-S score and region will be included as covariates in the MMRM and ANCOVA analyses, and for CGI-I responder analyses region will be included as a stratification factor.

12.3 Handling of Missing Data

For selected endpoints missing values will be imputed with the previous observed value (including Baseline) for that subject, i.e. last observation carried forward (LOCF).

12.4 Multiple Comparisons/Multiplicity

No adjustments will be made for multiple comparisons for comparisons across multiple endpoints and across multiple subgroups.

12.5 Examination of Subgroups

Treatment comparisons will be made with respect to exploratory efficacy and secondary safety variables separately for each subgroup by:

- Dementia related psychosis (DRP) (yes):

  - Subjects with dementia from any cause, and having either a Baseline NPI delusions or hallucinations score (frequency × severity) ≥ 2

- Primary cause of dementia due to Alzheimer's disease (AD) (yes):
  - Subjects with primary cause of dementia due to AD
- Alzheimer's disease psychosis (ADP) (yes)
  - Subjects with AD as the primary cause of dementia, and having either a Baseline NPI delusions or hallucinations score (frequency × severity) ≥ 2
- Other-dementia psychosis (yes)
  - Subjects with vascular dementia, frontotemporal dementia-spectrum disorders, or dementia with Lewy bodies, and having either a Baseline NPI delusions or hallucinations score (frequency × severity) ≥ 2

13 METHODS OF EFFICACY ANALYSES

13.1 Analysis of Efficacy Endpoints

Descriptive statistics for all exploratory efficacy endpoints listed in Section 12.1 will be tabulated by treatment group at scheduled timepoints. The summaries of the change from Baseline results will be presented by treatment group at scheduled timepoints. For change from Baseline values at each post-Baseline visit, LS means and standard errors (SE), the between-group difference in LS means with the corresponding 95% CI, p-value, and effect size will also be presented.

13.2 Mixed Model Repeated Measures (MMRM)

The MMRM method will be used for selected continuous exploratory endpoints, in which missing data is assumed to be missing at random. The dependent variable will be the change from Baseline. The independent variables in the model will include the fixed categorical effects of treatment group, region, visit, and treatment-by-visit interaction, as well as the continuous, fixed covariates of 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. Observed cases will be analyzed in the model and least squares (LS) means will be estimated using observed margins.

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

1) Heterogeneous Toeplitz

2) Heterogeneous compound symmetry
3) Heterogeneous autoregressive(1)
4) Toeplitz
5) Compound symmetry
6) Autoregressive(1)
7) Variance components

The first covariance structure that allows for convergence will be selected for the final model. For change from Baseline values at each post-Baseline visit, LS means, standard errors (SEs), and the between-group difference in LS means with the corresponding two-sided 95% confidence intervals (CIs), p-values, and effect sizes will also be presented. LS means ±SE over time for the change from Baseline values by treatment group will be presented in line plots.

At each visit, the effect size (Cohen's d) for the change from Baseline between the treatment groups will be calculated using the following formula:

$$Effect\ size = \frac{LS\ mean\ difference}{\sqrt{variance}}$$

The variance at a given visit will be obtained from the covariance matrix estimated for the MMRM model. The sign (+ or -) of the effect size will be chosen so that a positive value favors pimavanserin.

13.3 Analysis of Covariance (ANCOVA)

For those endpoints that are analyzed using ANCOVA (i.e., EQ-5D-5L VAS), the dependent variable will be the change from Baseline, and the independent variables in the model will include treatment group and region as factors, and Baseline score as a continuous covariate. LS means will be estimated using observed margins; the LSM, SE, LSM treatment difference, 95% CI of treatment difference, and p-value for treatment differences from the ANCOVA model will be reported. Observed cases excluding missing values, and the last observation carried forward (LOCF) method of imputation for missing data will be used. If necessary, the Baseline value will be carried forward.

13.4 Exploratory Efficacy Analyses

13.4.1 CGI-S

Observed CGI-S scores will be summarized using descriptive statistics at Baseline, Weeks 1, 2, 4, 6, and 8/ET. The change from Baseline values will be summarized descriptively at Weeks 1, 2, 4, 6, and Week 8/ET. The change from Baseline in CGI-S score will also be

analyzed using the MMRM model as described in Section 13.2, except that anti-dementia medication use at Baseline will be included in the model as a factor.

13.4.2 CGI-I

13.4.2.1 MMRM

Observed CGI-I scores will be summarized using descriptive statistics at Weeks 1, 2, 4, 6, and 8/ET. The CGI-I scores will also be analyzed using the MMRM model as described in Section 13.2, with the following modifications:

- the dependent variable is the observed score
- Baseline CGI-S score is used as the covariate
- Anti-dementia medication use at Baseline will be included in the model as a factor

13.4.2.2 Responder Analysis

A responder analysis will be conducted in which a CGI-I score of 1 (very much improved) or 2 (much improved) will be counted as a responder. At each post-Baseline timepoint the proportion of responders will be summarized by treatment group.

Two types of responder analyses will be performed:

- Observed case responder analysis, in which subjects with missing values at a given visit are excluded from the analysis for that visit, will be performed
- Missing values imputed as non-responders for the given visit

The adjusted difference in percent responders between treatment groups (pimavanserin group minus placebo group) using the Cochran-Mantel-Haenszel weighting scheme stratified by region (North America, Europe, or rest of world), p-value, and Newcombe's 95% CI will be presented.

13.4.3 SDI

Observed SDI scores will be summarized using descriptive statistics at Baseline, Week 4, and Week 8/ET. The change from Baseline values will be summarized descriptively at Week 4 and Week 8/ET. The change from Baseline in SDI score will also be analyzed using the MMRM model as described in Section 13.2.

13.4.4 EQ-5D-5L

13.4.4.1 VAS ANCOVA

The change from Baseline in the EQ-5D-5L VAS score will be summarized using an ANCOVA model with effects for region, treatment group, and Baseline score.

Observed EQ-5D-5L VAS scores will be summarized using descriptive statistics at Baseline and Week 8/ET. The change from Baseline values will be summarized descriptively at Week 8/ET. The change from Baseline in EQ-5D-5L VAS score will also be analyzed using OC and ANCOVA LOCF as specified in Section 13.3.

13.4.4.2 Descriptive System Categorical Analysis

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

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

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

The proportion of subjects in each PCHC category will be summarized descriptively at Week 8/ET using both OC and LOCF.

14 SAFETY ANALYSES

Safety variables include AEs, TEAEs, clinical laboratory variables, vital signs, body weight, BMI, physical examinations, electrocardiogram (ECG), C-SSRS, GCAS, MMSE, and ESRS-A variables. The safety summaries will be presented for the Safety Analysis Set using descriptive statistics. An MMRM model will be applied to the ESRS-A and MMSE total score endpoints. Safety variables will be summarized by treatment group.

14.1 Adverse Events

All Adverse events (AEs) will be coded using MedDRA version 20.0 or newer.

An AE will be considered a TEAE if it started on or 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 pre-dose findings will not be considered as TEAEs.

For subjects enrolling in the long-term open-label study ACP-103-047, any AE started before the first dose of open-label study drug in Study ACP-103-047 will be considered as TEAE in Study ACP-103-046 but not in study ACP-103-047. Similarly, any AE reported after the first dose of open-label study drug in ACP-103-047 will be reported in association with the ACP-103-047 study only.

14.1.1 Primary Safety Analyses

TEAEs are the primary measure for this study. For the following summaries, the event counts and the number and percentage of subjects reporting TEAEs will be tabulated by treatment group:

- TEAEs by SOC and preferred term (PT)
- Most frequently reported TEAEs by SOC and PT
  - PTs reported by ≥ 5% of subjects in either treatment group
- 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

For tabulations that include SOC and PT the display will be sorted alphabetically by SOC and then by descending subject frequency for the PTs based on counts in the combined treatment groups within each SOC. SOCs will not be included in the TEAEs by PT tabulation. This display will be sorted by descending subject frequency based on counts in the combined treatment groups.

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 AEs leading to study drug discontinuation or study termination and subjects who died (if any). In these listings, an indicator for TEAEs will also be included.

14.1.2 Assessment of the Impact of COVID-19 on Adverse Events

The relationship of selected AEs to COVID-19 will be assessed by investigator as detailed in the DMP Appendix B, for those subjects who were not unblinded as part of the first interim analysis. Each of the primary safety analyses described in Section 14.1.1 will be additionally summarized by relationship to COVID-19 (Not related to COVID-19 vs. Related to COVID-19).

Both the Not related to COVID-19 and the Related to COVID-19 AE tables will be summarized using 2 sets of subjects: all subjects in the Safety analysis set, and subjects in the Safety analysis set that were not included in IA-1, as a proxy for pre- and post-COVID-19 analyses.

14.2 Secondary Safety Analyses

The secondary endpoints for this study are change from Baseline to Week 8 in ESRS-A, and change from Baseline to Week 8 in MMSE.

14.2.1 ESRS-A

The ESRS-A total score and the 4 individual global CGI-S scores will be summarized using descriptive statistics at Baseline and Weeks 1, 2, 4, 6, and 8/ET. The change from Baseline scores will also be summarized at Weeks 1, 2, 4, 6, and 8/ET.

The ESRS-A total score change from Baseline will be summarized using MMRM as specified in Section 13.2. The model will include categorical fixed effects for region, treatment group, visit, treatment-by-visit interaction, as well as the continuous fixed covariates for Baseline score and Baseline score-by-visit interaction. The MMRM results for the pimavanserin and placebo groups will be tabulated by visit.

The individual item scores will be listed but not summarized.

14.2.2 MMSE

The MMSE total score will be summarized using descriptive statistics at Baseline, Weeks 1, 2, 4, 6, and 8/ET. The change from Baseline scores will also be summarized at Weeks 1, 2, 4, 6, and 8/ET.

The MMSE total score change from Baseline will be summarized using MMRM as specified in Section 13.2. The model will include categorical fixed effects for region, treatment group, visit, treatment-by-visit interaction, as well as the continuous fixed covariates for Baseline

score and Baseline score-by-visit interaction. The MMRM results for the pimavanserin and placebo groups will be tabulated by visit.

The individual item scores will be listed but not summarized.

14.3 Other Safety Analyses

14.3.1 Clinical Laboratory Variables

Clinical laboratory tests are performed at Screening, Baseline, Week 4, and Week 8/ET visits. In general, laboratory test results are collected from a central laboratory

Due to COVID-19 disruptions it is possible that some test results may be collected from a local laboratory. A separate eCRF will capture local laboratory results in order to facilitate medical monitoring of subject safety. Local laboratory results will not be included in any summary data analysis; they will, however, be included in data listings, as well as a separate local laboratory PCI listing. All results (central and local) will be displayed in Système International [SI] units.

14.3.2 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
    - Mg should only be performed at Screening
- Alaninine aminotransferase (ALT), aspartate aminotransferase (AST), gamma-glutamyl transpeptidase (GGT), alkaline phosphatase (ALP), total bilirubin (TBIL), lactate dehydrogenase (LDH)
- Glucose
- Albumin (ALB), total protein
- Prolactin
- Creatine kinase (CK)/creatine phosphokinase (CPK)
- Lipid panel
    - total cholesterol, HDL-cholesterol, triglycerides, LDL-cholesterol, cholesterol/HDL ratio, ratio, non-HDL-cholesterol, very low density lipoprotein cholesterol

- Magnesium (Mg), Vitamin B12, HbA1c, thyroid stimulating hormone (TSH) and reflex 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

14.3.3 Hematology

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 (Hgb), red blood cells (RBC), platelets, reticulocyte count

14.3.4 Urinalysis

Urinalysis tests include the following:

- Blood, RBCs, WBCs, 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 visit
    - If urine cannot be obtained in women of childbearing potential, a serum pregnancy test should be done in its place

14.3.5 Methods of Analysis for Clinical Laboratory Variables

14.3.5.1 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 scheduled post-Baseline visits. The change from Baseline values will also be summarized at scheduled post-Baseline visits.

The overall post-Baseline minimum and maximum observed and change from Baseline values will also be summarized. 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, visit, and treatment group.

For hemoglobin, hematocrit and uric acid, by-visit as well as overall post-Baseline minimum and maximum will be presented for each gender as well as for both genders combined.

14.3.5.2 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 and treatment group.

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 and treatment group.

For hemoglobin, hematocrit and uric acid, by-visit as well as overall post-Baseline minimum and maximum will be presented for each gender as well as for both genders combined.

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

14.3.5.4 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'). Central and local laboratory values will be presented separately.

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 (including Baseline) was observed. Central and local laboratory PCI listings will be presented separately.

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

14.3.6 Vital Signs

Vital signs are assessed at Screening, Baseline, Weeks 1, 2, 4, 6 and Week 8/ET visits.

Due to COVID-19 disruptions it may be necessary for vital signs to be collected outside of the clinic by persons other than study site staff. These results will not be included in any summary data analysis; they will, however, be presented in data listings, and in a separate PCI listing of vitals measured by persons other than study site staff.

14.3.6.1 Vital Signs Variables

Vital signs include height (only measured at Screening), weight, derived BMI, systolic blood pressure (SBP), diastolic blood pressure (DBP), pulse rate, temperature, and respiratory rate.

14.3.6.2 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.

Vital sign 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 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, 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 all PCI values will be provided (excluding those vitals collected by persons other than study site staff). There will be a separate PCI listing for vitals collected by persons other than study site staff. These listings will include all observations for those subjects and parameters for which at least one PCI value (including Baseline) was observed.

Table 5 Criteria for Potentially Clinically Important Vital Signs

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

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

14.3.7 Electrocardiogram (ECG)

In countries outside of the Czech Republic, ECG is performed at Screening, Baseline, Week 2, and Week 8/ET. At the clinical sites in the Czech Republic, ECG is performed at all visits. For ECG data summaries, the common scheduled visits across all clinical sites will be used (i.e., Baseline, Week 2, and Week 8/ET).

All tracings will be evaluated by a central reading laboratory, including any collected outside of the clinic due to COVID-19 disruptions when possible. 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.

14.3.7.1 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

14.3.7.2 ECG Methods of Analysis

14.3.7.2.1 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. 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.

14.3.7.2.2 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, 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; all post-Baseline values will be considered, including unscheduled and out-of-window values.

A listing of all PCI values will be provided for those subjects and parameters for which at least 1 PCI value (including Baseline) 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

14.3.7.2.3 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 given treatment group. 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 for the given treatment group.

14.3.7.2.4 Data Listings

All data, including machine-read Baseline results, will be listed. 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 (including Baseline) was observed.

14.3.8 Physical Examination

Physical examination is performed at Screening, Baseline, Week 4, and Week 8/ET visits.

Physical examination results (normal, abnormal, and not done) at Baseline, Week 4, and Week 8/ET will be summarized in a frequency table by body system and visit.

14.3.9 Suicidality

14.3.9.1 C-SSRS

The C-SSRS will be administered if the subject, in the Investigator's judgement, is able to reliably complete it. Otherwise, the GCAS will be administered and used thereafter in the study.

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 assessment within each treatment group.

14.3.9.2 GCAS

The number and percentage of subjects within each GCAS rating will be summarized using the clinician rating. The number and percentage of subjects reporting any post-Baseline rating of 3 or 4 will be tabulated. For these analyses, the denominator for calculating percentages is the number of subjects with at least one post-Baseline GCAS assessment within each treatment group.

Subject and study partner/caregiver ratings will not be summarized but will be included in data listings.

15 CLINICAL PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES

Not applicable.

16 UNBLINDED INTERIM ANALYSES

16.1 General Procedure

One or more unblinded IAs of data from subjects who have exited the study (completed with or without enrolling into the open-label extension, or terminated early) may be conducted in order to support safety and efficacy evaluations for regulatory submissions. Subjects who have exited the study will only be unblinded and analyzed if all other subjects in the same randomization block have also exited the study.

The analyses planned for an unblinded IA will align with the purpose of the regulatory submission. Study subjects, study partners/caregivers, Investigators, raters, and site personnel will remain blinded to subject treatment assignments until all subjects have completed the study and the clinical database is locked. The distribution of unblinded subject's treatment assignments within Acadia will be restricted. Further details regarding the restricted distribution of treatment assignments, IA results, and unblinding procedures will be specified in an IA Unblinding Plan.

16.2 Interim Analysis 1

The database lock for the first interim analysis (IA-1) occurred on December 13, 2019 (under version 1 of the SAP), and included data from 288 subjects whose last visit occurred on or before December 2, 2019. Data from these subjects may be used as a pre COVID-19

comparison group for selected safety and efficacy endpoints, in order to assess the potential impact of COVID-19 on the final study results.

17 DATA MONITORING/REVIEW COMMITTEE

An independent Data and Safety Monitoring Board (DSMB) will review interim and final safety data including AE and SAE data. The DSMB will be independent of the sponsor and the Investigators, and can recommend terminating the study early based on the safety review. An Independent Statistical Group (ISG), not affiliated with the Sponsor, will serve as a centralized communication gateway for data queries and additional data analyses/requests between the Sponsor and the DSMB.

For DSMB safety reviews, the treatment codes will be released to the ISG statistician who will be responsible for producing unblinded statistical outputs, and for providing these outputs to DSMB members using a secure method. The Sponsor and Investigators will remain blinded until the official unblinding of the database at the end of the study, except as described in Section 16. The outputs presented to DSMB members will include, but are not limited to, summaries of AEs, SAEs, deaths, laboratory measurements, ECGs, and vital signs. The DSMB will also be provided with SAEs and deaths reported in ACP-103-047, the open-label extension study for eligible subjects who complete ACP-103-046.

Additional details regarding the roles and responsibilities of the DSMB members and ISG, and planned frequency of meetings, are specified in the DSMB Charter.

In addition, blinded safety data will be monitored throughout the study and blinded aggregate safety reports will be produced and reviewed approximately quarterly.

18 COMPUTER METHODS

Statistical analyses will be performed using Version 9.4 (or newer) of SAS® software (SAS Institute Inc., Cary, North Carolina) on a suitably qualified and validated environment. Linear mixed effects models will be fit using the PROC MIXED procedure.

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

Added DRP, AD, ADP, and Other subgroups, statistical tests for treatment comparisons of efficacy and secondary safety variables, and data handling related to Ukraine geopolitical conflict.

20 REFERENCES

1. FDA Guidance on Conduct of Clinical Trials of Medical Products during COVID-19 Pandemic, March 2020; updated May 14, 2020.
2. EMA Guidance on the Management of Clinical Trials during the COVID-19 (Coronavirus) pandemic, March 2020.
3. Yan X., Su G. Stratified Wilson and Newcombe Confidence Intervals for Multiple Binomial Proportions, Statistics in Biopharmaceutical Research 2020; Vol. 2, No. 3, 329-335.

21 APPENDICES

21.1 Summary of Version Changes

Version No:	Document History Description of Update	Author(s)	Version Date
1.0	Original version	[redacted]	24 July 2019
2.0	Updated to include handling of analyses related to COVID-19	[redacted]	26 May 2020
3.0	In alignment with Protocol Amendment 7 changes were made to indicate that exploratory efficacy data from interim analyses conducted in this study may also be evaluated in order to support regulatory submissions.	[redacted]	06 March 2021
4.0	Add subgroup analyses, nominal p-values for selected outputs, added anti-dementia use as a factor in CGI-S and CGI-I MMRM models, and Ukraine data handling.	[redacted]	23 May 2022