Official Title: A Double-Blind, Placebo-Controlled Study to Examine the Safety and Efficacy of Pimavanserin for the Treatment of Agitation and Aggression in Alzheimer's Disease

NCT Number: NCT02992132

STATISTICAL ANALYSIS PLAN

Protocol No.:	ACP-103-032
Protocol Title:	A Double-Blind, Placebo-Controlled Study to Examine the Safety and Efficacy of Pimavanserin for the Treatment of Agitation and Aggression in Alzheimer's Disease
Protocol Version No. and Date	Amendment 3, 20NOV2017
Drug:	Pimavanserin
Sponsor:	ACADIA Pharmaceuticals Inc. [redacted]
SAP Version No. and Date	Version 1, 22NOV2017

SIGNATURE/APPROVAL PAGE

AUTHOR

[redacted]

22NOV 2017

Date

APPROVERS

[redacted]

28 NOV 2017

Date

[redacted]

22 NOV 2017

Date

[redacted]

22NOV2017

Date

[redacted]

29NOV2017

Date

Table of Contents

SIGNATURE/APPROVAL PAGE 2
TABLE OF CONTENTS 3
LIST OF TABLES 5
LIST OF FIGURES 5
ABBREVIATIONS 6
1 INTRODUCTION 8
2 OBJECTIVES 8
2.1 Primary Objective 8
2.2 Secondary Objective 8
2.3 Exploratory Objectives 8
3 STUDY DESIGN 9
3.1 General Study Design 9
3.2 Schedule of Assessments 9
3.3 Randomization 13
3.4 Blinding 13
3.5 Determination of Sample Size 13
4 ANALYSIS SETS 13
5 DATA HANDLING CONVENTIONS 14
5.1 General Data Reporting Conventions 14
5.2 Derived Variables 14
5.2.1 Cohen-Mansfield Agitation Inventory Long-form (CMAI) 15
5.2.2 Zarit Burden Interview (ZBI) 16
5.2.3 Alzheimer's Disease Cooperative Study – Activities of Daily Living Inventory (ADCS-ADL) 16
5.2.4 Neuropsychiatric Inventory – Clinician Rating Scale (NPI-C) 18
5.2.5 Modified Alzheimer's Disease Cooperative Study – Clinical Global Impression-Change (mADCS-CGIC) 20
5.2.6 Karolinska Sleepiness Scale (KSS) 20
5.2.7 Mini-Mental State Examination (MMSE) 21
5.2.8 Global Clinician Assessment of Suicidality (GCAS) 21
5.3 Analysis Visit Windows 21
5.3.1 Unscheduled Assessments 22
5.3.2 Multiple Measurements within Visit Windows 22

5.4 Missing or Incomplete Date for Last Dose of Study Drug 23
5.5 Missing or Incomplete Dates for Prior or Concomitant Medications 23
5.6 Missing or Incomplete Dates for Adverse Events 23
5.7 Missing Severity Assessment for Adverse Events 23
5.8 Missing Relationship to Study Drug for Adverse Events 24
5.9 Character Values of Clinical Laboratory Variables 24
6 SUBJECT DISPOSITION 24
7 PROTOCOL DEVIATIONS 25
8 DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS 25
9 MEDICAL HISTORY 26
10 EXTENT OF EXPOSURE AND TREATMENT COMPLIANCE 26
10.1 Exposure to Study drug 26
10.2 Measurement of Treatment Compliance 26
11 PRIOR, CONCOMITANT AND POST MEDICATION 27
12 EFFICACY ANALYSES 27
12.1 Efficacy Variables 27
12.1.1 Primary Efficacy Variable 27
12.1.2 Secondary Efficacy Variable 27
12.1.3 Exploratory Efficacy Variables 27
12.2 Adjustment for Covariates 28
12.3 Handling of Missing Data 28
12.4 Multiple Comparisons / Multiplicity 28
12.5 Examination of Subgroups 28
13 METHODS OF EFFICACY ANALYSES 29
13.1 Primary Efficacy Analysis 29
13.2 Secondary Efficacy Analysis 30
13.3 Exploratory Efficacy Analyses 30
13.4 Responder Analyses 32
14 SAFETY ANALYSES 33
14.1 Adverse Events 33
14.2 Clinical Laboratory Variables 34
14.3 Vital Signs 38
14.4 Electrocardiogram (ECG) 39
14.5 Physical Examination 40
14.6 Other Safety Variables 40
14.6.1 Suicidal Ideation and Behavior 40

15 CLINICAL PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES 40
16 INTERIM ANALYSIS 41
17 DATA MONITORING/REVIEW COMMITTEE 41
18 COMPUTER METHODS 41
19 CHANGES TO ANALYSES SPECIFIED IN PROTOCOL 41
20 REFERENCES 42
21 APPENDICES 42
21.1 Summary of Version Changes 42

List of Tables

Table 1 Schedule of Assessments 11
Table 2 ADCS-ADL Item Score Range 17
Table 3 NPI-C Domain Score 19
Table 4 Analysis Visit Windows 22
Table 5 Criteria for Potentially Clinically Important Laboratory Values – Hematology and Chemistry 37
Table 6 Criteria for Potentially Clinically Important Laboratory Values – Urinalysis 38
Table 7 Criteria for Potentially Clinically Important Vital Signs 39
Table 8 Criteria for Potentially Clinically Important ECG Values 40

List of Figures

Figure 1 Study Design 10

ABBREVIATIONS

AD	Alzheimer's disease
ADCS-ADL	Alzheimer's Disease Cooperative Study – Activities of Daily Living Inventory
AE	adverse event
ANCOVA	analysis of covariance
ATC	Anatomical Therapeutic Chemical
BLQ	below the limit of quantification
BMI	body mass index
bpm	beats per minute
CIs	confidence intervals
CMAI	Cohen-Mansfield Agitation Inventory
CT	computed tomography
ECG	electrocardiogram
eCRF	electronic case report form
EOS	end-of-study
ET	early termination
FAS	Full Analysis Set
GCAS	Global Clinician Assessment of Suicidality
KSS	Karolinska Sleepiness Scale
LOCF	last observation carried forward
LS	least squares
MMRM	mixed model for repeated measures
MMSE	Mini-Mental State Examination
mADCS-CGIC	modified Alzheimer's Disease Cooperative Study - Clinical Global Impression of Change
MedDRA	Medical Dictionary for Regulatory Activities
MRI	magnetic resonance imaging
msec	milliseconds
NIA-AA	National Institute on Aging-Alzheimer's Association
NPI-C	Neuropsychiatric Inventory-Clinician Rating Scale
OC	observed cases
PCI	potentially clinically important
PK	pharmacokinetic
QD	once daily
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
SE	standard error
SMC	Safety Monitoring Committee
SOC	system organ class
SSRI	selective serotonin reuptake inhibitor
TEAE	treatment-emergent adverse event
ZBI	Zarit Burden Interview

1 INTRODUCTION

This statistical analysis plan (SAP) provides a technical and detailed elaboration of the summaries of efficacy and safety data. Specifications for tables, figures, and listings are contained in a separate document.

Enrollment in the study was terminated early by the sponsor for business reasons not related to the safety of pimavanserin after 111 subjects were randomized. The last subject was randomized into the study on November 2, 2017. As a result, the study is not powered to definitively evaluate efficacy measures and the summaries will be descriptive in nature. No hypothesis testing is planned.

2 OBJECTIVES

The purpose of this study is to evaluate the safety and efficacy of pimavanserin at doses (free base) of 34 mg and 20 mg, compared with placebo in the treatment of agitation and aggression in subjects with probable Alzheimer's disease (AD).

2.1 Primary Objective

The primary objective of the study is to evaluate the efficacy of pimavanserin compared with placebo in the treatment of agitation and aggression after 12 weeks.

2.2 Secondary Objective

The secondary objective of the study is to evaluate the efficacy of pimavanserin compared with placebo on caregiver burden.

2.3 Exploratory Objectives

Exploratory objectives of the study are:

- To evaluate the efficacy of pimavanserin compared with placebo for the following:
  - Clinician's global assessment of treatment benefit
  - Other neuropsychiatric symptoms
  - Functional status
  - Sleep and daytime wakefulness
  - Cognition
  - Need for rescue medication
- To evaluate the safety and tolerability of pimavanserin
- To evaluate the plasma concentrations of pimavanserin and the major metabolite AC-279

3 STUDY DESIGN

3.1 General Study Design

This study will be conducted as a Phase 2, 12-week, randomized, double-blind, placebo-controlled, multicenter, outpatient study in subjects with a diagnosis of probable AD, according to the National Institute on Aging-Alzheimer's Association (NIA-AA) guidelines who have clinically significant agitation/aggression and who meet the following criteria:

- Mini-Mental State Examination (MMSE) score of 5 to 26 (inclusive) at the Screening Visit
- Neuropsychiatric Inventory-Clinician Rating Scale (NPI-C) combined agitation and aggression domain score of ≥14 at both the Screening and Baseline visits

The original planned sample size was approximately 432. For business reasons, not related to safety, the enrollment (randomization) of new subjects into the study was stopped after 111 subjects were randomized. The last subject was randomized into the study on November 2, 2017. On the first day of the randomized treatment phase (Baseline), eligible subjects were randomly assigned to receive pimavanserin 34 mg per day, pimavanserin 20 mg per day, or placebo daily in a 1:1:1 ratio, according to a computer-generated randomization schedule. The randomization was stratified according to geographic region (North America, Europe, or Rest of World).

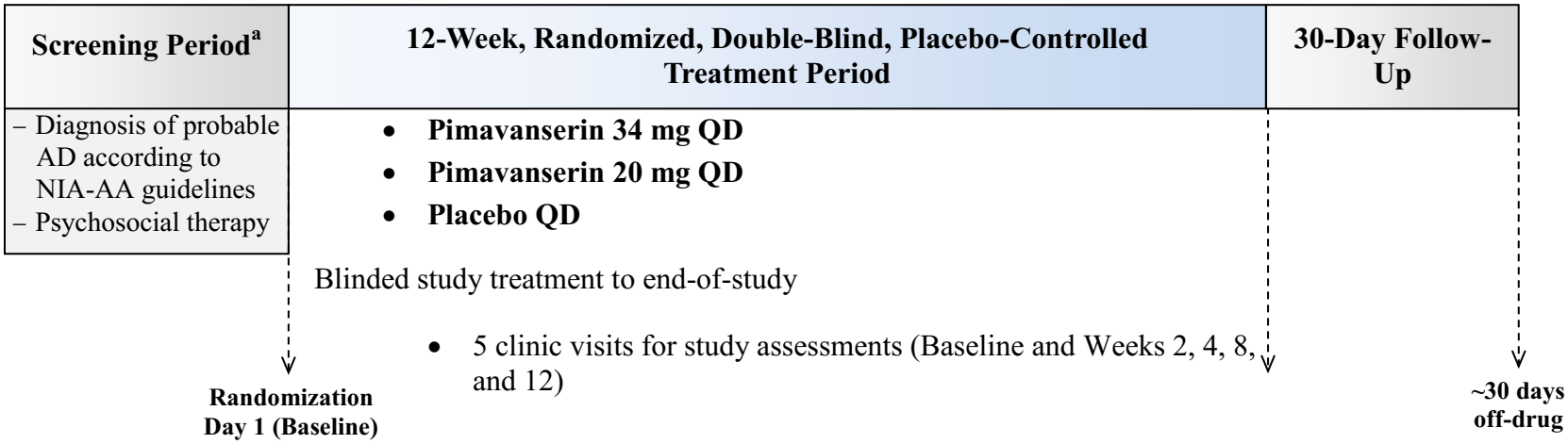
The duration of participation for individual subjects will be up to approximately 20 weeks. Each subject will participate in a 2- to 4-week screening period, a 12-week treatment period, and a 30-day safety follow-up (telephone call) for those subjects who discontinue prematurely from the study or who do not enroll in the open-label extension study (Study ACP-103-033). The end of the clinical trial will be when the last subject completes the last scheduled assessment (i.e., 30-day follow-up or enrolled in the extension study).

Figure 1 illustrates the study design.

3.2 Schedule of Assessments

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

Figure 1 Study Design



[a] For subjects taking antipsychotic medication, the screening period may be extended to 4 weeks if necessary for wash-out of at least 5 half-lives of the medication.

AD = Alzheimer's disease; NIA-AA = National Institute on Aging-Alzheimer's Association; QD = once daily

Table 1 Schedule of Assessments

Procedure	Screening Period	Treatment Period					Follow-Up
Visit Number		Visit 1[a]	Visit 2[a]	Visit 3[a]	Visit 4[a]	Visit 5[a] (EOS/ET)	Visit 6[a]
Visit Week (Day)	Week -2 to -4[b]	Baseline (1)	Week 2 (15)	Week 4 (29)	Week 8 (57)	Week 12 (85)	Week 16[c] (~115)
Allowable visit window (# days)	NA	NA	±3	±3	±3	±3	+7
Clinic (C) or Telephone (T) Visit	C	C	C	C	C	C	T
Informed consent	X						
Inclusion/exclusion criteria assessment	X	X[d]					
Medical history and demographics	X						
Alzheimer's disease history	X						
Physical examination	X					X	
Vital signs (including height[e] and weight)	X	X	X	X	X	X	
ECG[f]	X	X		X		X	
Clinical laboratory tests[g]	X	X		X		X	
Pregnancy test[h]	X	X		X		X	
PK sampling		X	X[i]	X		X	
CMAI		X	X	X	X	X	
ZBI		X	X	X	X	X	
NPI-C (Agitation and Aggression Domains)	X[j]	X	X	X	X	X	
NPI-C (All Remaining Domains)		X				X	
MMSE	X	X		X	X	X	
mADCS-CGIC		X	X	X	X	X	
KSS		X	X	X	X	X	
ADCS-ADL		X				X	
MRI or CT[k]	X						
Psychosocial therapy	X						
Assessment of concomitant medications/treatments	X	X	X	X	X	X	X
Assessment of adverse events[l]	X	X	X	X	X	X	X
GCAS	X	X	X	X	X	X	X
Randomization		X					
Dispense study drug		X		X	X		
Study drug administration at the clinic[m]		X					
Study drug accountability				X	X	X	

Abbreviations: AD = Alzheimer's Disease; ADCS-ADL = Alzheimer's Disease Cooperative Study-Activities of Daily Living Inventory; C = clinic visit; CMAI = Cohen-Mansfield Agitation Inventory; CT = computed tomography; ECG = electrocardiogram; EOS = end of study; ET = early termination; GCAS=Global Clinician Assessment of Suicidality; KSS = Karolinska Sleepiness Scale; MMSE = Mini-Mental State Examination; mADCS-CGIC = modified Alzheimer's Disease Cooperative Study – Clinical Global Impression of Change; MRI = magnetic resonance imaging; NA = not applicable; NPI-C = Neuropsychiatric Inventory-Clinician Rating Scale; PK = pharmacokinetic; ZBI = Zarit Burden Interview

a. Study visits are designated by weeks and have a ±3-day window (Visits 2 through 5) or a +7-day window (Visit 6) calculated from the Baseline Visit (Day 1). Clinic visits may be split over multiple days within the specified windows, if necessary.
b. The screening period must be at least 2 weeks long and no more than 4 weeks long.
c. For subjects who discontinue prematurely from the study or who do not enroll in the open-label extension study, a follow-up safety assessment will be conducted by telephone call approximately 30 days after the last dose of study drug.
d. NPI-C combined agitation and aggression domain score and wash-out of prohibited medications are the only eligibility assessments required at both Screening and Baseline (Day 1). All other eligibility criteria are to be established during Screening.
e. Height is assessed as part of the vital sign measurement only at the Screening Visit; weight will be assessed at every clinic visit as part of the vital sign measurements.
f. 12-lead ECG to be completed in triplicate at screening and single read at subsequent visits. ECGs can be performed any time before blood sampling or at least 30 minutes after blood sampling during clinic visits.
g. Urinalysis requirement not applicable to subjects who are unable to provide urine sample (e.g., incontinent subjects).
h. A serum pregnancy test should be performed at the initial screening visit. A urine pregnancy test should be performed at Baseline (Day 1), Week 4, and Week 12 (EOS/ET) for female subjects of childbearing potential.
i. Note: If collection at Week 2 does not occur, then collection should be attempted at Week 8.
j. Only the agitation and aggression domain will be completed at the Screening Visit.
k. MRI or CT must be obtained during or subsequent to diagnosis of probable AD, or during the screening period (prior to Baseline).
l. Any untoward medical occurrence that occurs after signing the informed consent form (i.e., during the screening period) should be recorded as an adverse event, even if dosing has not begun.
m. The first dose of study drug must be administered to the subject at the clinic only after all Baseline assessments are performed, including the blood sampling for PK assessment. If any Baseline assessments extend to an additional day, the first dose of study drug will be administered at completion of assessments on this day.

3.3 Randomization

On Day 1 of the treatment phase, eligible subjects who meet all inclusion and none of the exclusion criteria will receive pimavanserin 34 mg per day, pimavanserin 20 mg per day or daily placebo according to a computer-generated randomization schedule. Subjects will be randomized in a 1:1:1 ratio to 1 of 3 treatment groups (pimavanserin 34 mg, pimavanserin 20 mg, or placebo), and stratified according to geographic region (North America, Europe, or Rest of World).

3.4 Blinding

Treatment assignment will be double-blind such that neither the subjects, study partners/caregivers, sponsor personnel who oversee the study, nor the Investigator and study personnel will know which treatment is assigned to each subject.

3.5 Determination of Sample Size

The original planned sample size was approximately 432. For business reasons, not related to safety, the enrollment (randomization) of new subjects into the study was stopped after 111 subjects were randomized. The last subject was randomized into the study on November 2, 2017. Thus, this study is not powered to definitively evaluate efficacy measures and only descriptive summaries will be provided.

4 ANALYSIS SETS

Randomized Analysis Set

The Randomized Analysis Set will consist of all subjects who were randomized.

Subjects will be classified according to the randomized treatment assignment.

Safety Analysis Set

The Safety Analysis Set will consist of a subset of subjects in the Randomized Analysis Set who received at least one dose of study drug.

Subjects will be classified according to the actual treatment received.

Full Analysis Set

The Full Analysis Set will consist of a subset of subjects in the Safety Analysis Set who have both a Baseline value and at least 1 post-Baseline value for the CMAI total score.

Subjects will be classified according to the randomized treatment assignment.

Per-protocol Analysis Set

The Per-protocol Analysis Set will consist of a subset of subjects in the Full Analysis Set who do not have any protocol deviation which is considered to have substantial impact on the primary efficacy outcome, change from Baseline to Week 12 in the CMAI total score. The precise reasons for excluding subjects from the Per-protocol Analysis Set will be fully defined and documented *a priori* before the clinical database lock.

Subjects will be classified according to the randomized treatment assignment.

Pharmacokinetics Analysis Set

The Pharmacokinetics Analysis Set will consist of subjects with at least one measurable pimavanserin plasma concentration.

Subjects will be classified according to the actual treatment received.

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 confidence intervals will be 2-sided 95% confidence intervals.

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 with 1 decimal place.

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

5.2 Derived Variables

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

5.2.1 Cohen-Mansfield Agitation Inventory Long-form (CMAI)

The CMAI is assessed at Baseline, and Weeks 2, 4, 8 and 12/ET visits.

The CMAI is a 29-item scale designed to systematically assess agitation, rated on a 7-point (1-7) scale of frequency (1 = Never; 2 = Less than once a week; 3 = Once or twice a week; 4 = Several times a week; 5 = Once or twice a day; 6 = Several times a day; 7 = Several times an hour). Subjects are rated by their primary caregiver regarding the frequency with which they manifest physically aggressive, physically non-aggressive, and verbally agitated behaviors. The CMAI in this study is to be completed by interview of the caregiver. Ratings are inclusive of the 2 weeks prior to the administration of the scale.

In addition to the 7-point frequency scale, there are two other options for rating the behavior: 1) "8 - would occur if not prevented" (e.g., a person is physically restrained so he/she cannot pace), and 2) "9 - not applicable" (e.g., a non-verbal resident not being able to repeat sentences or questions, or a person who cannot walk or move a wheelchair not being able to pace, an amputated person not being able to kick). These two ratings are used only if the behavior really has never occurred in the past two weeks. If it has occurred, then the 1-7 point frequency scale should be used. These ratings of 8 and 9 are not included in the CMAI total score calculation. The CMAI has a range of 29-203 points, with higher scores indicating more severe agitation symptoms. If there are less than 6 missing items then the total score will be imputed by replacing each missing item with the mean (rounded to the nearest integer) of the non-missing values for that subject and timepoint. If there are more than 5 missing items then the total score will be missing.

In addition to the total score, behavioral subscales will be defined as follows (Rabinowitz et al, 2005):

- Aggressive Behavior: 12 Items (minimum = 12, maximum = 84)
- Sum of scores from items 3, 4, 7, 8, 9, 10, 11, 13, 14, 15, 21 and 25
- Physically Non-aggressive Behavior: 6 Items (minimum = 6, maximum = 42)
- Sum of scores from items 1, 2, 16, 22, 26, and 29
- Verbally Agitated Behavior: 4 Items (minimum = 4, maximum = 28)
  Sum of scores from items 5, 6, 18, and 19
- Hiding and Hoarding : 2 Items (minimum = 2, maximum =14)
  Sum of score from items 23 and 24

For the aggressive behavior subscale, if less than 3 items are missing then the aggressive behavior subscale score will be imputed by replacing the missing item with the mean of the non-missing values (rounded to the nearest integer) for that subject and timepoint within

aggressive behavior subscale. If more than 2 items are missing then the aggressive behavior subscale score will be missing.

For the physically non-aggressive subscale, if only 1 item is missing then the physically non-aggressive subscale score will be imputed by replacing the missing item with the mean of the non-missing values (rounded to the nearest integer) for that subject and timepoint within physically non-aggressive subscale. If more than 1 item is missing then the physically non-aggressive subscale score will be missing.

For the verbally agitated behavior and hiding and hoarding subscales, if any of the items are missing then the subscale score will be missing.

5.2.2 Zarit Burden Interview (ZBI)

The ZBI is assessed at Baseline, Weeks 2, 4, 8 and 12/ET visits.

The ZBI was designed to assess the stresses experienced by caregivers of patients with dementia. Caregivers are asked to respond to a series of 22 questions about the impact of the patient's disabilities on their life. For each item, caregivers are to indicate how often they felt that way (never, rarely, sometimes, quite frequently, or nearly always) (0-4). The total score of ZBI will range from 0 to 88, with higher scores denoting more stresses experienced by caregivers. If there are less than 5 missing items then the total score will be imputed by replacing each missing item with the mean (rounded to the nearest integer) of the non-missing values for that subject and timepoint. If there are more than 4 missing items then the total score will be missing.

If the informant of ZBI is a professional caregiver, their ZBI total scores will not be included in the summaries.

5.2.3 Alzheimer's Disease Cooperative Study – Activities of Daily Living Inventory (ADCS-ADL)

The ADCS-ADL is assessed at Baseline and Week 12/ET visits.

The ADCS-ADL is an inventory to assess activities of daily living (ADL) in subjects with AD. This is a caregiver-rated 23-item questionnaire, with sub-questions and sub-items for some items, including both basic (e.g. personal hygiene, eating, bathing) and instrumental (e.g. preparing a meal, using a telephone) ADLs. The recall period for this questionnaire is 4 weeks. The total score can range from 0 to 78, with lower scores indicating worse functioning.

A response of "DON'T KNOW" is not considered a missing value, rather it will be scored similarly to a "NO" response (i.e., that item score will be set to 0). The following table includes the item score range and the scoring rule. These rules will only apply in the analysis dataset; the raw data values will remain unchanged.

Table 2 ADCS-ADL Item Score Range

Item Number	Score Range	Note*
1-5, 8-12, 14, 17-19, 21, 22	0-3	For item 8, the score is sum of 8A, 8B, and 8C. The sub-question will be equal to 0 if the main or sub-question checks NO or DON'T KNOW. For item 18, the score is sum of 18A, 18B, and 18C. The sub-question will be equal to 0 if 1) the main or sub-question checks NO or DON'T KNOW; or 2) "patient is institutionalized" box is checked. For item 22, if NO or DON'T KNOW is checked then assign item score = 0. If YES is checked and one or more sub-item is checked, but the sub-question is not answered, then assign item score = 1. Otherwise, item score is equal to the non-missing sub-question score.
6	0-7	Sum of 6A (0-3; 0 if the main question checks NO or DON'T KNOW) and 6B (0-4).
7	0-5	
13, 15, 16, 23	0-4	For item 16, the score is sum of 16A (0-3; 0 if the main question checks NO or DON'T KNOW) and 16B (0-1; 0 if the main or sub-question checks NO or DON'T KNOW). For item 23, if NO or DON'T KNOW is checked then assign item score = 0. If YES is checked and one or more sub-item is checked, but the sub-question is not answered, then assign item score = 1. Otherwise, item score is equal to the non-missing sub-question score.
20	0-2	

* Item 7-21. If NO or DON'T KNOW is checked, then assign item score = 0. If YES is checked, the score of the item is sum of non-missing score(s) of the sub-question(s).

In addition to the total score, the following subscale scores will be defined:

- Basic subscale score: 6 Items (minimum = 0, maximum = 22)

  Sum of scores from items 1, 2, 3, 4, 5, 6
- Instrumental subscale score: 17 Items (minimum = 0, maximum = 56)
  Sum of scores from items 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23

After applying the preceding data handling rules, if missing values remain then the 23-item total, basic, and instrumental scores will be derived as follows:

For the 23-item total score, if the sum of the maximum possible scores among all of the missing items is less than or equal to 16 then the total score will be derived as 78 multiplied by (sum of observed item scores divided by the sum of maximum possible scores among observed items), rounded to the nearest integer. Otherwise, the total score will be missing.

For the basic subscale score, if the sum of the maximum possible scores among the missing items is less than or equal to 4 then the basic subscale score will be derived as 22 multiplied by (sum of observed basic item scores divided by the sum of the maximum possible scores among observed basic item scores), rounded to the nearest integer. If the sum of the maximum possible scores among the missing items is greater than 4 then the basic subscale score will be missing.

For the instrumental subscale score, if the sum of the maximum possible scores among the missing items is less than or equal to 11 then the instrumental subscale score will be derived as 56 multiplied by (sum of observed instrumental item scores divided by sum of the maximum possible scores among observed instrumental item scores), rounded to the nearest integer. If the sum of the maximum possible scores among the missing items is greater than 11 then the instrumental subscale score will be missing.

5.2.4 Neuropsychiatric Inventory – Clinician Rating Scale (NPI-C)

The complete NPI-C, which include agitation and aggression domains, will be administered at Baseline and Weeks 12/ET visits while only the NPI-C agitation and aggression domains will be administered at all other visits, i.e. Screening, and Weeks 2, 4, and 8 visits.

The NPI-C is a scale to assess the 14 neuropsychiatric symptoms (domains) in patients with dementia. For each domain, there are three NPI-C sections: caregiver interview, patient interview and clinician rating. The caregiver interview includes the frequency (0-4), severity (0-3) and distress (0-5) scores. The subject interview consists of the subject-reported frequency score (0-4). The clinician rating is a clinical impression severity score (0-3) based on all available interview information. For the caregiver interview, if a response of 0 is selected for the frequency score, then the severity and distress scores will be set to 0. For each domain, the domain score is the sum of the clinical impression severity scores within each domain, with higher scores denoting more serious symptoms. Before deriving any combined domain scores, if there are any missing items, the individual domain score will be imputed as follows:

For the delusions, hallucinations, aggression, elation/euphoria, aberrant motor disturbance, sleep disorders, appetite and eating disorders, and aberrant vocalizations domains, if only

1 item is missing then the respective domain score will be imputed by replacing the missing item with the mean of the non-missing items (rounded to the nearest integer) for that subject and timepoint within the domain. If more than 1 item is missing then the respective domain score will be missing.

For the agitation, dysphoria, anxiety, apathy/indifference, and irritability/lability domains, if up to 2 items are missing then the respective domain score will be imputed by replacing each missing item with the mean of the non-missing items (rounded to the nearest integer) for that subject and timepoint within the domain. If more than 2 items are missing then the respective domain score will be missing.

For the disinhibition domain, if up to 3 items are missing then the respective domain score will be imputed by replacing each missing item with the mean of the non-missing values (rounded to the nearest integer) for that subject and timepoint within the domain. If more than 3 items are missing then the respective domain score will be missing.

Table 3 NPI-C Domain Score

NPI-C Domain	# items	Domain Score Range	Maximum Missing Items Allowed for Deriving the Domain Score*
Delusions	8	0 – 24	1
Hallucinations	7	0 – 21	1
Agitation	13	0 – 39	2
Aggression	8	0 – 24	1
Dysphoria	13	0 – 39	2
Anxiety	14	0 – 42	2
Elation/Euphoria	6	0 – 18	1
Apathy/Indifference	11	0 – 33	2
Disinhibition	16	0 – 48	3
Irritability/Lability	12	0 – 36	2
Aberrant Motor Disturbance	9	0 – 27	1
Sleep Disorders	8	0 – 24	1
Appetite and Eating Disorders	9	0 – 27	1
Aberrant Vocalizations	8	0 – 24	1

* For each domain, the maximum number of missing items allowed for deriving the domain score is the nearest integer to but ≤ 20% of the total items in that domain. If there are any missing items, the individual domain score will be imputed first before calculating the combined domain scores.

The NPI-C combined agitation and aggression domain score is calculated as the sum of the agitation domain score and the aggression domain score. If either the agitation or aggression domain score is missing, the NPI-C combined agitation and aggression domain score will be missing.

The NPI-C combined delusions and hallucinations domain score is calculated as the sum of the delusions domain score and the hallucinations domain score. If either the delusions or hallucinations domain score is missing, the NPI-C combined delusions and hallucinations domain score will be missing.

The NPI-C combined dysphoria and apathy/indifference domain score is calculated as the sum of the dysphoria domain score and the apathy/indifference domain score. If either the dysphoria or apathy/indifference domain score is missing, the NPI-C combined dysphoria and apathy/indifference domain score will be missing.

The NPI-C total score is calculated as the sum of all 14 individual domain scores. The NPI-C total score ranges from 0 to 426, with higher scores denoting more serious symptoms. After applying the preceding data handling rules, if any of the domain scores remain missing, the total score will be missing.

5.2.5 Modified Alzheimer's Disease Cooperative Study – Clinical Global Impression-Change (mADCS-CGIC)

The mADCS-CGIC is assessed at Baseline, Weeks 2, 4, 8 and 12/ET visits.

The mADCS-CGIC scale will be used to allow the Investigator to determine the subject's overall clinical condition as it relates to their symptoms of agitation and aggression, and to address the clinical significance of changes from Baseline in other psychometric measures. The mADCS-CGIC interview will be performed by the Investigator or a medically qualified rater. After completion of the interview, the rater will be asked to rate the subject's symptoms of agitation and aggression relative to the Baseline interview, using a standardized 7-point scale (1 = marked improvement to 7 = marked worsening). Higher scores denote worsening or less improvement in agitation and aggression symptoms.

Missing mADCS-CGIC scores will not be imputed.

5.2.6 Karolinska Sleepiness Scale (KSS)

KSS is assessed at Baseline, Weeks 2, 4, 8 and 12/ET visits.

The KSS is a self-reported subjective measure of a subject's level of drowsiness. With the modified version, respondents must choose which of nine statements (1 = extremely alert to 9 = very sleepy, great effort to keep awake, fighting sleep) most accurately describes their level of sleepiness over a period of time, which for this study will be "on average over the previous week". Higher scores denote more drowsiness.

Missing KSS scores will not be imputed.

5.2.7 Mini-Mental State Examination (MMSE)

MMSE is assessed at Screening, Baseline, Weeks 4, 8 and 12/ET visits.

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. If there are less than or equal to 6 missing items then the total score will be imputed as the mean of the non-missing values multiplied by 30 and rounded to the nearest integer for that subject and timepoint. If there are more than 6 missing items then the total score will be missing.

5.2.8 Global Clinician Assessment of Suicidality (GCAS)

Global Clinician Assessment of Suicidality (GCAS) is assessed at Screening, Baseline, Weeks 2, 4, 8, 12/ET and Follow-up (if applicable) visits.

The GCAS will be used to assess the occurrence of treatment-emergent suicidal ideation and behavior. The GCAS is a clinician-rated, 5-point scale (0–4) 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 partner/caregiver rating, and a clinician rating. For a rating of 3 or 4 based on the clinician's assessment, the date of event will be recorded. At Screening visit lifetime suicidality and suicidality for the past 3 months will be assessed, at all other visits suicidality since the previous visit will be assessed.

Missing GCAS scores will not be imputed.

5.3 Analysis Visit Windows

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

Efficacy, safety, and PK assessments will be summarized by analysis visit as presented in Table 4 below.

Table 4 Analysis Visit Windows

Analysis Visit Name	Target Study Day[1]	Study Day Interval
Baseline	1	≤ 1
Week 2	15	2 to 21
Week 4	29	22 to 42
Week 8	57	43 to 70
Week 12	85	71 to 99
Follow-up	115	≥100

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

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, overall post-Baseline maximum and potentially clinically important values, all non-missing post-Baseline values should be considered, regardless of if the value is selected for the by-visit summaries. All results will be presented in data listings.

When ECG is collected in triplicate, the average of the triplicate will be considered as one assessment for the summaries.

5.4 Missing or Incomplete Date for Last Dose of Study Drug

In the Safety Analysis Set, if the last dose date of study drug is missing for a subject who completed or early terminated from the study, then the 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 the interim safety data review, if a subject is still ongoing, then this subject's last dose date will be imputed using the database extract date.

5.5 Missing or Incomplete Dates for Prior or Concomitant Medications

Missing or incomplete medication start or end dates will be imputed for the purpose of determining whether the medications are taken concomitantly (see Section 11 for definition). When the chronological order of a medication's utilization 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.

5.6 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.

5.7 Missing Severity Assessment for Adverse Events

If severity is missing for an AE starting prior to the date of the first dose of study drug, then a severity of "Mild" will be assigned. 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 presented 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 prior to the date of the first dose of study drug, then a causality of "Not Related" will be assigned. If the relationship to study drug is missing for an AE starting on or after the date of the first dose of study drug, then 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 summary due to, for example, a character string reported for a numerical 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

For subjects who participate in the screening phase but are not randomized (screen failures), their demographics information, screen failure reasons (the specific inclusion/exclusion criterion (or criteria) not met) 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.

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 by region and overall.

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

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

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 (significant) or minor (not significant) are detailed in the Study Deviation Rules Document.

A summary of the number and percentage of subjects with major protocol deviations for each deviation category will be presented by treatment group and overall. A listing of protocol deviations by site and subject will be provided.

8 DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS

Demographics and Baseline characteristics will be summarized by treatment group and overall for Randomized Analysis Set, Safety Analysis Set, Full Analysis Set, and Per-protocol Analysis Set using descriptive statistics. For the Full Analysis Set, summaries by region will be presented. Variables include age, sex, race, ethnicity, height, weight, body mass index (BMI), region, Baseline living situation category, Baseline caregiver category, Baseline NPI-C combined agitation and aggression domain score, Baseline MMSE total score, Baseline CMAI total score, Baseline ZBI total score, and concomitant selective serotonin reuptake inhibitor (SSRI) usage (yes/no).

The reported age reflects a subject's age at the informed consent date. Age, Baseline CMAI total score, Baseline MMSE total score, and Baseline NPI-C combined agitation and aggression domain score will be presented as both continuous and categorical variables. Age will be presented as ≤85 and >85 years old. Category of Baseline CMAI will be presented as <65 and ≥65. Category of Baseline NPI-C combined agitation and aggression will be presented as <22 and ≥22. Category of Baseline MMSE will be presented as ≤10, 11-20 and >20.

Alzheimer's disease history will be summarized by treatment group for Randomized Analysis Set, Safety Analysis Set, Full Analysis Set and Per-protocol Analysis Set using descriptive statistics. Variables include:

- Time (year) since diagnosis of probable Alzheimer's disease
- Duration (year) of symptoms of Alzheimer's disease
- Duration (year) of symptoms of agitation/aggression

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

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

9 MEDICAL HISTORY

Medical and surgical history data will be coded using Medical Dictionary for Regulatory activities (MedDRA) version 19.0 or newer. The subject incidence of relevant medical and surgical history data will be summarized by system organ class (SOC) and preferred term by treatment group and overall for Safety Analysis Set, Full Analysis Set, and Per-protocol Analysis Set. A subject will be counted only once per SOC or 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 will be provided.

10 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.

10.1 Exposure to Study drug

For each subject, duration of exposure to study drug will be calculated as (last dose date – first dose date + 1). Duration of exposure will be summarized as both continuous and categorical variables by treatment group and overall. For categorical presentation, the number and percentage of subjects in each of the following categories will be presented: <4 weeks (1 to 27 days), 4 to <8 weeks (28 days to 55 days), 8 to <12 weeks (56 days to 83 days), and ≥12 weeks (84 days or longer). Kaplan-Meier curves of duration on study drug will also be presented for each treatment group.

10.2 Measurement of Treatment Compliance

Study drug dosing compliance (in percentage) for a subject is defined as the total number of tablets actually taken 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 total number of tablets returned. The number of tablets expected to be taken is calculated as the duration of exposure (last dose date – first dose date + 1) multiplied by 2 (the planned number of tablets taken per day).

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

11 PRIOR, CONCOMITANT AND POST MEDICATION

For a subject, 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 after the first dose of study drug or with a start date between the dates of the first and last doses of study drug, inclusive. Any medication with a start date after the date of the last dose of study drug will be considered as post-treatment medication. Prior, concomitant, or post-treatment medications will be summarized separately.

Medications will be coded using WHO Drug Dictionary (WHODD) 2016 March 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 ATC category will be counted only once.

12 EFFICACY ANALYSES

All efficacy summaries will be performed using the planned treatment assignments based on the randomization schedule. No hypothesis testing is planned. Descriptive summaries of all efficacy endpoints will be provided. The summaries based on mixed model for repeated measures (MMRM) or analysis of covariance (ANCOVA) will include point estimates and confidence intervals.

12.1 Efficacy Variables

12.1.1 Primary Efficacy Variable

The primary efficacy endpoint is change from Baseline to Week 12 in the CMAI total score.

12.1.2 Secondary Efficacy Variable

The secondary efficacy endpoint is change from Baseline to Week 12 in the ZBI total score.

12.1.3 Exploratory Efficacy Variables

The exploratory efficacy endpoints include the following:

- mADCS-CGIC agitation score
- Change from Baseline in NPI-C combined agitation and aggression domain scores
- Change from Baseline in NPI-C individual agitation and aggression domain scores
- Change from Baseline to Week 12 in the ADCS-ADL score
- Change from Baseline to Week 12 in NPI-C total score

- Change from Baseline to Week 12 in NPI-C sleep disorders domain score
- Change from Baseline to Week 12 in NPI-C combined delusions and hallucinations domain scores
- Change from Baseline to Week 12 in NPI-C combined dysphoria and apathy/indifference domain scores
- Change from Baseline to Week 12 in individual NPI-C domain scores (other than agitation, aggression and sleep disorders)
- Change from Baseline in CMAI subscale scores
- Change from Baseline in KSS score
- Change from Baseline in the MMSE score
- The proportion of subjects taking any rescue medication

12.2 Adjustment for Covariates

For continuous variables summarized using MMRM or ANCOVA, the Baseline value of the endpoint being summarized (except mADCS-CGIC) and geographic region (North America, Europe, or Rest of World) will be included as covariates as described in Section 13.

12.3 Handling of Missing Data

The primary efficacy variable will be summarized assuming missing at random using an MMRM method. Total scores that are missing, after any imputation of individual missing items as described in Section 5.2.1, will not be imputed.

12.4 Multiple Comparisons / Multiplicity

No hypothesis testing is planned.

12.5 Examination of Subgroups

Subgroup analyses will be performed with respect to the primary and secondary efficacy variables. The subgroups will include the following:

- Geographic region: North America, Europe, or Rest of World
- Race: White or Non-white
- Sex: Female or Male
- Age group: ≤85 years or >85 years
- Baseline Living situation: Home or Assisted living/other
- Baseline Caregiver category: Professional caregiver or non- professional caregiver

- Baseline NPI-C combined agitation and aggression score: <22 or ≥22
- Baseline CMAI total scores: < 65 or ≥65
- Baseline MMSE scores: ≤10 or 11-20 or >20
- Concomitant selective serotonin reuptake inhibitor (SSRI) usage: Yes or No

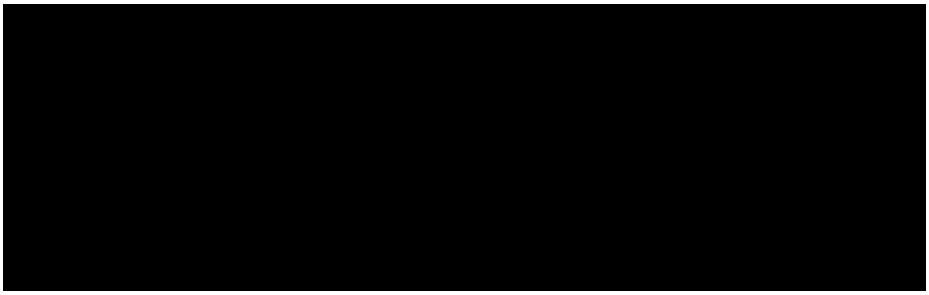
13 METHODS OF EFFICACY ANALYSES

13.1 Primary Efficacy Analysis

The primary efficacy endpoint is the change from Baseline to Week 12 in the CMAI total score. The primary endpoint will be summarized based on the Full Analysis Set.

The CMAI total score will be summarized using the Mixed Model Repeated Measures (MMRM) in the Full Analysis Set. The dependent variable will be the change from Baseline in the CMAI total score. The independent variables in the model will include the following: treatment group (pimavanserin 34 mg, pimavanserin 20 mg, or placebo), visit (Weeks 2, 4, 8, and 12), the treatment-by-visit interaction, the Baseline CMAI total score, the Baseline-by-visit interaction, and geographic region (North America, Europe, or Rest of World). An unstructured covariance matrix will be used and the Kenward-Roger approximation will be used to adjust the denominator degrees of freedom. Observed cases will be used in the model and the least squares (LS) means will be estimated using observed margins.

Sample SAS® code for the MMRM:



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

At each visit, the effect size (Cohen's d) for the change from Baseline between the treatment groups (each pimavanserin group versus placebo group) 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 an effect size favoring pimavanserin will be reported as a positive number and an effect size favoring placebo will be reported as a negative number.

Summary statistics for the CMAI total score (observed and change from Baseline) will be presented for all visits from Baseline through Week 12. 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% confidence interval and effect size will also be presented. In addition, LS mean ± SE over time for the change from Baseline values by treatment group will also be presented in line plots.

The MMRM summary as described above will also be repeated in the Per-protocol Analysis Set.

13.2 Secondary Efficacy Analysis

The secondary efficacy endpoint is the change from Baseline to Week 12 in the ZBI total score. The secondary endpoint will be summarized using an MMRM model in the Full Analysis Set similar to that described above for the primary endpoint in Section 13.1, except that the Baseline ZBI total score will be included in the MMRM model instead of the Baseline CMAI total score.

The MMRM summary as described above will also be repeated in the Per-protocol Analysis Set.

Results will be presented in summary table and line plots similar to the outputs described in Section 13.1.

13.3 Exploratory Efficacy Analyses

The Exploratory efficacy endpoints include the following:

- mADCS-CGIC agitation score
- Change from Baseline in NPI-C combined agitation and aggression domain scores
- Change from Baseline in NPI-C individual agitation and aggression domain scores

- Change from Baseline to Week 12 in the ADCS-ADL score
- Change from Baseline to Week 12 in NPI-C total score
- Change from Baseline to Week 12 in NPI-C sleep disorders domain score
- Change from Baseline to Week 12 in NPI-C combined delusions and hallucinations domain scores
- Change from Baseline to Week 12 in NPI-C combined dysphoria and apathy/indifference domain scores
- Change from Baseline to Week 12 in individual NPI-C domain scores (other than agitation, aggression and sleep disorders)
- Change from Baseline in CMAI subscale scores
- Change from Baseline in KSS score
- Change from Baseline in the MMSE score
- The proportion of subjects taking any rescue medication

The mADCS-CGIC agitation score at each timepoint will be summarized using an MMRM model with effects for treatment group (pimavanserin 34 mg, pimavanserin 20 mg, or placebo), visit (Week 2, Week 4, Week 8, and Week 12), the treatment-by-visit interaction, and geographic region (North America, Europe, or Rest of World). There is no Baseline value to include in the model. An unstructured covariance matrix will be used and the Kenward-Roger approximation will be used to adjust the denominator degrees of freedom. Observed cases will be used in the model and the LS means will be estimated using observed margins.

The change from Baseline to each post-Baseline timepoint in the NPI-C combined agitation and aggression domain scores, NPI-C individual agitation and aggression domain scores, CMAI subscale scores, and KSS will be summarized using an MMRM model similar to that described above for the primary endpoint in Section 13.1, except that the Baseline value of the endpoint being summarized will be included in the model instead of the Baseline CMAI total score.

The change from Baseline to Week 12 in the ADCS-ADL and NPI-C total score, combined delusions and hallucinations, combined dysphoria and apathy/indifference and individual domain scores (except the agitation and aggression domains) will be summarized using an analysis of covariance (ANCOVA) model with effects for treatment group (pimavanserin 34 mg, pimavanserin 20 mg, or placebo) and geographic region (North America, Europe, or Rest of World) as factors, and the Baseline value of the endpoint being summarized as a covariate. Missing values will be imputed with the previous value (including Baseline) for

that subject, i.e. last observation carried forward (LOCF). In addition, an observed-case summary, ignoring missing values, will also be presented.

At Week 12, the effect size (Cohen's d) for the change from Baseline between the treatment groups (each pimavanserin group versus placebo group) will be calculated using the following formula:

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

Where MSE is the mean squared error from the ANCOVA model. The sign (+ or -) of the effect size will be chosen so that an effect size favoring pimavanserin will be reported as a positive number and an effect size favoring placebo will be reported as a negative number.

The change from Baseline to each post-Baseline timepoint (Weeks 4, 8 and 12) in the MMSE will be summarized using an MMRM model in Full Analysis Set similar to that described above for the primary endpoint in Section 13.1 , except that the Baseline value of the MMSE will be included in the model instead of the Baseline CMAI total score. Also, the MMSE is not measured at Week 2.

Results will be presented in summary table similar to the outputs described in Section 13.1.

The use of rescue medication is determined from the Rescue Medications eCRF. The proportion of subjects taking any rescue medication during the treatment period for the Full Analysis Set will be summarized by treatment group.

Correlations among efficacy endpoints (observed cases) at Baseline (combining all 3 treatment groups) and for change from Baseline at Week 12 (combining all 3 treatment groups and by treatment group) will be assessed using Spearman's rank correlation coefficient (Spearman's rho) and presented in a correlation matrix. The 95% CIs for the Spearman's rank correlation coefficients will also be presented.

13.4 Responder Analyses

In the responder analyses, since CMAI is an interval scale that lacks a natural zero point (1 = Never), the percent change in CMAI total score will be calculated based on corrected scores after subtracting 29 points from the raw scores. For example, if a subject's Baseline CMAI total score is 79 and Week 12 CMAI total score is 29 (absent of all agitation behaviors), the percent change from Baseline to Week 12 in CMAI total score will be calculated as [(29– 29) – (79 – 29)] ÷ (79– 29) × 100% = -100%.

Multiple types of responders will be summarized within each treatment group at each post-Baseline visit:

- ≥20%, ≥30%, ≥50%, ≥75% and 100% reduction in the CMAI total score from Baseline
- ≥20%, ≥30%, ≥50%, ≥75% and 100% reduction in NPI-C combined agitation and aggression domain scores from Baseline
- ≥20%, ≥30%, ≥50%, ≥75% and 100% reduction in the individual agitation and aggression NPI-C domain scores from Baseline
- mADCS-CGIC score of 1 (marked improvement) or 2 (moderate improvement)

Other responder criteria based on point changes may be included based on examination of cumulative distribution functions and anchoring to the ADCS-CGIC.

For each of these responder analyses, the proportion of responders will be summarized by treatment group at each timepoint using observed cases and also with missing values imputed as non-response. The adjusted difference in percent responders between the treatment groups (pimavanserin group minus placebo group) using the weighting scheme of Cochran-Mantel-Haenszel and Newcombe's 95% CI stratified by geographic region will be presented. In addition, percent responders by treatment group at Week 12 will also be presented in bar charts.

14 SAFETY ANALYSES

The safety summaries will be presented based on the Safety Analysis Set using actual treatment. Safety variables include adverse events (AEs), clinical laboratory variables, vital signs, physical examinations, electrocardiogram (ECG) and Global Clinician Assessment of Suicidality (GCAS) variables. Safety variables will be summarized by treatment group using descriptive statistics.

14.1 Adverse Events

All adverse events will be coded using the MedDRA Version 19.0 or newer.

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

The event counts and the number and percentage of subjects reporting TEAEs in each treatment group will be tabulated by SOC and preferred term; and, by SOC, preferred term,

and maximum severity. 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 occurrence for the summarization by severity. In addition, the event counts and the number and percentage of subjects with TEAEs classified by the Investigators as related to the study drug, with most frequently reported TEAEs (preferred terms reported by ≥ 5% of subjects in any treatment group), with treatment-emergent serious AEs (TESAEs), with fatal AEs (i.e. events that cause death), and with TEAEs leading to discontinuation of study drug will be summarized by SOC and preferred term within each treatment group. These tables will be sorted alphabetically by SOC and then by descending subject frequency for the preferred terms (combined counts from all treatment groups) within each SOC.

The event counts and the number and percentage of subjects with any TEAEs will also be tabulated by preferred term without SOCs within each treatment group. This table will be sorted by descending subject frequency using combined counts from all treatment groups.

In addition, for the SAEs and death, the following estimated odds ratios and their 95% confidence intervals will be calculated using an exact logistic regression model with a single fixed effect for treatment group:

- Both active arms combined versus placebo;
- Pimavanserin 20 mg versus placebo;
- Pimavanserin 34 mg versus placebo.

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 preferred term. This listing will also include all relevant eCRF data associated with the event: e.g. date of onset, date resolved, date of first dose, date of last dose, severity, frequency, outcome, relationship to study drug, action taken with study drug, and required therapy. When a date is presented, the study day associated with the date will also be displayed. Separate listings will be presented for subjects with SAEs, subjects with AEs leading to discontinuation and subjects who died (if any). In these listings, an indicator for treatment-emergent events will also be included.

14.2 Clinical Laboratory Variables

Clinical laboratory tests are performed at Screening, Baseline, Weeks 4 and 12/ET visits.

- Chemistry serum tests include the following
  - Sodium (Na), potassium (K), chloride (Cl), phosphorus (P), calcium (Ca), carbon dioxide (CO_2), blood urea nitrogen (BUN), creatinine (CR), uric acid

- Alanine aminotransferase (ALT/SGPT), aspartate aminotransferase (AST/SGOT), gamma-glutamyl transpeptidase (GGT), alkaline phosphatase (ALP), total bilirubin (TBIL), lactate dehydrogenase (LDH)
- HbA1c, glucose
- Albumin (ALB), total protein
- Prolactin
- Creatine kinase (CK)/creatine phosphokinase (CPK)
- Lipid panel
    - Total cholesterol, HDL-cholesterol, triglycerides, LDL-cholesterol, cholesterol/HDL ratio, Non-HDL cholesterol
- Serum pregnancy test for women of childbearing potential
- Vitamin B12 assay (completed only at Screening)

- Hematology tests include the following:
  - Complete blood count (CBC) including:
    - White blood cell (WBC) count
    - Complete differential (relative and absolute)
    - Hematocrit (Hct), hemoglobin, red blood cells (RBC), platelets
    - Reticulocytes

- Urinalysis tests include the following:
  - Blood, RBCs, WBCs, protein, glucose, ketones, specific gravity, pH
  - Urine pregnancy test for women of childbearing potential.

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

Clinical laboratory values for hematology, chemistry and urinalysis will be summarized by treatment group using descriptive statistics at Baseline, Week 4, and Week 12 visits. The change from Baseline values will also be summarized by treatment group at the Week 4 and Week 12 visits. The overall minimum and maximum post-Baseline observed and change from Baseline values will also be summarized. For hemoglobin, hematocrit and uric acid, the above summaries will be presented for each gender as well as for both genders combined. For urinalysis with categorical results, the number and percentage of subjects will be tabulated by category at Baseline, Week 4, and Week 12 visits. For the categorical urinalysis by-visit

summary, the denominator is the total number of subjects with non-missing values for the given parameter, visit and treatment group.

The 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 (low), within (normal) and above (high) normal ranges at Week 4 and Week 12 visits relative to the same classification at the Baseline visit. The shifts from Baseline to overall post-Baseline minimum and overall post-Baseline maximum will also be presented. 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 summaries of shift to the overall post-Baseline minimum or maximum, 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, the shift summaries will be presented for each gender as well as for both genders combined.

Number and percentage of subjects with potentially clinically important (PCI) laboratory values at Week 4, Week 12 visits and overall post-Baseline will be summarized by treatment group and by Baseline status (all and within normal range) for selected parameters. PCI criteria are listed in Table 5 and Table 6. 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 laboratory 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 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 any PCI values will be presented in an additional listing.

Table 5 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 6 Criteria for Potentially Clinically Important Laboratory Values – Urinalysis

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

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

14.3 Vital Signs

Vital signs are assessed at Screening, Baseline, Weeks 2, 4, 8 and 12/ET visits.

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

Vital sign values will be considered PCI if they meet both the observed value criteria and the change from Baseline criteria listed in Table 7. The number and percentage of subjects with post-Baseline values that are PCI will be summarized by treatment group at each post-Baseline visit and for overall post-Baseline. For the overall post-Baseline summaries, all post-Baseline values will be considered, including unscheduled and out of window values. Subjects with multiple PCI values for a given parameter will be counted only once for that parameter. For the by-visit summary, the numerator for the percentage is the number of subjects with a post-Baseline PCI vital sign for the given parameter, visit and treatment group and the denominator is the number of subjects with non-missing values for the given parameter, visit and treatment group. For the overall post-Baseline summary, the numerator for the percentage is the number of subjects with at least 1 post-Baseline PCI vital sign for a given parameter and treatment group and the denominator is the number of subjects at least 1 post-Baseline vital sign for the given parameter and treatment group. Subjects with any PCI values will be presented in an additional listing.

Table 7 Criteria for Potentially Clinically Important Vital Signs

Vital Sign Parameter	Unit	Criteria[a]		
		Observed Value	And/Or	Change Relative to Baseline
Systolic blood pressure	mmHg	≥180	And	Increase of ≥20
		≤90	And	Decrease of ≥20
Diastolic blood pressure	mmHg	≥105	And	Increase of ≥15
		≤50	And	Decrease of ≥15
Pulse	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.

14.4 Electrocardiogram (ECG)

Electrocardiogram is performed at Screening, Baseline, Weeks 4 and 12/ET visits. All tracings will be evaluated by a central reading laboratory. ECG data summaries will be performed using the centrally evaluated data.

Observed values of ECG variables (heart rate, PR interval, RR interval, QRS interval, QT interval, and QTc intervals) and the changes from Baseline at each assessment timepoint will be summarized by treatment group at Baseline, Week 4, and Week 12 visits.

QTc intervals include QTcB (Bazett's formula) and QTcF (Fridericia's formula). 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 overall post-Baseline maximum:

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

Electrocardiogram variable values will be considered PCI if they meet or exceed the upper limit values listed in Table 8. 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 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 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 values will be provided.

Table 8 Criteria for Potentially Clinically Important ECG Values

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

14.5 Physical Examination

Physical examination is performed at Screening and Week 12/ET visits.

Physical examination results (normal, abnormal, and not done) at Screening and Week 12 will be summarized in a frequency table by treatment group, body system and visit.

14.6 Other Safety Variables

14.6.1 Suicidal Ideation and Behavior

GCAS is assessed at Screening, Baseline, Weeks 2, 4, 8, 12/ET and Follow-up (if applicable) visits.

The number and percentage of subjects for each GCAS rating (0-4) based on clinician's assessment will be tabulated by treatment group and visit. The number and percentage of subjects reporting any post-Baseline GCAS score of 3 or 4 based on clinician's assessment will also be tabulated for each treatment group.

15 CLINICAL PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES

For pimavanserin-treated subjects in the Pharmacokinetics Analysis Set, plasma concentration data for pimavanserin and its metabolite (AC-279) will be listed and summarized using descriptive summary statistics by visit and treatment group. Concentrations that are below the limit of quantification (BLQ) will be displayed as "BLQ" in the data listings and imputed as 0 for computing summary statistics.

16 INTERIM ANALYSIS

No interim analysis is planned in this study.

17 DATA MONITORING/REVIEW COMMITTEE

An independent Safety Monitoring Committee (SMC) will review interim safety data including data on AEs and SAEs. The SMC will be independent of the Sponsor and the Investigators and will be empowered to recommend stopping the study due to safety concerns. The roles and responsibilities of SMC members and planned frequency of meetings are detailed in the SMC Charter.

An independent statistician (and/or programmer) not affiliated with the Sponsor will produce unblinded statistical outputs and provide these outputs to SMC members using a secure method. The Sponsor and the Investigators will remain blinded until the official unblinding of the database at the end of the study. The outputs presented to SMC members will include but not limited to summaries of enrollment and disposition, demographics and Baseline characteristics, medical and AD histories, concomitant medications, study drug exposure, all adverse events (including deaths, SAE and AEs leading to discontinuation), vital signs, laboratory test results, and ECG parameters. Subject profiles, boxplots of Baseline and most extreme post-Baseline values for clinical laboratory data, and listings of AEs and potentially clinically important laboratory and QTcF results will also be provided to SMC members for their review.

18 COMPUTER METHODS

All data summaries will be performed using Version 9.3 (or newer) of SAS® (SAS® Institute, Inc., Cary, North Carolina) on a suitably qualified and validated environment.

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

19 CHANGES TO ANALYSES SPECIFIED IN PROTOCOL

None.

20 REFERENCES

Rabinowitz J, Davidson M., De Deyn PP., Katz I., Brodaty H., Cohen-Mansfield J. (2005). "Factor Analysis of the Cohen-Mansfield Agitation Inventory in Three Large Samples of Nursing Home Patients With Dementia and Behavioral Disturbance" The American Journal of Geriatric Psychiatry; 13(11): 991–998.

21 APPENDICES

21.1 Summary of Version Changes

Version No:	Document History Description of Update	Author	Effective Date
		[redacted]	