Official Title: A Phase 3, randomized, double-blind, placebo-controlled study to evaluate the efficacy and safety of adjunctive pimavanserin for the treatment of schizophrenia

NCT Number: NCT02970292

STATISTICAL ANALYSIS PLAN

Protocol No.:	ACP-103-034
Protocol Title:	A Phase 3, Randomized, Double-Blind, Placebo-Controlled Study to Evaluate the Efficacy and Safety of Adjunctive Pimavanserin for the Treatment of Schizophrenia
Drug:	Pimavanserin
Sponsor:	ACADIA Pharmaceuticals Inc. [redacted]
Version No. and Date	Version 1.0, 01 June 2018

SIGNATURE/APPROVAL PAGE

AUTHOR

[redacted]

01 JUN 2018

Date

Senior Director, Biostatistics, Biometrics
ACADIA Pharmaceuticals Inc.

APPROVERS

[redacted]

04 JUN 2018

Date

Senior Vice President, Clinical Development and Chief Medical Officer
ACADIA Pharmaceuticals Inc.

[redacted]

04 JUN 2018

Date

Vice President, Clinical Development
ACADIA Pharmaceuticals Inc.

[redacted]

01 JUNE 2018

Date

Vice President, Biometrics
ACADIA Pharmaceuticals Inc.

[redacted]

01 JUN 2018

Date

Executive Director, Biostatistics and SAS Programming
ACADIA Pharmaceuticals Inc.

TABLE OF CONTENTS

TABLE OF CONTENTS 3
LIST OF TABLES 6
LIST OF FIGURES 6
ABBREVIATIONS 7
1 INTRODUCTION 9
2 OBJECTIVES 9
2.1 Primary Objective 9
2.2 Secondary Objective 9
3 STUDY DESIGN 9
3.1 General Study Design 9
3.2 Schedule of Assessments 12
3.3 Randomization 15
3.4 Blinding 15
3.5 Determination of Sample Size 15
4 ANALYSIS SETS 15
5 DATA HANDLING CONVENTIONS 16
5.1 General Data Reporting Conventions 16
5.2 Derived Variables 17
5.2.1 Positive and Negative Syndrome Scale (PANSS) 17
5.2.2 Clinical Global Impressions – Severity (CGI-S) Scale 19
5.2.3 Clinical Global Impressions – Improvement (CGI-I) Scales 20
5.2.4 Personal and Social Performance Scale (PSP) 20
5.2.5 Drug Attitude Inventory (DAI-10) 20
5.2.6 Karolinska Sleepiness Scale (KSS) 21
5.2.7 Calgary Depression Scale for Schizophrenia (CDSS) 21
5.2.8 36-Item Short Form Health Survey (SF-36) 21
5.2.9 Columbia-Suicide Severity Rating Scale (C-SSRS) 24
5.2.10 Abnormal Involuntary Movement Scale (AIMS) 25
5.2.11 Barnes Akathisia Scale (BARS) 26
5.2.12 Simpson Angus Extrapyramidal Side Effect Scale (SAS) 26
5.3 Analysis Visit Windows 26
5.3.1 Unscheduled Assessments 27

5.3.2 Multiple Measurements within Visit Windows 27
5.4 Missing or Incomplete Date for Last Dose of Study Drug 27
5.5 Missing or Incomplete Dates for Prior or Concomitant Medications 28
5.6 Missing or Incomplete Dates for Adverse Events 28
5.7 Missing Severity Assessment for Adverse Events 28
5.8 Missing Relationship to Study Drug for Adverse Events 28
5.9 Character Values of Clinical Laboratory Variables 28
5.10 Duplicate Subjects 29
6 SUBJECT DISPOSITION 29
7 PROTOCOL DEVIATIONS 30
8 DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS 30
9 MEDICAL HISTORY 31
10 EXTENT OF EXPOSURE AND TREATMENT COMPLIANCE 31
10.1 Exposure to Study drug 31
10.2 Measurement of Treatment Compliance 32
11 PRIOR, CONCOMITANT AND POST-TREATMENT MEDICATION 32
12 EFFICACY ANALYSES 32
12.1 Efficacy Variables 33
12.2 Adjustment for Covariates 33
12.3 Handling of Missing Data 34
12.4 Multiple Comparisons / Multiplicity 34
12.5 Examination of Subgroups 34
13 METHODS OF EFFICACY ANALYSES 35
13.1 Primary Efficacy Analysis 35
13.1.1 Primary Analysis using MMRM in the Full Analysis Set 35
13.1.2 Sensitivity Analysis using MMRM in the Per-Protocol Analysis Set 37
13.1.3 Sensitivity Analysis using ANCOVA with Missing Data Imputed Using Pattern-Mixture Models with Placebo-Based Multiple Imputation 37
13.1.4 Sensitivity Analysis using Van Elteren Test in the Full Analysis Set 39
13.1.5 Complete Case Analysis using ANCOVA in the Full Analysis Set 40
13.2 Key Secondary Efficacy Analysis 40
13.3 Other Secondary Efficacy Analyses 40
13.3.1 Responder Analysis 40

13.3.2 Other Secondary Efficacy Analysis for Continuous Variables 42
13.4 Exploratory Efficacy Analyses 43
14 SAFETY ANALYSES 43
14.1 Adverse Events 43
14.2 Clinical Laboratory Variables 44
14.3 Vital Signs 48
14.4 Electrocardiogram (ECG) 49
14.5 Physical Examination 50
14.6 Other Safety Variables 51
14.6.1 Suicidality 51
14.6.2 Extrapyramidal Symptom Measures 51
15 CLINICAL PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES 52
16 UNBLINDED INTERIM ANALYSIS 53
17 DATA MONITORING/REVIEW COMMITTEE 54
18 COMPUTER METHODS 54
19 CHANGES TO ANALYSES SPECIFIED IN PROTOCOL 55
20 REFERENCES 56
21 APPENDICES 57
21.1 Summary of Version Changes 57

LIST OF TABLES

Table 1 Schedule of Events and Assessments....13
Table 2 SF-36 Re-Coding Rules for Individual Items....22
Table 3 SF-36 Scale Components....23
Table 4 Analysis Visit Windows....27
Table 5 Criteria for Potentially Clinically Important Laboratory Values – Hematology and Chemistry....47
Table 6 Criteria for Potentially Clinically Important Laboratory Values - Urinalysis....48
Table 7 Criteria for Potentially Clinically Important Vital Signs....49
Table 8 Criteria for Potentially Clinically Important ECG Values....50

LIST OF FIGURES

Figure 1 Schematic of Study Design....11

ABBREVIATIONS

AE	adverse event
AIMS	Abnormal Involuntary Movement Scale
ANCOVA	analysis of covariance
ATC	Anatomical Therapeutic Chemical
BARS	Barnes Akathisia Rating Scale
BLQ	below the limit of quantification
BMI	body mass index
CDSS	Calgary Depression Scale for Schizophrenia
CGI-I	Clinical Global Impression – Improvement scale
CGI-S	Clinical Global Impression – Severity scale
CMH	Cochran-Mantel-Haenszel
C-SSRS	Columbia-Suicide Severity Rating Scale
DAI-10	10-item Drug Attitude Inventory
DSMB	Data and Safety Monitoring Board
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
ECG	electrocardiogram
eCRF	electronic case report form
EOS	End-of-Study
ET	Early Termination
IRT	interactive response technology
KSS	Karolinska Sleepiness Scale
LOCF	last observation carried forward
MAR	missing at random
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	mixed model for repeated measures
MNAR	missing not at random
msec	milliseconds
OC	observed cases
PANSS	Positive and Negative Syndrome Scale
PCI	potentially clinically important
PD	pharmacodynamic
PK	pharmacokinetic
PMM	pattern mixture model
PSP	Personal and Social Performance scale
QTcB	QT interval corrected for heart rate using Bazett's formula

QTcF	QT interval corrected for heart rate using Fridericia's formula
SAE	serious adverse event
SAP	statistical analysis plan
SAS	Simpson-Angus Extrapyramidal Side Effects Scale
SAS®	Statistical Analysis System
SD	standard deviation
SE	standard error
SF-36	36-item Short Form Health Survey
SOC	system organ class
TEAE	treatment-emergent adverse event

1 INTRODUCTION

This statistical analysis plan (SAP) provides a technical and detailed elaboration of the statistical analyses of efficacy and safety data as described in the final study protocol Amendment 2 dated 30MAR2017. Specifications for tables, figures, and listings are contained in a separate document. Statistical analyses for population pharmacokinetic (PK) and PK/pharmacodynamics (PD) modeling will be presented in a separate report and therefore will not be included in this SAP.

For the Czech Republic, a country-specific protocol amendment (Amendment 2-CZ finalized 31 January 2018) specifies additional electrocardiogram (ECG) procedures to be conducted only at clinical sites within the Czech Republic. For subjects enrolled at these sites, ECGs will be measured at all visits instead of the schedule given in protocol Amendment 2.

2 OBJECTIVES

2.1 Primary Objective

The primary objective of the study is to evaluate the efficacy of adjunctive pimavanserin compared with adjunctive placebo in the treatment of schizophrenia.

2.2 Secondary Objective

The secondary objectives of the study are:

- To evaluate the safety and tolerability of adjunctive pimavanserin compared to adjunctive placebo in the treatment of schizophrenia
- To characterize the pharmacokinetics (PK) and pharmacodynamics (PD) of adjunctive pimavanserin for the treatment of schizophrenia

3 STUDY DESIGN

3.1 General Study Design

This study will be conducted as a Phase 3, 6-week, randomized, double-blind, placebo-controlled, multicenter, multinational outpatient study in subjects with schizophrenia with an inadequate response to current antipsychotic treatment. Schizophrenia is defined by the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), confirmed by a customized module of the Structured Clinical Interview for DSM-5, Clinical Trials Version (SCID-5-CT).

This study will enroll approximately 380 subjects (190 subjects per treatment group) with schizophrenia and an inadequate response to antipsychotic medications across approximately

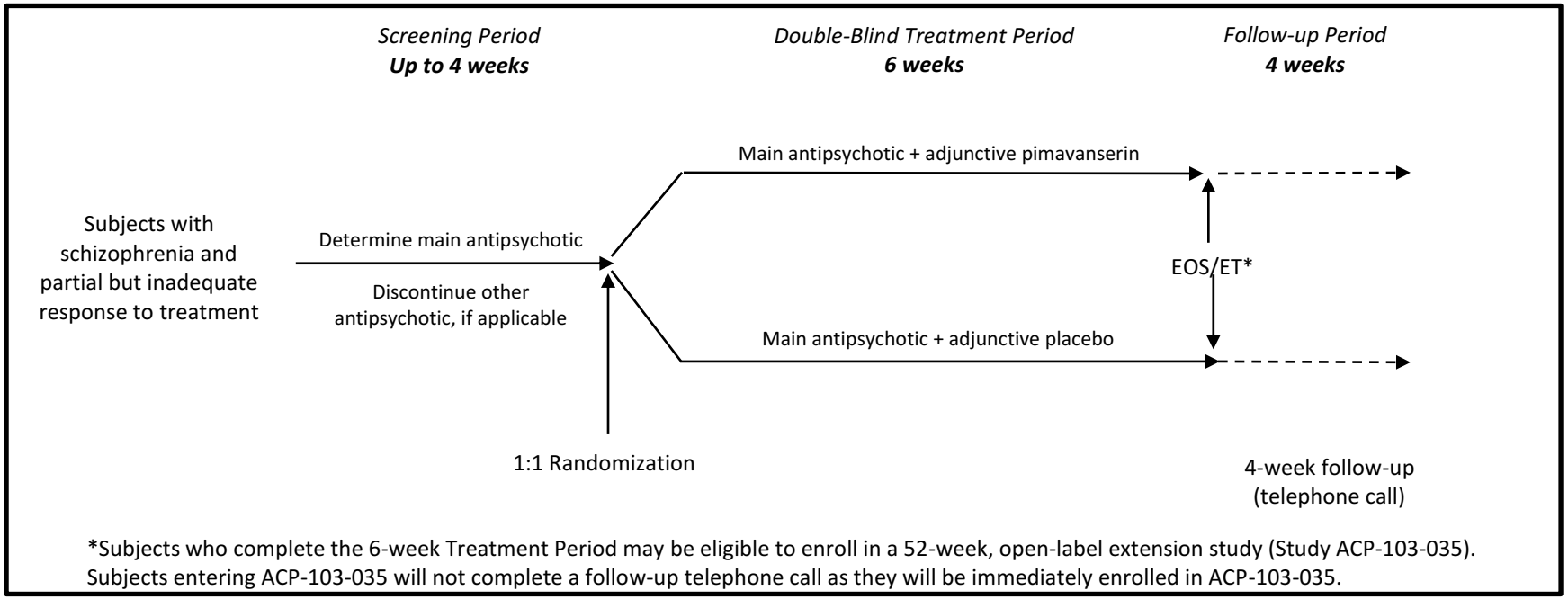
70 study sites globally. On the first day of the randomized treatment phase (Baseline), eligible subjects will be randomly assigned to receive adjunctive pimavanserin 20 mg (free base) per day or placebo daily in a 1:1 ratio, according to a computer-generated randomization schedule. The randomization will be stratified according to geographic region (North America, Europe, or rest of world). Subjects will participate in the study for up to 14 weeks, including a Screening Period of up to 4 weeks, a 6-week Treatment Period, and a 4-week safety follow-up (telephone call) for those subjects who discontinue prematurely from the study or who do not enroll in the 52-week, open-label, extension study (Study ACP-103-035).

Study drug will remain at 20 mg/day dose level for the first week of treatment. Study drug will be administered under double-blind conditions throughout the Treatment Period. The designated main antipsychotic medication will be continued at a stable dose.

At the Weeks 1, 2, and 3 visits, the daily dose of adjunctive pimavanserin may remain at 20 mg or it may be either increased to 34 mg (for symptom improvement) or decreased to 10 mg daily (if the 20 mg dose is not well tolerated). After the Week 3 visit, no study drug dose changes may be made. Clinic visits occurring after Baseline will be conducted at Weeks 1, 2, 3, 4, 5, and 6 (End-of-Study [EOS]/Early Termination [ET] visit).

Figure 1 illustrates the study design.

Figure 1 Schematic of Study Design



3.2 Schedule of Assessments

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

Table 1 Schedule of Events and Assessments

Period	**Screening Period**		**Treatment Period**						**Follow-Up[a]**
Visit[b]	**1**	**2 (Baseline)**	**3**	**4**	**5**	**6**	**7**	**8 (EOS/ET)**	**Telephone call**
Day or Week	**(Days -28 to -7)**	**Day 1/Week 0**	**Week 1**	**Week 2**	**Week 3**	**Week 4**	**Week 5**	**Week 6**	**Week 10**
Allowable visit window (# days)			**±3**	**±3**	**±3**	**±3**	**±3**	**±3**	**+7**
Informed consent[c]	X								
Inclusion/exclusion criteria	X	X							
Medical history[d] and demographics	X								
Schizophrenia disease history	X								
SCID-5-CT customized module	X								
Interview by an independent clinician[e]	X								
Physical examination[f]	X	X						X	
Vital signs	X	X	X	X	X	X	X	X	
Height and weight[g]	X	X						X	
12-lead ECG[h]	X	X				X		X	
Clinical laboratory tests[i]	X	X						X	
Confirmation of main antipsychotic[j]	X								
Pregnancy test[k]	X	X						X	
PK blood draws[l]		X	X		X			X	
Pharmacogenomic blood draws[m]		X							
PANSS and IQ-PANSS	X	X	X	X	X	X	X	X	
CGI-S	X	X	X	X	X	X	X	X	
CGI-I			X	X	X	X	X	X	
PSP		X						X	
CDSS	X	X						X	
SF-36		X						X	
DAI-10		X						X	
KSS		X	X	X	X	X	X	X	
C-SSRS[n]	X	X	X	X	X	X	X	X	
AIMS		X	X		X			X	
BARS		X	X		X			X	
SAS		X	X		X			X	
Assessment of concomitant medications	X	X	X	X	X	X	X	X	X
Assessment of adverse events	X	X	X	X	X	X	X	X	X
Dispense study drug		X	X[o]	X[o]	X[o]	X[o]	X[o]		
Study drug accountability			X	X	X	X	X	X	

Footnotes for Table 1 on next page:

Abbreviations: AIMS = Abnormal Involuntary Movement Scale; BARS = Barnes Akathisia Rating Scale; CDSS = Calgary Depression Scale for Schizophrenia; CGI-I = Clinical Global Impression-Improvement; CGI-S = Clinical Global Impression-Severity; C-SSRS = Columbia Suicide Severity Rating Scale; DAI-10 = 10-item Drug Attitude Inventory; ECG = electrocardiogram; EOS = End-of-Study; ET = Early Termination; HCV = hepatitis C virus; HIV = human immunodeficiency virus; IQ-PANSS = Informant Questionnaire for the Positive and Negative Syndrome Scale; KSS = Karolinska Sleepiness Scale; PANSS = Positive and Negative Syndrome Scale; PK = pharmacokinetic; PSP = Personal and Social Performance scale; SAS = Simpson-Angus Extrapyramidal Side Effects Scale; SCID-5-CT = Structured Clinical Interview for DSM-5, Clinical Trials Version; SF-36 = 36-item Short Form Health Survey.

[a] For subjects who discontinue prematurely from the study or who do not enroll in the 52-week,open-label, extension study (ACP-103-035), a safety follow-up telephone call will occur approximately 4 weeks after the last dose of study drug.

[b] Study visits are designated by weeks and have a window, calculated from the Baseline visit, of ±3 days. The window for the Follow-up visit is +7 days.

[c] To participate in the optional pharmacogenomic component of the study, subjects must sign a separate pharmacogenomic informed consent form, indicating their willingness to participate. The pharmacogenomic ICF may be signed Baseline or at any time during the study. The subject's caregiver must provide written agreement prior to any Screening procedures being performed indicating their agreement to participate in the study in the caregiver role.

[d] Medical history is to include a history of tobacco and nicotine use. A review of any history of HIV, hepatitis B, or HCV will also be performed.

[e] A structured telemedicine interview of the subject by an independent clinician, is to be performed during the Screening period. The interview will be conducted by video and will not be recorded.

[f] Physical examinations may include a review of body systems (pelvic and/or urogenital examination may be deferred, unless the Investigator deems this to be clinically indicated).

[g] Height will only be measured at the Screening visit.

[h] A single 12-lead ECG will be performed in triplicate at Screening. Single 12-lead ECG recordings will be performed at Baseline, Week 4, and at the Week 6 (EOS/ET) visit. A single 12-lead ECG can be performed any time before blood sampling or at least 30 minutes after blood sampling during clinic visits.

[i] To include hematology, serum chemistry, prolactin levels, urinalysis, and urine drug screen (note: additional laboratory studies [in addition to scheduled timepoints shown in the table] for a given subject may be repeated at any time throughout the Treatment Period, at the discretion of the Investigator). It is preferable but not required that subjects be in a fasted state (e.g., fasting for approx. 10 hours) before the blood sample for clinical chemistry is obtained.

[j] Blood samples for measurement of the following will be obtained at Screening only: glycosylated hemoglobin (HbA1c), thyroid stimulating hormone (TSH), and main antipsychotic detection. Measurement of a full thyroid panel will be conducted only if the TSH value is outside of the laboratory reference range.

[k] A serum pregnancy test will be completed at the Screening visit for all female subjects; urine pregnancy tests will be completed at the Baseline and Week 6 visits for all female subjects.

[l] At the Screening visit, a PK sample will be collected for the presence or absence of the subject's main antipsychotic. At each subsequent timepoint, a PK sample will be collected for pimavanserin, the metabolite AC-279, and the main antipsychotic. The Baseline PK sample should be collected pre-dose. When possible, an additional PK sample will be collected from subjects who experience an SAE or an AE leading to discontinuation, as soon as possible after the occurrence of that event. For all PK samples (scheduled and unscheduled), the dates and times of administration of the last 3 doses of both study drug and the main antipsychotic should be recorded. For samples collected from subjects who experience an SAE or an AE leading to discontinuation, the date and time of the last dose prior to the SAE or AE should also be recorded.

[m] A blood sample will be collected from subjects who give informed consent for the pharmacogenomic component of the study. The blood draw should be completed pre-dose. A sample collected at a later timepoint does not constitute a protocol violation and would not require protocol amendment.

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

[o] Subjects are to return unused study drug and all kit materials at each subsequent visit; a new kit will be dispensed at each identified visit.

3.3 Randomization

Eligible subjects will be randomized into 1 of 2 treatment groups (adjunctive pimavanserin or adjunctive placebo) in a 1:1 ratio using an interactive response technology (IRT) system. The randomization will be stratified by geographic region (North America, Europe, or rest of world). The assignments will be based on a pre-generated permuted-block randomization schedule.

3.4 Blinding

This is a double-blind study. 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 adjunctive pimavanserin and adjunctive placebo treatments. Neither the subjects nor the study personnel at the clinical sites will know which treatment is administered to each subject.

3.5 Determination of Sample Size

The primary efficacy endpoint is the change from Baseline to Week 6 in the Positive and Negative Syndrome Scale (PANSS) total score.

Let Δ be the difference in the mean change from Baseline to Week 6 in the PANSS total score between the adjunctive pimavanserin and the adjunctive placebo groups:

The null hypothesis for the primary efficacy endpoint is: $\Delta = 0$

The alternative hypothesis for the primary efficacy endpoint is: $\Delta \neq 0$

Assuming the true difference in the mean change in the PANSS total score from Baseline to Week 6 is 6.0 points between the adjunctive pimavanserin group and the adjunctive placebo group, and the common standard deviation is 17.0 points, 171 evaluable subjects per treatment group will provide at least 90% power to detect a difference between the adjunctive pimavanserin group and the adjunctive placebo group at a significance level of 0.05, using a 2-sided t-test.

Adjusting for a potential non-evaluable rate of up to 10%, approximately 380 subjects (190 subjects per treatment group) will be randomized.

4 ANALYSIS SETS

<u>Randomized Analysis Set</u>

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

Note that if there are duplicate subjects (see Section 5.10 for details), only the identity under the first enrollment will be included in the Randomized Analysis Set.

Subjects will be classified according to the randomized treatment assignment.

Safety Analysis Set

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

Subjects will be classified according to the actual treatment received.

Full Analysis Set

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

Subjects will be classified according to the randomized treatment assignment.

Per-protocol Analysis Set

The Per-protocol Analysis Set will consist of a subset of subjects in the Full Analysis Set who are at least 80% compliant (based on drug accountability) and without any protocol deviation which is considered to have substantial impact on primary efficacy outcome. The precise reasons for excluding subjects from the Per-protocol Analysis Set will be fully defined and documented *a priori* before the clinical database lock.

Subjects will be classified according to the randomized treatment assignment.

Pharmacokinetics Analysis Set

For pimavanserin and AC-279 plasma concentration summaries, the Pharmacokinetics Analysis Set will consist of subjects with at least one measurable pimvanserin plasma concentration.

Subjects will be classified according to the actual treatment received.

5 DATA HANDLING CONVENTIONS

All data collected in the study will be listed.

5.1 General Data Reporting Conventions

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

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

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

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.

5.2 Derived Variables

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

5.2.1 Positive and Negative Syndrome Scale (PANSS)

The PANSS is assessed at Screening, Baseline, Weeks 1, 2, 3, 4, 5, and Week 6/ET visits.

The PANSS is a 30-item scale used to evaluate the presence, absence, and severity of schizophrenia symptoms. Items are scored over the past week (7 days) on the following 7-point scale: 1 = absent; 2 = minimal; 3 = mild; 4 = moderate; 5 = moderate-severe; 6 = severe; 7 = extreme. The PANSS total score can range from a minimum of 30 to a maximum of 210, with higher scores denoting more severe symptoms.

The following are the 3 PANSS subscales and the PANSS items that define each subscale:

- Positive Scale: 7 Items (minimum = 7, maximum = 49)
  - P1 Delusions
  - P2 Conceptual disorganization
  - P3 Hallucinatory behavior
  - P4 Excitement
  - P5 Grandiosity
  - P6 Suspiciousness/persecution
  - P7 Hostility

- Negative Scale: 7 Items (minimum = 7, maximum = 49)
  - N1 Blunted affect
  - N2 Emotional withdrawal
  - N3 Poor rapport
  - N4 Passive/apathetic social withdrawal
  - N5 Difficulty in abstract thinking
  - N6 Lack of spontaneity and flow of conversation
  - N7 Stereotyped thinking
- General Psychopathology Scale: 16 Items (minimum = 16, maximum = 112)
  - G1 Somatic concern
  - G2 Anxiety
  - G3 Guilt feelings
  - G4 Tension
  - G5 Mannerisms and posturing
  - G6 Depression
  - G7 Motor retardation
  - G8 Uncooperativeness
  - G9 Unusual thought content
  - G10 Disorientation
  - G11 Poor attention
  - G12 Lack of judgment and insight
  - G13 Disturbance of volition
  - G14 Poor impulse control
  - G15 Preoccupation
  - G16 Active social avoidance

The following are the 5 Marder factors and the PANSS items that define each factor:

- Negative Symptoms: 7 Items (minimum = 7, maximum = 49)
  - N1 Blunted affect
  - N2 Emotional withdrawal
  - N3 Poor rapport
  - N4 Passive/apathetic social withdrawal
  - N6 Lack of spontaneity and flow of conversation
  - G7 Motor retardation
  - G16 Active social avoidance
- Positive Symptoms: 8 Items (minimum = 8, maximum = 56)
  - P1 Delusions
  - P3 Hallucinatory behavior
  - P5 Grandiosity
  - P6 Suspiciousness/persecution
  - N7 Stereotyped thinking
  - G1 Somatic concern
  - G9 Unusual thought content

  - G12 Lack of judgment and insight

- Disorganized Thought: 7 Items (minimum = 7, maximum = 49)
  - P2 Conceptual disorganization
  - N5 Difficulty in abstract thinking
  - G5 Mannerisms and posturing
  - G10 Disorientation
  - G11 Poor attention
  - G13 Disturbance of volition
  - G15 Preoccupation

- Uncontrolled Hostility/Excitement: 4 Items (minimum = 4, maximum = 28)
  - P4 Excitement
  - P7 Hostility
  - G8 Uncooperativeness
  - G14 Poor impulse control

- Anxiety/Depression: 4 Items (minimum = 4, maximum = 28)
  - G2 Anxiety
  - G3 Guilt feelings
  - G4 Tension
  - G6 Depression

For each of the subscales (positive, negative, or general), if more than 1 item score is missing, then that particular subscale score and the PANSS total score will be missing. When there is only 1 item missing for a subscale, then the missing single item will be imputed using the average of the non-missing item scores for that subscale, subject and timepoint, rounded to the nearest integer. After the missing item(s) is imputed within the subscale(s), then the PANSS total score and Marder factor scores will be computed without further imputation. A Marder factor score will be missing if any item score is missing for that factor, after the imputations within the subscales were applied.

5.2.2 Clinical Global Impressions – Severity (CGI-S) Scale

The CGI-S is assessed at Screening, Baseline, Weeks 1, 2, 3, 4, 5, and Week 6/ET visits.

The CGI-S is a clinician-rated, 7-point scale that is designed to rate the severity of the subject's illness, in this case schizophrenia, at the time of assessment, making use of the clinician's judgment and past experience with subjects who have the same disorder. The 7-point scores are: 1 = normal, not ill; 2 = minimally ill; 3 = mildly ill; 4 = moderately ill; 5 = markedly ill; 6 = severely ill; 7 = among the most severely ill. Higher scores denote more severe schizophrenia. The licensed version used in Study 034 is a blend of Guy's (Guy W; Early Clinical Drug Evaluation Unit Assessment Manual for Psychopharmacology; US Department of Health, Education, and Welfare; National Institute of Mental Health; 1976) and Haro's (Haro JM; The Clinical Global Impression-schizophrenia scale: a simple instrument to measure the diversity of symptoms present in schizophrenia; 2003) versions.

Missing CGI-S scores will not be imputed.

5.2.3 Clinical Global Impressions – Improvement (CGI-I) Scales

The CGI-I is assessed at Weeks 1, 2, 3, 4, 5, and 6/ET visits.

The CGI-I is a clinician-rated, 7-point scale that is designed to assess how much the subject's illness has improved or worsened relative to study Baseline. The 7-point scores are: 1 = very much improved; 2 = much improved; 3 = minimally improved; 4 = no change; 5 = minimally worse; 6 = much worse; 7 = very much worse. Higher scores denote worse or less improvement in schizophrenia.

Missing CGI-I scores will not be imputed.

5.2.4 Personal and Social Performance Scale (PSP)

The PSP is assessed at Baseline and Week 6/ET visits.

The PSP is a validated 100-point (1 to100) single-item rating scale to assess the psychosocial functioning of subjects with schizophrenia. Ratings are based on four main areas: (a) socially useful activities, including work and study; (b) personal and social relationships, (c) self-care; and (d) disturbing and aggressive behaviors. The time period assessed is "past month" and does not take into account suicidal risk and behavior. Higher scores denote better psychosocial functioning: scores of 71-100 reflect only mild difficulties; 31-70 reflect manifest disabilities of various degrees; 1-30 reflect functioning so poor that intensive support or supervision is needed.

Missing PSP scores will not be imputed.

5.2.5 Drug Attitude Inventory (DAI-10)

The DAI-10 is assessed at Baseline and Week 6/ET visits.

The DAI-10 contains 6 items (1, 3, 4, 7, 9, and 10) that a subject who is fully adherent to the prescribed medication would answer as "True" and 4 items (2, 5, 6, and 8) that a subject who is fully adherent to the prescribed medication would answer as "False." A correct answer is scored +1 and an incorrect answer is scored -1. The total score is the sum of pluses and minuses, which can range from -10 to 10 in increments of 2. A positive total score indicates a positive subjective response (adherent) and a negative total score indicates a negative subjective response (non-adherent). Higher scores denote better adherence.

Missing DAI-10 item scores will not be imputed. The total score will be missing if any item score is missing.

5.2.6 Karolinska Sleepiness Scale (KSS)

The KSS is assessed at Baseline, Weeks 1, 2, 3, 4, 5, and Week 6/ET visits.

The KSS is a self-reported subjective measure of a subject's level of drowsiness. Respondents must choose statements that most accurately describe their level of sleepiness over the past few minutes or, with the modified version used in this study, over an average period of time. In Study 034, sleepiness level during the last week (7 days) is reported. Scoring is based on a 9-point verbally anchored scale ranging from 1 (extremely alert) to 9 (very sleepy, great effort to keep awake, fighting sleep). Higher scores denote more drowsiness. In Study 034, the KSS will be administered by a trained interviewer.

Missing KSS scores will not be imputed.

5.2.7 Calgary Depression Scale for Schizophrenia (CDSS)

The CDSS is assessed at Screening, Baseline, and Week 6/ET visits.

The CDSS is a 9-item scale that was developed specifically to assess the level of depression in schizophrenia. It was originally developed to differentiate depressive symptoms from negative symptoms. Items are scored over the past 2 weeks (14 days) on the following 4-point scale: 0 = absent; 1 = mild; 2 = moderate; 3 = severe. The total score is the sum of the 9 item scores which can range from 0 to 27, with higher scores denoting more severe depression.

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

5.2.8 36-Item Short Form Health Survey (SF-36)

The SF-36 is assessed at Baseline and Week 6/ET visits.

The SF-36 is a 36-item survey that measures the overall health status of a subject. The SF-36 assesses eight health concepts during the past 4 weeks: physical functioning, bodily pain, role limitations due to physical health problems, role limitations due to personal or emotional problems, emotional well-being, social functioning, energy/fatigue, and general health perceptions. It also includes a single item that provides an indication of perceived change in health compared to one year ago. In Study 034, the RAND 36-Item Health Survey (Version 1.0) is used.

Scoring the SF-36 is a 2-step process. First, pre-coded numeric values are re-coded per the scoring key given in Table 2. Note that all items are scored so that a high score defines a more favorable health state and less disability. In addition, each item is scored on a 0 to 100 range so that the lowest and highest possible scores are 0 and 100, respectively. Scores

represent the percentage of total possible score achieved, with higher scores denoting higher level of functioning. In step 2, items in the same scale are averaged together to create the 8 scale scores. Table 3 lists the items averaged together to create each scale. Items that are left blank (missing data) are not taken into account when calculating the scale scores. Hence, scale scores represent the average for all items in the scale that the respondent answered.

Table 2 SF-36 Re-Coding Rules for Individual Items

Item Number	Original Response Category[1]	Re-Coded Value
2, 20, 34, 36	1	100
	2	75
	3	50
	4	25
	5	0
3, 4, 5, 6, 7, 8, 9, 10, 11, 12	1	0
	2	50
	3	100
13, 14, 15, 16, 17, 18, 19	1	0
	2	100
23, 26, 27, 30	1	100
	2	80
	3	60
	4	40
	5	20
	6	0
24, 25, 28, 29, 31	1	0
	2	20
	3	40
	4	60
	5	80
	6	100
32, 33, 35	1	0
	2	25
	3	50
	4	75
	5	100
1	1	100
	2	85
	3	60
	4	25
	5	0
21	1	100
	2	88

Table 2 SF-36 Re-Coding Rules for Individual Items (Continued)

Item Number	Original Response Category[1]	Re-Coded Value
	3	64
	4	42
	5	24
	6	0
22	1 (Item 21 = 1 or .)	100
	1 (Item 21 = 2 to 6)	80
	2 (Item 21 = .)	75
	2 (Item 21 ^= .)	60
	3 (Item 21 = .)	50
	3 (Item 21 ^= .)	40
	4 (Item 21 = .)	25
	4 (Item 21 ^= .)	20
	5	0

[1] Pre-coded response choices as printed in the questionnaire.

Table 3 SF-36 Scale Components

Scale	Items
PHYFUN10: Physical functioning	3, 4, 5, 6, 7, 8, 9, 10, 11, 12
ROLEP4: Role limitations due to physical health	13, 14, 15, 16
ROLEE3: Role limitations due to emotional problems	17, 18, 19
ENFAT4: Energy/fatigue	23, 27, 29, 31
EMOT5: Emotional well-being	24, 25, 26, 28, 30
SOCFUN2: Social functioning	20, 32
SFPAIN2: Pain	21, 22
SFGENH5: General health	1, 33, 34, 35, 36

To enable meaningful comparisons across scales, the 8 scale scores are converted to norm-based z-scores based on 1998 General Population statistics:

PF_Z = (PHYFUN10 - 82.96845) / 23.83795 ;
RP_Z = (ROLEP4 - 77.93107) / 35.34865 ;
RE_Z = (ROLEE3 - 83.10276) / 31.64149 ;
EN_Z = (ENFAT4 - 56.99917) / 21.12677 ;
EM_Z = (EMOT5 - 75.21913) / 17.60698 ;
SF_Z = (SOCFUN2 - 83.56494) / 23.02758 ;
BP_Z = (SFPAIN2 - 70.22865) / 23.35310 ;
GH_Z = (SFGENH5 - 70.10060) / 21.35900 ;

In addition, physical and mental health composite scores will also be computed using the individual-scale z-scores:

Physical health composite scores AGPHYS_Z
= (PF_Z * 0.42402) + (RP_Z * 0.35119) + (BP_Z * 0.31754) +
(GH_Z * 0.24954) + (EM_Z * -.22069) + (RE_Z * -.19206) +
(SF_Z * -.00753) + (EN_Z * 0.02877);

Mental health composite scores AGMENT_Z
= (PF_Z * -.22999) + (RP_Z * -.12329) + (BP_Z * -.09731) +
(GH_Z * -.01571) + (EM_Z * 0.48581) + (RE_Z * 0.43407) +
(SF_Z * 0.26876) + (EN_Z * 0.23534);

The 2 composite scores will not be calculated if any individual scale score is missing.

Finally, the z-scores will be transformed to t-scores for data analyses which will be detailed in Section 13.4.

```
AGPHYS_T = 50 + (AGPHYS_Z * 10);
AGMENT_T = 50 + (AGMENT_Z * 10);
PF_T = 50 + (PF_Z * 10) ;
RP_T = 50 + (RP_Z * 10) ;
RE_T = 50 + (RE_Z * 10) ;
EN_T = 50 + (EN_Z * 10) ;
EM_T = 50 + (EM_Z * 10) ;
SF_T = 50 + (SF_Z * 10) ;
BP_T = 50 + (BP_Z * 10) ;
GH_T = 50 + (GH_Z * 10) ;
```

5.2.9 Columbia-Suicide Severity Rating Scale (C-SSRS)

The C-SSRS is assessed at Screening, Baseline, Weeks 1, 2, 3, 4, 5, and Week 6/ET visits.

The C-SSRS monitors changes in suicidal thinking and behavior over time, in order to determine risk. Four constructs are measured: the severity of ideation, the intensity of ideation, behavior, and lethality.

The C-SSRS Baseline/Screening version will be completed at the Screening visit and the version assessing information since the last visit will be completed at all following visits (including the Baseline visit). The C-SSRS results for each subject should be reviewed by the Investigator at each visit. If at any time the C-SSRS results for a given subject reveal potential suicidality, then the Investigator should assess the clinical significance of such results. If a clinically significant risk of suicidality is identified for a subject, then the Investigator should discontinue the subject and implement appropriate treatment.

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

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

Missing C-SSRS item scores will not be imputed.

5.2.10 Abnormal Involuntary Movement Scale (AIMS)

The AIMS is assessed at Baseline, Weeks 1, 3, and Week 6/ET visits.

The AIMS is a rating scale that was designed to measure involuntary movements known as tardive dyskinesia. The AIMS has a total of 12 items rating involuntary movements of various areas of the subject's body.

- Items 1 to 7 assess the severity of dyskinesia (orofacial, extremity and truncal movements) are rated on a 5-point scale of severity: 0 (none), 1 (minimal; may be extreme normal), 2 (mild), 3 (moderate), and 4 (severe).
- Items 8 and 9 assess the overall severity and incapacitation, and are also rated on a 5-point scale of severity: 0 (none, normal), 1 (minimal), 2 (mild), 3 (moderate), and 4 (severe).
- Item 10 assesses the subject's level of awareness of the movements with associated distress and is rated on a 5-point scale: 0 (no awareness), 1 (aware, no distress), 2 (aware, mild distress), 3 (aware, moderate distress), and 4 (aware, severe distress).
- Items 11 and 12 refer to dental status and the responses are yes (scored as 1) or no (scored as 0).

The AIMS total score is the sum of the 12 item scores which can range from 0 to 42, with higher scores denoting more severe dyskinesia symptoms.

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

5.2.11 Barnes Akathisia Scale (BARS)

The BARS is assessed at Baseline, Weeks 1, 3, and Week 6/ET visits.

The BARS is a 4-item, physician-administered scale that assesses the severity of drug-induced akathisia. Items 1 to 3 assess the objective presence and frequency of akathisia, the subjective awareness of restlessness, and the subjective distress related to restlessness. These 3 items are rated on a 4-point scale from 0 to 3 and the total score is the sum of these 3 item scores, which can range from 0 to 9. Additionally, there is a global clinical assessment of akathisia which is rated on a 6-point scale from 0 to 5. For total or global clinical assessment scores, higher scores denote more severe akathisia symptoms.

Missing BARS item scores will not be imputed. The total score will be missing if any non-global item score is missing.

5.2.12 Simpson Angus Extrapyramidal Side Effect Scale (SAS)

The SAS is assessed at Baseline, Weeks 1, 3 and Week 6/ET visits.

The SAS is a 10-item physician-administered scale commonly used for the assessment of parkinsonian movement disorder related to psychiatric drug treatment. One item on the SAS measures gait/hypokinesia; 6 items measure rigidity (arm-dropping, shoulder shaking, elbow rigidity, wrist rigidity or fixation of position, head rotation, and akathisia); and 3 items measure glabella tap, tremor, and salivation, respectively. The grade of severity of each item is rated using a 5-point scale from 0 to 4. The SAS total score is the sum of the 10 items, which can range from 0 to 40, with higher scores denoting more severe parkinsonian symptoms.

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

5.3 Analysis Visit Windows

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

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

Table 4 Analysis Visit Windows

Analysis Visit Name	Target Study Day[1]	Study Day Interval
Baseline	1	≤1
Week 1	8	2 to 11
Week 2	15	12 to 18
Week 3	22	19 to 25
Week 4	29	26 to 32
Week 5	36	33 to 39
Week 6	43	40 to 50
Follow-up	71	≥51

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

5.3.1 Unscheduled Assessments

Both scheduled and unscheduled assessments, including early termination visits, will be considered for planned timepoint analyses. All assessments will be presented in data listings.

5.3.2 Multiple Measurements within Visit Windows

If more than one assessment falls within a given window then the assessment closest to the target study day will be selected for the by-visit analyses. In these analyses, if two 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 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, the average of the triplicate will be considered as one assessment for the analyses.

5.4 Missing or Incomplete Date for Last Dose of Study Drug

In the Safety Analysis Set, if the last dose date of study drug is missing for a subject who completed or early terminated from the study, then the date of the end-of-study/early termination visit will be used in the calculation of treatment duration. For the incomplete last dose date of the study drug, the imputation algorithms will be detailed in the analysis dataset

specification document. The missing or incomplete dates will be displayed in the data listings as reported on the eCRF rather than the imputed dates.

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

5.5 Missing or Incomplete Dates for Prior or Concomitant Medications

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

5.6 Missing or Incomplete Dates for Adverse Events

Missing or incomplete adverse event (AE) start dates will be imputed for the purpose of determining whether the AEs are treatment-emergent (see Section 14.1 for definition). When the chronological order of an AE onset relative to the study drug treatment period is unclear due to missing or incomplete date(s), the AE will be considered as treatment-emergent. The imputation algorithms will be detailed in the analysis dataset specification document. The missing or incomplete dates will be displayed in the data listings as reported on the eCRF rather than the imputed dates.

5.7 Missing Severity Assessment for Adverse Events

If 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 presented in data listings.

5.8 Missing Relationship to Study Drug for Adverse Events

If the relationship to study drug is missing for an AE starting 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.

5.9 Character Values of Clinical Laboratory Variables

If the reported value of a clinical laboratory variable cannot be used in a statistical analysis due to, for example, a character string reported for a numeric variable, an appropriately

determined coded value will 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.

5.10 Duplicate Subjects

Duplicate-subject screening will be performed throughout the schizophrenia program to identify whether there are individuals who are randomized more than once into the double-blind schizophrenia studies ACP-103-034, -038 or -039. Confirmed duplicates will only be included for statistical analyses under the subject number and study to which they were first randomized. Data collected under other subject number(s) or study (studies) will be listed but will not be analyzed or summarized. Case narratives will be provided for duplicate subjects.

6 SUBJECT DISPOSITION

For subjects who participate in the screening phase but are not randomized (screen failures), their demographics information (including age, sex, and primary race), screen failure reasons (the specific inclusion/exclusion criterion (or criteria) not met or other 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 one subject may be deemed ineligible for multiple inclusion/exclusion criteria and may be allowed to rescreen with the permission of the Medical Monitor, provided the screen failure was due to a temporary condition that subsequently resolved.

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

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

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

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 Project Instruction document.

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

8 DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS

Demographics and baseline characteristics will be summarized by treatment group and overall for Randomized Analysis Set, Safety Analysis Set, Full Analysis Set, and Per-Protocol Analysis Set using descriptive statistics. For the Full Analysis Set, summaries by region will also be presented. Variables include age, sex, primary race (subjects of multi-racial background can only identify/select one primary race on eCRF, or choose “other” and specify), ethnicity, height, weight, body mass index (BMI), region, current smoking status, highest education level, marital status, employment status, Baseline PANSS total score, and Baseline CGI-S.

The reported age reflect a subject’s age at the informed consent date. Age, Baseline PANSS total score and Baseline CGI-S will be presented as both continuous and categorical variables. Age categories will be presented as 18 to 35 and >35 years old. Baseline PANSS will be presented as <90 (lower symptom severity) or ≥90 (higher symptom severity). Baseline CGI-S categories will be presented as scores 1 to 7 as well as 4 (lower disease severity) or ≥5 (higher disease severity).

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

- Age (years) at diagnosis of schizophrenia disease
- Age (years) when received first antipsychotic medication for schizophrenia
- Duration (years) of schizophrenia disease
- Time (years) since first antipsychotic treatment
- Treated with clozapine in the past (yes or no)
- Number of hospitalizations for treatment of schizophrenia (0, 1 to 5, 6 to 10, or >10)
- Time (years) since last hospitalization

- Current main background antipsychotic medication
- Duration (months) of current main background antipsychotic medication
- Ever had suicidal ideation or behavior (yes or no)
- Had suicidal ideation or behavior within past 6 months (yes or no)

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

Additional information will be listed, including the name of first antipsychotic medication and the dose and frequency of the current main background antipsychotic medication. For subjects who had received clozapine, date of last clozapine dose and the dose level will also be listed. For subjects who had been hospitalized before, date of last hospitalization and treatment received in the hospital will also be listed.

9 MEDICAL HISTORY

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

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

10.1 Exposure to Study drug

For each subject, the duration of exposure to study drug (last dose date – first dose date + 1), cumulative dose (first multiply the number of doses taken by the dose level for each kit utilized, then sum the results from all kits), and average daily dose (cumulative dose in mg divided by duration of exposure in days) will be calculated and summarized by treatment group. Duration of exposure will be summarized as both continuous and categorical variables. For categorical presentation, the number and percentage of subjects in each of the following categories will be presented: <1week (1 to 6 days), 1 to <2 weeks (7 to 13 days), 2 to <3 weeks (14 to 20 days), 3 to <5 weeks (21 to 34 days), and ≥5 weeks (35 days or

longer). Kaplan-Meier curves of duration on study drug will also be presented for each treatment group. In addition, whether subject had any dose change (yes vs. no), highest dose level (20 mg or 34 mg), lowest dose level (10 mg or 20 mg), and last dose level (10 mg, 20 mg, or 34 mg) will also be summarized by category and treatment group.

The pimavanserin dose levels are expressed as free base.

10.2 Measurement of Treatment Compliance

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

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

11 PRIOR, CONCOMITANT AND POST-TREATMENT MEDICATION

For a subject, prior medication is defined as any medication with the start and stop dates prior to the date of the first dose of study drug. Concomitant medication is defined as any medication that is taken while this subject is also treated with the study drug. Any medication with a start date after the date of the last dose of study drug will be considered as post-treatment medication. Prior, concomitant, or post-treatment medications will be summarized separately. Medications will be coded using WHO Drug Dictionary (WHODRUG-DDE-B2) 2016 March or newer version. The number and percentage of subjects taking each drug class (ATC Level 3) and medication preferred term will be tabulated by treatment group and overall for Safety Analysis Set. Multiple medication usage by a subject in the same category will be counted only once.

12 EFFICACY ANALYSES

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

12.1 Efficacy Variables

Primary Efficacy Endpoint

The primary efficacy endpoint is the change from Baseline to Week 6 in the PANSS total score.

Key Secondary Efficacy Endpoint

The key secondary efficacy endpoint is the change from Baseline to Week 6 in the CGI-S score.

Other Secondary Efficacy Endpoints

Other secondary efficacy endpoints are the following:

- CGI-I score at Week 6
- Proportion of PANSS responders (≥20% and ≥30% reduction in PANSS total score) at Week 6
- Proportion of CGI-I responders (CGI-I score of 1 or 2) at Week 6
- Change from Baseline to Week 6 in PSP score
- Change from Baseline to Week 6 in DAI-10 score
- Change from Baseline to Week 6 in KSS score
- Change from Baseline to Week 6 in PANSS subscale scores (positive scale, negative scale, and general psychopathology scale)

Exploratory endpoints are the following:

- Change from Baseline to Week 6 in PANSS Marder factor scores (negative symptoms, positive symptoms, disorganized thought, uncontrolled hostility/excitement, and anxiety/depression factors)

  See Section 5.2.1 for the PANSS items included in each Marder factor.
- Change from Baseline to Week 6 in CDSS score
- Change from Baseline to Week 6 in SF-36 score
- Change from Baseline to Week 6 in PANSS individual item scores

12.2 Adjustment for Covariates

For continuous variables (except CGI-I) analyzed using the mixed model for repeated measures (MMRM) or analysis of covariance (ANCOVA), the Baseline value of the endpoint

being analyzed and geographic region (North America, Europe, or rest of world) will be included as covariates as described in Section 13. For CGI-I, the Baseline CGI-S score and geographic region will be included as covariates in the MMRM analysis.

12.3 Handling of Missing Data

The primary analysis of the primary efficacy variable will be performed assuming missing at random (MAR) using an MMRM method. Total scores that are missing, after any imputation of individual missing items as described in Section 5.2.1, will not be imputed. Sensitivity analyses of the primary efficacy variable will be performed as described in Section 13.

12.4 Multiple Comparisons / Multiplicity

A hierarchical testing procedure will be used to control the type 1 error rate across the primary and key secondary endpoints. That is, if there is no evidence to show the superiority of the adjunctive pimavanserin treatment over the adjunctive placebo with respect to the primary efficacy endpoint at the 2-sided significance level of 0.05, no formal testing for the key secondary efficacy endpoint will be performed. If the primary endpoint test is not significant at the 0.05 level then the key secondary endpoint is also not significant at the 0.05 level.

12.5 Examination of Subgroups

Treatment comparisons will be made with respect to the primary and key secondary efficacy variables using the MMRM analysis described in Section 13.1.1 separately for each subgroup by:

- region (North America, Europe, or rest of world)
- age group (18 to 35 or >35 years old)
- sex (male or female)
- primary race (white, black or African American, or other)
- main background antipsychotic medication (aripiprazole, asenapine, brexpiprazole, cariprazine, lurasidone, olanzapine, or risperidone)
- duration of schizophrenia (≤5 or >5 years)
- Baseline smoking status (smoker or non-smoker)
- Baseline BMI (<25, 25 to 30, or >30)
- Baseline symptom severity measured by PANSS total score (<90 or ≥90)
- Baseline disease severity measured by CGI-S score (4 or >5)

The LS mean differences with corresponding 95% CIs from the subgroups will also be graphically presented in forest plots for the primary efficacy variable and for the key secondary efficacy variable.

13 METHODS OF EFFICACY ANALYSES

13.1 Primary Efficacy Analysis

The primary endpoint is the change from Baseline to Week 6 in the PANSS total score. The primary analysis will be based on the Full Analysis Set. The Randomized Analysis Set and Per-Protocol Analysis Set will be used for sensitivity analyses.

Let Δ be the difference in the mean change from Baseline to Week 6 in the PANSS total score between the adjunctive pimavanserin and adjunctive placebo groups:

The null hypothesis for the primary efficacy endpoint is: $\Delta = 0$

The alternative hypothesis for the primary efficacy endpoint is: $\Delta \neq 0$

13.1.1 Primary Analysis using MMRM in the Full Analysis Set

The PANSS total score will be analyzed using mixed model repeated measures (MMRM) in the Full Analysis Set. The dependent variable will be the change from Baseline in the PANSS total score. The independent variables in the model will include the following: treatment group (adjunctive pimavanserin or adjunctive placebo), visit (Week 1, Week 2, Week 3, Week 4, Week 5, and Week 6), treatment-by-visit interaction, Baseline-by-visit interaction, geographic region (North America, Europe, or rest of world), and the Baseline PANSS total score (continuous covariate). 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 the least squares (LS) means will be estimated using observed margins.

[redacted]

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

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

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

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

Summary statistics for the PANSS total score (observed and change from baseline) will be presented for all visits from Baseline through Week 6. For change from Baseline values at each post-Baseline visit, LS means and standard errors (SE), the between-group difference in LS means with the corresponding 95% confidence interval, p-value and effect size will also be presented. In addition, LS mean ± SE over time for the change from Baseline values by treatment group will also be presented in line plots.

The hypothesis testing will be performed based on the difference in LS means at Week 6 and will be tested at an alpha level of 0.05 (2-sided). The treatment-group comparisons at each of the other timepoints (Weeks 1, 2, 3, 4, and 5) using the same MMRM model will be considered exploratory.

The LS mean differences with corresponding 95% CIs from the primary and sensitivity analyses (see following sections for details regarding the sensitivity analyses) will also be graphically presented in a forest plot.

Since the treatment dose level can be adjusted during the first 3 weeks in the study, descriptive summary statistics for the PANSS total score (observed and change from baseline) will also be presented by pimavanserin last dose level for all visits from Baseline through Week 6. In addition, the LS mean changes in PANSS total score from Baseline with corresponding 95% CIs by visit and last dose level will also be estimated using a MMRM including the following independent variables: last dose level (adjunctive placebo, adjunctive pimavanserin 10mg, adjunctive pimavanserin 20mg, or adjunctive pimavanserin 34mg), visit (Week 1, Week 2, Week 3, Week 4, Week 5, and Week 6), treatment-by-visit interaction, Baseline-by-visit interaction, geographic region (North America, Europe, or rest of world), and the Baseline PANSS total score (continuous covariate). Like the primary MMRM, 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 the least squares (LS) means will be estimated using observed margins.

13.1.2 Sensitivity Analysis using MMRM in the Per-Protocol Analysis Set

The same MMRM model as described in Section 13.1.1 will be performed in the Per-Protocol Analysis Set to assess the consistency of the conclusions from different analysis sets. Results will be presented in a summary table similar to the outputs described in Section 13.1.1.

13.1.3 Sensitivity Analysis using ANCOVA with Missing Data Imputed Using Pattern-Mixture Models with Placebo-Based Multiple Imputation

Because it is not possible to be certain that the data are missing at random (MAR), sensitivity analyses will be conducted so that the robustness of the primary endpoint with regard to missing data can be assessed, as recommended by the National Research Council report on missing data (National Research Council, 2010) and the European Medicines Agency Guidance on missing data (European Medicines Agency, 2010). Therefore, models that assume data is missing not at random (MNAR) will be used to test the robustness of the conclusions from MMRM results.

In this sensitivity analysis the assumption is that the adjunctive pimavanserin subjects who discontinue early will display the same trajectory of disease as the adjunctive placebo subjects who remain on study. Formally, this means that subjects who drop out will be assumed to have correlation with future visits similar to subjects in the adjunctive placebo arm. Thus, missing observations in the adjunctive placebo group are imputed assuming MAR, while missing values adjunctive in the adjunctive Pimavanserin group are imputed assuming MNAR. This approach may be described as a pattern mixture model (PMM) as defined by Molenberghs and Kenward (2007), with the patterns in this case defined by treatment group and time of withdrawal.

The PMM approach will be implemented using multiple imputation (MI) using the following 3 steps:

1. The posterior mean and covariance estimates from the SAS® MI procedure using the available non-missing placebo data and a random number seed of 103034 will be utilized to impute missing data in the PANSS total score. The imputed values will be constrained to be within the range of 30 to 210 and the imputed values will not be rounded.

[redacted]

[redacted]

2. The change from baseline in the PANSS total score will then be calculated and analyzed by visit for each of the fully imputed datasets using an ANCOVA model with treatment group and region as factors and the Baseline value as a covariate.

3. The treatment LS mean differences will be averaged and the associated standard errors will be summarized based on within-imputation and between-imputation variance using the SAS® MIANALYZE procedure, to yield a combined estimate for treatment effect with its associated 95% CI and p-value for each visit.

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

$$Effect\ size = \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\left(\frac{LS\ mean\ difference}{standard\ error\ of\ LS\ mean\ difference}\right)$$

Where n_1 and n_2 are the sample size in the adjunctive pimavanserin and the adjunctive placebo treatment groups at a given visit after missing data have been imputed.

The LS mean difference and its standard error will be obtained from the SAS® MIANALYZE procedure. The sign (+ or -) of the effect size will be chosen so that a positive value favors pimavanserin.

This sensitivity analysis will be performed in the Full Analysis Set and Randomized Analysis Set.

Results will be presented in a summary table similar to the outputs described in Section 13.1.1.

13.1.4 Sensitivity Analysis using Van Elteren Test in the Full Analysis Set

In order to assess the sensitivity to the normality and missing at random assumptions, a nonparametric stratified Wilcoxon rank sum test with modified ridit scores (van Elteren test) will be performed on the primary endpoint at Week 6, stratified by the geographic region (North America, Europe, or rest of world), using the Full Analysis Set.

This analysis will consist of two steps. In step 1, subjects will be ranked first according to the timepoint of the last observed PANSS total change score and then by the magnitude of the last observed PANSS total change score. In other words, subjects who have the change from Baseline in PANSS total score at Week 6 will be ranked higher (better) than subjects with missing Week 6 data. Among the subjects with missing Week 6 data, the rank will be based on the timepoint that the subject's last change from Baseline in PANSS total score is recorded: the later, the better. For subjects with the same last PANSS assessment timepoint, the rank will be based on the magnitude of the change score at the last assessment timepoint: the larger the reduction from Baseline, the better. In step 2, the rank scores computed in step 1 will be analyzed using the van Elteren test, stratified by region.

In addition to the p-value from the van Elteren test, the stratified Mann-Whitney estimator for the probability of a better outcome for the adjunctive pimavanserin treatment group compared to the adjunctive placebo treatment group and its 95% CI will also be presented. The summary table will also present the probability and its corresponding 95% CI for each region separately.

The stratified Mann-Whitney estimator will be calculated as follows:

For each region, the Mann-Whitney estimator for the probability of a better outcome (based on the rank scores from step 1 above) for the adjunctive pimavanserin treatment group compared to the adjunctive placebo treatment group will be calculated from Somers' D (which can be obtained from the SAS® FREQ procedure):

$$MW = (1 + D)/2$$

The stratified Mann-Whitney estimator for all regions will then be calculated as the weighted average across the regions:

$$StratMW = \frac{\sum_{h=1}^{3} w_h MW_h}{\sum_{h=1}^{3} w_h}$$

Where $w_h = \frac{n_{h1} n_{h2}}{n_{h1} + n_{h2} + 1}$;

and n_{h1} and n_{h2} are the sample size in the adjunctive pimavanserin and the adjunctive placebo treatment groups within region h.

13.1.5 Complete Case Analysis using ANCOVA in the Full Analysis Set

The observed cases excluding missing values will be analyzed using an ANCOVA model with treatment group and region as factors and the Baseline value as a covariate.

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

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

Where MSE is the mean squared error from the ANCOVA model. The sign (+ or -) of the effect size will be chosen so that a positive value favors pimavanserin.

13.2 Key Secondary Efficacy Analysis

The key secondary endpoint is the change from Baseline to Week 6 in the CGI-S score. The change from Baseline in the CGI-S score will be analyzed using an MMRM model similar to that described in Section 13.1.1 for the primary endpoint, except that the Baseline CGI-S score will be included in the model instead of the Baseline PANSS total score. The treatment comparison will be based on the difference in LS means at Week 6 and will be tested at an alpha level of 0.05 (2-sided) using the Full Analysis Set.

Let Δ be the difference in the mean change from Baseline to Week 6 in the CGI-S score between the adjunctive pimavanserin and adjunctive placebo groups:

The null hypothesis for the key secondary efficacy endpoint is: $\Delta = 0$

The alternative hypothesis for the key secondary efficacy endpoint is: $\Delta \neq 0$

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

Results will be presented in a summary table and line plots similar to the outputs described in Section 13.1.1.

13.3 Other Secondary Efficacy Analyses

13.3.1 Responder Analysis

Three types of responders will be counted within each treatment group at each post-Baseline visit:

- ≥ 20% reduction in the corrected PANSS total score (subtracting 30 points from the raw scores; see details below) from Baseline

- ≥ 30% reduction in the corrected PANSS total score (subtracting 30 points from the raw scores; see details below) from Baseline
- CGI-I score of 1 (very much improved) or 2 (much improved)

Note that since PANSS is an interval scale that lacks a natural zero point, the percent change in PANSS total score will be calculated based on corrected scores after subtracting 30 points from the raw scores (Leucht et al., 2007; Leucht et al., 2009; Obermeier et al., 2010; Obermeier et al., 2011). For example, if a subject's Baseline PANSS total score is 80 and Week 6 PANSS total score is 30 (absent of all symptoms), the percent change from Baseline to Week 6 in PANSS total score will be calculated as [(30 – 30) – (80 – 30)] ÷ (80 – 30) × 100% = -100%.

For each responder analysis at each timepoint, the proportion of responders will be summarized by treatment group. The treatment groups will be compared using a Cochran-Mantel-Haenszel (CMH) test stratified by geographic region (North America, Europe, or rest of world). The adjusted difference in percent responders between the treatment groups (adjunctive pimavanserin group minus adjunctive placebo group) using the weighting scheme of CMH and Newcombe's 95% CI will also be presented.

At any given visit, the subjects with missing values will be imputed as non-responders for that visit. Using this imputation method, the proportion of responders (≥ 20% reduction in the corrected PANSS total score, ≥ 30% reduction in the corrected PANSS total score, or with CGI-I score of 1 or 2) at Week 6 by treatment group will also be presented in a bar chart.

In addition, an observed-case responder analysis (subjects with missing values at a given visit are excluded from the analysis for that visit) will also be performed.

For change in PANSS total score from Baseline, the cumulative distribution function at Week 6 for the observed cases will be plotted by treatment group.

The proportion of subjects who were worsened or no change, based on CGI-I evaluation, will be summarized by treatment group for each visit. Worsening or no change is defined as having a CGI-I score of 4 (no change), 5 (minimally worse), 6 (much worse), or 7 (very much worse) at a given visit. The treatment groups will be compared using a similar CMH test as described above for the responder analyses. Results will be presented using 2 different missing-handling methods: (1) subjects with missing CGI-I at a given visit are considered as worsened or no change for that visit, and (2) observed-case analysis (subjects with missing values at a given visit are excluded from the analysis for that visit).

13.3.2 Other Secondary Efficacy Analysis for Continuous Variables

The change from Baseline to each post-Baseline timepoint in the PANSS subscores (positive scale, negative scale, and general psychopathology scale) and the KSS score will be analyzed using an MMRM model similar to that described above for the primary endpoint, except that the Baseline value of the endpoint being analyzed will be included in the model instead of the Baseline PANSS total score.

The CGI-I score at each post-Baseline timepoint will be analyzed using an MMRM model. The dependent variable will be the CGI-I score. The independent variables in the model will include the following: treatment group (adjunctive pimavanserin or adjunctive placebo), visit (Week 1, Week 2, Week 3, Week 4, Week 5, and Week 6), treatment-by-visit interaction, Baseline-by-visit interaction, geographic region (North America, Europe, or rest of world) and the Baseline CGI-S score (continuous covariate).

To assess whether there is a linear correlation between the change in PANSS total score from Baseline and CGI-I rating, the change in PANSS total score from Baseline will also be summarized by CGI-I rating and treatment group for all post-baseline visits. The mean $\pm$ SE change in PANSS total score from Baseline by CGI-I rating will also be plotted for Week 6 observed cases.

The change from Baseline to Week 6 in the PSP and DAI-10 scores will be analyzed using an analysis of covariance (ANCOVA) model with treatment group (adjunctive pimavanserin or adjunctive placebo) and geographic region (North America, Europe, or rest of world) as factors, and the Baseline value of the endpoint being analyzed as a covariate. Missing values will be imputed with the previous value (including Baseline) for that subject, i.e. last observation carried forward (LOCF). In addition, an observed-case analysis, ignoring missing values, will also be performed.

At Week 6, the effect size (Cohen's d) for the change from Baseline between the treatment groups will be calculated using the following formula:

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

Where MSE is the mean squared error from the ANCOVA model. The sign (+ or -) of the effect size will be chosen so that a positive value favors pimavanserin.

Results will be presented in summary tables similar to the outputs described in Section 13.1.1.

13.4 Exploratory Efficacy Analyses

The change from Baseline to each post-Baseline timepoint in the PANSS Marder factor scores (negative symptoms, positive symptoms, disorganized thought, uncontrolled hostility/excitement, and anxiety/depression factors) and the PANSS individual item scores will be analyzed using an MMRM model similar to that described above for the primary endpoint, except that the Baseline value of the endpoint being analyzed will be included in the model instead of the Baseline PANSS total score.

The change from Baseline to Week 6 in the CDSS and the normalized SF-36 scale scores and composite scores (physical functioning, bodily pain, role limitations due to physical health problems, role limitations due to personal or emotional problems, emotional well-being, social functioning, energy/fatigue, and general health perceptions scale scores; physical health composite score, mental health composite score) will be analyzed using an analysis of covariance (ANCOVA) model with treatment group (adjunctive pimavanserin or adjunctive placebo) and geographic region (North America, Europe, or rest of world) as factors, and the Baseline value of the endpoint being analyzed as a covariate. Missing values will be imputed using LOCF. In addition, an observed-case analysis, ignoring missing values, will also be performed.

Results will be presented in summary tables similar to the outputs described in Section 13.1.1.

Correlations among efficacy endpoints (observed cases) will be assessed using Spearman's rank correlation coefficient (Spearman's rho) and presented in a correlation matrix. The 95% CIs for the Spearman's rank correlation coefficients will also be presented.

14 SAFETY ANALYSES

The safety analysis will be performed based on the Safety Analysis Set using actual treatment. Safety variables include adverse events (AEs), clinical laboratory variables, vital signs, body weight, BMI, physical examinations, electrocardiogram (ECG), C-SSRS, AIMS, BARS and SAS variables.

14.1 Adverse Events

All Adverse events (AEs) will be coded using MedDRA version 19.0 or newer.

An AE (classified by preferred term) will be considered a treatment-emergent AE (TEAE) if it started after first study dose administration and no later than last study dose date + 30. AEs reported on Day 1 based on Baseline (pre-dose) findings (e.g. clinically significantly

abnormal vital signs, laboratory test results, or electrocardiogram parameters) will not be considered as TEAEs.

The event counts and the number and percentage of subjects reporting TEAEs in each treatment group will be tabulated by SOC and preferred term; and, by SOC, preferred term, and maximum severity. If more than 1 AE occurs with the same preferred term for the same subject, the subject will be counted only once for that preferred term using the most severe occurrence for the summarization by severity. In addition, the event counts and the number and percentage of subjects with TEAEs classified by the Investigators as related to the study drug, with most frequently reported TEAEs (preferred terms reported by ≥ 5% of subjects in either treatment group), with treatment-emergent serious AEs (TESAEs), with fatal AEs (i.e. events that cause death), and with TEAEs leading to discontinuation of study drug will be summarized by SOC and preferred term within each treatment group. The display in these tables will be sorted alphabetically by SOC and then by descending subject frequency for the preferred terms (combined counts from both treatment groups) within each SOC.

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

An AE listing by subject will display all events, including those which are not treatment-emergent, and will include the verbatim term in addition to the MedDRA SOC and preferred term. This listing will also include all relevant eCRF data associated with the event: e.g. date of onset, date resolved, date of first dose, date of last dose, last prescribed 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. Separate listings will be presented for subjects with SAEs, subjects with AEs leading to discontinuation and subjects who died (if any). In these listings, an indicator for treatment-emergent events will also be included.

14.2 Clinical Laboratory Variables

Clinical laboratory tests are performed at Screening, Baseline and Week 6/ET visits.

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

- Serum chemistry tests include the following:
  - Sodium (Na) , potassium (K), chloride (Cl), phosphorus (P), calcium (Ca), carbon dioxide (CO_2), blood urea nitrogen (BUN), creatinine (Cr), uric acid
  - Alanine aminotransferase (ALT/SGPT), aspartate aminotransferase (AST/SGOT), gamma-glutamyl transpeptidase (GGT), alkaline phosphatase (ALP), total bilirubin (TBIL), lactate dehydrogenase (LDH), glucose
  - Albumin (ALB), total protein
  - Creatine kinase (CK)/creatine phosphokinase (CPK)
  - Prolactin
  - Lipid panel
    - Total cholesterol, high-density lipoprotein (HDL)-cholesterol, triglycerides, low-density lipoprotein(LDL)-cholesterol, cholesterol/HDL ratio, non-HDL cholesterol
  - Screening only:
    - HbA1c
    - Thyroid stimulating hormone (TSH); full thyroid panel only if TSH value is outside of the laboratory reference range (full thyroid panel will be listed but not summarized)
    - Presence or absence of the identified main antipsychotic in the plasma will also be assessed at Screening (will be listed but not summarized)
- Urinalysis tests include the following:
  - Occult blood, leukocyte esterase, protein, glucose, ketones, specific gravity, pH

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').

Clinical laboratory values for hematology, chemistry and urinalysis (red blood cells, white blood cells, specific gravity and pH) will be summarized by treatment group using descriptive statistics at Baseline and Week 6 visits. The change from Baseline values will also be summarized by treatment group at the Week 6 visit. The overall minimum and maximum post-Baseline observed and change from Baseline values will also be summarized. For hemoglobin, hematocrit and uric acid, the above summaries will be presented for each gender as well as for both genders combined. For urinalysis with categorical results (blood, protein, glucose, and ketones), the number and percentage of subjects will be tabulated by category at Baseline and Week 6, and the denominator is the number of subjects with non-missing values for the given parameter, visit and treatment group.

Laboratory values (except the ones that were only assessed at screening) will also be summarized in shift tables by treatment group, to determine the number and percentage of

subjects with values classified as below (low), within (normal), and above (high) normal ranges at Week 6, overall post-Baseline minimum and overall post-Baseline maximum, relative to the same classification at the Baseline visit. For the by-visit shift summary, the denominator is the number of subjects with non-missing values at Baseline and the given visit for the given parameter and treatment group. For the shift to the overall post-Baseline minimum or maximum, the denominator is the number of subjects with non-missing Baseline and at least 1 post-Baseline value for the given parameter and treatment group. For hemoglobin, hematocrit and uric acid, the shift summaries will be presented for each gender as well as for both genders combined.

Number and percentage of subjects with potentially clinically important laboratory values (PCI) at Week 6 and overall post-Baseline will be summarized by treatment group for selected parameters. PCI criteria are listed in Table 5 and Table 6. For the overall post-Baseline summaries of PCI values, all post-Baseline values will be considered, including unscheduled and out of window values. Subjects with multiple PCI values for a given parameter will be counted only once for that parameter. For the by-visit summary, the numerator for the percentage is the number of subjects with a post-Baseline PCI value for the given parameter, visit and treatment group, and the denominator is the number of subjects with non-missing values for the given parameter, visit and treatment group. For the overall post-Baseline summary, the numerator for the percentage is the number of subjects with at least 1 post-Baseline PCI laboratory value for the given parameter and treatment group, and the denominator is the number of subjects with at least 1 post-Baseline laboratory value for the given parameter and treatment group. For hemoglobin, hematocrit and uric acid, the count and percentage of subjects with PCI values will be presented for each gender as well as for both genders combined. Subjects with any PCI values will be presented in an additional listing.

Table 5 Criteria for Potentially Clinically Important Laboratory Values – Hematology and Chemistry

Analyte	Conventional Unit	Low PCI Criteria	High PCI Criteria	SI Unit	Low PCI Criteria	High PCI Criteria
Hematology (whole blood)						
Hemoglobin (male)	g/dL	<11	>18	g/L	<110	>180
Hemoglobin (female)	g/dL	<10	>17	g/L	<100	>170
Hematocrit (male)	%	<30	>55	L/L	<0.3	>0.55
Hematocrit (female)	%	<30	>50	L/L	<0.3	>0.5
Leukocyte (White Blood Cell Count)	x 10^3/uL	≤2.8	≥15	x 10^9/L	≤2.8	≥15
Neutrophils	x 10^3/uL	≤1.5	No upper limit	x 10^9/L	≤1.5	No upper limit
Platelet Count	x 10^3/uL	≤75	≥700	10^9/L	≤75	≥700
Chemistry (serum or plasma)						
ALT (SGPT)	U/L	No lower limit	≥3 X ULN	U/L	No lower limit	≥3 X ULN
AST (SGOT)	U/L	No lower limit	≥3 X ULN	U/L	No lower limit	≥3 X ULN
Total Bilirubin	mg/dL	No lower limit	≥1.5 ULN	umol/L	No lower limit	≥1.5 ULN
BUN	mg/dL	No lower limit	≥30.0	mmol/L	No lower limit	≥10.71
Creatine Kinase (CK)	U/L	No lower limit	≥3 ULN	U/L	No lower limit	≥3 ULN
Sodium	mEq/L	≤125	≥155	mmol/L	≤125	≥155
Potassium	mEq/L	≤3.0	≥5.5	mmol/L	≤3.0	≥5.5
Calcium, total	mg/dL	<8.0	>11.0	mmol/L	<2.0	>2.75
Lactate Dehydrogenase (LDH)	U/L	No lower limit	≥3 X ULN	U/L	No lower limit	≥3 X ULN
Alkaline Phosphatase	U/L	No lower limit	≥3 X ULN	U/L	No lower limit	≥3 X ULN
Uric acid (male)	mg/dL	No lower limit	≥10.5	umol/L	No lower limit	≥624.75
Uric acid (female)	mg/dL	No lower limit	≥8.5	umol/L	No lower limit	≥505.75
Albumin	g/dL	≤2.6	≥6.0	g/L	≤26	≥60
Total Protein	g/dL	≤5.0	≥10.0	g/L	≤50	≥100
Chloride	mEq/L	≤85	≥120	mmol/L	≤85	≥120
Glucose (random)	mg/dL	≤45.1	≥200.0	mmol/L	≤2.48	≥11
Serum Creatinine	mg/dL	Not Applicable	>1.5 ULN	umol/L	Not Applicable	>1.5 ULN
Triglycerides	mg/dL	Not Applicable	>300	mmol/L	Not Applicable	>3.39
Gamma-Glutamyl Transferase (GGT)	U/L	Not Applicable	≥3 ULN	U/L	Not Applicable	≥3 ULN

Table 6 Criteria for Potentially Clinically Important Laboratory Values - Urinalysis

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

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

14.3 Vital Signs

Vital signs are assessed at Screening, Baseline, Weeks 1, 2, 3, 4, 5 and Week 6/ET visits.

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

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

Table 7 Criteria for Potentially Clinically Important Vital Signs

Vital Sign Parameter	Unit	Criteria[a]		
		Observed Value	And/Or	Change Relative to Baseline
Systolic blood pressure (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 (ECG)

Electrocardiogram is performed at Screening, Baseline, Week 4 and Week 6/ET visits.

All tracings will be evaluated by a central reading laboratory. At the Baseline visit, the machine-read results will also be recorded. All data, including the machine-read Baseline results, will be listed. ECG data summaries will be performed using the centrally evaluated data.

Observed values of ECG variables (heart rate, PR interval, RR interval, QRS interval, QT interval, and QTc intervals) and the changes from baseline at each assessment timepoint will be summarized by treatment group. QTc intervals include QTcB (Bazett's formula) and QTcF (Fridericia's formula).

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

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

For cardiologist's interpretations, the number and percentage of subjects with ECG results that are determined as normal or abnormal will be summarized at scheduled visits. The overall post-baseline worst interpretation will also be summarized (i.e. if a subject has one or more post-baseline ECG results that is/are considered as abnormal, this subject will be counted in the abnormal category). Cardiologist's interpretations will also be summarized in a shift table to determine the number and percentage of subjects with ECG results classified as normal or abnormal at scheduled post-Baseline visits relative to the same classification at the Baseline visit. The shifts from Baseline to overall post-Baseline worst interpretation will also be presented. For the by-visit shift summary, the denominator is the number of subjects with non-missing cardiologist's interpretation at Baseline and the given visit for the given treatment group. For the summaries of shift from Baseline to the overall post-Baseline worst interpretation, the denominator is the number of subjects with non-missing Baseline and at least 1 post-Baseline cardiologist's interpretation for the given treatment group.

Electrocardiogram variable values will be considered PCI if they meet or exceed the upper limit values listed in Table 8. The number and percentage of subjects with post-baseline PCI values will be summarized by treatment group at each post-Baseline visit and for overall post-Baseline. For the by-visit summary, the numerator for the percentage is the number of subjects with a post-Baseline PCI ECG for the given parameter, visit and treatment group, and the denominator is the number of subjects with non-missing values for the given parameter, visit and treatment group. For the overall post-Baseline summary, the numerator for the percentage is the number of subjects with at least 1 post-Baseline PCI ECG for the given parameter and treatment group, and the denominator is the number of subjects with at least 1 post-Baseline ECG value for the given parameter and treatment group. A listing of all subjects with any PCI values will be provided.

Table 8 Criteria for Potentially Clinically Important ECG Values

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

14.5 Physical Examination

Physical examination is performed at Screening, Baseline, and Week 6/ET visits.

Physical examination results (normal, abnormal, and not done) at Baseline and Week 6 will be summarized in a frequency table by treatment group, body system and visit.

14.6 Other Safety Variables

14.6.1 Suicidality

The C-SSRS is assessed at Screening, Baseline, Weeks 1, 2, 3, 4, 5, and Week 6/ET visits.

The event counts and the number and percentage of subjects reporting any post-Baseline suicidal ideation (wish to be dead; non-specific active suicidal thoughts; active suicidal ideation with any methods (not plan) without intent to act; active suicidal ideation with some intent to act, without specific plan; active suicidal ideation with specific plan and intent), suicidal behavior (preparatory acts or behavior; aborted attempt; interrupted attempt; actual attempt; suicide), or suicidality (any suicidal ideation or behavior) will be tabulated for each treatment group.

14.6.2 Extrapyramidal Symptom Measures

14.6.2.1 Abnormal Involuntary Movement Scale (AIMS)

The AIMS is assessed at Baseline, Weeks 1, 3 and Week 6/ET visits.

The AIMS total score will be summarized by treatment group using descriptive statistics at Baseline, Weeks 1, 3 and 6 visits. The change from Baseline scores will also be summarized by treatment group at Weeks 1, 3 and 6 visits.

In addition, the number and percentage of subjects with dyskinesia will be summarized by treatment group at each visit and for overall post-Baseline. Dyskinesia is defined as having a score of 3 or more on any of the first 7 AIMS items or a score of 2 or more on any two of the first 7 AIMS items. If there are multiple assessments performed within the same visit window, all assessments will be evaluated using the above criteria. The tabulations will be presented for subjects who have at least 1 AIMS assessment as well as for a subset of these subjects who do not have dyskinesia at Baseline.

The individual item scores will be listed but not summarized.

14.6.2.2 Barnes Akathisia Scale (BARS)

The BARS is assessed at Baseline, Weeks 1, 3 and Week 6/ET visits.

The BARS total score and the Global Clinical Assessment of Akathisia score will be summarized by treatment group using descriptive statistics at Baseline, Weeks 1, 3 and 6 visits. The change from Baseline scores will also be summarized by treatment group at Weeks 1, 3 and 6 visits.

In addition, the number and percentage of subjects with akathisia will be summarized by treatment group at each visit and for overall post-Baseline. Akathisia is defined as having a

Global Clinical Assessment of Akathisia score ≥2. If there are multiple assessments performed within the same visit window, all assessments will be evaluated using the above criteria. The tabulations will be presented for subjects who have at least 1 BARS assessment as well as for a subset of these subjects who do not have akathisia at Baseline.

The individual item scores will be listed but not summarized.

14.6.2.3 Simpson Angus Scale (SAS)

The SAS is assessed at Baseline, Weeks 1, 3 and Week 6/ET visits.

The SAS total score will be summarized by treatment group using descriptive statistics at Baseline, Weeks 1, 3 and 6 visits. The change from Baseline scores will also be summarized by treatment group at Weeks 1, 3 and 6 visits.

The number and percentage of subjects with Parkinsonism will be summarized by treatment group at each visit and for overall post-Baseline. Parkinsonism is defined as having a SAS total score >3. If there are multiple assessments performed within the same visit window, all assessments will be evaluated using the above criteria. The tabulations will be presented for subjects who have at least 1 SAS assessment as well as for a subset of these subjects who do not have Parkinsonism at Baseline.

The individual item scores will be listed but not summarized.

15 CLINICAL PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES

PK samples for pimavanserin, its metabolite (AC-279), and the main background antipsychotic medications are collected pre-dose at Baseline, Weeks 1, 3 and Week 6/ET visits.

For pimavanserin-treated subjects, plasma concentration for pimavanserin and AC-279 will be listed. For pimavanserin-treated subjects in the Pharmacokinetics Analysis Set, plasma concentration for pimavanserin and AC-279 will be summarized using descriptive statistics by visit and the last dose level received before the PK sample is drawn. Note that if there is a dose adjustment at Week 1 or Week 3, the dose level of the PK sample will reflect the previous dose level before the adjustment since the PK sample is collected before the dose adjustment. Concentrations that are below the limit of quantification (BLQ) will be displayed as “BLQ” in the data listings and imputed as 0 for computing summary statistics.

Plasma concentration data for the background main antipsychotics will be listed.

If data allow, population PK and PK/PD analyses will be performed to further characterize the PK profile and exposure response relationship of pimavanserin and its metabolite using measures of safety, and efficacy parameters. The results of population PK and PK/PD

modeling will be presented in a separate report. Pimavanserin and AC-279 plasma concentration data will remain blinded until the unblinding of the clinical database at the end of the study.

16 UNBLINDED INTERIM ANALYSIS

One interim analysis to assess futility based on the primary endpoint will be conducted by an independent statistician when 152 subjects (40% of the planned sample size) complete the Week 6 visit or discontinue from the study before the data extract date. The independent statistical group not affiliated with the Sponsor or the DSMB will compute the predictive power as detailed below and will communicate to the DSMB. The DSMB chair will communicate to the Sponsor designated contact person in writing whether the study met the futility criterion or not (a binary outcome "met futility criterion" or "did not meet futility criterion"). In order to protect the blinding and the integrity of this study, only the binary outcome will be reported to the Sponsor without any specific results. The actual computational output will be securely stored by the independent statistical group and retained until the study is unblinded after the final database lock.

For futility evaluation, the predictive power will be computed based on the interim data using this formula (Jennison and Turnbull 2000):

$$Predictive\ Power = \Phi\left(\frac{-Z_k\sqrt{I_F} - z_{1-\alpha/2}\sqrt{I_k}}{\sqrt{I_F - I_k}}\right)$$

Where

Φ is the cumulative distribution function of the standard normal distribution;

Z_k is the t statistic computed from the interim observed data at time k for the treatment group comparison (adjunctive pimavanserin minus adjunctive placebo) of the change from Baseline in PANSS total scores at Week 6 using the MMRM model specified in Section 13.1.1;

$z_{1-\alpha/2}$ is the standard normal value for the test with a 1-sided type 1 error rate of $\alpha/2 = z_{0.975}$;

I_k is the interim information level = $\frac{1}{\sigma^2\left(\frac{1}{n_{1k}}+\frac{1}{n_{2k}}\right)}$;

I_F is the final information level = $\frac{1}{\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$;

σ^2 is the pooled variance from the interim observed data obtained from the covariance matrix estimated for MMRM;

n_{1k} and n_{2k} are the interim sample size in the adjunctive pimavanserin and the adjunctive placebo treatment groups at time k ;

n_1 and n_2 are the final sample size in the adjunctive pimavanserin and the adjunctive placebo treatment groups.

The predictive power assesses the probability that the study will be able to demonstrate a significant difference in favor of pimavanserin (for the primary endpoint) at the final analysis, given the interim data. At the interim analysis, the study may be terminated if the predictive power is less than 0.2. This study will not be stopped early for superior efficacy and therefore the type 1 error rate will not be inflated. Investigators and Sponsor personnel will remain blinded throughout the study.

17 DATA MONITORING/REVIEW COMMITTEE

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

An independent statistician (and/or programmer) not affiliated with the Sponsor will produce unblinded 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 will include but are not limited to summaries of enrollment and disposition, demographics and baseline characteristics, medical and schizophrenia histories, concomitant medications, study drug exposure, all treatment-emergent adverse events (including deaths, SAE and AEs leading to discontinuation), vital signs, laboratory test results, and ECG parameters. Subject profiles, boxplots of Baseline and most extreme post-Baseline values for clinical laboratory data, and listings of AEs and potentially clinically important laboratory and QTcF results will also be provided to DSMB members for their review.

18 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.

19 CHANGES TO ANALYSES SPECIFIED IN PROTOCOL

The protocol stated that a hierarchical testing procedure will be used to control the Type 1 error rate across the primary and secondary endpoints. As indicated in this SAP, the hypothesis testing for the primary and key secondary efficacy endpoints will be tested in a hierarchical (sequential) order. For other secondary endpoints, nominal p-values for the treatment-group comparisons will be reported.

20 REFERENCES

Jennison C and Turnbull BW (2000). *Group Sequential Methods with Applications to Clinical Trials.* Chapman & Hall/CRC: New York, NY; 210-213.

Leucht S, et al. Defining 'response' in antipsychotic drug trials: recommendations for the use of scale-derived cutoffs. *Neuropsychopharmacology.* 2007; 32: 1903-1910.

Leucht S, et al. Definitions of response and remission in schizophrenia: recommendations for their use and their presentation. *Acta Psychiatr Scand.* 2009; 119 (Suppl. 438): 7-14.

Obermeier M, et al. Should the PANSS be rescaled? *Schizophrenia Bulletin.* 2010; 36(3): 455-460.

Obermeier M, et al. Is the PANSS used correctly? A systematic review. *BMC Psychiatry.* 2011; 11:113.

21 APPENDICES

21.1 Summary of Version Changes

Version No:	Document History Description of Update	Author(s)	Version Date
1.0	Original version	[redacted]	01 June 2018