Official Title: A Double-blind, Placebo-controlled, Relapse Prevention Study of Pimavanserin for the Treatment of Hallucinations and Delusions Associated With Dementia-related Psychosis

NCT Number: NCT03325556

Document Date: 26 Jul 2019

STATISTICAL ANALYSIS PLAN

Protocol No.:	ACP-103-045
Protocol Title:	A Double-blind, Placebo-controlled, Relapse Prevention Study of Pimavanserin for the Treatment of Hallucinations and Delusions Associated With Dementia-related Psychosis
Drug:	Pimavanserin
Sponsor:	ACADIA Pharmaceuticals Inc. [redacted]
Version No. and Date	Version 1.0, 26JUL2019

SIGNATURE/APPROVAL PAGE

AUTHOR
See electronic signature page _____ _____ [redacted] Date Associate Director, Biostatistics ACADIA Pharmaceuticals Inc.

APPROVERS
See electronic signature page _____ _____ [redacted] Date Vice President, Biometrics ACADIA Pharmaceuticals Inc.

Signature Page for ACP-103-045 SAP

Approve	[redacted] 29-Jul-2019 15:57:16 GMT+0000

Approve	[redacted] 29-Jul-2019 16:05:45 GMT+0000

Signature Page for ACP-103-045 SAP

TABLE OF CONTENTS

TABLE OF CONTENTS 4
LIST OF TABLES 8
LIST OF FIGURES 8
ABBREVIATIONS 9
1 INTRODUCTION 11
2 STUDY OBJECTIVES AND ENDPOINTS 11
2.1 Primary Objective 11
2.1.1 Primary Endpoint 11
2.2 Key Secondary Objective 11
2.2.1 Key Secondary Endpoint 11
2.3 Exploratory Objectives 11
2.3.1 Exploratory Endpoints 12
2.4 Safety Objective 12
2.4.1 Safety Endpoints 12
2.5 Pharmacokinetic Objectives 13
2.5.1 Pharmacokinetic Endpoints 13
3 STUDY DESIGN 13
3.1 General Study Design 13
3.2 Randomization 21
3.3 Blinding 21
3.4 Determination of Sample Size 21
4 ANALYSIS SETS 22
4.1 Intent-to-treat (ITT) Analysis Set 22
4.2 Open-label Safety Analysis Set 22
4.3 Double-blind Safety Analysis Set 22
4.4 Per-protocol Analysis Set 23
4.5 Open-label Pharmacokinetic Analysis Set 23
4.6 Double-blind Pharmacokinetic Analysis Set 23
5 DATA HANDLING CONVENTIONS 23
5.1 General Data Reporting Conventions 23

5.1.1 Open-label Period 24
5.1.2 Double-blind Period 24
5.2 Derived Variables....25
5.2.1 Open-label Baseline....25
5.2.2 Double-blind Baseline 25
5.2.3 Time from Randomization to Relapse in the Double-blind Period 25
5.2.4 Time from Randomization to Discontinuation from the Double-blind Period for Any Reason....26
5.2.5 Scale for the Assessment of Positive Symptoms-Hallucinations + Delusions 27
5.2.6 Clinical Global Impressions – Severity Scale 27
5.2.7 Clinical Global Impressions – Improvement Scale 28
5.2.8 Zarit Burden Interview 28
5.2.9 Karolinska Sleepiness Scale 28
5.2.10 EQ-5D-5L Proxy Version 1 29
5.2.11 Global Clinician Assessment of Suicidality 29
5.2.12 Mini-Mental State Examination 30
5.2.13 Extrapyramidal Symptom Rating Scale-Abbreviated 30
5.3 Analysis Visit Windows....31
5.3.1 Open-label Period 31
5.3.2 Double-blind Period 32
5.3.3 Unscheduled Assessments 32
5.3.4 Multiple Measurements within Visit Windows 33
5.4 Data Handling Conventions....33
5.4.1 Missing or Incomplete Dates for Prior or Concomitant Medications 34
5.4.2 Missing or Incomplete Dates for Adverse Events 34
5.4.3 Missing Severity Assessment for Adverse Events....34
5.4.4 Missing Relationship to Study Drug for Adverse Events....34
5.4.5 Character Values of Clinical Laboratory Variables....35
6 SUBJECT DISPOSITION....35
6.1 Analysis for the Open-label Period 35
6.2 Analysis for the Double-blind Period....36
6.3 Screen Failure 36
7 PROTOCOL DEVIATIONS....37
8 DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS 37
8.1 Demographics 37
8.2 Disease Characteristics 38

9 MEDICAL HISTORY 38
10 PRIOR, CONCOMITANT AND POST-TREATMENT MEDICATION 38
10.1 Prior Medication 39
10.2 Concomitant Medication 39
10.2.1 Analysis for the Open-label Period 39
10.2.2 Analysis for the Double-Blind Period 39
10.3 Analysis for Post-Treatment Medication 39
11 EXTENT OF EXPOSURE AND TREATMENT COMPLIANCE 40
11.1 Exposure to Study Drug 40
11.1.1 Analysis for the Open-label Period 40
11.1.2 Analysis for the Double-blind Period 40
11.2 Study Drug Compliance 41
11.2.1 Analysis for the Open-label Period 41
11.2.2 Analysis for the Double-blind Period 41
12 EFFICACY ANALYSES 41
12.1 Efficacy Variables 42
12.1.1 Primary Endpoint 42
12.1.2 Key Secondary Endpoint 42
12.1.3 Exploratory Study Endpoints 42
12.2 Adjustment for Covariates 42
12.3 Handling of Informative Censoring 43
12.4 Multiple Comparisons/Multiplicity 43
12.5 Examination of Subgroups 43
12.5.1 Double-Blind Period 43
12.5.2 Open-Label Period 44
13 METHODS OF EFFICACY ANALYSES 44
13.1 Primary Efficacy Analysis 44
13.1.1 Deviations From Non-Informative Censoring 46
13.1.2 Sensitivity Analyses 48
13.2 Key Secondary Efficacy Analysis 49
13.2.1 Deviations From Non-Informative Censoring 51
13.2.2 Sensitivity Analysis 51
13.3 Exploratory Efficacy Analysis 52
13.3.1 Analysis for the Double-blind Period 52

13.3.1.1 Mixed Model Repeated Measures 52
13.3.1.2 ANCOVA 53
13.3.1.3 Treatment Comparisons 53
13.3.2 Analysis for the Open-label Period 54
13.3.3 SAPS-H+D 54
13.3.4 SAPS Hallucinations Domain 55
13.3.5 SAPS Delusions Domain 55
13.3.6 CGI-S Dementia-Related Psychosis 56
13.3.7 Responder Analysis 56
13.3.8 ZBI Score 56
13.3.9 KSS Score 56
13.3.10 CGI-I Dementia-Related Psychosis 57
13.3.11 EQ-5D VAS 57
13.3.12 EQ-5D 5L Descriptive System 57
13.3.13 Pareto Classification of Health Change 58
14 SAFETY ANALYSES 58
14.1 Adverse Events 59
14.2 Clinical Laboratory Variables 60
14.2.1 Analysis for the Open-label Safety Analysis set 62
14.2.2 Analysis for the Double-blind Safety Analysis Set 63
14.3 Vital Signs 66
14.3.1 Analysis for the Open-label Safety Analysis Set 66
14.3.2 Analysis for the Double-blind Safety Analysis Set 67
14.4 Electrocardiogram 68
14.4.1 Analysis for the Open-label Safety Analysis Set 68
14.4.2 Analysis for the Double-blind Safety Analysis Set 69
14.5 Physical and Neurological Examinations 71
14.5.1 Analysis for the Open-label Safety Analysis Set 71
14.5.2 Analysis for the Double-blind Safety Analysis Set 71
14.6 Global Clinician Assessment of Suicidality 71
14.6.1 Analysis for the Open-label Safety Analysis Set 71
14.6.2 Analysis for the Double-blind Safety Analysis Set 72
14.7 Mini-Mental State Examination 72
14.7.1 Analysis for the Open-label Safety Analysis Set 72
14.7.2 Analysis for the Double-blind Safety Analysis Set 72
14.8 Extrapyramidal Symptom Rating Scale - Abbreviated 73

14.8.1 Analysis for the Open-label Safety Analysis Set 73
14.8.2 Analysis for the Double-blind Safety Analysis Set 73
15 CLINICAL PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES 73
15.1 Analysis for the Open-label Pharmacokinetic Analysis Set 74
15.2 Analysis for the Double-blind Pharmacokinetic Analysis Set 74
16 UNBLINDED INTERIM EFFICACY ANALYSIS 74
16.1 Primary Efficacy Endpoint 75
16.2 Key Secondary Efficacy Endpoint 77
16.2.1 Overrunning of Relapse Events at Interim Analysis 78
16.2.2 Analysis Following a Sequential Test 78
17 DATA AND SAFETY MONITORING BOARD 78
18 INDEPENDENT ADJUDICATION COMMITTEE 79
19 COMPUTER METHODS 79
20 CHANGES TO ANALYSES SPECIFIED IN PROTOCOL 79
21 REFERENCES 80
22 APPENDICES 81
Appendix 1 Summary of Version Changes 81

LIST OF TABLES

Table 1 Schedule of Assessments for Screening and Open-label Period 17
Table 2 Schedule of Assessments for Double-blind Period 19
Table 3 Analysis Visit Windows for the Open-label Period 32
Table 4 Analysis Visit Windows for the Double-Blind Period 32
Table 5 Criteria for PCI Laboratory Values – Hematology and Chemistry 65
Table 6 Criteria for PCI Laboratory Values - Urinalysis 66
Table 7 Criteria for PCI Vital Signs 68
Table 8 Criteria for PCI ECG Values 71
Table 9 O'Brien-Fleming Stopping Boundary Information 75

LIST OF FIGURES

Figure 1 Schematic of Study Design 14

ABBREVIATIONS

AE	adverse event
ATC	Anatomical Therapeutic Chemical
BMI	body mass index
CI	confidence interval
CGI-I	Clinical Global Impression-Improvement
CGI-S	Clinical Global Impression-Severity
DB	double-blind
DSMB	Data and Safety Monitoring Board
eCRF	electronic case report form
ECG	electrocardiogram
EOT	end-of-treatment
FDA	Food and Drug Administration
FTD	frontotemporal dementia
GCAS	Global Clinician Assessment of Suicidality
IAC	Independent Adjudication Committee
ISG	Independent Statistical Group
ITT	intent-to-treat
KSS	Karolinska Sleepiness Scale
LLOQ	lower limit of quantification
LS	least squares
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed model repeated measures
MMSE	Mini-Mental State Examination
msec	milliseconds
OL	open-label
PCI	potentially clinically important
PK	pharmacokinetic
QTcF	QT interval corrected for heart rate using Fridericia's formula
SAE	serious adverse event
SAP	statistical analysis plan
SAPS-H+D	Scale for Assessment of Positive Symptoms-Hallucinations + Delusions
SAS®	Statistical Analysis System
SD	standard deviation

SE	standard error
SOC	system organ class
TEAE	treatment-emergent adverse event
ZBI	Zarit Burden Interview

1 INTRODUCTION

This statistical analysis plan (SAP) provides a technical and detailed elaboration of the statistical analyses of efficacy and safety data as described in Protocol Amendment 1 dated 16 August 2018. Specifications for tables, figures, and listings are contained in a separate document.

2 STUDY OBJECTIVES AND ENDPOINTS

2.1 Primary Objective

The primary objective is to evaluate relapse prevention in subjects with dementia-related psychosis treated with pimavanserin compared to placebo.

2.1.1 Primary Endpoint

The primary endpoint for this study will be time from randomization to relapse in the double-blind period.

2.2 Key Secondary Objective

The key secondary objective is to evaluate the time to discontinuation of the study for any reason in subjects with dementia-related psychosis treated with pimavanserin compared to placebo.

2.2.1 Key Secondary Endpoint

The key secondary endpoint for this study will be time from randomization to discontinuation from the double-blind period for any reason.

2.3 Exploratory Objectives

The exploratory objectives are to evaluate the benefit of pimavanserin compared to placebo in the following domains in subjects with dementia-related psychosis who have been stabilized on pimavanserin therapy:

- symptoms of hallucinations and delusions
- clinical global impression
- caregiver burden

- daytime sleepiness
- quality of life

2.3.1 Exploratory Endpoints

The exploratory endpoints for this study are as follows:

- Change from double-blind Baseline on the following:
    - Scale for the Assessment of Positive Symptoms-Hallucinations+Delusions (SAPS-H+D) score
    - SAPS Hallucinations domain score
    - SAPS Delusions domain score
    - Clinical Global Impression-Severity (CGI-S)-dementia-related psychosis
    - Zarit Burden Interview (ZBI) score
    - Karolinska Sleepiness Scale (KSS) score
    - EQ-5D-5L score
    - Clinical Global Impression-Improvement (CGI-I)-dementia-related psychosis

2.4 Safety Objective

The safety objective for this study is to evaluate the safety and tolerability of pimavanserin compared to placebo in subjects with dementia-related psychosis who have been stabilized on pimavanserin therapy.

2.4.1 Safety Endpoints

The safety endpoints for this study are as follows:

- Treatment-emergent adverse events (TEAEs)
- Serious adverse events (SAEs)
- Withdrawals due to adverse events (AEs)
- Potentially clinically important changes in other safety assessments
- Global Clinician Assessment of Suicidality (GCAS) score
- Mini-Mental State Examination (MMSE) score
- Extrapyramidal Symptom Rating Scale A (ESRS-A) score

2.5 Pharmacokinetic Objectives

The pharmacokinetic (PK) objectives for this study are as follows:

- To characterize the PK of pimavanserin in subjects with dementia-related psychosis
- To assess the pharmacokinetic/pharmacodynamics (PK/PD) relationship with safety and efficacy endpoints in subjects with dementia-related psychosis

2.5.1 Pharmacokinetic Endpoints

The PK endpoints for this study are as follows:

- Plasma concentration of pimavanserin and AC-279
- Pimavanserin PK parameters using a population PK approach
- PK/PD using appropriate PK/PD analysis methods

3 STUDY DESIGN

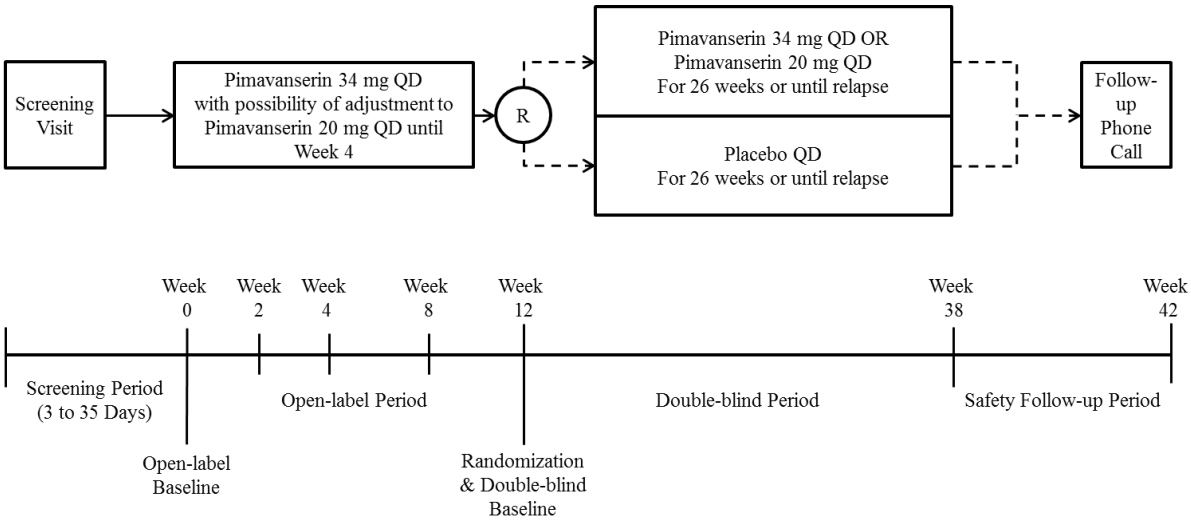
3.1 General Study Design

This is a Phase 3, prospective, double-blind, placebo-controlled, multicenter, relapse prevention study in subjects with dementia-related psychosis. Approximately 95 study sites will participate in this study.

The duration of participation for individual study subjects will be up to 47 weeks, consisting of a screening period of up to 5 weeks; an open-label (OL) period of 12 weeks; a double-blind (DB) period of up to 26 weeks; and an approximately 4-week safety follow-up period.

The study discontinuation date is defined as the day on which the Sponsor notifies sites that the study is ending (for example, in the event of a positive interim analysis or after 75 adjudicated relapse events). The study completion date is defined as the day on which the last subject completes the last scheduled assessment (i.e., safety follow-up). A study schematic is provided in Figure 1.

Figure 1 Schematic of Study Design



Screening Period (3-35 Days)

During the screening period, subjects will be assessed for study eligibility and prohibited medications will be discontinued when medically appropriate. Subjects and partner/caregivers will also receive a standardized psychosocial therapy training.

Investigators should not withdraw a subject's prohibited medication for the purpose of enrolling them into the study. Medications should be discontinued only if it is deemed clinically appropriate to do so.

Open-label Period (12 Weeks)

During the open-label period, eligible subjects will begin receiving pimavanserin 34 mg once daily (QD) beginning at Week 0 (open-label Baseline). Subsequent clinic visits will occur at Weeks 2, 4, 8, and 12. Dose adjustments are permitted at scheduled or unscheduled visits until Week 4. After Week 4, the subject's dose will remain fixed at either 34 mg or 20 mg once daily.

To enter the double-blind period, eligible subjects must meet the following response criteria at Weeks 8 and 12:

- Subject experiences a ≥30% reduction (improvement) from Week 0 (open-label Baseline) on the Scale for the Assessment of Positive Symptoms – Hallucinations + Delusions (SAPS-H+D) Total Score AND

- Subject has a Clinical Global Impression – Improvement (CGI-I) score of 1 (very much improved) or 2 (much improved), relative to Week 0 (open-label Baseline)

A subject who does not meet response criteria at Weeks 8 and 12 will be withdrawn from study drug and enter the safety follow-up period of the study.

Pharmacokinetic samples will be collected at Weeks 0, 8, and 12. Pharmacokinetic samples will also be collected, if possible, at any ET visit or the visit immediately following any SAE or following an AE leading to discontinuation.

Double-blind Period (up to 26 Weeks)

Randomization will occur at the double-blind Baseline visit (Week 12). Subjects will be randomly assigned 1:1 to continue their pimavanserin dose (34 mg or 20 mg) or matching placebo.

The protocol-defined relapse criteria for dementia-related psychosis are designed to identify subjects with an impending relapse or relapse of psychosis:

- Subject experiences a ≥30% increase (worsening) from Week 12 (double-blind Baseline) on the SAPS-H+D Total Score AND has a CGI-I score of 6 (much worse) or 7 (very much worse), relative to the double-blind Baseline; OR
- Subject is treated with an antipsychotic (other than pimavanserin) for dementia-related delusions and/or hallucinations; OR
- Subject stops study drug or withdraws from study for lack of efficacy, (as reported by the subject or study partner/caregiver) or the Investigator discontinues study drug due to lack of efficacy; OR
- Subject is hospitalized for worsening dementia-related psychosis

Relapse criteria will be assessed weekly for the first 2 weeks after randomization (Weeks 13 and 14), every 2 weeks until Week 26, and every 4 weeks through Week 38. Relapse criteria may also be evaluated at unscheduled visits. Any subject who meets any of the relapse criteria after randomization will be withdrawn from study drug and enter the safety follow-up period of the study.

All subjects who discontinue the study between randomization and Week 38 should complete an ET visit, if possible. The Independent Adjudication Committee (IAC) will adjudicate early termination and relapse events that occurred before the study discontinuation date. Early termination and relapse events that occur in the double-blind period after the study discontinuation date will not be adjudicated.

An interim and final analysis will be planned after 38 and 75 adjudicated relapse events have occurred, respectively. The study may be terminated early if the interim analysis results establish the efficacy and safety of pimavanserin for the treatment of hallucinations and delusions associated with dementia-related psychosis.

The Data Safety Monitoring Board (DSMB) will monitor, review, and evaluate the unblinded safety data on a periodic basis. In addition, the DSMB will review the results of the pre-planned unblinded interim analysis conducted by the Independent Statistical Group.

Pharmacokinetic samples will be collected at Weeks 13, 22, and 38/EOT. Pharmacokinetic samples will also be collected, if possible, at any ET visit or the visit immediately following any SAE or following an AE leading to discontinuation.

Safety Follow-up Period (Approximately 4 Weeks)

Approximately 4 weeks after the last dose of study drug, subjects will have a safety follow-up telephone call visit.

The Sponsor will provide investigative sites with 3 months of after-study assistance to transition subjects to standard of care therapy after their participation in the study.

Table 1 Schedule of Assessments for Screening and Open-label Period

	Screening	Open-label Period				
Visit week	**-5 to 0**	**0**	**2**	**4**	**8**	**12[a]**
Allowable visit window (# days)			**±3**	**±3**	**±3**	**+7**
Visit Number	**1**	**2**	**3**	**4**	**5**	**6**
Informed consent	X					
Inclusion/exclusion criteria assessment	X	X				
Medical history and demographics	X					
Mini-Mental State Examination	X	X	X	X	X	X
MRI or CT[b]	X					
Psychosocial therapy training	X					
Physical and neurological examinations	X	X				X
Vital signs and weight	X	X	X	X	X	X
Height	X					
12-lead electrocardiogram[c]	X	X	X			X
Clinical laboratory tests	X	X				X
Pregnancy test[d]	X	X				X
SAPS-H+D	X	X	X	X	X	X
Clinical Global Impression-Improvement[e]			X	X	X	X
Clinical Global Impression-Severity	X	X	X	X	X	X
Zarit Burden Interview[f]		X				X
Karolinska Sleepiness Scale		X				X
EQ-5D-5L		X				X
Assessment for response criteria					X	X
Assessment for concomitant medications	X	X	X	X	X	X
Global Clinician Assessment of Suicidality	X	X	X	X	X	X
Extrapyramidal Symptom Rating Scale A		X				X
Assessment for adverse events	X	X	X	X	X	X
Pharmacokinetic sample collection[g]		X			X	X
Pharmacogenomic sample collection[h]		X				
Drug dispensation		X	X	X	X	X[i]
Drug return and accountability			X	X	X	X
Randomization						X

Footnotes for Table 1 on next page

Abbreviations: CT=computed tomography; MRI=magnetic resonance imaging;
SAPS-H+D=Scale for the Assessment of Positive Symptoms, Hallucinations and Delusions subscales

Note: All efforts should be made to complete visits in a single day. In the event of a split visit, the CGI-S/CGI-I and the SAPS H+D must be completed on the same day (pertinent for all visits).

Note: When a subject is found to be appropriate for discontinuation from the trial, the visit during which this is determined should immediately be converted to an ET visit (as described in Table 2) and all procedures and assessments done at the ET visit (Week 38/EOT; Visit 17/EOT) should be completed (additional details provided in Section 4.4 of the protocol). Safety follow-up should be completed as scheduled.

For additional details on study procedures, see Section 6 of the protocol.

[a] Visit 6 (Week 12) will serve as both the final visit of the open-label period and the Baseline visit of the double-blind period (Table 2).

[b] A non-contrast brain MRI or non-contrast head CT will be completed if the subject has not had a CT or MRI scan completed (a) within the past 3 years AND (b) during or subsequent to the onset of dementia.

[c] The ECG will be completed in triplicate at Visit 1 (Screening). A single ECG will be completed at all other visits.

[d] A pregnancy test (serum at Visit 1 and urine at all other visits) is only required for women of child-bearing potential. If urine cannot be obtained in women of child-bearing potential, a serum pregnancy test should be done in its place.

[e] The CGI-I should be scored relative to Visit 2, the subject's open-label Baseline.

[f] The ZBI should only be administered to study partners/caregivers who are family members.

[g] Pharmacokinetic samples will also be collected, if possible, at any ET visit (Table 2) or the visit immediately following any SAE or following an AE leading to discontinuation.

[h] A separate informed consent (and assent, if applicable) must be given for the pharmacogenomic component of the study. This consent may be obtained at any time during the study. If informed consent is given in time for sample collection at Visit 2, a pre-dose sample should be collected at Visit 2. If informed consent for pharmacogenomics is not given in time for sample collection at Visit 2, a sample may be collected any time after informed consent for pharmacogenomics is given.

[i] Blinded study drug will be dispensed after randomization at Visit 6 (Week 12).

Table 2 Schedule of Assessments for Double-blind Period

Visit week	13	14	16	18	20	22	24	26	30	34	38/EOT	42
Allowable visit window (# days)	±3	±3	±3	±3	±3	±3	±3	±3	±7	±7	+7	+3
Visit Number	7	8	9	10	11	12	13	14	15	16	17/ET	18
Visit Type (Clinic [C] or Telephone [T])	C	C	C	C	T	C	T	C	C	C	C	T
Mini-Mental State Examination	X	X	X	X		X		X	X	X	X	
Physical and neurological examinations						X					X	
Vital signs and weight	X	X	X	X		X		X	X	X	X	
12-lead electrocardiogram				X				X			X	
Clinical laboratory tests				X				X			X	
Pregnancy test[a]								X			X	
SAPS-H+D	X	X	X	X		X		X	X	X	X	
Clinical Global Impression-Improvement[b]	X	X	X	X		X		X	X	X	X	
Clinical Global Impression-Severity	X	X	X	X		X		X	X	X	X	
Zarit Burden Interview[c]				X				X			X	
Karolinska Sleepiness Scale				X				X			X	
EQ-5D-5L				X				X			X	
Assessment for protocol-defined relapse criteria	X	X	X	X	X	X	X	X	X	X	X	
Assessment for concomitant medications	X	X	X	X	X	X	X	X	X	X	X	X
Global Clinician Assessment of Suicidality	X	X	X	X	X	X	X	X	X	X	X	
Extrapyramidal Symptom Rating Scale A	X	X	X	X		X		X	X	X	X	
Assessment for adverse events	X	X	X	X	X	X	X	X	X	X	X	X
Pharmacokinetic sample collection[d]	X					X					X	
Drug dispensation		X	X	X		X		X	X	X		
Drug return and accountability	X[e]	X	X	X		X		X	X	X	X	

Footnotes for Table 2 on next page

Abbreviations: EOT=end-of-treatment; ET=early termination; SAPS-H+D=Scale for the Assessment of Positive Symptoms, Hallucinations and Delusions subscales

For additional details on study procedures, see Section 6 of the protocol.

Note: All efforts should be made to complete visits in a single day. In the event of a split visit, the CGI-S/CGI-I and the SAPS H+D must be completed on the same day (pertinent for all visits).

Note: When a subject is found to be appropriate for discontinuation from the trial, the visit during which this is determined should immediately be converted to an ET visit (as described in Table 2) and all procedures and assessments done at the ET visit (Week 38/EOT; Visit 17/EOT) should be completed (additional details provided in Section 4.4 of the protocol). Safety follow-up should be completed as scheduled.

[a] A urine pregnancy test is only required for women of child-bearing potential. If urine cannot be obtained in women of child-bearing potential, a serum pregnancy test should be done in its place

[b] The CGI-I should be scored relative to Visit 6 (Week12), the subject's double-blind Baseline.

[c] The ZBI should only be administered to study partners/caregivers who are family members.

[d] Pharmacokinetic samples will also be collected, if possible, at the visit immediately following any SAE or following an AE leading to discontinuation.

[e] Only drug accountability will be completed at this visit.

3.2 Randomization

At Week 12 (Visit 6), subjects who meet response criteria will be randomized in a 1:1 ratio to either continue receiving pimavanserin at the current dose level or to receive matched placebo, using an interactive response technology system. Randomization will be stratified by:

- Most likely dementia subtype or most prominent cause of dementia (defined as designated dementia subtype):
    - Alzheimer's disease or frontotemporal dementia (FTD) spectrum disorders
    - Vascular dementia
    - Parkinson's disease dementia or dementia with Lewy bodies
- Region (North America, Western Europe, Eastern Europe, Asia Pacific, Latin America)

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

3.3 Blinding

All subjects will receive open-label pimavanserin (34 mg or 20 mg) during the open-label period of the study.

Treatment assignment during the double-blind period 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.

3.4 Determination of Sample Size

This study will enroll approximately 356 subjects in order to randomize approximately 178 subjects with dementia-related psychosis who meet response criteria at Weeks 8 and 12 of the open-label period.

The sample size calculation was based on the following assumptions: a placebo relapse event rate of 60% over 26 weeks; a pimavanserin relapse event rate of 35% over 26 weeks (hazard ratio = 0.47); a dropout rate of 25% over 26 weeks; an overall two-sided alpha level of 0.05; use of a one-sided (0.025) O'Brien-Fleming stopping boundary to adjust for a single interim

analysis that will be performed when one half of the total planned number of post-randomization relapse events have occurred; and a power of 90%. The total number of post-randomization relapse events required at the final analysis is 75 and the calculated sample size is 89 in each of the two treatment groups (giving a total estimate of 178 subjects). Study enrollment will be closed when approximately 178 subjects have been randomized in the double-blind period or when 75 adjudicated post-randomization relapse events have occurred. In the event that the randomization rate is lower than anticipated and/or the relapse event rate is lower than anticipated, the number enrolled and randomized may be increased up to a maximum of 400 randomized subjects in order to observe 75 post-randomization relapse events. If fewer than 75 post-randomization relapse events have been observed after 400 randomized subjects have completed the study or terminated early, the study will end. An interim efficacy analysis will be performed, and the study may be terminated early if the interim analysis results meet prespecified stopping criteria.(see Section 16)

4 ANALYSIS SETS

4.1 Intent-to-treat (ITT) Analysis Set

The ITT Analysis Set will include all randomized subjects. Subjects will be analyzed based on their randomized treatment. The ITT Analysis Set will be used for analyses of the double-blind efficacy data.

4.2 Open-label Safety Analysis Set

The Open-label Safety Analysis set will include all subjects who received at least one dose of open-label study drug (pimavanserin). The Open-label Safety Analysis Set will be used for analyses of the open-label data.

4.3 Double-blind Safety Analysis Set

The Double-blind Safety analysis set will include all subjects who received at least one dose of double-blind study drug (pimavanserin or placebo). Subjects will be analyzed based on the treatment that they actually received. The Double-blind Safety Analysis Set will be used for all analyses of the double-blind safety data.

4.4 Per-protocol Analysis Set

The Per-protocol Analysis Set will consist of those subjects in the ITT Analysis Set without any protocol deviations which could potentially have a substantial impact on the primary efficacy outcome. The precise reasons for excluding subjects from the Per-protocol Analysis Set will be fully defined and documented prior to the clinical database lock. Subjects will be analyzed based on their randomized treatment.

4.5 Open-label Pharmacokinetic Analysis Set

The Open-label Pharmacokinetic (PK) Analysis Set includes all subjects who have at least one measurable pimavanserin plasma concentration collected in the open-label period.

4.6 Double-blind Pharmacokinetic Analysis Set

The Double-blind PK Analysis Set includes all subjects who have at least one measurable pimavanserin plasma concentration collected in the double-blind period.

5 DATA HANDLING CONVENTIONS

All data collected in the study will be listed.

5.1 General Data Reporting Conventions

In general, listings will be sorted by randomized treatment group (placebo, pimavanserin), enrolled but not randomized, screen failure (screened but not enrolled), and subject ID, and all relevant data will be presented. Data from screen failures will be included in listings where available.

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. For pharmacokinetic data, the geometric mean will also be provided. 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.

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.

Unless specified otherwise, all statistical tests will be 2-sided hypothesis tests and all CIs will be 2-sided 95% CIs. Two-sided P-values will generally be presented to 4 decimal places; values less than 0.0001 will be presented as <0.0001.

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

5.1.1 Open-label Period

In general, dates associated with open-label data will be accompanied on the listings by the study day relative to the date of first dose of open-label pimavanserin. Open-label study day 1 is the date of first dose of open-label pimavanserin.

For assessments occurring on or after the open-label study day 1, open-label study day is calculated for a given subject as

assessment date – first open-label dose date + 1.

For assessments occurring prior to the open-label study day 1, open-label study day is calculated for a given subject as

assessment date – first open-label dose date.

Unless specified otherwise, the following summaries will be provided:

1. all subjects combined;
2. by randomization status (not randomized and randomized);
3. by final dose level (20 and 34 mg) of open-label pimavanserin administered during the open-label period.

5.1.2 Double-blind Period

In general, dates associated with double-blind data points will be accompanied on the listings by the study day relative to the date of first dose of double-blind study drug. Double-blind study day 1 is the date of first dose of double-blind study drug.

For assessments occurring on or after the double-blind study day 1, double-blind study day is calculated for a given subject as

assessment date – first double-blind dose date + 1.

For assessments occurring prior to the double-blind study day 1, double-blind study day is calculated for a given subject as

assessment date – first double-blind dose date.

Unless specified otherwise, the following summaries will be provided:

1. by treatment group (placebo and pimavanserin);
2. by dose level (20 and 34 mg) of double-blind pimavanserin administered during the double-blind period for the pimavanserin group.

5.2 Derived Variables

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

5.2.1 Open-label Baseline

Open-Label Baseline is defined as the last non-missing assessment, including those from repeated and unscheduled measurements, before the date of first dose of open-label pimavanserin.

5.2.2 Double-blind Baseline

Double-Blind Baseline is defined as the last non-missing assessment, including those from repeated and unscheduled measurements, on or before the date of first dose of double-blind study drug.

5.2.3 Time from Randomization to Relapse in the Double-blind Period

For subjects who experience a relapse event, the time to relapse will be calculated as

date of relapse – date of randomization + 1.

For the primary efficacy endpoint the date of relapse is determined by the IAC.

Subjects who complete the 26-week double-blind period, those who discontinue early, and those who are ongoing at the time of the applicable analysis cutoff, without experiencing a relapse event, will be censored at the time of the last applicable SAPS-H+D assessment. See Section 13.1 for further details regarding the analysis cutoffs, and Section 13.2.2 for censoring times for sensitivity analyses. The censored follow-up time will be calculated as:

date of last applicable SAPS-H+D assessment – date of randomization + 1.

Subjects who have not experienced a relapse event and do not have any post-baseline SAPS-H+D assessments in the double-blind period, will be censored at the time of randomization. The censored follow-up time will have a value of 1 day.

5.2.4 Time from Randomization to Discontinuation from the Double-blind Period for Any Reason

For subjects who experience a relapse event, the time to discontinuation will be calculated as:

date of relapse – date of randomization + 1.

For subjects who don't experience a relapse event and who discontinue early from the double-blind period, prior to the study discontinuation date, for any reason other than termination of the study by the sponsor, the time to discontinuation will be calculated as:

date of last dose – date of randomization + 1.

Subjects who don't experience a relapse event and who complete the 26-week double-blind period will be censored at the date of last dose. The censored follow-up time will be calculated as:

date of last dose – date of randomization + 1.

All subjects who are ongoing in the 26-week double-blind period at the time of study discontinuation by the sponsor will be censored at the study discontinuation date. The censored follow-up time will be calculated as

study discontinuation date – date of randomization + 1.

All subjects who are ongoing in the 26-week double-blind period at the time of the database cutoff for the interim analysis will be censored at the database cutoff date. The censored follow-up time will be calculated as

date of database cutoff – date of randomization + 1.

5.2.5 Scale for the Assessment of Positive Symptoms-Hallucinations + Delusions

The Scale for the Assessment of Positive Symptoms-Hallucinations + Delusions (SAPS-H+D) is assessed at Screening, Baseline (Week 0), and Weeks 2, 4, 8, and 12 during the open-label period, and at Weeks 13, 14, 16, 18, 22, 26, 30, 34, and 38 during the double-blind period.

The SAPS was designed to measure positive psychotic symptoms. Positive symptoms include hallucinations, delusions, abnormalities in language and behavior, and disordered thought processes. The SAPS-H+D subscales will be administered in this study. The Hallucinations and Delusions subscales consist of 20 items, including 2 global ratings of severity for hallucinations (H7) and delusions (D13). Each of the 20 items is scored on a 6-point scale (0=none, 1=questionable, 2=mild, 3=moderate, 4=marked, 5=severe). The SAPS-H+D total score is the sum of the 20 item scores with a possible range of 0 to 100. Higher scores denote more severe symptoms. The hallucinations score (SAPS-H) is the sum of the 7 hallucinations item scores and the delusions score (SAPS-D) is the sum of the 13 delusions item scores.

Missing SAPS-H+D item scores will not be imputed. The total score will be missing if any item score is missing. Similarly, the SAPS-H and SAPS-D scores will be missing if any component item score is missing.

5.2.6 Clinical Global Impressions – Severity Scale

The Clinical Global Impressions – Severity (CGI-S) scale is assessed at Screening, Baseline (Week 0), and Weeks 2, 4, 8, and 12 during the open-label period, and at Weeks 13, 14, 16, 18, 22, 26, 30, 34, and 38 during the double-blind period.

The CGI-S scale is a clinician-rated, 7-point scale that is designed to rate the severity of the subject's hallucinations and delusions at the time of assessment using the Investigator's judgment and past experience with subjects who have the same disorder (i.e., dementia-related psychosis). The possible 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 patients.

Missing CGI-S scores will not be imputed.

5.2.7 Clinical Global Impressions – Improvement Scale

The Clinical Global Impressions – Improvement (CGI-I) scale is assessed at Weeks 2, 4, 8, and 12 during the open-label period, and at Weeks 13, 14, 16, 18, 22, 26, 30, 34, and 38 during the double-blind period.

The CGI-I scale is a clinician-rated, 7-point scale that is designed to rate the improvement in the subject's hallucinations and delusions at the time of assessment, relative to the symptoms at Baseline (relative to open-label Baseline for response criteria and relative to double-blind Baseline for relapse criteria). The possible scores are 1=very much improved, 2=much improved, 3=minimally improved, 4=no change, 5=minimally worse, 6=much worse, 7=very much worse.

Missing CGI-I scores will not be imputed.

5.2.8 Zarit Burden Interview

The Zarit Burden Interview (ZBI) is assessed at Baseline (Week 0) and Week 12 during the open-label period, and at Weeks 18, 26, and 38 during the double-blind period.

The ZBI was designed to assess the stress experienced by caregivers of patients with dementia. The ZBI will be administered as an interview. The interview consists of 22 statements reflecting how people sometimes feel when taking care of another person. The statements are phrased as questions for a caregiver who is a family member to indicate how often they feel the way described in the statement. Responses are Never, Rarely, Sometimes, Quite Frequently, and Nearly Always. When the study caregiver is not a family member or when the study caregiver changes, this scale will not be completed. Each of the responses is scored from 0 to 4. The ZBI total score is the sum of the 22 item scores with a possible range of 0 to 88. Higher scores denote greater stress.

Missing scores will not be imputed. The total score will be missing if any item score is missing.

5.2.9 Karolinska Sleepiness Scale

The Karolinska Sleepiness Scale (KSS) is assessed at Baseline (Week 0) and Week 12 during the open-label period, and at Weeks 18, 26, and 38 during the double-blind period.

The KSS is a self-reported subjective measure of a subject's level of drowsiness. Scoring is based on a 9-point verbally anchored scale going from "1 = extremely alert" to "9 = very sleepy, great effort to keep awake, fighting sleep".

Missing KSS scores will not be imputed.

5.2.10 EQ-5D-5L Proxy Version 1

The EQ-5D-5L Proxy Version 1 is assessed at Baseline (Week 0) and Week 12 during the open-label period, and at Weeks 18, 26, and 38 during the double-blind period.

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. For this version, the study partner/caregiver (the proxy) is asked to rate subject's health-related quality of life in their (the proxy's) opinion. The EQ-5D consists of the EQ-5D-5L descriptive system and the EQ-5D Visual Analogue scale (EQ-5D VAS).

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

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

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

5.2.11 Global Clinician Assessment of Suicidality

The Global Clinician Assessment of Suicidality (GCAS) is assesse d at Screening, Baseline (Week 0), and Weeks 2, 4, 8, and 12 during the open-label period, and at Weeks 13, 14, 16, 18, 20, 22, 24, 26, 30, 34, and 38 during the double-blind period.

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 (only the clinician rating is entered in the database). At Screening visit, 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.

Missing GCAS scores will not be imputed.

5.2.12 Mini-Mental State Examination

The Mini-Mental State Examination (MMSE) is assessed at Screening, Baseline (Week 0), and Weeks 2, 4, 8, and 12 during the open-label period, and at Weeks 13, 14, 16, 18, 22, 26, 30, 34, and 38 during the double-blind period.

The MMSE consists of an 11-domain, 30-point questionnaire that is used to quantitatively assess cognition. The MMSE is being used in this study to screen for cognitive impairment and as a safety measure. Total score of MMSE is calculated as the sum of the 11 area scores. Lower scores denote more severe cognitive impairment.

Missing scores will not be imputed. The total score will be missing if any area score is missing.

5.2.13 Extrapyramidal Symptom Rating Scale-Abbreviated

The Extrapyramidal Symptom Rating Scale-Abbreviated (ESRS-A) is assessed at Baseline (Week 0) and Week 12 during the open-label period, and at Weeks 13, 14, 16, 18, 22, 26, 30, 34, and 38 during the double-blind period.

ESRS-A, derived from the original ESRS, was designed as a more concise instrument for assessing the four types of drug-induced movement disorders Parkinsonism, akathisia, dystonia, and dyskinesia. 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 symptoms of a movement disorder.

Missing ESRS-A item scores of the 4 subscales will not be imputed. The ESRS-A total score will be missing if any item score of the 4 subscales is missing.

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). The score ranges from 0 to 5, with higher scores indicating more severe movement disorders. Missing CGI-S score will not be imputed.

5.3 Analysis Visit Windows

5.3.1 Open-label Period

Efficacy, safety and pharmacokinetic assessments for the open-label (OL) period will be summarized by analysis visit as presented in Table 3. For exploratory efficacy endpoints and MMSE, assessments more than 14 days after last dose of study drug will be assigned an analysis visit but will not be selected for the by-visit analyses.

Table 3 Analysis Visit Windows for the Open-label Period

Study Visit	Analysis Visit	Target OL Study Day[a]	Study Day Interval
Week 0	OL Baseline	1	≤1
Week 2	OL Week 2	15	2 to 21
Week 4	OL Week 4	29	22 to 42
Week 8	OL Week 8	57	43 to 70
Week 12	OL Week 12	85	71 to 98 (or day of first dose of double-blind study drug[b])

[a]Derivation of open-label study day is described in Section 5.1.1. Open-label study day 1 is the date of first dose of open-label pimavanserin.
[b]For subjects who are randomized in the double-blind period.

5.3.2 Double-blind Period

Efficacy, safety, and pharmacokinetic assessments for the double-blind (DB) period will be summarized by analysis visit as presented in Table 4. For exploratory efficacy endpoints and MMSE, assessments more than 14 days after last dose of study drug will be assigned an analysis visit but will not be selected for the by-visit analyses.

Table 4 Analysis Visit Windows for the Double-Blind Period

Study Visit	Analysis Visit	Target DB Study Day[a]	Study Day Interval
Week 12	DB Baseline	1	≤1
Week 13	DB Week 1	8	2 to 11
Week 14	DB Week 2	15	12 to 21
Week 16	DB Week 4	29	22 to 35
Week 18	DB Week 6	43	36 to 49
Week 20	DB Week 8	57	50 to 62
Week 22	DB Week 10	71	63 to 77
Week 24	DB Week 12	85	78 to 91
Week 26	DB Week 14	99	92 to 112
Week 30	DB Week 18	127	113 to 140
Week 34	DB Week 22	155	141 to 168
Week 38	DB Week 26	183	169 to 197

[a]Derivation of double-blind study day is described in Section 5.1.2. Double-blind study day 1 is the date of first dose of double-blind study drug.

5.3.3 Unscheduled Assessments

Unscheduled assessments, including the assessments at early termination visits, will be considered for planned time point analyses. All assessments will be presented in data listings.

5.3.4 Multiple Measurements within Visit Windows

For endpoints other than the exploratory efficacy endpoints and MMSE, if more than one assessment falls within a study visit then the assessment closest to the target study day will be selected for the by-visit analyses. For exploratory efficacy endpoints and MMSE, if more than one assessment falls within a study visit and each of them were collected within 14 days of last dose of study drug then the assessment closest to the target study day will be selected for the by-visit analyses. In these analyses, if two eligible assessments are equidistant from the target day then the chronologically last assessment will be used. Exceptions may be made for incomplete assessments, in which case, more complete assessments may be given priority. Details are provided in a separate programming conventions document.

For both open-label and double-blind safety analysis where the most extreme values should be selected (e.g. overall post-baseline minimum, overall post-baseline maximum, and potentially clinically important values), all non-missing post-baseline values should be considered, regardless of whether the value is selected for the by-visit summaries. All results will be presented in data listings.

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

5.4 Data Handling Conventions

In the Open-label and Double-blind 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 with respect to the open-label or double-blind period, 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 data summarization at the time of the database cutoff for the interim analysis, if a subject is still ongoing, then this subject’s last dose date will be imputed using the database cutoff date.

5.4.1 Missing or Incomplete Dates for Prior or Concomitant Medications

Missing or incomplete medication start or end dates will be imputed for the purpose of determining whether the medications are taken concomitantly (see Section 10 for definition). When the chronological order of medication use relative to the study drug treatment period is unclear due to missing or incomplete date(s), the medication will be considered as concomitant. The imputation algorithms will be detailed in the 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.

5.4.2 Missing or Incomplete Dates for Adverse Events

Missing or incomplete adverse event (AE) start dates will be imputed for the purpose of determining whether the AEs are treatment-emergent (see Section 14.1 for definition). When the chronological order of an AE onset relative to the study drug treatment period is unclear due to missing or incomplete date(s), the AE will be considered as treatment-emergent. The imputation algorithms will be detailed in the 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 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, and the actual values will be used in data listings.

5.4.4 Missing Relationship to Study Drug for Adverse Events

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

5.4.5 Character Values of Clinical Laboratory Variables

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

6 SUBJECT DISPOSITION

The number and percentage of sites that screened at least 1 subject, enrolled at least 1 subject, and randomized at least 1 subject will be summarized by region and all regions combined. The number and percentage of subjects screened, unique subjects screened, subjects enrolled, and subjects randomized will be summarized by region and all regions combined.

6.1 Analysis for the Open-label Period

Subject disposition for the open-label period will be summarized for the Open-label Analysis Set. The number and percentage of subjects who were randomized to the double-blind period, completed 12 weeks of open-label period, discontinued, or are ongoing at the time of database cutoff will be presented. The number and percentage of subjects meeting the following discontinuation reasons will be tabulated:

- Lack of response in open-label;
- Adverse event;
- Death;
- Non-compliance with study drug;
- Used of prohibited medications;
- Physician decision;
- Pregnancy;
- Protocol violation;
- Study terminated by Sponsor;
- Subject withdrew consent;
- Lost to follow-up;
- Other.

6.2 Analysis for the Double-blind Period

Subject disposition for the double-blind period will be summarized for the ITT Analysis Set, Per-protocol Analysis Set, and Double-blind Safety Analysis Set. The number and percentage of subjects who completed the study, discontinued, or are ongoing at the time of database cutoff will be presented. The number and percentage of subjects meeting the following discontinuation reasons will be tabulated:

- Relapse of hallucinations and delusions associated with dementia-related psychosis;
- Adverse event;
- Death;
- Non-compliance with study drug;
- Used of prohibited medications;
- Physician decision;
- Pregnancy;
- Protocol violation;
- Study terminated by Sponsor;
- Subject withdrew consent;
- Lost to follow-up;
- Other.

The number and percentage of subjects in the ITT Analysis Set who are excluded from the Per-protocol Analysis Set will be presented in a summary table by exclusion reason.

6.3 Screen Failure

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.

7 PROTOCOL DEVIATIONS

Protocol deviations will be reviewed periodically over the course of the study. The review process, definition of the deviation categories, and the classification of a deviation as major or minor are detailed in the Protocol Deviation Management Plan.

A protocol deviation for a given subject that occurs on or after the date of first dose of double-blind study drug will be assigned to double-blind period, otherwise it will be assigned to the open-label period. A summary of the number and percentage of subjects with major protocol deviations for each deviation category will be presented. A listing of protocol deviations by site and subject will also be provided.

Major protocol deviations occurring in the open-label period will be summarized for the Open-label Safety Analysis Set. Summaries for the ITT Analysis Set will include major protocol deviations from both open-label and double-blind study periods.

8 DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS

8.1 Demographics

Demographic variables will be summarized using descriptive statistics for the Open-label Safety Analysis Set, ITT Analysis Set, Double-blind Safety Analysis Set, and Per-protocol Analysis Set. 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, BMI, region, subject living situation, and partner/caregiver relationship.

The reported age reflects a subject’s age at the informed consent date. Age will be presented as both continuous and categorical variables. Age categories will be presented as <65, 65 – 74, 75 – 84, and ≥85 years old.

8.2 Disease Characteristics

Disease characteristics will be summarized descriptively for the Open-label Safety Analysis Set, ITT Analysis Set, Double-blind Safety Analysis Set, and Per-protocol Analysis Set.

Dementia subtype: Alzheimer's disease, FTD spectrum disorders, vascular dementia, Parkinson's disease dementia, and dementia with Lewy bodies; and dementia severity will be presented as a categorical variable. Open-label and double-blind Baseline MMSE total score, SAPS-H+D total score, ZBI score, KSS score, EQ-5D VAS score, and CGI-S score will be presented as continuous variables. In addition, open-label and double-blind Baseline CGI-S score will be presented as categorical variables using categories of 1 to 7.

Neuropsychiatric history will be summarized descriptively for continuous and categorical variables.

9 MEDICAL HISTORY

Medical history will be summarized descriptively for the Open-label Safety Analysis Set, ITT Analysis Set, Double-blind Safety Analysis Set, and Per-protocol Analysis Set.

Medical 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. 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 will be provided.

10 PRIOR, CONCOMITANT AND POST-TREATMENT MEDICATION

Medications will be coded using WHO Drug Dictionary (WHODrug-DDE-B2) 2017 March or newer version. The number and percentage of subjects taking each drug class (ATC Level 3) and medication preferred term will be tabulated. Multiple medication usage by a subject in the same category will be counted only once. Prior, concomitant, and post-treatment medications will be summarized separately.

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 open-label pimavanserin. Summaries of prior medications will be provided for the Open-label Safety Analysis Set and Double-blind Safety Analysis Set.

10.2 Concomitant Medication

10.2.1 Analysis for the Open-label Period

Concomitant medication during the open-label period is defined as any medication with a start date prior to the date of first dose of open-label pimavanserin and continuing after the date of first dose of open-label pimavanserin or with a start date between the date of the first dose of open-label pimavanserin and before the date of the first dose of double-blind study drug for subjects randomized into the double-blind period or last dose in the open-label period for subjects not randomized, inclusive. Concomitant medications during the open-label period will be summarized for the Open-label Safety Analysis Set.

10.2.2 Analysis for the Double-Blind Period

Concomitant medication during the double-blind period is defined as any medication with a start date prior to the date of first dose of double-blind study drug and continuing after the date of first dose of double-blind study drug or with a start date between the dates of the first and last doses of double-blind study drug, inclusive. Concomitant medications during the double-blind period will be summarized for the Double-blind Safety Analysis Set.

10.3 Analysis for Post-Treatment Medication

Post-treatment medication is defined as any medication with a start date after the date of the last dose of open-label pimavanserin for non-randomized subjects and after the date of the last dose of double-blind study drug for randomized subjects. Summaries of post-treatment medications will be provided for the Open-label Safety Analysis Set.

11 EXTENT OF EXPOSURE AND TREATMENT COMPLIANCE

11.1 Exposure to Study Drug

The duration of exposure during the open-label period and the double-blind period for a given subject will be calculated separately. Duration of exposure to study drug will be summarized descriptively as both continuous and categorical variables for the open-label period and the double-blind period.

11.1.1 Analysis for the Open-label Period

Duration of exposure during the open-label period for a given subject is calculated as follows:

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

For the categorical presentation of duration of exposure during the open-label period, the number and percentage of subjects in each of the following categories will be displayed: <1 week, 1 to <2 weeks, 2 to <4 weeks, 4 to <8 weeks, 8 to <12 weeks, and ≥12 weeks. Summaries of duration of exposure will be provided for the Open-label Safety Analysis Set. Kaplan-Meier curves of duration of exposure will also be provided.

11.1.2 Analysis for the Double-blind Period

Duration of exposure during the double-blind period for a given subject is calculated as follows:

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

For the categorical presentation of duration of exposure during the double-blind period, the number and percentage of subjects in each of the following categories will be displayed: <1 week, 1 to <2 weeks, 2 to <4 weeks, 4 to <8 weeks, 8 to <12 weeks, 12 to <18 weeks, 18 to <26 weeks, and ≥26 weeks. Summaries of duration of exposure will be provided for the Double-blind Safety Analysis Set. In addition, Kaplan-Meier curves of duration of exposure will be provided.

11.2 Study Drug Compliance

Study drug dosing compliance (in percentage) for the open-label period and the double-blind period for a given subject will be calculated separately. Study drug compliance will be summarized descriptively as both continuous and categorical variables for the open-label period and the double-blind period. 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%.

11.2.1 Analysis for the Open-label Period

Study drug compliance during the open-label (OL) period for a given subject is calculated as follows:

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

Summaries of compliance will be provided for the Open-label Safety Analysis Set.

11.2.2 Analysis for the Double-blind Period

Study drug compliance (in percentage) during the double-blind (DB) period for a given subject is calculated as follows:

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

Summaries of compliance will be provided for the Double-blind Safety Analysis Set.

12 EFFICACY ANALYSES

Efficacy analyses of the primary and key secondary endpoints for the double-blind period will be performed on the ITT Analysis Set and the Per-protocol Analysis Set. Efficacy analyses of the exploratory efficacy endpoints for double-blind period will be performed on the ITT Analysis Set. The following summaries will be provided:

1. by treatment group (placebo and pimavanserin);
2. by dose level (20 and 34 mg) of double-blind pimavanserin administered during the double-blind period for the pimavanserin group. No formal comparisons between the dose levels (20 and 34 mg) will be performed.

12.1 Efficacy Variables

12.1.1 Primary Endpoint

The primary endpoint is the time from randomization to relapse in the double-blind period.

12.1.2 Key Secondary Endpoint

The key secondary endpoint is time from randomization to discontinuation from the double-blind period for any reason (other than termination of the study by the sponsor).

12.1.3 Exploratory Study Endpoints

Change from double-blind Baseline on the following:

- Scale for the Assessment of Positive Symptoms-Hallucinations+Delusions (SAPS-H+D) score
- SAPS Hallucinations domain score
- SAPS Delusions domain score
- Clinical Global Impression-Severity (CGI-S)-dementia-related psychosis
- Zarit Burden Interview (ZBI) score
- Karolinska Sleepiness Scale (KSS) score
- EQ-5D VAS score and EQ-5D-5L descriptive system
- Clinical Global Impression-Improvement (CGI-I)-dementia-related psychosis

12.2 Adjustment for Covariates

For time-to-event variables analyzed using the Cox regression model, region and designated dementia subtype (as defined by randomization strata) will be included as covariates as described in Section 13.

For continuous variables (except CGI-I) analyzed using a mixed model for repeated measures (MMRM) or an analysis of covariance (ANCOVA), the Baseline value of the endpoint being analyzed, region, and designated dementia subtype will be included as covariates as described in Section 13.3. For CGI-I, the Baseline CGI-S score, region, and designated dementia subtype will be included as covariates in the MMRM and ANCOVA analyses.

12.3 Handling of Informative Censoring

The primary time-to-event analysis using the Cox regression model assumes the censoring times are independent of the event times (i.e., the censoring is non-informative). The robustness of the primary analysis results to certain deviations from non-informative censoring will be evaluated using a reference-based multiple imputation method (Lu et. al.,2015). Further details are provided in Section 13.1.1.

12.4 Multiple Comparisons/Multiplicity

A hierarchical testing procedure will be used to control the overall type I error rate for the primary and key secondary efficacy endpoints. Hypothesis testing will occur in the following sequential order:

1. Primary endpoint
2. Key secondary endpoint

Adjustment of the analyses of the primary and key secondary efficacy endpoints to account for the interim efficacy analysis is described in Section 16.

12.5 Examination of Subgroups

12.5.1 Double-Blind Period

Treatment comparisons will be made with respect to the primary and key secondary efficacy variables using the Cox regression described in Sections 13.1 and 13.2 for the following subgroups:

- Region (North America, Western Europe, Eastern Europe, Asia Pacific, Latin America)
- Designated dementia subtype (Alzheimer's disease or FTD spectrum disorders, Vascular dementia, Parkinson's disease dementia or dementia with Lewy bodies)
- Parkinson's disease dementia (yes)
- Alzheimer's disease (yes)
- FTD spectrum disorders (yes)

- Dementia with Lewy bodies (yes)
- Age group (<75 or ≥75 years old)
- Sex (male or female)
- Primary race (white or non-white)
- Screening dementia severity (mild, moderate, severe)
- OL Baseline antidementia medication (memantine and/or acetylcholinesterase inhibitors) use (yes, no)
- Antipsychotic use within 14 days of screening (yes, no).
- OL Baseline MMSE (<18, ≥18)

For each subgroup, Kaplan-Meier curves of time to relapse and time to discontinuation from the double-blind period for any reason (other than termination of the study by the sponsor) will be presented for each treatment group. The hazard ratios with corresponding 95% CIs obtained from the Cox regression model for each subgroup will be presented in forest plots for the primary and key secondary efficacy variables.

12.5.2 Open-Label Period

For each subgroup listed above, the number and percentage of subjects in the Open-label Safety Analysis Set who met the response criteria for randomization will be summarized.

SAPS-H+D, SAPS-H, and SAPS-D assessments from the open-label period for the Open-label Safety Analysis Set will be descriptively summarized (Section 13.3.2) by designated dementia subtype.

13 METHODS OF EFFICACY ANALYSES

13.1 Primary Efficacy Analysis

The primary efficacy endpoint is the time from randomization to relapse in the double-blind period. The hypotheses are stated in terms of the hazard ratio, $HR = \lambda_{PIM} / \lambda_{PBO}$, where λ_{PIM} and λ_{PBO} denote the hazard rates of relapse for the pimavanserin group (20 mg and 34 mg combined) and the placebo group, respectively.

The null hypothesis for the primary efficacy endpoint is: HR = 1

The alternative hypothesis for the primary efficacy endpoint is: HR ≠ 1

The primary efficacy analysis will be based on adjudicated relapse events determined by the IAC for the ITT Analysis Set. If the study is stopped at the interim analysis, the primary efficacy analysis will be based on the interim analysis dataset (see Section 16 for additional details). If the study is not stopped at the interim analysis, the primary analysis will be based on all data collected up until the study discontinuation date; data collected after the study discontinuation date will not be used in the primary analysis.

The treatment effect will be measured by the HR. The pimavanserin group includes all subjects taking pimavanserin irrespective of the dose (20 mg or 34 mg). The time from randomization to relapse in the double-blind period will be compared between treatment groups using a Cox regression model with effects for treatment group, designated dementia subtype (Alzheimer's disease or FTD spectrum disorders, Vascular dementia, Parkinson's disease dementia or dementia with Lewy bodies), and region (levels: North America, Western Europe, Eastern Europe, Asia Pacific, Latin America) as covariates in the model. A robust sandwich-type variance method will be applied to compute the variance of the estimated HR (Lin et. al., 1989), which can be computed as $\mathbf{D'D}$ where $\mathbf{D}$ is the matrix of DFBETA residuals. The Cox exact discrete method will be used to compute the partial likelihood function to account for tied relapse times. The estimated HR, SE, 95% confidence interval, and p-value obtained from the Cox regression model will be presented. The algorithm for censoring of subjects who discontinued early, completed the study without having experienced a relapse event, or were ongoing at the time of database cutoff is described in Section 5.2.3. The number and percentage of subjects having the following possible reasons for censoring will be tabulated for each treatment group:

- Completed Week 26 without relapse;
- Prematurely discontinued prior to Week 26, for reasons other than termination of the study by the sponsor;
- Ongoing at the time of database cutoff (for the interim analysis only);
- Termination of the study by the sponsor (for the final analysis only).

Kaplan-Meier curves of time from randomization to relapse will be presented for each treatment group. In addition, Kaplan-Meier estimate of median survival (relapse free time) and probability estimates at the double-blind Weeks 4, 8, 12, 18, and 26 along with pointwise 95% confidence intervals will be provided. The confidence intervals will be calculated using a log-log transformation and Greenwood's variance formula.

The number and percentage of subjects meeting at least one of the following relapse criteria will be tabulated for each treatment group:

- Experienced a ≥30% increase (worsening) on the SAPS-H+D total score AND has a CGI-I score of 6 (much worse) or 7 (very much worse), relative to the double-blind Baseline;
- Treated with an antipsychotic (other than pimavanserin) for dementia-related delusions and/or hallucinations;
- Stopped study drug or withdrew from study for lack of efficacy, (as reported by the subject or study partner/caregiver) or the Investigator discontinues study drug due to lack of efficacy;
- Hospitalized for worsening dementia-related psychosis.

The number and percentage of subjects meeting each of the relapse criteria will be tabulated. A subject who meets more than one of the relapse criteria will be counted once for each of the relapse criteria that were met.

13.1.1 Deviations From Non-Informative Censoring

One of the assumptions underlying the primary analysis is that the censoring mechanism is non-informative. In order to assess the sensitivity of the primary analysis to deviations from this assumption, a reference-based multiple imputation method will be used (Lu et. al., 2015). Subjects who don't relapse but who discontinue from the study prior to the Week 26 assessment, for any reason other than termination of the study by the sponsor, will have time to relapse values imputed conditionally on their observed censored values. Censoring due to completion of the 26 week double-blind period, study being terminated by the sponsor or ongoing at the time of the interim analysis datacut (for the interim analysis only), will be treated as non-informative and no imputation will be performed for these subjects. The

imputation algorithm will be based on the assumption that the risk of relapse (i.e. hazard rate) for pimavanserin subjects who prematurely discontinue is the same as that for placebo subjects. If the imputed event time exceeds the applicable cutoff date for a planned analysis, or the date corresponding to Week 26 for the subject, whichever comes first, the time-to-event will be right censored at the earlier of the two dates. Random number seeds of 201938 and 202075 will be used for the posterior draw of model parameters from the imputation model and for imputation of censored event times, respectively. A total of 1000 imputed datasets will be created.

The imputation model will be a Cox regression model with piecewise exponential baseline hazards fit to the ITT Analysis Set with treatment, designated dementia subtype, and region as covariates. A noninformative improper prior for the baseline hazard parameters and an improper uniform prior for the regression coefficients will be assumed. For the piecewise constant baseline hazards, the post-baseline follow-up time axis will be partitioned into 2 disjoint intervals, where the interval cut-points will be such that there will be a similar amount of relapse events in each interval. Sampling of model parameters will set the burn-in to 200 and every 10th sample will be kept. [redacted]

When imputing censored event times, posterior samples for the treatment parameter will be set to 0 so that the assumption that the risk of relapse (i.e. hazard rate) for pimavanserin subjects who prematurely discontinue is the same as that for placebo subjects. Additional details of how censored event times will be imputed can be found in Lu et. al. (2015). If the imputation model cannot be fit due to sparse covariate data, this will be documented and the imputation model will be modified to include only treatment as a covariate.

For each of the 1000 imputed datasets, a Cox regression model with effects for treatment group, designated dementia subtype, and region will be used to obtain an estimate of the log

hazard ratio and associated standard error. These estimates will be combined using Rubin's rule to obtain an overall hazard ratio, standard error, 95% confidence interval, and p-value (Rubin, 1976).

13.1.2 Sensitivity Analyses

If the study is stopped at the interim analysis, the following sensitivity analyses will be performed:

- Time to relapse based on adjudicated relapse events using data collected up to the study discontinuation date (ITT analysis set). Subjects without any such relapse event will be censored at the last SAPS-H+D assessment prior to the study discontinuation date.
- Time to relapse based on adjudicated relapse events occurring up to the study discontinuation date, plus any additional investigator determined relapse events that occur after the study discontinuation date (ITT analysis set). Subjects without any such relapse event will be censored at the last SAPS-H+D assessment.
- Time to relapse based on investigator determined relapse events using data collected up to the study discontinuation date (ITT analysis set). Subjects without any such relapse event will be censored at the last SAPS-H+D assessment prior to the study discontinuation date.
- Time to relapse based on adjudicated relapse events using the interim analysis dataset and using the Per-protocol analysis set. Subjects without any such relapse event will be censored at the last SAPS-H+D assessment.

If the study is not stopped at the interim analysis, the following sensitivity analyses will be performed:

- Time to relapse based on adjudicated relapse events occurring up to the study discontinuation date, plus any additional investigator determined relapse events that occur after the study discontinuation date (ITT analysis set). Subjects without any such relapse event will be censored at the last SAPS-H+D assessment.

- Time to relapse based on investigator determined relapse events using data collected up to the study discontinuation date (ITT analysis set). Subjects without any such relapse event will be censored at the last SAPS-H+D assessment prior to the study discontinuation date.
- Time to relapse based on adjudicated relapse events occurring up to the study discontinuation date using the Per-protocol analysis set. Subjects without any such relapse event will be censored at the last SAPS-H+D assessment prior to the study discontinuation date.

The same Cox regression model as described in Section 13.1 will be used for each of the sensitivity analyses. The results will be presented in a summary table similar to that described for the primary efficacy analysis. The estimated HRs with corresponding 95% CIs from the primary efficacy analysis and the sensitivity analyses will be presented in a forest plot.

13.2 Key Secondary Efficacy Analysis

The key secondary endpoint is time from randomization to discontinuation from the double-blind period for any reason (other than termination of the study by the sponsor). The hypotheses are stated in terms of the hazard ratio, $HR = \lambda_{PIM} / \lambda_{PBO}$, where λ_{PIM} and λ_{PBO} denote the hazard rates of discontinuation from the double-blind period for any reason (other than termination of the study by the sponsor) for the pimavanserin group and the placebo group, respectively.

The null hypothesis for the key secondary endpoint is: HR = 1

The alternative hypothesis for the key secondary endpoint is: HR ≠ 1

The key secondary efficacy endpoint will be analyzed using the same Cox regression model described for the primary efficacy endpoint. The null hypothesis for the key secondary efficacy endpoint cannot be rejected unless the null hypothesis for the primary efficacy endpoint has been rejected. Adjustment of stopping boundaries for the key secondary endpoint analysis to account for multiple endpoints and group sequential testing is described in Section 16.

The estimated HR, SE, 95% confidence interval, and p-value obtained from the Cox regression model will be presented. The algorithm for censoring of subjects who completed

the 26-week double-blind period without experiencing a relapse event, or who administratively exited the study due to study termination by the sponsor, or who were ongoing at the time of database cutoff is described in Section 5.2.4. The number and percentage of subjects having the following possible reasons for censoring will be tabulated for each treatment group:

- Completed Week 26 without relapse;
- Termination of the study by the sponsor (for the final analysis only);
- Ongoing at the time of the interim database cutoff (for the interim analysis only).

Kaplan-Meier curves of time from randomization to discontinuation from the double-blind period for any reason (other than termination of the study by the sponsor) will be presented for each treatment group. In addition, Kaplan-Meier probability estimates of median survival and at the double-blind Weeks 4, 8, 12, 18, and 26 along with pointwise 95% confidence intervals will be provided. The confidence intervals will be calculated using a log-log transformation and Greenwood's variance formula.

The number and percentage of subjects meeting the following discontinuation reasons for the double-blind period will be tabulated for each treatment group:

- Relapse of hallucinations and delusions associated with dementia-related psychosis adjudicated by the IAC;
- Adverse event;
- Death;
- Non-compliance with study drug;
- Used of prohibited medications;
- Physician decision;
- Protocol violation;
- Pregnancy;
- Subject withdrew consent;
- Lost to follow-up;
- Other.

13.2.1 Deviations From Non-Informative Censoring

For the key secondary endpoint, censoring can only occur among those subjects who complete the 26 week double-blind period or who discontinue due to the study being terminated by the sponsor. This censoring is non-informative and therefore sensitivity analyses to address informative censoring are not needed.

13.2.2 Sensitivity Analysis

Sensitivity analysis for the key secondary efficacy endpoint will be performed on the Per-protocol Analysis Set. The results will be presented in a summary table similar to that described for the key secondary efficacy analysis.

13.3 Exploratory Efficacy Analysis

13.3.1 Analysis for the Double-blind Period

Exploratory efficacy variables SAPS-H+D, SAPS Hallucinations domain, SAPS Delusions domain, CGI-S, ZBI, KSS, EQ-5D VAS, and EQ-5D-5L descriptive system evaluated in the double-blind period will be analyzed for the ITT Analysis Set. For each exploratory efficacy variable other than the EQ-5D-5L descriptive system, observed and change from double-blind Baseline values at each analysis visit will be summarized using descriptive statistics by treatment group (placebo and pimavanserin). In addition, the descriptive summaries will be provided by pimavanserin dose level (20 and 34 mg) for the double-blind pimavanserin group.

For each exploratory efficacy variable other than the EQ-5D-5L descriptive system, treatment comparisons will be made between the placebo group and the pimavanserin group (20 mg and 34 mg combined) using a MMRM and an ANCOVA. No adjustment for multiple comparisons will be made among the exploratory efficacy variables. At each post-baseline analysis visit, LS means and standard errors (SE), the corresponding 95% confidence interval and p-value will be presented. In addition, plots of LS mean ± SE versus analysis visit will be provided.

13.3.1.1 Mixed Model Repeated Measures

For each exploratory efficacy variable other than the EQ-5D-5L descriptive system, the change from baseline will be analyzed using mixed model repeated measures (MMRM) method with missing data assumed to be missing at random. The dependent variable will be the change from double-blind Baseline. Assessments more than 14 days after last dose of study drug will be excluded from the analysis. The independent variables in the model will include treatment group (placebo or pimavanserin), designated dementia subtype, region, analysis visit, treatment group by analysis visit interaction, double-blind Baseline score, and double-blind Baseline score by analysis visit interaction. 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 analyzed in the model and least squares (LS) adjusted 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:

- Heterogeneous Toeplitz;
- Heterogeneous compound symmetry;
- Heterogeneous autoregressive(1);
- Toeplitz;
- Compound symmetry
- Autoregressive(1);
- Variance components.

The first covariance structure that allows for convergence will be selected for the final model.

13.3.1.2 ANCOVA

For each exploratory efficacy variable other than the EQ-5D-5L descriptive system, the change from baseline will be analyzed using an analysis of covariance (ANCOVA) at each post-baseline analysis visit and using the last observation carried forward (LOCF) method of imputation for missing data. The LOCF method will carry forward the double-blind Baseline value if necessary; assessments more than 14 days after last dose of study drug will not be carried forward. The dependent variable will be the change from double-blind Baseline. The independent variables in the model will include treatment group (placebo or pimavanserin), designated dementia subtype, region, and double-blind Baseline score. Least squares (LS) adjusted means will be estimated using observed margins.

13.3.1.3 Treatment Comparisons

For each exploratory efficacy variable other than the EQ-5D-5L descriptive system, the treatment comparisons will be based on the difference in LS means obtained from the statistical model (MMRM or ANCOVA) at each post-baseline analysis visit for the respective exploratory efficacy variable and will be tested at a 2-sided alpha level of 0.05.

Summary statistics for the observed and change from Baseline values will be presented for all analysis visits. For the change from double-blind Baseline values at each post-baseline analysis visit, LS means and standard errors (SE), the between-group difference in LS means

with the corresponding 95% confidence interval, p-value and effect size will be presented. In addition, LS mean ± SE over time for the change from Baseline values by treatment group will be presented in line plots.

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

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

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

13.3.2 Analysis for the Open-label Period

Exploratory variables SAPS-H+D, SAPS Hallucinations domain, SAPS Delusions domain, CGI-S, ZBI, KSS, EQ-5D VAS, and EQ-5D-5L descriptive system assessed in the open-label period will be summarized for the Open-label Safety Analysis Set.

For each exploratory variable other than the EQ-5D-5L descriptive system, the observed and change from open-label Baseline values at each analysis visit will be summarized using descriptive statistics. The following summaries will be provided:

1. all subjects combined;
2. by randomization status (not randomized and randomized);
3. by final dose level (20 and 34 mg) of open-label pimavanserin.

The mean change from open-label Baseline to each analysis visit and 95% CI will be provided for the following exploratory endpoints: SAPS-H+D, SAPS Hallucinatins domain, SAPS Delusions domain, CGI-S, ZBI, KSS and EQ-5D VAS.

13.3.3 SAPS-H+D

For the double-blind period, observed SAPS-H+D total score will be summarized using descriptive statistics at double-blind Baseline, double-blind Weeks 1, 2, 4, 6, 10, 14, 18, 22, and 26. The change from double-blind Baseline values will also be summarized at the double-blind Weeks 1, 2, 4, 6, 10, 14, 18, 22, and 26. The change from double-blind Baseline

in SAPS-H+D total score will be analyzed using the MMRM and ANCOVA models as described in Sections 13.3.1.1, 13.3.1.2, and 13.3.1.3.

For the open-label period, observed SAPS-H+D total score will be summarized using descriptive statistics at open-label Baseline, open-label Weeks 2, 4, 8, and 12. The change from open-label Baseline values will also be summarized at the open-label Weeks 2, 4, 8, and 12.

13.3.4 SAPS Hallucinations Domain

For the double-blind period, observed SAPS Hallucinations domain score will be summarized using descriptive statistics at the double-blind Baseline and Weeks 1, 2, 4, 6, 10, 14, 18, 22, and 26. The change from double-blind Baseline values will also be summarized at the double-blind Weeks 1, 2, 4, 6, 10, 14, 18, 22, and 26. The change from double-blind Baseline in SAPS Hallucinations domain score will be analyzed using the MMRM and ANCOVA models as described in Sections 13.3.1.1, 13.3.1.2, and 13.3.1.3.

For the open-label period, observed SAPS Hallucinations domain score will be summarized using descriptive statistics at the open-label Baseline, Weeks 2, 4, 8, and 12. The change from open-label Baseline values will also be summarized at the open-label Weeks 2, 4, 8, and 12.

13.3.5 SAPS Delusions Domain

For the double-blind period, observed SAPS Delusions domain score will be summarized using descriptive statistics at the double-blind Baseline and Weeks 1, 2, 4, 6, 10, 14, 18, 22, and 26. The change from double-blind Baseline values will also be summarized at the double-blind Weeks 1, 2, 4, 6, 10, 14, 18, 22, and 26. The change from double-blind Baseline in SAPS Delusions domain score will be analyzed using the MMRM and ANCOVA models as described in Sections 13.3.1.1, 13.3.1.2, and 13.3.1.3.

For the open-label period, observed SAPS Delusions domain score will be summarized using descriptive statistics at the open-label Baseline and Weeks 2, 4, 8, and 12. The change from open-label Baseline values will also be summarized at the open-label Weeks 2, 4, 8, and 12.

13.3.6 CGI-S Dementia-Related Psychosis

For the double-blind period, observed CGI-S score will be summarized using descriptive statistics at the double-blind Baseline and Weeks 1, 2, 4, 6, 10, 14, 18, 22, and 26. The change from double-blind Baseline values will also be summarized at the double-blind Weeks 1, 2, 4, 6, 10, 14, 18, 22, and 26. The change from double-blind Baseline in CGI-S score will be analyzed using the MMRM and ANCOVA models as described in Sections 13.3.1.1, 13.3.1.2, and 13.3.1.3.

For the open-label period, observed CGI-S score will be summarized using descriptive statistics at the open-label Baseline and Weeks 2, 4, 8, and 12. The change from open-label Baseline values will also be summarized at the open-label Weeks 2, 4, 8, and 12.

13.3.7 Responder Analysis

For the open-label period, the number and percentage of subjects in the Open-label Safety Analysis Set that met the following response criteria (≥30% reduction from open-label Baseline on SAPS-H+D Total Score and CGI-I score of 1 or 2, relative to open-label Baseline) will be provided for Week 2, Week 4, Week 8, and Week 12, and at both Weeks 8 and 12.

13.3.8 ZBI Score

For the double-blind period, observed ZBI scores will be summarized using descriptive statistics at the double-blind Baseline and Weeks 6, 14, and 26. The change from double-blind Baseline values will also be summarized at the double-blind Weeks 6, 14, and 26. The change from double-blind Baseline in ZBI score will be analyzed using the MMRM and ANCOVA models as described in Sections 13.3.1.1, 13.3.1.2, and 13.3.1.3.

For the open-label period, observed ZBI score will be summarized using descriptive statistics at the open-label Baseline and Week 12. The change from open-label Baseline values will also be summarized at the open-label Week 12.

13.3.9 KSS Score

For the double-blind period, observed KSS scores will be summarized using descriptive statistics at double-blind Baseline and Weeks 6, 14, and 26. The change from double-blind

Baseline values will also be summarized at the double-blind Weeks 6, 14, and 26. The change from double-blind Baseline in KSS score will be analyzed using the MMRM and ANCOVA models as described in Sections 13.3.1.1, 13.3.1.2, and 13.3.1.3.

For the open-label period, observed KSS scores will be summarized using descriptive statistics at open-label Baseline and open-label Week 12. The change from open-label Baseline values will also be summarized at the open-label Week 12.

13.3.10 CGI-I Dementia-Related Psychosis

For the double-blind period, observed CGI-I scores will be summarized using descriptive statistics at double-blind Weeks 1, 2, 4, 6, 10, 14, 18, 22, and 26. The CGI-I score will be analyzed using the MMRM and ANCOVA models as described in Sections 13.3.1.1, 13.3.1.2, and 13.3.1.3. For both the MMRM and ANCOVA models, the covariate double-blind Baseline score is the double-blind Baseline CGI-S score.

For the open-label period, observed CGI-I scores will be summarized using descriptive statistics at open-label Weeks 2, 4, 8, and 12.

13.3.11 EQ-5D VAS

For the double-blind period, observed EQ-5D VAS scores will be summarized using descriptive statistics at double-blind Baseline and double-blind Weeks 6, 14, and 26. The change from double-blind Baseline values will also be summarized at the double-blind Weeks 6, 14, and 26. The change from double-blind Baseline in QD-5D VAS score will be analyzed using the MMRM and ANCOVA models as described in Sections 13.3.1.1, 13.3.1.2, and 13.3.1.3.

For the open-label period, observed EQ-5D VAS scores will be summarized using descriptive statistics at open-label Baseline and Week 12. The change from open-label Baseline scores will also be summarized at the open-label Week 12.

13.3.12 EQ-5D 5L Descriptive System

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

For the double-blind period, EQ-5D-5L descriptive system will be summarized at double-blind Baseline and double-blind Weeks 6, 14, and 26.

For the open-label period, EQ-5D-5L descriptive system will be summarized at open-label Baseline and open-label Week 12.

13.3.13 Pareto Classification of Health Change

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 each post-Baseline visit, 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

For the double-blind period, the proportion of subjects in each PCHC category will be summarized descriptively at the double-blind Weeks 6, 14, and 26.

For the open-label period, the proportion of subjects in each PCHC category will be summarized descriptively at the open-label Week 12.

14 SAFETY ANALYSES

The safety endpoints for this study are as follows:

- Treatment-emergent adverse events (TEAEs)
- Serious adverse events (SAEs)
- Withdrawals due to adverse events (AEs)
- Potentially clinically important changes in other safety assessments
- Global Clinician Assessment of Suicidality (GCAS) score

- Mini-Mental State Examination (MMSE) score
- Extrapyramidal Symptom Rating Scale A (ESRS-A) score

For the open-label period, safety data will be summarized for the Open-label Safety Analysis Set. The following summaries will be provided:

1. all subjects combined;
2. by randomization status (not randomized and randomized);
3. by final dose level (20 and 34 mg) of open-label pimavanserin administered during the open-label period.

For the double-blind period, safety data will be summarized for the Double-blind Safety Analysis Set. The following summaries will be provided:

1. by treatment group (placebo and pimavanserin)
2. by dose level (20 and 34 mg) of pimavanserin administered during the double-blind period for the pimavanserin group.

14.1 Adverse Events

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

An AE (classified by preferred term) will be considered a treatment emergent AE (TEAE) if the onset date was on or after the date of first dose of open-label pimavanserin and up to 30 days past the last dose of study drug. AEs reported on Day 1 based on open-label Baseline (pre-dose) findings (e.g. clinically significantly abnormal vital signs, laboratory test results, or electrocardiogram parameters) will not be considered as TEAEs.

Each TEAE will be assigned to either the open-label or double-blind period based on the onset date. For randomized subjects, if the onset date was on or after the date of first dose of double-blind study drug, the TEAE will be assigned to the double-blind period. All other TEAE will be assigned to the open label period.

The event counts and the number and percentage of subjects reporting TEAEs 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, 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 overall for the open-label period and ≥5% of subjects in either treatment group for the double-blind treatment period), with treatment emergent serious AEs, 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. The display in these tables will be sorted alphabetically by SOC and then by descending subject frequency for the preferred terms using combined counts from both 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. This table will be sorted by descending subject frequency using combined counts from all subjects.

An AE listing by subject will display all events in the open-label period and in the double-blind period, 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. study period, date of onset, date resolved, study dose level at AE onset, severity, frequency, outcome, relationship to study drug, action taken with study drug, and required therapy. When a date is presented, the study day associated with the date will also be displayed. If the onset date is prior to, or during, the open-label period, the study day will be calculated relative to the date of first dose of open-label pimavanserin; if the onset date is during the double-blind period, the study day will be calculated relative to the date of first dose of double-blind study drug. Separate listings will be presented for subjects with SAEs, subjects with AEs leading to discontinuation and subjects with fatal AEs (if any). In these listings, an indicator for treatment-emergent events will also be included.

14.2 Clinical Laboratory Variables

Clinical chemistry serum tests include the following:

- Sodium (Na), potassium (K), chloride (Cl), phosphorus (P), calcium (Ca), carbon dioxide (CO2), blood urea nitrogen (BUN), creatinine (CR), uric acid

- Alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma-glutamyl transpeptidase (GGT), alkaline phosphatase (ALP), total bilirubin (TBIL), lactate dehydrogenase (LDH)
- Magnesium (Mg) should only be performed at Visit 1 (Screening)
- Vitamin B12 should only be performed at Visit 1 (Screening)
- HbA1c should only be performed at Visit 1 (Screening), Visit 6, and Visit 17/ET
- Glucose
- Albumin (ALB), total protein
- Prolactin
- Thyroid stimulating hormone (TSH) and free T4
    - TSH should be performed only at Visit 1 (Screening)
    - Free T4 will be measured only if TSH is abnormal
- Lipid Panel
    - Total cholesterol, high-density lipoprotein (HDL)-cholesterol, triglycerides, low-density lipoprotein(LDL)-cholesterol, cholesterol/HDL ratio, non-HDL cholesterol

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
    - Reticulocyte count

Urinalysis tests include the following:

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

All laboratory test results (including urine drug screen) are from a central laboratory and will be listed. The listings will include date and study day of collection. All units will be displayed

in Système International [SI] units. Out of range values will be flagged in the data listings (e.g. 'L' or 'H').

14.2.1 Analysis for the Open-label Safety Analysis set

Clinical laboratory tests are performed at Screening, Baseline and Week 12 visits during the open-label period. Clinical laboratory values for hematology, chemistry and urinalysis (specific gravity and pH) will be summarized using descriptive statistics at Baseline and Week 12 visits. The change from Baseline values will also be summarized at the Week 12 visit.

The overall minimum and maximum post-baseline observed and change from Baseline values during the open-label period 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 and Week 12 visit, and the denominator is the number of subjects with non-missing values for the given parameter and visit.

Laboratory values (except the ones that were only assessed at screening) will also be summarized in shift tables, to determine the number and percentage of subjects with values classified as below (low), within (normal), and above (high) normal ranges at Week 12 visit, overall post-baseline minimum and overall post-baseline maximum during the open-label period. For the Week 12 visit shift summary, the denominator is the number of subjects with non-missing values at Baseline and Week 12 visit for the given parameter. For the 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 during the open-label period for the given parameter. For hemoglobin, hematocrit and uric acid, the shift summaries will be presented for each gender as well as for both genders combined.

The number and percentage of subjects with potentially clinically important (PCI) laboratory values at Week 12 visit and overall post-baseline during the open-label period will be summarized for selected parameters. PCI criteria for hematology and chemistry are listed in Table 5 and PCI criteria for urinalysis are listed in Table 6. For the overall post-baseline summaries of PCI values, all post-baseline values during the open-label period will be considered, including unscheduled and out of window values. Subjects with multiple PCI

values for a given parameter will be counted only once for that parameter. For the by-visit summary, the numerator for the percentage is the number of subjects with a post-baseline PCI value for the given parameter and visit, and the denominator is the number of subjects with non-missing values for the given parameter and visit. For the overall post-baseline summary, the numerator for the percentage is the number of subjects with at least 1 post-baseline PCI laboratory value during the open-label period for the given parameter, and the denominator is the number of subjects with at least 1 post-baseline laboratory value during the open-label period for the given parameter. For hemoglobin, hematocrit and uric acid, the count and percentage of subjects with PCI values will be presented for each sex as well as for both sexes combined. Subjects with any post-baseline PCI values will be presented in an additional listing.

14.2.2 Analysis for the Double-blind Safety Analysis Set

In addition to the open-label assessments at Screening, Baseline, and Week 12 visits, clinical laboratory tests are performed at Weeks 18, 26, and 38 visits during the double-blind period. Summaries for hematology, chemistry and urinalysis for the Double-Blind Safety Analysis Set will be similar to those described above for the Open-Label Safety Analysis Set and will include the following:

- Observed values and change from open-label baseline by visit including both open-label and double-blind visits; overall minimum and maximum for the open-label period; overall minimum and maximum for the double-blind period.
- Observed values and change from double-blind baseline by visit for the double-blind period; overall minimum and maximum for the double-blind period.
- Shifts from open-label baseline by visit including both open-label and double-blind visits; overall minimum and maximum for the open-label period; overall minimum and maximum for the double-blind period.
- Shifts from double-blind baseline by visit for the double-blind period; overall minimum and maximum for the double-blind period.

- Incidence of PCI values by visit including both open-label and double-blind visits; open-label overall post-baseline; double-blind overall post-baseline, based on the OL baseline.
- Incidence of PCI values by visit for the double-blind period; double-blind overall post-baseline, based on the DB baseline.

Urinalysis tests with categorical results will be summarized (number and percent of subjects in each category) by visit including both open-label and double-blind visits.

Table 5 Criteria for PCI 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 (WBC 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
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	NA	>1.5 ULN	umol/L	NA	>1.5 ULN
Triglycerides	mg/dL	NA	>300	mmol/L	NA	>3.39
GGT	U/L	NA	≥3 ULN	U/L	NA	≥3 ULN

NA = not applicable

Table 6 Criteria for PCI Laboratory Values - Urinalysis

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

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

14.3 Vital Signs

14.3.1 Analysis for the Open-label Safety Analysis Set

Vital signs are assessed at Screening, Baseline, and Weeks 2, 4, 8, and 12 during the open-label period.

Respiration rate, systolic blood pressure, diastolic blood pressure, and pulse rate, weight, height (only at Screening), and derived BMI will be summarized using descriptive statistics at Baseline and Weeks 2, 4, 8, and 12 during the open-label period. The change from Baseline values will also be summarized at the Weeks 2, 4, 8 and 12.

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 PCI vital sign values at Weeks 2, 4, 8, and 12, and overall post-baseline during the open-label period will be summarized for selected parameters. For the overall post-Baseline summaries, all post-Baseline values during the open-label period 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 value during the open-label period for the given parameter and visit, and the denominator is the number of subjects with non-missing values for the given parameter and visit. For the overall post-baseline summary, the numerator for the percentage is the number of subjects with at least 1 post-baseline PCI vital sign value during the open-label period for the given parameter, and the denominator is the number of subjects with at least 1 post-baseline vital sign value during the open-label period for the given parameter. Subjects with any post-baseline PCI values will be presented in an additional listing.

14.3.2 Analysis for the Double-blind Safety Analysis Set

In addition to the open-label assessments at Screening, Baseline and Weeks 2, 4, 8, and 12, vital signs are assessed at Weeks 13, 14, 16, 18, 22, 26, 30, 34, and 38 during the double-blind period.

Vital sign summaries for the Double-Blind Safety Analysis Set will be similar to those described above for the Open-Label Safety Analysis Set and will include the following:

- Observed values and change from open-label baseline by visit including both open-label and double-blind visits; overall minimum and maximum for the open-label period; overall minimum and maximum for the double-blind period.
- Observed values and change from double-blind baseline by visit for the double-blind period; overall minimum and maximum for the double-blind period.
- Incidence of PCI values by visit including both open-label and double-blind visits; open-label overall post-baseline; double-blind overall post-baseline.
- Incidence of PCI values by visit for the double-blind period; double-blind overall post-baseline.

Table 7 Criteria for PCI 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.

14.4 Electrocardiogram

14.4.1 Analysis for the Open-label Safety Analysis Set

All ECGs will be centrally read; the interpretation by the central cardiologist is considered the official interpretation. 12-lead electrocardiogram (ECG) assessments are performed at Screening, Baseline, Weeks 2 and 12 during the open-label period.

ECG results (ventricular rate, PR, QRS, QT, and QTcF intervals) will be summarized using descriptive statistics at Baseline, Weeks 2 and 12 during the open-label period. The change from Baseline values will also be summarized at the Weeks 2 and 12.

QTcF (msec) observed and change from Baseline values during the open-label period will also be classified into the following categories:

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

The number and percentage of subjects in each category will be summarized at each visit and for the overall post-baseline maximum during the open-label period.

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 PCI ECG values at Weeks 2 and 12, and overall post-baseline maximum during the open-label period will be

summarized for selected parameters. For the by-visit summary, the numerator for the percentage is the number of subjects with a post-baseline PCI ECG value during the open-label period for the given parameter and visit, and the denominator is the number of subjects with non-missing values during the open-label period for the given parameter and visit. For the overall post-Baseline summary, the numerator for the percentage is the number of subjects with at least 1 post-baseline PCI ECG value during the open-label period for the given parameter, and the denominator is the number of subjects with at least 1 post-baseline ECG value during the open-label period for the given parameter. Subjects with any post-baseline PCI values will be presented in an additional listing.

Overall ECG interpretation at Baseline, Weeks 2 and 12 and overall post-baseline worst interpretation during the open-label period will be summarized descriptively. Within each summary time frame a subject is counted only once in the category with the worst interpretation ("abnormal, clinically significant" is worse than "abnormal, not clinically significant" is worse than "normal"). For overall post-Baseline worst interpretation, all post-baseline interpretations during the open-label period will be considered, including unscheduled and out-of-window interpretations. For the by-visit summary, the denominator for the percentage is the number of subjects with non-missing interpretations during the open-label period at the given visit. For the overall post-Baseline summary the denominator is the number of subjects with at least one post-baseline interpretation during the open-label period.

Overall ECG interpretation results will also be summarized in shift tables, to determine the number and percentage of subjects with results classified as normal and abnormal at Weeks 2 and 12, and overall post-Baseline worst interpretation during the open-label period. For the by-visit shift summary, the denominator is the total number of subjects with non-missing results at Baseline and the given visit during the open-label period. For the overall post-Baseline summary the denominator is the number of subjects with a non-missing Baseline and at least one post-baseline result during the open-label period.

14.4.2 Analysis for the Double-blind Safety Analysis Set

In additional to the open-label assessments at Screening, Baseline and Weeks 2 and 12, ECG assessments are performed at Weeks 18, 26, and 38 during the double-blind period.

ECG summaries for the Double-Blind Safety Analysis Set will be similar to those described above for the Open-Label Safety Analysis Set and will include the following:

- Observed values and change from open-label baseline by visit including both open-label and double-blind visits.
- Observed values and change from double-blind baseline by visit for the double-blind period.
- Categorical analysis of QTcF Observed Values and Change from open-label Baseline by Visit including both open-label and double-blind visits; overall maximum for the open-label period; overall maximum for the double-blind period.
- Categorical analysis of QTcF Observed Values and Change from double-blind Baseline by Visit for the double-blind period; overall maximum for the double-blind period.
- Incidence of PCI values by visit including both open-label and double-blind visits; open-label overall post-baseline; double-blind overall post-baseline.
- Incidence of PCI values by visit for the double-blind period; double-blind overall post-baseline.
- Overall interpretation by visit including both open-label and double-blind visits; overall worst interpretation for the open-label period; overall worst interpretation for the double-blind period.
- Overall interpretation shift from open-label baseline by visit including both open-label and double-blind visits; overall worst interpretation for the open-label period; overall worst interpretation for the double-blind period.
- Overall interpretation shift from double-blind baseline by visit for the double-blind period; overall worst interpretation for the double-blind period.

Table 8 Criteria for PCI ECG Values

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

14.5 Physical and Neurological Examinations

14.5.1 Analysis for the Open-label Safety Analysis Set

Physical examinations and neurological examinations (cranial nerves, motor, sensory, reflexes, gait, and coordination), are performed at Screening, Baseline, and Week 12 during the open-label period.

Physical and neurological examination results (normal, abnormal, and not done) will be summarized in a frequency table by body/neurological system at Baseline and Week 12.

14.5.2 Analysis for the Double-blind Safety Analysis Set

In addition to the open-label assessments at Screening, Baseline and Week 12, physical examinations and neurological examinations (cranial nerves, motor, sensory, reflexes, gait, and coordination), are performed at Weeks 22 and 38 during the double-blind period. Summaries for the Double-blind Safety Analysis Set will include both the open-label and double-blind periods.

Physical and neurological examination results (normal, abnormal, and not done) during the open-label and double-blind periods will be summarized in a frequency table by body/neurological system.

14.6 Global Clinician Assessment of Suicidality

14.6.1 Analysis for the Open-label Safety Analysis Set

Global Clinician Assessment of Suicidality (GCAS) scores are measured at Screening, Baseline, and Weeks 2, 4, 8, and 12 visits during the open-label period. The number and percentage of subjects for each GCAS rating (0-4) will be tabulated by visit. The number and

percentage of subjects reporting any post-Baseline GCAS score of 3 or 4 during the open-label period will also be tabulated.

14.6.2 Analysis for the Double-blind Safety Analysis Set

In addition to the open-label assessments at Screening, Baseline and Weeks 2, 4, 8, and 12 during the open-label period, CGAS is assessed at Weeks 13, 14, 16, 18, 20, 22, 24, 26, 30, 34, 38 during the double-blind period. Summaries for the Double-blind Safety Analysis Set will include both the open-label and double-blind periods.

The number and percentage of subjects for each GCAS rating (0-4) will be tabulated by visit. The number and percentage of subjects reporting any post-Baseline GCAS score of 3 or 4 during the open-label period and during the double-blind period will also be tabulated .

14.7 Mini-Mental State Examination

14.7.1 Analysis for the Open-label Safety Analysis Set

Mini-Mental State Examination (MMSE) scores are measured at Screening, Baseline, and Weeks 2, 4, 8, and 12 during the open-label period.

Observed MMSE scores will be summarized using descriptive statistics at Baseline, Weeks 2, 4, 8, and 12. The change from Baseline scores at Weeks 2, 4, 8, and 12 values will also be summarized.

14.7.2 Analysis for the Double-blind Safety Analysis Set

In addition to the open-label assessments at Screening, Baseline and Weeks 2, 4, 8, and 12, MMSE scores are measured at Weeks 13, 14, 16, 18, 22, 26, 30, 34, and 38 during the double-blind period. Summaries for the Double-blind Safety Analysis Set will include both the open-label and double-blind study periods.

Observed MMSE values and change from open-label baseline will be summarized by visit for both open-label and double-blind visits.

Observed MMSE values and change from double-blind baseline will be summarized by visit for the double-blind visits. In addition, treatment comparisons will be made between the placebo group and the pimavanserin group (20 mg and 34 mg combined) for change from

double-blind baseline using the MMRM and an ANCOVA model respectively described in Sections 13.3.1.1 and 13.3.1.2.

14.8 Extrapyramidal Symptom Rating Scale - Abbreviated

14.8.1 Analysis for the Open-label Safety Analysis Set

Extrapyramidal Symptom Rating Scale - Abbreviated (ESRS-A) scores are measured at Baseline and Week 12 during the open-label period.

Observed ESRS-A total scores and the four individual global CGI-S scores will be summarized using descriptive statistics at Baseline and Week 12. The change from Baseline values will also be summarized at Week 12.

14.8.2 Analysis for the Double-blind Safety Analysis Set

In addition to the open-label assessments at Baseline and Week 12, ESRS-A scores are measured at Weeks 13, 14, 16, 18, 22, 26, 30, 34, and 38 during the double-blind period. Summaries for the Double-blind Safety Analysis Set will include both the open-label and double-blind study periods.

ESRS-A total scores and the four individual global CGI-S scores, along with change from open-label baseline, will be summarized by visit for both open-label and double-blind visits.

ESRS-A total scores and the four individual global CGI-S scores, along with change from double-blind baseline, will be summarized by visit for the double-blind visits.

15 CLINICAL PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES

The PK endpoints for this study are as follows:

- Plasma concentration of pimavanserin and AC-279
- Pimavanserin PK parameters using a population PK approach

Pimavanserin and AC-279 plasma concentration data collected in the double-blind period will not be available to the Sponsor until the unblinding of the clinical database at the end of the study.

If data allow, population PK and PK/PD analyses may be performed to further characterize the PK profile and exposure response relationship of pimavanserin and AC-279 using

measures of safety, and efficacy parameters. The results of population PK and PK/PD modeling will be presented in a separate report.

Concentrations that are below the limit of quantitation (LLOQ) will be displayed as "LLOQ" in the data listings and imputed as 0 for computing summary statistics.

15.1 Analysis for the Open-label Pharmacokinetic Analysis Set

Pharmacokinetic samples are collected at Baseline, and Weeks 8 and 12 during the open-label period, and also at any early termination visit or the visit immediately following any SAE or an AE leading to discontinuation in the open-label period.

Plasma concentrations for pimavanserin and AC-279 will be summarized by visit using descriptive statistics for the Open-label Pharmacokinetic Analysis Set.

15.2 Analysis for the Double-blind Pharmacokinetic Analysis Set

In addition to the open-label assessments at Baseline, and Weeks 8 and 12 during the open-label period, pharmacokinetic samples are collected at double-blind Weeks 13, 22, and 38 of the double-blind period, and also at any early termination visit or the visit immediately following any SAE or an AE leading to discontinuation in the double-blind period. Summaries for the Double-blind Pharmacokinetic Analysis Set will include both the open-label and double-blind study periods.

Plasma concentrations for pimavanserin and AC-279 will be summarized by visit using descriptive statistics for both open-label and double-blind visits.

16 UNBLINDED INTERIM EFFICACY ANALYSIS

The database cutoff date for the interim analysis is defined as the date that the adjudication of the 38th relapse event is complete (complete is defined as when the data has been entered by the responsible reviewer, source data verified by the independent adjudication committee project manager, and entry restricted to further modification). Double-blind discontinuations that occur on or before the database cutoff date will be adjudicated and included in the interim analysis. The ITT Analysis Set for the interim analysis will include all subjects randomized on or before the database cutoff date.

The interim analysis will be based on the interim analysis dataset, created from the interim analysis data extract. Data occurring after the database cutoff date will not be included in the interim analysis dataset; however, data that occurs [e.g., from a visit] on or before the database cutoff date but are entered into the clinical database in the period between the database cutoff date and the date of the interim analysis data extract would be included in the interim analysis datasets. The final database will be locked after all data are included at study termination. If the study is stopped at the interim analysis, sensitivity analyses investigating the impact of overrun events (See Section 16.2.2) will be based on datasets created from the final database. If the study is not stopped at the interim analysis, the primary analysis of the primary and key secondary efficacy endpoints will be based on the final database using all data collected up until the study discontinuation date.

16.1 Primary Efficacy Endpoint

An interim analysis evaluating efficacy based on the primary efficacy endpoint will be conducted by an independent statistical group (ISG) when at least 38 adjudicated relapse events determined by the IAC have been accrued. The ISG is not affiliated with the Sponsor or the DSMB. The ISG statistician will disseminate the unblinded interim results at the closed session of the DSMB meeting. In order to protect the blinding and the integrity of this study, the actual computational output will be securely stored by the ISG and retained until the study is unblinded after the final database lock.

Table 9 displays the O'Brien-Fleming stopping boundary information for the group sequential design having one interim analysis if it is conducted with exactly 38 relapse events and a final analysis to be conducted with exactly 75 relapse events.

Table 9 O'Brien-Fleming Stopping Boundary Information

			Stopping Boundary: P-value Scale	
Stage	Relapse Events	Proportion	1-sided	2-sided
Interim	38	0.5067	0.0027	0.0054
Final	75	1.0000	0.0239	0.0479

In the event that not exactly 38 relapse events have accrued at the interim analysis, the O'Brien-Fleming stopping boundary will be recalculated based on the actual number of

relapse events accrued at the interim analysis. The decision to stop or continue the study will be made based on this recalculated boundary. More specifically, let Z_1 be the Wald test statistics evaluated at the time of interim analysis and let Z_2 be the Wald test statistics at the time of the final analysis. The test statistics are constructed such that a negative value favors the study drug. Let n be the number of relapse events observed at interim time and N be the total number of events and $t = \frac{n}{N}$ be the information fraction for the interim analysis. At the trial design stage, the plan is to take $n = 38, N = 75$ and $t = \frac{38}{75}$ which implies that the correlation between Z_1 and Z_2 is $\sqrt{t}$. Let c be a constant such that

$$P\left(Z_1 < \frac{c}{\sqrt{t}}\right) + P\left(Z_1 \geq \frac{c}{\sqrt{t}}, Z_2 < c\right) = 0.025.$$

The one-sided P-value scale boundary for the interim analysis is $p_1^{(1)} = \Phi\left(\frac{c}{\sqrt{t}}\right)$ and the one-sided P-value scale boundary for final analysis is $p_2^{(1)} = \Phi(c)$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function.

At the trial implementation stage, the observed number of relapse events might be different from the planned one. Denote the actual observed number of events at the interim by n^* and let $t^* = \frac{n^*}{75}$ be the observed information fraction at the time of interim analysis.

Let c^* be the constant satisfying the following equation

$$P\left(Z_1 < \frac{c^*}{\sqrt{t^*}}\right) + P\left(Z_1 \geq \frac{c^*}{\sqrt{t^*}}, Z_2 < c^*\right) = 0.025,$$

where the updated correlation between Z_1 and Z_2 is $\sqrt{t^*}$. Then the updated one-sided P-value scale boundaries for the interim and final analyses are given by $p_1^{(1)*} = \Phi\left(\frac{c^*}{\sqrt{t^*}}\right)$ and $p_2^{(1)*} = \Phi(c^*)$.

At the interim analysis, if the observed one-sided p-value from the Cox regression analysis is less than the corresponding lower stopping boundary p-value $p_1^{(1)*}$, then the null hypothesis will be rejected with the conclusion that the study demonstrates superiority of pimavanserin compared to placebo. Conversely, if the observed one-sided p-value is greater than or equal to the lower stopping boundary p-value, the null hypothesis will not be rejected and the study

recruitment will continue until 75 adjudicated relapse events have been accrued for the final analysis. At the final stage, if the observed p-value is less than the corresponding lower stopping boundary p-value $p_2^{(1)*}$, then the null hypothesis will be rejected with the conclusion that the study demonstrates superiority of pimavanserin compared to placebo. In the event that not exactly 75 relapse events have accrued at the final analysis, the O'Brien-Fleming stopping boundary will be recalculated based on the actual number of relapse events accrued at the interim and final analyses and the information fraction. This recalculated boundary will be used to determine superiority of pimavanserin compared to placebo. More specifically, let

$\alpha_1 = \Phi\left(\frac{c^*}{\sqrt{t^*}}\right)$ denote the actual type I error spent at the interim analysis time. Let $\tilde{N}$ be the observed number of relapse events at the final analysis and $\tilde{t} = \frac{n^*}{\tilde{N}}$. The one-sided P-value scale boundary for the final analysis is given by $\Phi(\tilde{c})$ where the $\tilde{c}$ satisfies the following equation

$$\alpha_1 + P\left(Z_1 \geq \frac{c^*}{\sqrt{t^*}},\ \ Z_2 < \tilde{c}\right) = 0.025,$$

and where the correlation between Z_1 and Z_2 is $\sqrt{\tilde{t}}$.

16.2 Key Secondary Efficacy Endpoint

Testing of the key secondary endpoint will be conducted at most once either at the interim analysis or the final analysis provided that the primary endpoint reaches statistical significance. The test statistics are constructed such that a negative value favors the study drug.

An O'Brien-Fleming stopping boundary will also be used for the key secondary endpoint. Let Y_1 and Y_2 be the Wald statistics for testing the key secondary efficacy endpoint at the interim and final analysis. The one-sided P-value scale boundary for the key secondary efficacy endpoint at the interim analysis is given by

$$p_1^{(2)*} = \alpha_1,$$

i.e. the same type I error spent at the interim analysis for the primary efficacy endpoint.

At the time of the final analysis, the one-sided P-value scale boundary is $p_2^{(2)*} = \Phi(d)$ where d satisfies the following equation

$$\alpha_1 + P\left(Y_1 \geq \frac{c^*}{\sqrt{t^*}}, Y_2 < d\right) = 0.025,$$

and where the correlation between Y_1 and Y_2 is given by $\sqrt{\frac{m}{M}}$ and m is the observed number of discontinuation events from the double-blind period at the time of interim analysis and M is the total number of discontinuation events at the final analysis.

16.2.1 Overrunning of Relapse Events at Interim Analysis

If the decision to stop the study for efficacy at the interim analysis is triggered, an overrun in termination events may occur if double-blind termination events occur after the database cutoff date for the interim analysis. For the primary efficacy endpoint defined by adjudicated relapse events, the overrun includes events that occur between the database cutoff date and the study discontinuation date. These events will be adjudicated by the Independent Adjudication Committee and will be included in the sensitivity analysis. Early terminations and relapse events that occur after the study discontinuation date will not be adjudicated and thus will not be included in this sensitivity analysis. Sensitivity analysis of overrun events described in Section 13.1.2 will be considered supportive.

16.2.2 Analysis Following a Sequential Test

To account for the fact that a sequential design is used, the stage-wise ordering of the sample space will be used to compute an adjusted estimate for the hazard ratio, 95% confidence interval, and p-value following rejection of the null hypothesis at the final analysis, if the study was not stopped early at the interim analysis (Tsiatis, et. al., 1984).

17 DATA AND SAFETY MONITORING BOARD

An independent Data and Safety Monitoring Board (DSMB) will review interim safety data including data on AEs and SAEs. The DSMB will be independent of the Sponsor and the Investigators and will be empowered to recommend stopping the study due to safety concerns. In addition, the DSMB will review the results of the pre-planned unblinded interim

analyses conducted by the ISG. The roles and responsibilities of DSMB members and ISG and planned frequency of meetings are detailed in the DSMB Charter.

The ISG will produce unblinded interim statistical outputs and provide these outputs to DSMB 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 DSMB members by the independent statistician will be consistent with what is described in the DSMB Charter.

18 INDEPENDENT ADJUDICATION COMMITTEE

An Independent Adjudication Committee (IAC) will perform validation of protocol-defined relapse events. The validation is based on review of predefined clinical data related to: 1) a relapse (identified by the Investigator) occurring before the study discontinuation date and 2) all early termination events which occur in the double-blind period before the study discontinuation date. Early termination and relapse events that occur in the double-blind period after the study discontinuation date will not be adjudicated.

The events will be reviewed by the IAC in a blinded manner.

Predefined clinical data consist of copies of source documents collected and delivered by the investigational sites. The IAC will work in accordance with written guidelines included in the IAC Charter describing in detail the composition, tasks, responsibilities and work processes of the committee. Only events determined to be relapse events by the IAC will be used in the primary efficacy analysis (Section 13.1).

19 COMPUTER METHODS

Statistical analyses will be performed using Version 9.4 (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 statistical analyses will follow appropriate standard operating procedures.

20 CHANGES TO ANALYSES SPECIFIED IN PROTOCOL

There are no planned deviations from the specified analyses in Protocol Amendment 1, dated 16 August, 2018.

21 REFERENCES

1. Lin, DY, Wei LJ. The robust inference for the proportional hazards model. Journal of the American Statistical Association 1989;84:1074-1078.

2. Rubin DB. Inference and missing data. Biometrika 1976;63:581:592.

3. Tsiatis AA, Rosner GL, Mehta CR. Exact confidence interval following a group sequential test. Biometrics 1984;40:797-803.

4. ICH Guidance for Industry: E14 Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs, 2005.

5. Lu K, et.al. Comparison Between Two Controlled Multiple Imputation Methods for Sensitivity Analyses of Time-to-Event Data With Possibly Information Censoring. Statistics in Biopharmaceutical Research. 2015; 7, 199-213.

22 APPENDICES

Appendix 1 Summary of Version Changes

Version No:	Document History Description of Update	Author(s)	Version Date