[redacted]

Statistical Analysis Plan: 15-AVP-786-202 Study for Residual Schizophrenia

Study Title:	A Phase 2, Multicenter, Randomized, Double-Blind, Placebo-Controlled Study to Assess the Efficacy, Safety, and Tolerability of AVP-786 (Deuterium Modified Dextromethorphan Hydrobromide/Quinidine Sulfate) as an Adjunctive Therapy in Patients with Residual Schizophrenia
Study Number:	15-AVP-786-202
Study Phase:	Phase 2
Sponsor:	Avanir Pharmaceuticals Inc. 30 Enterprise, Suite 400 Aliso Viejo, California 92656 (949) 389-6700
Version:	Final
Date:	15 August 2017

NCT Number: NCT02477670
This NCT number has been applied to the document for purposes of posting on Clinicaltrials.gov

1 TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

2 SIGNATURE PAGE

Study Title: A Phase 2, Multicenter, Randomized, Double-Blind, Placebo-Controlled Study to Assess the Efficacy, Safety, and Tolerability of AVP-786 (Deuterium Modified Dextromethorphan Hydrobromide/Quinidine Sulfate) as an Adjunctive Therapy in Patients with Residual Schizophrenia

Study Number: 15-AVP-786-202

Prepared by: [redacted] Date: 8/15/2017

Approved by [redacted] Date: 8/15/2017

Approved by [redacted] Date: 8/15/2017

Approved by [redacted] Date: 8/15/2017

Approved by [redacted] Date: 8/15/2017

Approved by: [redacted] Date: 8/15/2017

3 LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

AA	Apathy avolition
AE	Adverse event
AIMS	Abnormal Involuntary Movements Scale
ALT/SGPT	Alanine aminotransferase/serum glutamic-pyruvic transaminase
ANCOVA	Analysis of covariance
AST/SGOT	Aspartate aminotransferase/serum glutamic-oxaloacetic transaminase
AUC	Area under the concentration-time curve
AVP-786	Deuterated [d6]-dextromethorphan hydrobromide and quinidine sulfate combination
BAS	Barnes Akathisia Scale
BID	Twice-daily
BP	Blood pressure
bpm	Beats per minute
BUN	Blood urea nitrogen
C_{max}	Maximum plasma concentration
CDSS	Calgary Depression Scale for Schizophrenia
CFBL	Change from Baseline
CGI-C	Clinical Global Impression of Change
CGI-S	Clinical Global Impression of Severity of Illness
CK	Creatine kinase
C-SSRS	Columbia Suicide Severity Rating Scale
[redacted]	[redacted] Subject Database
d6-DM	Deuterated (d6)-dextromethorphan hydrobromide
DM	Dextromethorphan hydrobromide
DSMB	Data and Safety Monitoring Board
DSM-IV-TR	Diagnostic and Statistical Manual of Mental Disorders, 4th edition, Text Revision
ECG	Electrocardiogram
eCRF	Electronic case report form
EEfRT	Effort Expenditure for Reward Task
FDA	US Food and Drug Administration
GGT	Gamma-glutamyl transferase
HbA1c	Glycosylated hemoglobin
HR	Heart rate
ICF	Informed Consent Form
ITT	Intent-to-Treat
LDH	Lactate dehydrogenase
LOCF	Last observation carried forward

M.I.N.I.	Mini International Neuropsychiatric Interview
MATRICS	Measurement and Treatment Research to Improve Cognition in Schizophrenia
MCCB	MATRICS Consensus Cognitive Battery
MedDRA	Medical Dictionary for Regulatory Activities
mg	Milligram(s)
mITT	Modified Intent-to-Treat
MMRM	Mixed-effects model for repeated measures
NSA-16	16-Item Negative Symptom Assessment
NSA-4	4-Item Negative Symptom Assessment
OLS	Ordinary least squares
PANSS	Positive and Negative Syndrome Scale
PCS	Potentially clinically significant
PD	Pharmacodynamic(s)
PGIC	Patient Global Impression of Change
pH	Potential hydrogen
PK	Pharmacokinetic(s)
PP	Per-protocol
PT	Preferred Term
Q	Quinidine (sulfate)
QD	Once daily
QRS	The Q-R-S complex from an ECG tracing
QT	QT interval from an ECG tracing
QTc	QT interval corrected for heart rate
QTcF	QT interval corrected for heart rate using the Fridericia's formula
SAE	Serious adverse event
SAP	Statistical analysis plan
SAS	Simpson Angus Scale for Extrapyramidal Symptoms
SAS®	Statistical Analysis Software
SGA	Second generation antipsychotic
SOC	System Organ Class
SNRI	Selective serotonin-norepinephrine reuptake inhibitor
SPCD	Sequential parallel comparison design
SSRI	Selective serotonin reuptake inhibitor
SUR	Seemingly unrelated regression
TEAE	Treatment-emergent adverse event
TFLs	Tables, figures, and listings
ULN	Upper limit of normal
US	United States of America
WHO	World Health Organization

4 INTRODUCTION

This document provides a detailed description of the statistical methods and procedures to be implemented during the analysis of Avanir Pharmaceuticals Incorporated Protocol 15-AVP-786-202 (Amendment 4, Version 5), dated 25 January 2016.

5 TRIAL OBJECTIVES

The objectives of this 12-week study are to evaluate the efficacy, safety, and tolerability of AVP-786, a combination of deuterated-dextromethorphan hydrobromide (d6-DM) and quinidine sulfate (Q), as an adjunctive therapy compared with placebo in patients with residual schizophrenia on negative and positive symptoms and cognitive function.

5.1 Primary Objectives

The primary study objective is to evaluate the efficacy of AVP-786 compared to placebo for the treatment of negative symptoms associated with schizophrenia, as measured by the 16-item Negative Symptom Assessment (NSA-16) total score.

5.2 Secondary Objectives

The secondary objectives of this study are to evaluate the efficacy, safety, and tolerability of AVP-786 compared to placebo in patients with residual schizophrenia.

- Clinical Efficacy will be measured by the Positive and Negative Syndrome Scale (PANSS) total score and PANSS subscales; NSA-16 factor domains, global score, individual items, and NSA-4; proportion of patients with a reduction of 20% or greater in the PANSS total score; MCCB composite score, CGI-S, CGI-C, and PGIC scores; CDSS; and EEfRT.

- Safety and tolerability will be measured by reported AEs, SAEs, physical examinations, vital signs, weight, pregnancy tests, clinical laboratory assessments, 12-lead ECGs, as well as C-SSRS, AIMS, BAS, and SAS scales.

6 STUDY DESIGN CONSIDERATIONS

6.1 Study Design and Population

This is a prospective, multicenter, randomized, double-blind, placebo-controlled study with a duration of 12 weeks, consisting of 2 consecutive double-blind treatment stages, each 6 weeks in duration (Stage 1 and Stage 2). The study utilizes a Sequential Parallel Comparison Design (SPCD).

Table 1. Overview of Study Duration

	Screening	Stage 1	Stage 2	Telephone Follow-Up
Duration in Days	Day -28 to -7	Day 1 to 42	Day 43 to 85	Day 86 to 90

Approximately 120 patients will participate in the study at roughly 15 research centers in the United States. Eligible patients for this study must be adult outpatients (between 18 and 60 years of age) who are diagnosed with schizophrenia, are clinically stable, in a residual (non-acute) phase of illness, and will have been treated with a second generation antipsychotic (SGA) agent for at least 3 months (90 days) and on a stable dose for 1 month (30 days) prior to the Screening Visit. The diagnosis of residual schizophrenia will be based on the Diagnostic and Statistical Manual of Mental Disorders, 4th edition, Text Revision (DSM-IV-TR).

Initial randomization will occur at baseline and will mark the start of Stage 1. Patients will be randomized in a [redacted] ratio (AVP-786 to placebo).

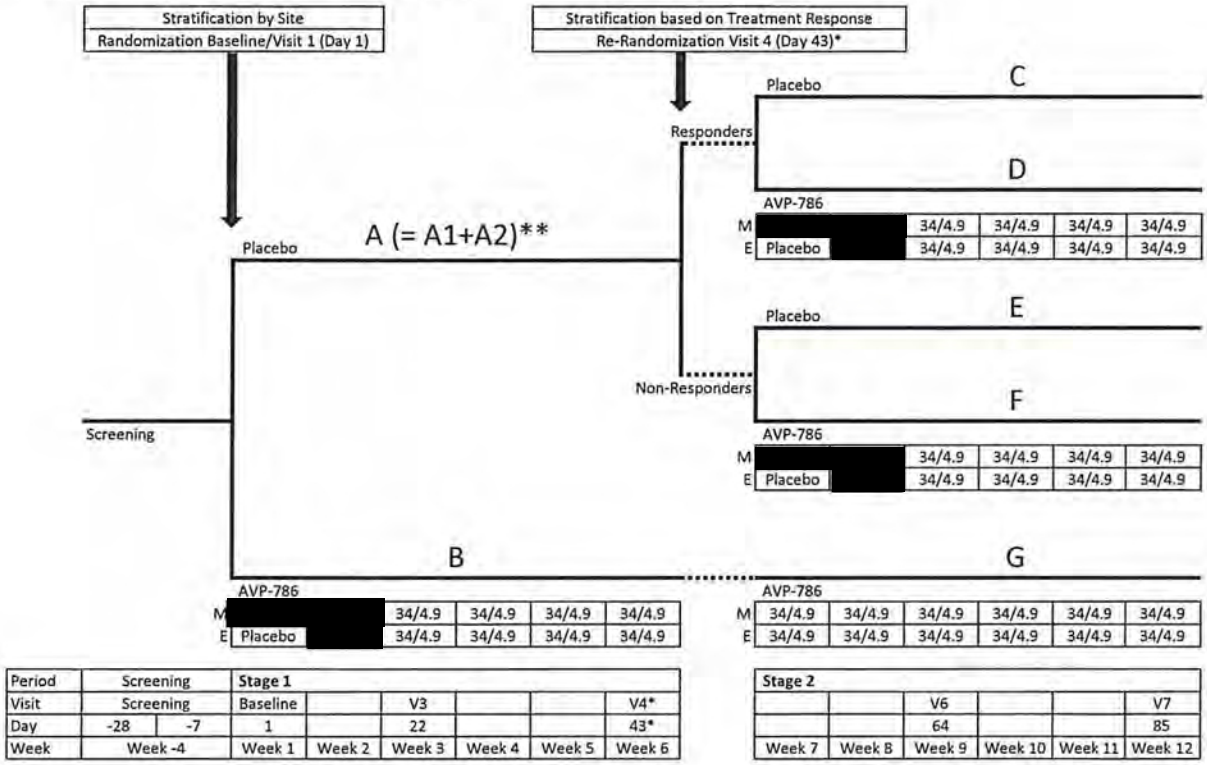
Patients who have completed Stage 1 are eligible to participate in Stage 2 of the study, which is also 6 weeks long. Stage 2 starts at Week 7 and will continue to the end of Week 12. Patients who received placebo in Stage 1 will be further stratified into 2 subgroups ("responders" and "non-responders") based on their response to the [redacted] from the Baseline total score constitutes a response. Each subgroup will be re-randomized to receive AVP-786 or placebo in a [redacted] ratio. Patients who received AVP-786 in Stage 1 will continue to receive AVP-786 for the entire 6-week duration of Stage 2.

Clinic visits will occur at Screening, Baseline/Visit 1 (Day 1), Week 2/Visit 2 (Day 15), Week 3/Visit 3 (Day 22), Week 6/Visit 4 (Day 43), Week 8/Visit 5 (Day 57), Week 9/Visit 6 (Day 64), and Week 12/Visit 7/Early Termination (Day 85). A schematic of the study along with the visit schedule and treatment segments is shown in Figure 1. The letters for each treatment segment are described below and will be used throughout the Statistical Analysis Plan (SAP) for clarity:

- A: A = A1 + A2, where A1 and A2 are the Stage 1 data for patients who are randomized to Placebo/Placebo and Placebo/AVP-786, respectively.
- B: Stage 1 data for patients randomized to AVP-786 in Stage 1
- C: Stage 2 data for patients who were placebo 'responders' in Stage 1 and re-randomized to placebo in Stage 2
- D: Stage 2 data for patients who were placebo 'responders' in Stage 1 and re-randomized to AVP-786 in Stage 2
- E: Stage 2 data for patients who were placebo 'non-responders' in Stage 1 and re-randomized to placebo in Stage 2
- F: Stage 2 data for patients who were placebo 'non-responders' in Stage 1 and re-randomized to AVP-786 in Stage 2
- G: Stage 2 AVP-786 data for patients randomized to AVP-786 in Stage 1.

Figure 1. Study Schematic (With Treatment Segments)

Protocol 15-AVP-786-202 (Schizophrenia Study)



Period	Screening		Stage 1					
Visit	Screening		Baseline		V3			V4*
Day	-28	-7	1		22			43*
Week	Week -4		Week 1	Week 2	Week 3	Week 4	Week 5	Week 6

Stage 2					
		V6			V7
		64			85
Week 7	Week 8	Week 9	Week 10	Week 11	Week 12

Study medication (active or placebo) will be administered as 1 capsule in the morning and 1 capsule in the evening approximately 12 hours apart.
M: Morning Dose (in mg d6-DM/mg Q)
E: Evening Dose (in mg d6-DM/mg Q)
*: Visit 4 (Day 43) stratification is based on treatment response criteria followed by Re-Randomization [redacted].
**: A1, A2 are the Stage 1 data for patients who are randomized to Placebo/Placebo and Placebo/AVP-786, respectively.

6.2 Study Drug Administration

Study medication will be administered orally BID throughout the study.

Patients randomized to receive active drug will start with [redacted] mg of d6-DM and [redacted] mg of Q (AVP-786 [redacted]) once a day (QD) for the first 7 days of the study in the morning and matching placebo in the evening. Starting on Day 8, these patients will receive AVP-786 [redacted] BID until their next scheduled visit and starting at Visit 2 (Day 15) their dose will be escalated to AVP-786 34/4.9 BID for the duration of the study.

Patients randomized to receive placebo in Stage 1 will take placebo BID for the entire stage.

Patients who are re-randomized to placebo will continue to receive placebo BID for the entire 6-week duration of Stage 2.

Patients who are re-randomized to AVP-786 at the beginning of Stage 2 will receive AVP-786 in the same dose escalating manner as patients in Stage 1.

[redacted]

7 STUDY ENDPOINTS

7.1 Scales and Questionnaires

7.1.1 16-Item and 4-Item Negative Symptom Assessments (NSA-16, NSA-4)

The NSA-16 total score is the primary efficacy measure used in this study, and measures the presence, severity, and range of negative symptoms associated with schizophrenia. The NSA-16 is performed at Screening, Baseline/Visit 1 (Day 1), Visit 3 (Day 22), Visit 4 (Day 43), Visit 6 (Day 64), and Visit 7/Early Termination (Day 85).

The NSA-16 consists of 16 items, and uses a 5-factor model to describe negative symptoms:

Factor Domain	NSA-16 Item Numbers	Score Range
Communication	1, 2, 3, 4	4 – 24
Emotion/Affect	5, 6, 7	3 – 18
Social Involvement	8, 9, 10	3 – 18
Motivation	11, 12, 13, 14	4 – 24
Retardation	15, 16	3 – 18
Total	1 – 16	16 – 96

Each item is rated on a 6-point Likert scale from 1 to 6, with detailed descriptions of anchors at each point on the scale. The severity scores can be interpreted as follows:

1 = The behavior being assessed is not reduced or absent as compared to a healthy young human.

2 = The behavior being assessed is minimally reduced, significance is questionable.

3 = The behavior being assessed is mildly reduced; it might only be noted as reduced by a trained rater, but he/she notes a definite reduction.

4 = The behavior being assessed is moderately reduced; the reduction should be obvious to an untrained rater.

5 = The behavior being rated is markedly reduced; this behavior is easily observable and definitely interferes with the subject's functioning.

6 = The behavior being assessed is severely reduced or entirely absent; it is glaring and markedly interferes with functioning.

Furthermore, each item admits a score of 9 if the item is not ratable. The NSA-16 total score is defined as the sum of all 16 items excluding any items receiving a score of 9. If more than 3 items are not ratable, the total score cannot be calculated and will be regarded as missing; otherwise, the total score will be calculated as 16 times the arithmetic mean of the scores of all ratable items, rounded to the nearest tenth. Therefore, the minimum score is 16 and the maximum score is 96, with a higher score indicating greater clinical severity of symptoms.

Additionally, two global ratings are included: a global negative symptoms rating, which is scored on the basis of the overall impression of negative symptoms in the patient; and a global level of functioning, which represents the clinician's overall assessment of the patient's level of functioning, relative to a healthy young adult. Both are scored on the following scale:

1 = No evidence of this symptom/impairment

2 = Minimal evidence of this symptom/impairment

3 = Mild evidence of this symptom/impairment

4 = Moderate evidence of this symptom/impairment, apparent to the casual observer

5 = Marked evidence of this symptom/impairment, readily apparent to the casual observer

6 = Severe symptom/impairment, not only obvious but has marked impact on functioning

7 = Extremely severe symptom/impairment that is incapacitating for the subject.

The NSA-4 is a subset of the NSA-16, using the following items: (1) restricted speech quantity; (2) emotion: reduced range; (3) reduced social drive; (4) reduced interests. Therefore, the NSA-4 total score ranges from 4 to 24.

In addition to the NSA-16 total score as the primary efficacy endpoint, the NSA-16 individual factor domain scores, global negative symptoms/functioning rating, individual item scores, and NSA-4 total score are secondary efficacy endpoints. For each factor domain, scores will be calculated only if all items in that domain are rated; otherwise, it will be regarded as missing.

7.1.2 Positive and Negative Syndrome Scale (PANSS)

The PANSS is a 30-item clinical scale to assess positive and negative symptoms of schizophrenia, and is performed at Screening, Baseline/Visit 1 (Day 1), Visit 3 (Day 22), Visit 4 (Day 43), Visit 6 (Day 64), and Visit 7/Early Termination (Day 85).

Each item is rated on the following scale:

1 = **Absent**.

2 = **Minimal**: questionable or subtle or suspected pathology, or may allude to the extreme end of the normal range.

3 = **Mild**: presence is clearly established but not pronounced, and interferes little in day-to-day functioning.

4 = **Moderate**: characterizes a symptom which, though representing a serious problem, either occurs only occasionally or intrudes on daily life only to a moderate extent.

5 = **Moderate to Severe**: indicates marked manifestations that distinct impact on one's functioning but are not all-consuming and usually can be contained at will.

6 = **Severe**: represents gross pathology that is present very frequently, proves highly disruptive to one's life, and often calls for direct supervision.

7 = **Extremely Severe**: refers to the most serious level of psychopathology, whereby the manifestations drastically interfere in most or all major life functions, typically necessitating close supervision and assistance in many areas.

Therefore, the PANSS total score range is 30 to 210. Items can be grouped into the following subscales with the corresponding score ranges:

Subscale	Item Numbers	Score Range
Negative Subscale	N1 – N7	7 – 49
Positive Subscale	P1 – P7	7 – 49
General Psychopathology Subscale	G1 – G16	16 – 112
Prosocial Factors	G16, N2, N4, N7, P3, P6	6 – 42
Marder Negative Factors	N1 – N4, N6, G7, G16	7 – 49
Excitement Component	P4, P7, G4, G8, G14	5 – 35

The PANSS total score (N1 – N7, P1 – P7, G1 – G16), and the PANSS subscales (positive, negative, general psychopathology, prosocial factors, Marder negative factors, excitement component) are secondary efficacy endpoints.

Additionally, the proportion of patients with a reduction of 20% or greater in the PANSS total score will be calculated, and is a secondary efficacy endpoint.

7.1.3 Clinical Global Impression (CGI) Scales

The CGI was developed to provide a brief, stand-alone assessment of the clinician's view of the patient's global functioning prior to and after initiating a study medication. The CGI provides an overall clinician-determined summary measure that takes into account all available information, including knowledge of the patient's history, psychosocial circumstances, symptoms, behavior, and the impact of the symptoms on the patient's ability to function. The following CGI scores are secondary efficacy endpoints, and each is rated on a scale of 1 to 7, with a score of 0 indicating no assessment was performed.

7.1.3.1 Clinical Global Impression – Severity (CGI-S)

The CGI-S is a 7-point scale that requires the clinician to rate the severity of the patient's illness at the time of assessment, relative to the clinician's past experience with patients with the same diagnosis. Ratings range on the following scale:

0 = Not assessed

1 = Normal, not at all ill

2 = Borderline ill

3 = Mildly ill

4 = Moderately ill

5 = Markedly ill

6 = Severely ill

7 = Among the most extremely ill patients.

The CGI-S will be performed at Baseline/Visit 1 (Day 1), Visit 4 (Day 43), and Visit 7/Early Termination (Day 85). A score of 0 will be treated as missing data.

7.1.3.2 Clinical Global Impression – Change (CGI-C)

The CGI-C is a 7-point scale that requires the clinician to rate the change of the patient's condition at the time of assessment, relative to the clinician's past experience with the patient's condition at admission. Ratings range from:

0 = Not assessed

1 = Very much improved

2 = Much improved

3 = Minimally improved

4 = No Change

5 = Minimally worse

6 = Much worse

7 = Very much worse.

The CGI-C will be performed at Visit 4 (Day 43) to reflect changes from Baseline/Visit 1 (Day 1); and Visit 7/Early Termination (Day 85) to reflect changes from both Visit 4 (Day 43) and from Baseline/Visit 1 (Day 1). A score of 0 will be treated as missing data.

7.1.3.3 Patient Global Impression – Change (PGIC)

The PGIC is a 7-point patient-rated scale used to assess treatment response. Ratings range from:

0 = Not assessed

1 = Very much improved

2 = Much improved

3 = Minimally improved

4 = No Change

5 = Minimally worse

6 = Much worse

7 = Very much worse.

The PGIC will be performed at Visit 4 (Day 43) and Visit 7/Early Termination (Day 85), in both instances reflecting changes from Baseline/Visit 1 (Day 1). A score of 0 will be treated as missing data.

7.1.4 Calgary Depression Scale for Schizophrenia (CDSS)

The CDSS is a 9-item scale derived from the Hamilton Depression Scale (Ham-D), that is designed to assess depression specifically in patients with schizophrenia. Each item on the scale is scored as: 0 (absent); 1 (mild); 2 (moderate); 3 (severe). The CDSS score is obtained by adding each of the item scores; therefore ranges from 0 to 27. The CDSS will be performed at Screening, Baseline/Visit 1 (Day 1), Visit 3 (Day 22), Visit 4 (Day 43), Visit 6 (Day 64), and Visit 7/Early Termination (Day 85). The CDSS total score is a secondary efficacy endpoint.

7.1.5 Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) Consensus Cognitive Battery (MCCB)

The MCCB assesses cognitive change in trials of cognitive-enhancing agents in schizophrenia, and includes 10 tests that measure 7 cognitive domains:

1. Speed of processing
2. Attention/vigilance
3. Working memory
4. Verbal learning
5. Visual learning
6. Reasoning and problem solving
7. Social cognition

The MCCB evaluation will be performed at Screening, Baseline/Visit 1 (Day 1), Visit 4 (Day 43), and Visit 7/Early Termination (Day 85). Alternate versions of the battery will be used at different visits to decrease the learning confound.

The composite score (*T*-score) is the secondary efficacy endpoint, with higher values indicating better cognitive performance.

7.1.6 Effort Expenditure for Reward Task (EEfRT)

The EEfRT is a multi-trial computerized task in which participants are given an opportunity on each trial to choose between 2 tasks that differ in difficulty level and are associated with varying levels of monetary reward. This task examines willingness to expend physical effort (button pressing) in pursuit of varying amounts of monetary reward.

The ratio of hard task choices to the total number of trials, as well as the difference in proportion of hard-task choices for high effort vs. low effort trials will be the outcome measures for negative symptoms, and are secondary efficacy endpoints. The EEfRT evaluation will be performed at Baseline/Visit 1 (Day 1), Visit 4 (Day 43), and Visit 7/Early Termination (Day 85).

[redacted]

[redacted]

7.1.8 Mini International Neuropsychiatric Interview (M.I.N.I.) Version 6.0.0

The M.I.N.I. is an interview developed by psychiatrists and clinicians for psychiatric disorders and has an administration time of approximately 15 minutes.

The interview is divided into 15 categories, which are listed below:

- Major depressive episode
- Suicidality
- Manic and hypomanic episodes
- Panic disorder
- Agoraphobia
- Social phobia
- Obsessive-compulsive disorder
- Posttraumatic stress disorder
- Alcohol dependence/abuse
- Substance dependence/abuse (non-alcohol)
- Psychotic disorders and mood disorder with psychotic features
- Anorexia nervosa
- Bulimia nervosa
- Generalized anxiety disorder
- Antisocial personality disorder

Based on the answers, each category will be classified into a timeframe (e.g., current, past, etc.). Missing values will not be imputed. The M.I.N.I. is performed at Screening for the purpose of assessing whether the patient meets diagnostic criteria for schizophrenia residual type. Results will be presented in the data listings.

7.1.9 Columbia Suicide Severity Rating Scale (C-SSRS)

The C-SSRS is a low-burden measure of the spectrum of suicidal ideation and behavior and consists of a clinical interview, providing a summary of both suicidal ideation and behavior that will be used during treatment to monitor for clinical worsening.

The C-SSRS evaluation is a safety endpoint and will be performed at Screening, Baseline/Visit 1 (Day 1), Visit 3 (Day 22), Visit 4 (Day 43), Visit 6 (Day 64), and Visit 7/Early Termination (Day 85).

7.1.10 Simpson Angus Scale for Extrapyramidal Symptoms (SAS)

The SAS is a 10-item evaluation used to assess pseudoparkinsonism. Each item is graded on a 5-point severity scale; thus the total score ranges from 0 to 40, with higher scores indicating greater severity of extrapyramidal symptoms.

The SAS evaluation is a safety endpoint and will be performed at Baseline/Visit 1 (Day 1), Visit 4 (Day 43), and Visit 7/Early Termination (Day 85).

7.1.11 Barnes Akathisia Scale (BAS)

The BAS is an assessment of the presence and frequency of akathisia (a movement disorder characterized by restlessness and a compulsion to move), as well as its global severity. The objective item of the scale measures the clinician's assessment of the patient's restlessness. The subjective portion of the scale measures both the patient's awareness of restlessness and the distress arising from such awareness. Each item is rated on a 4-point scale from 0 to 3, and the total score is the sum of the three item scores, thus yielding a total score ranging from 0 to 9. The Global Clinical Assessment of Akathisia is rated on a 5-point scale from 0 to 4. Higher scores indicate greater severity of symptoms.

The BAS evaluation is a safety endpoint and will be performed at Baseline/Visit 1 (Day 1), Visit 4 (Day 43), and Visit 7/Early Termination (Day 85).

7.1.12 Abnormal Involuntary Movements Scale (AIMS)

The AIMS is a 12-item scale to assess dyskinesia. Items related to the severity of orofacial, extremity, and trunk movements, global judgment about incapacitation, and patient awareness are rated using a 5-point scale (0 = None, 1 = Minimal, 2 = Mild, 3 = Moderate, 4 = Severe). Two items related to dental status are scored using "yes/no" responses.

The BAS evaluation will be performed at Baseline/Visit 1 (Day 1), Visit 4 (Day 43), and Visit 7/Early Termination (Day 85).

7.2 Efficacy Endpoints

The primary efficacy endpoint is the change from baseline (CFBL) in the NSA-16 total score.

Secondary endpoints are the CFBL (or actual scores for the CGI-C and PGIC) in the scores of the assessments listed below (and described above):

- PANSS total score, PANSS subscales (positive, negative, general psychopathology, Marder negative factors, excitement component, and prosocial factors)
- NSA-16 factor domains, global symptom/functioning score, individual items, and NSA-4
- MCCB composite score
- CGI-S, CGI-C, and PGIC scores
- CDSS total score
- EEfRT score

In addition, the following secondary endpoint will be calculated:

- Proportion of patients with a reduction of 20% or greater in the PANSS total score

7.3 Safety Endpoints

Safety and tolerability measurements will include: adverse events (AEs), vital signs, weight, urine pregnancy tests, clinical laboratory assessments, resting 12-lead ECGs, Columbia Suicide Severity Rating Scale (C-SSRS), Abnormal Involuntary Movements Scale (AIMS), Barnes Akathisia Scale (BAS), and Simpson Angus Scale for Extrapyramidal Symptoms (SAS).

7.4 Pharmacokinetic, [redacted] Parameters

Plasma concentrations of d6-DM, its metabolites, Q, and antipsychotic medications will be assessed at Baseline/Visit 1 (Day 1, post-dose), Visit 4 (Day 43), and Visit 7/Early Termination (Day 85). Corresponding Cmax and AUC values will be estimated.

[redacted]

8 ANALYSIS POPULATIONS

A summary table containing the number of patients in each of the populations described below will be provided.

8.1 Modified Intent-to-Treat (mITT) Population

The mITT Population will be used to analyze treatment efficacy according to the SPCD design. Due to the study design, the patients included in the population are determined separately for Stage 1 and Stage 2, although the Stage 2 group will be a subset of the Stage 1 group. The population is defined below:

- Stage 1: All patients randomized in Stage 1 who had at least one post-baseline NSA-16 total score efficacy assessment in Stage 1.
- Stage 2: All patients who were re-randomized into Stage 2 (regardless of Stage 1 treatment group) and had at least one NSA-16 total score efficacy assessment in Stage 2 (after Week 6/Visit 4)

When implementing changes in Protocol Amendment 4, the IRT vendor who carried out the IVRS found that 14 Stage 1 Placebo patients were re-randomized for Stage 2 at Week 3 instead of Week 6. These 14 patients with randomization error are excluded from the mITT analysis population.

[redacted]

8.3 Safety Population

The safety population will be used for all safety analyses. It includes all patients who received at least one dose of study medication.

[redacted]

9 OVERALL STATISTICAL CONSIDERATIONS

9.1 Definition of Baseline

Baseline is generally defined as the last assessment prior to the first dose of study drug, but will vary depending on the analysis, population, treatment group, and parameter. See below for specific considerations:

- For the primary efficacy analysis and secondary efficacy analyses on the mITT or PP populations, Stage 1 Baseline is the last assessment prior to first dose (or prior to randomization date, for patients who did not take study medication), which is typically Day 1. Stage 2 Baseline is the Visit 4 (Week 6) assessment (re-randomization visit). Note that Stage 2 Baseline only applies to patients who were randomized to placebo in Stage 1 and re-randomized in Stage 2.

[redacted]

- Baseline for descriptive safety analyses are described below:
  - For patients receiving AVP-786 or placebo for the entire study duration, baseline is the last non-missing assessment prior to the first dose of study drug.
  - For patients randomized to placebo then re-randomized to AVP-786, Stage 1 baseline is the last non-missing assessment to first dose of study drug and Stage 2 baseline is the last non-missing assessment occurring after Day 1 and prior to re-randomization at Week 6.

9.2 Definition of End of Treatment

End of treatment is defined as the last value for a given patient, whenever it occurred (including Stage 1 values). This terminology will not be used if referring to analysis that is done by study stage.

9.3 Change from Baseline

Change from baseline (CFBL) will always be calculated as (post-baseline – baseline). It will be calculated for patients with both a baseline and post-baseline value as applicable.

If a baseline value has not been recorded for a parameter, then CFBL will not be calculated for that parameter, and the patient will be excluded from CFBL analysis.

Percent CFBL, when needed, is the CFBL divided by the baseline value multiplied by 100%. Patients with a value of 0 at baseline cannot have percent CFBL calculated.

9.4 Visit Windows

Data at scheduled visits will be assigned to analysis visits as defined in the Visit Window tables below, to ensure that all visits have the potential to be included in the summaries.

Visit windows will be used to classify unscheduled and early termination visits. If 2 or more visits occur within the same analysis window, the value closest to the target day will be summarized. If the assessments are the same distance from the target day, the latest one will be used.

Separate visit window tables are provided based on how frequently the assessments were done. Each table is also divided into patients who have advanced to Stage 2 and those that have not.

Table 2. Visit Windows for Assessments Done at Baseline and Weeks 2, 3, 6, 8, 9, 12

Stage/Visit	Target Day	Study Days (Patients Advancing to Stage 2)		Study Days (Patients Not Advancing to Stage 2)	
Stage 1		**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**
Week 2 (Visit 2)	15	2	18	2	18
Week 3 (Visit 3)	22	19	ReR – 2	19	32
Week 6 (Visit 4)	43	ReR – 1	ReR	33	day of last visit
Stage 2					
Stage 2 Baseline		ReR – 1	ReR	N/A	N/A
Week 8 (Visit 5)	57	ReR + 1	60	N/A	N/A
Week 9 (Visit 6)	64	61	74	N/A	N/A
Week 12 (Visit 7)	85	75	day of last visit	N/A	N/A

N/A = Not applicable. ReR = re-randomization visit day.

Table 3. Visit Windows for Assessments Done at Baseline and Weeks 3, 6, 9, and 12

Stage/Visit	Target Day	Study Days (Patients Advancing to Stage 2)		Study Days (Patients Not Advancing to Stage 2)	
Stage 1		**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**
Week 3 (Visit 3)	22	2	ReR – 2	2	32
Week 6 (Visit 4)	43	ReR – 1	ReR	33	day of last visit
Stage 2					
Stage 2 Baseline		ReR – 1	ReR	N/A	N/A
Week 9 (Visit 6)	64	ReR + 1	74	N/A	N/A
Week 12 (Visit 7)	85	75	day of last visit	N/A	N/A

N/A = Not applicable. ReR = re-randomization visit day.

Table 4. Visit Windows for Assessments Done at Baseline, Week 6, and Week 12 Only

Stage/Visit	Target Day	Study Days (Patients Advancing to Stage 2)		Study Days (Patients Not Advancing to Stage 2)	
Stage 1		**Lower Bound**	**Upper Bound**	**Lower Bound**	**Upper Bound**
Week 6 (Visit 4)	43	2	ReR	2	day of last visit
Stage 2					
Stage 2 Baseline		ReR – 1	ReR	N/A	N/A
Week 12 (Visit 7)	85	ReR + 1	day of last visit	N/A	N/A

N/A = Not applicable. ReR = re-randomization visit day.

Table 2 applies to Vital Signs, Weight, C-SSRS; Table 3 applies to Pregnancy test, NSA-16, PANSS, CDSS; and Table 4 applies to ECG, Chemistry/Hematology/Urinalysis, SGA plasma concentrations, Study drug plasma concentrations, MCCB, CGI-S, CGI-C, PGIC, EEfRT, AIMS, BAS, SAS, [redacted].

9.5 Handling of Missing Data

Missing data will be handled differently depending on the parameter and analysis. Note that analyses done on 'observed cases' will not follow any imputation rules below. See below for considerations:

- Missing baseline values will not be imputed in any situation.
- For SPCD [redacted] efficacy analyses (excluding observed cases analyses), missing data will be imputed by Last Observation Carried Forward (LOCF) within a study stage. For example, missing assessments from Week 12 (Visit 7) can be imputed with the value from Week 9 (Visit 6) but not from Week 3 (Visit 2). Note that if Week 3 or Week 6 data are missing for an mITT patient, then baseline will be carried forward. Likewise, Stage 2 baseline will be carried forward if Week 9 data are missing for patients with Stage 2 data. No imputation will be done for patients without data in Stage 2.

[redacted]

- For any safety assessments of patients without a Week 12 safety endpoint assessment, the last non-missing post-baseline observation will be used as the end of treatment value.
- Missing post-baseline values for by-visit safety data will be summarized using the Visit Windows from Section 9.4. If a value is not available within a given window, no imputation will be done.
- Missing data for AE relationship will be imputed as "Related."
- Rules for partial dates are described in Appendix 3. These will apply to concomitant medications when applicable.

9.6 Treatment Misallocations

Efficacy data will be summarized "as randomized," meaning the patient will be summarized under the treatment they were randomized to, regardless of what treatment was actually received.

Safety data will be summarized "as treated," meaning the patient will be summarized based on the treatment that was actually received.

9.7 Summary Statistics

Quantitative displays will be summarized using descriptive statistics. The mean, standard deviation, median, minimum, and maximum will be provided. Decimal precision will be based on the mean value. The median contains the same number of decimal places as the mean, the standard deviation contains one more decimal place, and the minimum and maximum contain one less decimal place. The mean will typically have one more decimal place than the raw values but if necessary, decimal precision for the mean will be provided in the table specifications.

9.8 Data Listings

Unless otherwise specified, data listings will be provided on observed values and calculated PK parameters. Efficacy listings for the primary and secondary endpoints will include analysis flags and/or imputations to show which values were used in the summary tables. [redacted] Listings will be provided to serve as support for all summary tables or figures.

9.9 Graphical Displays

Supporting figures may be used for some efficacy, PK or safety analyses in addition to the summary tables. Details regarding the content, layout, and structure of figures will be provided in the table specifications.

9.10 Hypothesis Testing

All statistical testing will be 2-sided and performed at the 0.05 significance level.

9.11 Multiplicity of Secondary Endpoints

No adjustments will be made for testing multiple secondary outcome measures. Since it is possible that some significant results could occur by chance alone, undue consideration will not be given to isolated significant differences; rather, interpretations will be made based on patterns of significant differences and their consistency with the primary endpoint analysis.

10 EFFICACY STATISTICAL ANALYSIS METHODS

The study schematic in Figure 1 can be referenced to identify treatment groups that are described in the subsequent efficacy sections.

10.1 Primary Efficacy: SPCD MMRM (Mixed-Model Repeated Measures)

The primary efficacy endpoint (CFBL in NSA-16 total score), will be analyzed using a weighted test statistic with the treatment effects in each stage estimated by likelihood-based mixed-model repeated measures (MMRM) analysis on the observed data (Chen et al, 2011). This model will be run on the mITT population, which will include observed data from

Stage 1 mITT patients and from the Stage 1 Placebo Non-Responder Subset in Stage 2. Visit windows will be applied for unscheduled or early termination visits. The null hypothesis to be tested is that there will be no difference in the CFBL in NSA-16 total score between AVP-786 and placebo in Stage 1 and Stage 2.

Per the SPCD, separate MMRMs will be run using data from Stage 1 and Stage 2. Stage 1 data (CFBL to Week 6) will contain data for all mITT Stage 1 patients and will compare study segments A and B from the study schematic. Stage 2 data (change from Week 6 to end of treatment) will be used for patients in the Stage 1 Placebo Non-responders of the Stage 2 mITT population and will use study segments E and F. Stage 1 treatment effect will be estimated by Week 6 treatment difference and the Stage 2 treatment effect will be estimated by the Week 12 treatment difference. The model will include terms for treatment, visit, treatment-by-visit interaction, baseline NSA-16 value and baseline-by-visit interaction. An unstructured covariance matrix (UN) will be used. However, if there are convergence problems, analyses with a first-order autoregressive covariance structure (AR[1]) or the compound symmetry covariance structure (CS) will be used. The SAS model statement is provided below:

[redacted]

The parameter estimates at Week 6 (Stage 1) and Week 12 (Stage 2) will be used in each stage to come up with a combined weighted test statistic, Z_{MMRM} that is described in Chen, et al (2011). The formula for the test statistic is given below:

$$Z_{\text{MMRM}} = \frac{w\,\hat{\theta}^{(1)} + (1-w)\,\hat{\theta}^{(2)}}{\sqrt{w^2\,\widehat{\text{Var}}(\hat{\theta}^{(1)}) + (1-w)^2\,\widehat{\text{Var}}(\hat{\theta}^{(2)})}}$$

where a weight of $w =$ [redacted] is used for Stage 1, which means a weight of $1 - w =$ [redacted] is used for Stage 2. The estimated treatment effects $\hat{\theta}^{(1)}, \hat{\theta}^{(2)}$ and squared standard errors $\widehat{\text{Var}}(\hat{\theta}^{(1)})$, $\widehat{\text{Var}}(\hat{\theta}^{(2)})$ of each treatment stage are obtained directly from the model output. The test statistic will be used to come up with a 2-sided p-value for the hypothesis test, shown below:

```
p_value=2*(1-probnorm(abs(Z_MMRM)));
```

[redacted]

[redacted]

[redacted]

10.3 Secondary Efficacy Analysis

The following secondary endpoints will be analyzed in this study, using the methodologies described in the following subsections:

- PANSS total score
- PANSS subscales (positive, negative, general psychopathology, Marder negative factors, excitement component, pro-social factors)
- NSA-16 factor domains, global symptom/functioning score, individual items, global negative symptoms rating, and NSA-4 total score
- Proportion of patients with a reduction of 20% or greater in the PANSS total score
- MCCB composite score
- CGI-S, CGI-C, and PGIC scores
- CDSS score
- EEfRT score [redacted]

10.3.1 Secondary SPCD MMRM and SPCD OLS ANCOVA

The CFBL for all quantitative secondary endpoints listed above (except the CGI-C and the PGIC) will be analyzed using the OLS ANCOVA method as described in Section 10.2.1. The mITT population (Stage 1 plus the Stage 1 Placebo Non-Responders for Stage 2) will be used for these analyses. LOCF within a study stage will be used for missing values at Week 6 or Week 12. In addition, the secondary endpoints listed above measured by NSA-16, PANSS and CDSS (which were assessed at Baseline, Week 3, 6, 9 and 12) will also be analyzed using the SPCD MMRM method as described in Section 10.1.

For the CGI-C and the PGIC, an ANCOVA model with treatment as a factor and baseline NSA-16 total score as a covariate will be run. Note that baseline NSA-16 total score depends on the study stage and is defined in Section 9.1. All model considerations will be the same as the primary parameters as described in Section 10.1.

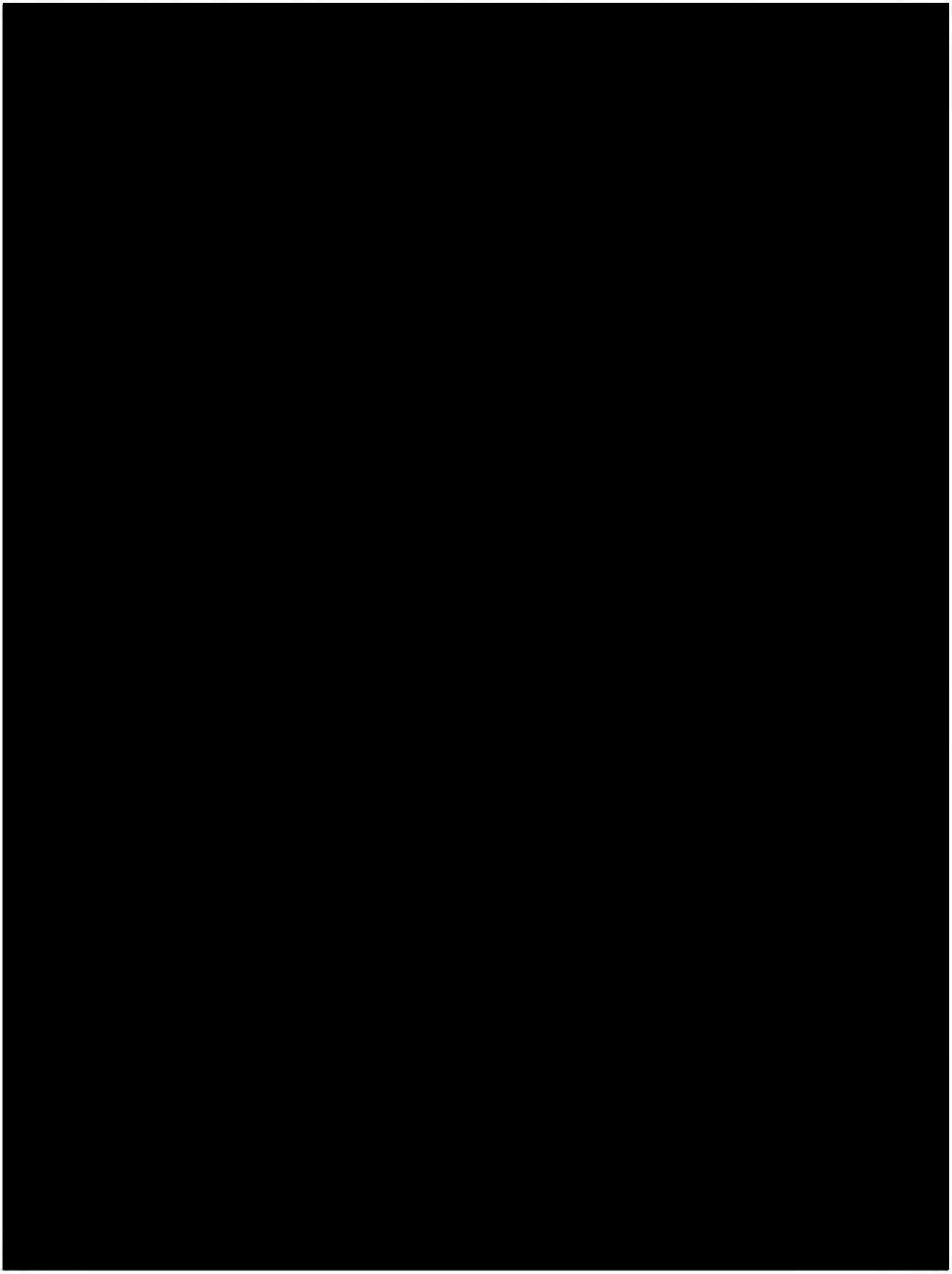
10.3.2 PANSS Response Analysis

The number and percent of patients who have favorable treatment responses according to the PANSS will be summarized at Week 6 and Week 12 using the mITT population (Stage 1 plus the Stage 1 Placebo Non-responders in Stage 2). The following category will be used to classify patients:

- Response: patients with at least a 20% reduction in the PANSS total score.

The number and percent of responders will be provided by stage and treatment group. LOCF will be used for patients missing data. Overall Stage 1 and 2 treatment differences will be tested via SPCD 1 degree of freedom (DOF) score test assuming Stage 2 and 1 treatment effect ratio $\rho=1$ (A. Ivanova, et al, 2011). In addition, the treatment effect will also be tested at each visit by Chi-square test or Fisher's Exact (if the expected cell counts are < 5).

[redacted]



[redacted]

[redacted]

[redacted]

11 SAFETY STATISTICAL ANALYSIS METHODS

Safety will be assessed through the analysis of AEs, physical and neurological examinations, vital signs, clinical laboratory assessments, ECGs, the C-SSRS, the SAS, the BAS, and the AIMS. All Safety analyses will be completed on the Safety Population.

Unless otherwise specified, safety analyses that include summaries of number and percent (e.g., AEs) will be displayed using the following treatment groups:

- Placebo/Placebo: patients receiving Placebo/Placebo during the study (study segment A1, C and E). Note that patients randomized to Placebo/AVP-786 but dropped out in Stage 1 (part of A2) are not included in this population. Instead, these will be summarized under the 'All Placebo' treatment group.
- AVP-786/AVP-786: patients receiving AVP-786 for the entire duration of the study (B and G).
- Placebo/AVP-786: patients who switched from placebo to AVP-786. This group will be further divided into data that occurred while on placebo (A) and data that occurred while on AVP-786 (D and F).
- All Placebo: This includes data from the stages when patients received placebo (A, C, and E).
- All AVP-786: This includes data from the stages when patients received AVP-786 (B, D, F, and G).

[redacted]

All Placebo and All AVP-786 treatment groups summarize the safety information for their corresponding treatment group under 6 week or 12 week treatment exposure in either Stage 1, Stage 2, or both.

For quantitative summaries (e.g., ECGs, labs), the All Placebo and All AVP-786 groups will not be included. As noted in Section 9.1, baseline for safety assessments (excluding shift tables) is the last non-missing value prior to taking study drug at baseline. For those switching from placebo to AVP-786, baseline is the last non-missing value prior to taking AVP-786.

11.1 Adverse Events

Adverse event tables (except the AE overview table) will only include summaries of TEAEs. Treatment-emergent adverse events are defined as AEs which first occur, or worsen, after the first dose of study medication and within 30 days after the permanent discontinuation of the study medication (i.e., first dose date $\leq$ AE start date $\leq$ last dose date + 30 days).

An overview table containing the number and percent of the following will be included:

- Number of total AEs, TEAEs, and deaths
- Incidence of patients with at least one TEAE, drug-related TEAE, serious AE (SAE), drug-related SAE, non-serious TEAEs, and death
- Incidence of patients who discontinued due to TEAE, drug-related TEAE, SAE, and drug-related SAE
- Incidence of deaths and deaths due to drug-related TEAE

Treatment-emergent adverse events will be summarized by SOC and PT, descending frequency of PT, by maximum severity, by dose at onset, and by age group (< 45, ≥ 45). Summaries of PTs will also be done for those occurring in at least 5% of patients in any treatment or study stage. In addition, summaries of patients taking beta blockers at baseline will be summarized as described in Section 10.4.6.

Adverse events leading to discontinuation will be summarized by SOC and PT. Drug-related AEs will be summarized by SOC and PT and in descending frequency of PT.

Time to onset for common TEAEs (as defined below) will be summarized descriptively for each of these events. In addition, summary stats for the duration and percentage of total study days will be provided for each AE. The number and percent of patients with recurrences will also be given.

Serious adverse events will be summarized by SOC and PT and will include a summary of drug-related events.

Below are the rules to follow for AE summaries:

- For patients who took AVP-786/AVP-786 or Placebo/Placebo: If a patient has multiple AEs within the same SOC or PT, the patient will only be counted once within a level of Medical Dictionary for Regulatory Activities (MedDRA).
- For patients who switched from placebo to AVP-786: If the same AE starts in both study stages, it should be counted under both placebo and AVP-786.
- A drug-related AE is defined as an AE with an assigned relationship of "possibly related," "related," or missing.
- When assessing severity, if a patient has 2 TEAEs within a study stage, the TEAE with the worst severity will be chosen. AEs with missing severity will be excluded from summaries of AE by severity.
- For dose at onset summaries, patients are counted only once within a dose, but can be counted in multiple dose groups if the same AE occurred while on different doses.
- A common TEAE is defined as a TEAE an incidence of $\geq 3\%$ in the All AVP-786 treatment group AND ≥ 2 times the incidence of the All Placebo treatment group.
- Time to onset will be calculated in days as (AE start date – first dose date). For patients who switched from placebo to AVP-786, the following rules will apply:
  - If an AE occurs when the patient is on AVP-786, time to onset is calculated as (AE start date – re-randomization date). This includes the time on AVP-786 only.
  - If the same AE occurs while the patient is on placebo and then again when the patient is on AVP-786, the AE will be counted in both groups, and first dose date will be defined as the first dose of placebo or first dose of AVP-786, depending on which treatment the patient was on when the AE occurred.
- Duration of AE is generally defined as (AE end date – AE start date + 1). Duration will be calculated for placebo and AVP-786 separately. Below are some additional considerations for AE duration:
  - If the patient has an AE on placebo that has not ended when the patient takes the last dose of placebo (either due to switching to AVP-786 or end of study), AE end date is defined as the last dose of placebo.
  - If the patient has an AE on AVP-786 that has not ended when the patient ends the study, AE end date is defined as the last dose of study drug.
  - If the same AE occurs more than once while the patient is on the same treatment, duration will be the sum of the individual AE durations.
- For a given patient, percentage of total study days is defined as total duration (as defined above) divided by (last dose date – first dose date + 1) × 100%.
- Recurrence is defined as a new report of the same TEAE with a new AE start date within a given treatment. An AE that occurs for a patient on placebo and then again on AVP-786 will not be considered a recurrent event.

AEs will be coded using MedDRA version 18.1.

11.2 Clinical Laboratory Assessments

Clinical labs will be reported for hematology, chemistry and urinalysis. Labs are collected at Screening, Week 6 and Week 12, and will be summarized by visit, by stage and for the Placebo/Placebo and AVP-786/AVP-786 groups as described below.

The following parameters will be summarized descriptively through CFBL and percent CFBL for numeric values. If a parameter is categorical, it will be listed only.

- Chemistry: calcium, magnesium, phosphorus, glucose, sodium, potassium, chloride, carbon dioxide, blood urea nitrogen (BUN), serum creatinine, uric acid, albumin, total bilirubin, alkaline phosphatase, lactate dehydrogenase (LDH), aspartate aminotransferase/serum glutamic oxaloacetic transaminase (AST/SGOT), alanine aminotransferase/serum glutamic pyruvic transaminase (ALT/SGPT), creatine kinase (CK), gamma-glutamyl transferase (GGT), triglycerides, total protein, total cholesterol, and glycosylated hemoglobin (HbA1c, at Screening and Visit 7).
- Hematology: red blood cell (RBC) count, hemoglobin, hematocrit, white blood cell (WBC) count, neutrophils, bands, lymphocytes, monocytes, eosinophils, basophils, platelet count, and morphology
- Urinalysis: pH, specific gravity, protein, glucose, ketones, bilirubin, urobilinogen, nitrites, leucocytes, and blood

Out-of-range values will be assessed through shift tables. Each lab value will be assessed as low, normal or high based on the normal ranges provided by the central lab. Frequencies of each combination of shifts will be provided by treatment group.

Shift tables will be created by stage and for the Stage 1 Placebo, AVP-786, Stage 2 Placebo (Stage 1 Placebo re-randomized to Placebo) and Stage 2 AVP-786 (Stage 1 Placebo re-randomized to AVP-786). In Stage 1, AVP-786 will be compared to placebo and in Stage 2, the placebo patients who were re-randomized to either AVP-786 or placebo will be compared. Baseline in all shift tables is the last assessment prior to first dose in each stage. Last post-baseline visit for Stage 1 is the last assessment prior to or on the Week 6 visit, while the last post-baseline visit in Stage 2 and the Placebo/Placebo, AVP-786/AVP-786 groups is the last assessment, whenever it occurred. [redacted]

Potentially clinically significant (PCS) tables will also be used to summarize out-of-range values. The PCS values are found in Table 5. The number and percent of patients meeting the criteria below will be summarized by treatment group. Summaries will be given for any time post-baseline. The denominator for the percentages will be the number of patients who had a post-baseline assessment for each parameter.

Table 5. Lab PCS Criteria

Laboratory Parameter	Unit	Low PCS Criteria	High PCS Criteria	Laboratory Parameter	Unit	Low PCS Criteria	High PCS Criteria
Chemistry							
Albumin	g/L	≤ 26	≥ 60	GGT	U/L	None	≥ 60
Alkaline Phosphatase	U/L	None	≥ 3×ULN	Glucose	mmol/L	≤ 2.775	≥ 11.1
ALT (SGPT)	U/L	None	≥ 3×ULN	LDH	U/L	None	≥ 3×ULN
AST (SGOT)	U/L	None	≥ 3×ULN	Magnesium	mmol/L	< 0.37	> 1.23
Bilirubin	µmol/L	None	≥ 1.5×ULN	Phosphate	mmol/L	≤ 0.4522	> 3.88
BUN	mmol/L	None	≥ 10.71	Potassium	mmol/L	≤ 3.0	≥ 5.5
Calcium	mmol/L	≤ 1.75	≥ 3.0	Protein	g/L	≤ 50	≥ 100
Carbon Dioxide	mmol/L	≤ 9	> 40	Sodium	mmol/L	≤ 130	≥ 155
Chloride	mmol/L	≤ 85	≥ 120	Triglycerides	mmol/L	None	> 3.39
Cholesterol	mmol/L	None	≥ 7.77	Uric acid (Male)	µmol/L	None	≥ 624.54
Creatine Kinase	U/L	None	≥ 3×ULN	Uric acid (Female)	µmol/L	None	≥ 505.58
Creatinine	µmol/L	None	> 132.6	HbA1c	%	None	6.5%
Hematology							
Hemoglobin	g/L	< 100	> 180	Monocytes	$\times 10^9$/L	None	> 1
Hematocrit	proportion of 1.0	< 0.3	> 0.5	Monocytes/ Leukocytes	%	None	≥ 15
Basophils	$\times 10^9$/L	None	> 0.3	Neutrophils/ Leukocytes	%	≤ 15	None
Eosinophils/ Leukocytes	%	None	≥ 10	Leukocytes	$\times 10^9$/L	≤ 2.8	≥ 16
Lymphocytes	$\times 10^9$/L	≤ 0.5	> 4	Erythrocytes	$\times 10^{12}$/L	≤ 2.5	≥ 7.0
Lymphocytes/ Leukocytes	%	≤ 10	≥ 60	Platelet Count	$\times 10^9$/L	≤ 100	≥ 700

Over the course of the study, there may be some lab tests performed that were not mentioned in the protocol. These tests will not be summarized but will be included in the listings and flagged as non-protocol tests.

11.3 ECGs

Electrocardiograms will be assessed by a central reader and will be recorded at the following visits:

- Screening (triplicate readings)
- Baseline/Visit 1 (Day 1) (pre-dose)
- Baseline/Visit 1 (Day 1) (2-3 hours post-dose)
- Visit 4/Week 6 (Day 43) (pre-dose)

- Visit 4/Week 6 (Day 43) (2-3 hours post-dose)
- Visit 7/Early Termination (Day 85) (pre-dose)
- Visit 7/Early Termination (Day 85) (2-3 hours post-dose)

The following quantitative parameters will be reported by the central reader: heart rate, PR interval, QRS duration, QT interval (uncorrected), and QT interval with Fridericia's correction (QTcF). Change from baseline and percent change from baseline will be calculated for each parameter and summarized by the treatment groups mentioned in Section 11. Note that for patients who switched from placebo to AVP-786, baseline is defined as described in Section 9.1.

In addition, since ECGs are recorded pre- and post-dose at Baseline, Week 6 and Week 12, change from pre- to post-dose will be summarized at these visits.

PR interval and QTcF will be further investigated through PCS tables, for which the criteria are found in Table 6 below. The number and percent of patients meeting the criteria below will be summarized by treatment group. Summaries will be given for both overall (i.e., any time post-baseline) and by visit. For QTcF, males and females will be assessed separately. Patients will be included in all categories for which they qualify. For criteria on the "Actual" values, the denominator for the percentages is the number of patients who had a post-baseline assessment for each parameter. For criteria on the change, the denominator is the number of patients who had a baseline and post-baseline assessment.

Table 6. ECG PCS Criteria

ECG Parameter	Sex	Actual or Change	PCS Criteria
PR Interval (msec)	Both	Actual	> 200 to ≤ 220, > 220 to ≤ 250, > 250
QTcF (msec)	Males	Actual	> 450 to ≤ 480, > 480 to ≤ 500, > 500
	Females	Actual	> 460 to ≤ 485, > 485 to ≤ 500, > 500
	Both	Change from baseline (increase)	≥ 30, ≥ 60

ECG overall interpretations will be summarized by the number and percent that were normal or abnormal. The interpretations by the cardiologist (i.e., central ECG, iCardiac) will be used for these summaries. For females, iCardiac uses QTcF > 460 msec to ≤ 485 msec for PCS criteria and this will be flagged in corresponding tables and listings. The listings will provide all interpretations and corresponding details.

11.4 Vital Signs

Vital signs will be assessed at all visits. Orthostatic blood pressure is taken at Screening and supine/semi-recumbent blood pressure is taken at all subsequent visits. All

supine/semi-recumbent measurements are recorded twice, so the mean of the 2 measurements will be taken and used for all summaries mentioned below. The following parameters recorded in the supine/semi-recumbent positions will be summarized: systolic blood pressure (SBP), diastolic blood pressure (DBP), heart rate. Weight will also be summarized. These parameters will be summarized through CFBL and percent CFBL in similar fashion as the ECG parameters. For heart rate, SBP and DBP, box plots by visit will be presented for Stage 1 [redacted] safety population.

Vital signs will also be assessed through PCS criteria, which are given in Table 7. Patients will be counted if they meet the criteria below at any time post-baseline. The definition of baseline is consistent with those for ECG parameters. The denominators are the number of patients with both a baseline and post-baseline assessment.

Table 7. Vital Sign PCS Criteria

Vital Sign Parameter	High Values	Low Values
SBP (mmHg)	> 180 AND ≥ 20 increase from baseline	≤ 90 AND ≥ 20 decrease from baseline
DBP (mmHg)	≥ 105 AND ≥ 15 increase from baseline	≤ 50 AND ≥ 15 decrease from baseline
Pulse (bpm)	≥ 120 AND ≥ 15 increase from baseline	≤ 50 AND ≥ 15 decrease from baseline
SBP and pulse	SBP ≥ 10 increase from baseline AND pulse ≥ 5 increase from baseline	Not Applicable
DBP and pulse	DBP ≥ 5 increase from baseline AND pulse ≥ 5 increase from baseline	Not Applicable

All vital sign parameters will be included in the listings.

11.5 Physical and Neurological Exams

Physical and neurological exams are assessed at Screening and will be listed.

11.6 C-SSRS

Scoring for the C-SSRS will be analyzed using the following indicators:

- Ideation severity: each of the 5 questions regarding type of ideation (yes/no)
- Intensity of most severe ideation: a sum of the 5 intensity items
- Suicidal behavior types: each of the 4 suicidal behavior types (yes/no)
- Suicidal behavior: suicidal behavior question
- Actual lethality: actual lethality question (0-5 scale) on most lethal attempt
- Potential lethality: potential lethality question (0-2 scale) on most lethal attempt

The individual questions listed above that have yes/no responses will be summarized by treatment group and visit using number and percent.

Intensity of most or severe ideation will be summarized descriptive by treatment group and visit.

Items that contain an ordinal response will be summarized both through descriptive statistics and number and percent.

All C-SSRS data, including individual text descriptions (included as part of some questions), will be included in the listings.

11.7 Simpson Angus Scale for Extrapyramidal Symptoms (SAS)

The individual items, as well as the total score, will be summarized. Descriptive statistics will be provided by treatment group and visit for males and females separately, as well as both sexes overall. For each item, a shift table of the number and percent of patients in each score (0 to 4) from Baseline to Week 6 and Baseline to Week 12 will be summarized overall and for males and females separately.

11.8 Barnes Akathisia Scale (BAS)

The objective assessment, subjective awareness, subjective distress, and global clinical assessment, as well as the total score, will be summarized. Descriptive statistics will be provided by treatment group and visit for males and females separately, as well as both sexes overall. Shift tables of the number and percent of patients of the total score and the global clinical assessment will be presented for Baseline to Week 6 and Baseline to Week 12, for males and females separately as well as both sexes overall.

11.9 Abnormal Involuntary Movement Scale (AIMS)

The following table represents the scores to be calculated and summarized:

Subscale/Score	Item Numbers	Score Range
Orofacial Movements	1, 2, 3, 4	0 – 16
Extremity Movements	5, 6	0 – 8
Trunk Movements	7	0 – 4
Global Judgment	8, 9, 10	0 – 12
Dental Status	11, 12	Yes/No
Observed Movements	1 – 7	0 – 28
Overall Severity	8	0 – 4

Descriptive statistics will be provided by treatment group and visit for males and females separately, as well as both sexes overall. Shift tables of the number and percent of patients will be presented for Baseline to Week 6 and Baseline to Week 12 for the overall severity, for males and females separately as well as both sexes overall.

12 ADDITIONAL STATISTICAL ANALYSIS METHODS

12.1 Patient Disposition

Counts of patient enrollment will provide the number of patients screened along with the reason for screen failures. A summary of randomized treatment group by responder status will be provided. Patient status will be summarized by the following:

- Randomized in Stage 1
- Took study medication
- Discontinued during each stage
- Re-randomized to Stage 2
- Completed study

Counts will be provided by Stage 1 randomized treatment group based on the study stage. Primary reasons for discontinuation will be provided based on the number of patients in the treatment group and stage.

An overall number of patients by stage in the mITT, [redacted] and Safety populations will be provided.

12.2 Demographics and Baseline Characteristics

Demographics will be summarized by randomized treatment group and overall for the mITT, and Safety populations. The following characteristics will be summarized.

- Sex (Male, Female)
- Race (White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Other)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino)
- Age
- Age group ($< 55, \geq 55$)
- Height
- Weight
- Body mass index

A summary of baseline characteristics will include baseline values of all primary and secondary endpoints, in addition to height, weight, and body mass index. These will be summarized by randomized treatment group. Additional tables of primary and secondary endpoints for placebo non-responders and placebo responders will be created. Stage 2 baseline will be used for these patients.

For categorical parameters, the denominators for the percentages are the number of patients who had the parameter assessed.

12.3 Medical History and Disease Characteristics

Medical history will be summarized by SOC and PT based on Stage 1 randomization and will be presented in the listings. Summary tables will be provided for the frequency and percentage of patients by the following subgroups:

- MCCB score < 30 versus ≥ 30
- Baseline concomitant benzodiazepine, SSRI, and SSNI medication use (Yes/No)

Additionally, descriptive statistics for the number of years since the onset of schizophrenia, and the number of years since the onset of residual schizophrenia, will be presented. In both cases, the number of years will be counted relative to the patient's randomization date in Stage 1.

12.4 Exposure and Treatment Compliance

Duration of exposure will be summarized quantitatively using the number of days on study medication for each patient, displayed by treatment group. Summaries will be done for Placebo/Placebo, AVP-786/AVP-786, and each arm of the Placebo/AVP-786 group. Duration for the Placebo/Placebo and AVP-786/AVP-786 groups will be calculated as (last dose date – first dose date + 1). For Stage 1 Placebo/AVP-786 (placebo portion), duration will be calculated from first dose until the day prior to re-randomization, i.e., (re-randomization date – first dose date). For Stage 2 (AVP-786 portion), duration will be calculated from re-randomization to last dose date, i.e., (last dose date – re-randomization date + 1).

An additional summary of exposure will be provided by dose taken for patients who took AVP-786 at any time during the study. The number of days at the specified dose level will be summarized. Treatment groups will be AVP-786/AVP-786 and Placebo/AVP-786. Since the dosing regimen increases after one week on treatment but a visit does not occur for 2 weeks after starting treatment, it is assumed that each dose level is taken for 7 days, as specified in the protocol. Doses summarized will be [redacted] mg QD, [redacted] mg BID and 34 mg BID. The dosing regimen can be found in Figure 1.

Overall treatment compliance will be calculated as a percentage using the total number of capsules that were dispensed and returned. Patients will be grouped into categories of $< 80\%$, 80% to 120%, $> 120\%$. Counts will be summed over the visits for each patient to calculate an overall compliance value.

Compliance will be summarized as described above and through descriptive statistics for Placebo/Placebo, AVP-786/AVP-786, and each arm of the Placebo/AVP-786 group. The number of doses taken and number of doses should have taken are necessary to calculate compliance. The calculation for number of doses should have taken will be slightly different for patients who were randomized vs. those who were not re-randomized, as well as between Stage 1 and Stage 2. These differences are provided in the table below. The 3 steps for calculating compliance are shown here:

1. Doses taken: dispensed amount – returned amount
2. Doses should have taken:

Stage	Patients Re-randomized	Patients Not Re-randomized
Stage 1	2 × (re-randomization day – first dose day)	2 × (last dose day – first dose day) + 1
Stage 2	2 × (last dose day – re-randomization day) +1	N/A
Overall	2 × (last dose day – first dose day) + 1	2 × (last dose day – first dose day) + 1

3. Compliance = (doses taken / doses should have taken) × 100%

For the calculations in table above, the formulas are based on the assumption that patients take 2 capsules per day except for the last day, in which they only take 1. One exception is for the Stage 1 calculation of patients who were re-randomized. For these patients, 1 is not subtracted because they will take 2 capsules on the last day of Stage 1.

The number of capsules taken for kits that were not returned will be imputed to be the number of capsules that were dispensed.

12.5 Prior and Concomitant Medications

The number and percent of prior and concomitant medications will be provided by the treatment groups used for the safety analyses. Prior medication is defined as medication with a stop date prior to first dose date. Concomitant medication is defined as a medication with a start date on or before the last dose date AND a stop date on or after first dose date. Medication that is ongoing at the time of first dose is also considered concomitant. Partial dates will be imputed using rules in Appendix 3.

Prior and concomitant medications will be coded using WHO Drug Dictionary (version September 2015). Summaries will be provided by anatomical therapeutic chemical classification (ATC) and preferred term.

12.6 Analysis of Pharmacokinetics and Pharmacodynamics

[redacted]

13 ADDITIONAL STATISTICAL CONSIDERATIONS

13.1 Data Safety Monitoring Board (DSMB)

There is no DSMB for this study.

13.2 Changes from Protocol

When implementing changes in protocol amendment 4, the IRT vendor who carried out the IVRS found that 14 Stage 1 Placebo patients were re-randomized for Stage 2 at Week 3 instead of Week 6. These 14 patients with randomization error are excluded from the mITT and ITT analysis population although they satisfied protocol mITT and ITT population criteria. Their safety data are summarized under safety population and their efficacy data are included in the corresponding data listings.

14 REFERENCES

15 APPENDICES

Appendix 1 Schedule of Assessments and Procedures

			Stage 1				Stage 2					
PROCEDURE	Visit	Screening	Baseline Visit 1	Phone[7]	Visit 2[2]	Visit 3[1]	Visit 4[2]	Phone[7]	Visit 5[2]	Visit 6[1]	Visit 7/ET[2,4]	Post Exit Phone[7]
	Study Day	Day -28 to -7	Day 1	Day 8	Day 15	Day 22	Day 43	Day 50	Day 57	Day 64	Day 85	Day 86-90
	Study Week	Week -4 to -1		Week 1	Week 2	Week 3	Week 6	Week 7	Week 8	Week 9	Week 12	
Informed Consent		X										
[redacted]		X									X	
Medical History		X										
M.I.N.I. Exam		X										
Inclusion and Exclusion		X	X									
Randomization / (re-randomization)			X				X					
Physical Examination		X										
Resting 12-lead ECG		X[6]	X[6]				X[6]				X[6]	
Chemistry, Hematology, and Urinalysis[3]		X					X				X	
Pregnancy Test [3]		X	X			X	X			X	X	
[redacted]			X[5]									
Plasma Antipsychotic Levels		X					X				X	
PK and [redacted]			X[5]				X[5]				X[5]	
Review of Adverse Events			X	X	X	X	X	X	X	X	X	X
Concomitant Medications		X	X	X	X	X	X	X	X	X	X	X
Record Vital Signs/Weight[9]		X	X		X	X	X		X	X	X	
NSA-16		X	X			X	X			X	X	
PANSS		X	X			X	X			X	X	
MCCB[10]		X	X				X				X	
CGI-S			X				X				X	
CDSS		X	X			X	X			X	X	
CGI-C							X				X	
PGIC							X				X	
RDoC Task (EEfRT)			X				X				X	
Side Effects Scales (AIMS, BAS, SAS)			X				X				X	
C-SSRS		X	X		X	X	X		X	X	X	
[redacted]												
Dispense Study Medication			X			X	X			X		
Dose in Clinic			X		X	X	X		X	X	X	
Review and/or Return Unused Study Medication				X[8]	X	X	X	X[8]	X	X	X	

1. Visits 3 and 6 have a +3 day window.
2. Visits 2, 4, 5, and 7 have a ±3 day window.
3. Urinary (beta-hCG) test will be performed for all females regardless of childbearing potential (serum beta-hCG at Screening only). Fasting glucose and lipids will be measured at Screening, Visit 4, and Visit 7/ET.
4. Final visit or Early Termination Visit for patients who withdraw prior to study completion.
5. PK blood draws will be taken 2-3 hours post-dose for Baseline, Visit 4, and Visit 7/ET. [redacted]
6. Electrocardiogram will be performed prior to dosing and 2-3 hours post-dosing for Visits 1 and 4. Post-dose only for Visit 7. Triplicate ECGs at Screening only.
7. Telephone call will have a +3 day window. Post study exit calls will be daily for 5 days to assess any changes in health (AEs)/concomitant medications.
8. Patient will be asked if they have taken their medications as directed during telephone calls on Days 8 and 50.
9. Height will be measured at Screening only.
10. The MCCB should be conducted at approximately the same time of day (+/- 2 hours) and preferably in the AM. After the Screening Visit, patients who use sedatives/hypnotics or benzodiazepine medications on a prn basis should not take any of these medications the day of, or the day before, the assessment of cognitive function by MCCB. Patients who are on stable dose regimens of sedatives/hypnotics or benzodiazepine medications should take their medication as prescribed.

[redacted]

[redacted]

Appendix 3 Prior and Concomitant Medication Start Date Imputation

Imputation Rules for Partial Dates (D = day, M = month, Y = year)

Parameter	Missing	Additional Conditions	Imputation
Start date for con meds	D only	M and Y same as M and Y of first dose of study drug	Date of first dose of study drug
		M and/or Y not same as date of first dose of study drug	First day of month
	M and D	Y same as Y of first dose of study drug	Date of first dose of study drug
		Y not same as Y of first dose of study drug	Use Jan 01 of Y
	M, D, and Y	None - date completely missing	Day prior to date of first dose of study drug
Stop date for con meds	D only	M and Y same as M and Y of last dose of study drug	Date of last dose of study drug
		M and/or Y not same as date of last dose of study drug	Last day of month
	M and D	Y same as Y of last dose of study drug	Date of last dose of study drug
		Y not same as Y of last dose of study drug	Use Dec 31 of Y
	M, D, and Y	None - date completely missing and NOT ongoing	Date of last dose of study drug

Note: In all cases, if an estimated start date is after a complete stop date, use the first day of the stop date month.
Similarly, if the estimated stop date is before a complete or imputed start date, use the last day of the start day month.

[redacted]

Appendix 5 List of Selective Serotonin-norepinepherine Reuptake Inhibitors (SNRIs)

Preferred Term	Trade Names
Venlafaxine	Efexor, Effexor
Sibutramine	Meridia, Reductil
Duloxetine	Cymbalta, Ariclaim, Xeristar, Yentreve
Atomoxetine	Strattera
Desvenlafaxine	Pristiq
Milnacipran	Savella, Ixel, Dalcipran, Toledomin
Levomilnacipran	Fetzima