Statistical Analysis Plan

Sponsor:	Otsuka Pharmaceutical Development & Commercialization, Inc.
Protocol No:	17-AVP-786-205
Protocol Version No./ Date	Protocol Amendment 5/ 23 March 2020 and Addendum Amendment 1/31AUG2021
Title	A Phase 2, multicenter, randomized, double-blind, placebo-controlled study to assess the efficacy, safety, and tolerability of AVP-786 (deudextromethorphan hydrobromide [d6-DM]/quinidine sulfate [Q]) for the treatment of neurobehavioral disinhibition including aggression, agitation, and irritability in patients with traumatic brain injury (TBI)
CRF Version No./ Date	5.0 5-JUN-2020
ICON/PRA Project ID:	AVA78625-786205
SAP Version No./ Date:	2.2 18-Aug-2022

Approvals

Sponsor	
Sponsor Name:	Otsuka Pharmaceutical Development & Commercialization, Inc.
Global Clinical Development / Title:	PPD , MD / PPD
Signature/ Date:	
Biostatistics/ Title:	PPD , PhD / PPD
Signature/ Date:	
Biostatistics/ Title:	PPD
Signature/ Date:	
ICON/PRA	
Biostatistician/ Title:	PPD , Ph.D. / PPD
Signature/ Date:	

Table of Contents

1.0 Introduction

This statistical analysis plan (SAP) describes the statistical methods to be used during the reporting and analyses of data collected under Otsuka Pharmaceutical Development & Commercialization, Inc. Protocol 17-AVP-786-205, a Phase 2, multicenter, randomized, double-blind, placebo-controlled, sequential parallel comparison design (SPCD) study to assess the efficacy, safety, and tolerability of AVP-786 (deudextromethorphan hydrobromide [d6-DM]/quinidine sulfate [Q]) for the treatment of neurobehavioral disinhibition including aggression, agitation, and irritability in patients with traumatic brain injury (TBI).

This SAP should be read in conjunction with the study protocol and case report forms (CRF).

2.0 Study Objectives

The study objectives and corresponding endpoints are listed in the following table.

Table 1 Study Objectives and Endpoints

Objective (s)	Endpoint (s)
Primary Efficacy	
To evaluate the clinical efficacy of one dose of AVP-786 42.63/4.9 mg twice daily compared with placebo, for treatment of neurobehavioral disinhibition including aggression, agitation, and irritability in patients with traumatic brain injury (TBI). using the Neuropsychiatric Inventory- Clinician Rating Scale 3 (NPI-C-3)	Change from Baseline to Week 6 (Stage 1) and from Week 6 to Week 12 (Stage 2) in the composite of the Clinical Impression severity scores on the Neuropsychiatric Inventory-Clinician (NPI-C) rating subscales of Aggression, Agitation, and Irritability/Lability (NPI-C-3)
Secondary Efficacy	
To evaluate additional measures of TBI-related neurobehavioral disinhibition including aggression, agitation, and irritability	Mean Change from Baseline to Week 6 (Stage 1) and from Week 6 to Week 12 (Stage 2) for the following measures: • NPI-C subscales (Aggression, Agitation, Irritability/Lability, and Disinhibition) • mCGI-S • PGI-S Mean score at Week 6 and Week 12 for the following measures: • mCGI-C • PGI-C
Safety	
To evaluate the safety and tolerability of one dose of AVP-786 42.63/4.9 mg twice daily compared with placebo, for treatment of neurobehavioral disinhibition including aggression, agitation, and irritability in patients with TBI	Treatment-emergent adverse events (TEAEs), physical and neurological examinations, vital signs, clinical laboratory assessments, resting 12-lead electrocardiograms (ECGs), and the Sheehan Suicidality Tracking Scale (S-STS).

2.1 Summary of Study Design

This is a Phase 2, multicenter, randomized, double-blind, placebo-controlled, sequential parallel comparison design (SPCD) study to evaluate the efficacy, safety, and tolerability of AVP-786

(deudextromethorphan hydrobromide [d6-DM]/quinidine sulfate [Q]) for the treatment of neurobehavioral disinhibition including aggression, agitation, and irritability in patients with TBI. Patients with mild, moderate or severe TBI neurobehavioral disinhibition including aggression, agitation, and irritability that persists following non-penetrating brain injury with score of ≥ 4 (moderately ill) on the mCGI-S scale at Screening and Baseline visits, and a score of ≥ 4 on the Agitation/Aggression or Irritability/Lability subscales of the Neuropsychiatric Inventory (NPI) scale at Screening and Baseline visits – who met other eligibility criteria will be randomly assigned (1:1) to AVP-786-42.63/4.9 BID or matching Placebo. For eligibility, responses to Agitation/Aggression and Irritability/Lability subscales of the NPI will be derived from the NPI-C at screening and baseline.

Patients will be enrolled in the study for approximately 17 weeks, with an approximate 4-week screening period, 2 consecutive double-blind treatment stages (Stage 1 and Stage 2) of 6 weeks duration each, and a 1-week follow up period (phone call). The study utilizes a SPCD with a re-randomization for Stage 1 placebo patients. Table 2 provides an overview of the study duration for each stage:

Table 2 Overview of Study Duration

Sequence	**Screening**	**Stage 1**	**Stage 2**	**Follow-Up Phone Call[a]**
Duration	4 Weeks (28 Days)	6 weeks (Day 1 to Week 6)	6 weeks (Week 7-12)	1-week after last dose of study medication (Week 13)

a: Any previously reported and not resolved AE, and any newly reported AE will be followed-up for up to 30 days after the last dose of study medication.

Approximately 150 patients will be enrolled at approximately 40 centers in the United States (US). Patients will attend clinic visits at Screening, Baseline (Day 1), and on Visit 3 (Week 1), Visit 4 (Week 6), Visit 5 (Week 9) and Visit 6 (Week 12).

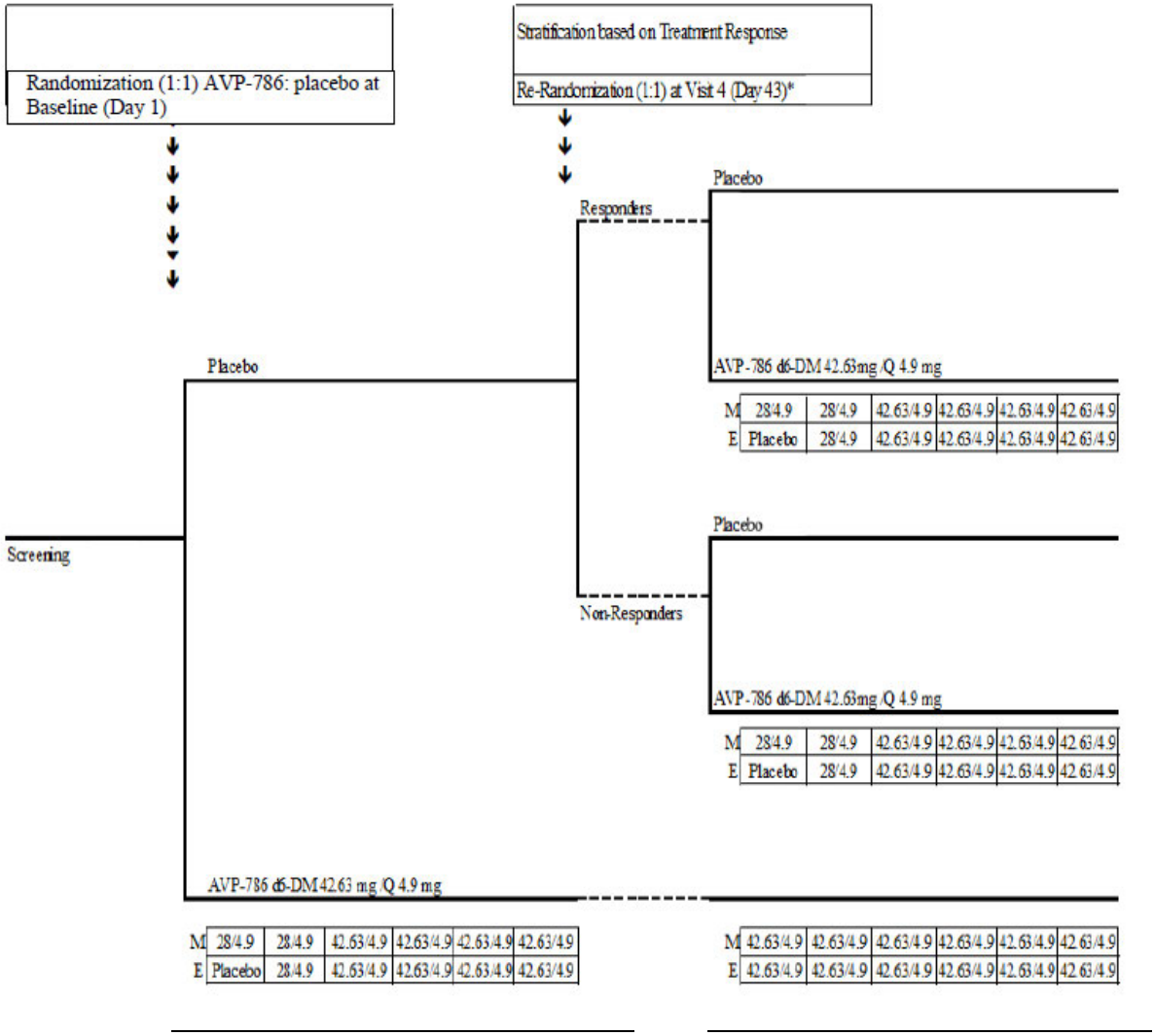
A schematic of the study along with the treatment segments is shown in Figure 1. Study procedures will be performed at each visit as outlined in the Schedule of Evaluations and Visits. The schedule of events is provided in Table 3 Schedule of Events and Visits.

The letters for each treatment segment are described below and will be used throughout the 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 in Stage 1
- B: Data for patients randomized to AVP-786
- 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

Figure 1 Study Schematic



Period	Screening		Stage 1					
Visit	Screening		Baseline	V2		V3		
Day	-28	-1	1 7	8 14	15 21	22 28	29 35	36 42
Week	Week -4		Week 1	Week 2	Week 3	Week 4	Week 5	Week 6

Stage 2					
V4*			V5		V6
43* 49	50 56	57 63	64 71	72 78	79 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 (1:1).
**:A1, A2 are the Stage 1 data for patients who are randomized to Placebo/Placebo and Placebo/AVP-786, respectively.

2.2 Randomization

The protocol describes that upon entry into the study (after ICF is signed at screening), all patients will be assigned a 6-digit patient number. The first 3 digits consist of the center number. The last 3 digits will be assigned sequentially starting with 001. This 6-digit number is the main identifier for patients. Eligible patients will be randomized to receive AVP-786 or matching placebo capsules on Day 1 in a double-blind manner according to a randomization scheme devised by Otsuka or its representative and managed within an Interactive Web Response(IWRS). The randomization will be stratified by study site.

The following information describes the unblinded randomization procedure.

Stage 1:

Eligible patients will be randomized to receive AVP-786 capsules or matching placebo capsules in a 1:1 ratio within each stratum on Day 1 (Baseline, Stage 1) in a double-blind manner according to a randomization scheme-devised by Otsuka or its representative and managed within an IWRS.

Stage 2:

Re-randomization will occur for patients who were assigned to placebo in Stage 1. The patient number will not be re-assigned; it will remain the same in both stages of the study.

- Patients who receive placebo in Stage 1 will be further stratified into 2 subgroups ("responders" and "non-responders") based on their treatment responses and by site. Patients who meet the responder criteria described in section 2.3 will be considered "responders". Patients who do not meet these criteria will be considered "non-responders". Patients within each placebo subgroup will be re-randomized in a 1:1 (active:placebo) ratio within each stratum to receive either AVP-786 or placebo.
- For statistical purposes, patients who receive placebo and drop out prior to the completion of Stage 1 will also be assigned a re-randomization treatment in the same manner as the other placebo patients. Their "responder" and "non-responder" status will be based on their treatment response using measurements at their early termination visit.
- Patients who receive AVP-786 in Stage 1 will not be re-randomized and will continue to receive the same dose of AVP-786 BID for the entire 6-week duration of Stage 2.

2.3 Responder Status

In this study, patients will be stratified as responders or non-responders at the end of Stage 1. Responders are those patients who meet both of the following criteria:

1. Modified Clinical Global Impression of Severity (mCGI-S) score is ≤ 3 at Visit 4 (Week 6) or Early Termination Visit and
2. Neuropsychiatric Inventory composite score (NPI-C-3) score as Visit 4 (Week 6) or Early Termination Visit has decreased by ≥ 25% from baseline, i.e.

$$\frac{\text{NPI-C-3}_{\text{Total}}(\text{Visit 4}) - \text{NPI-C-3}_{\text{Total}}(\text{Baseline})}{\text{NPI-C-3}_{\text{Total}}(\text{Baseline})} \times 100\% \leq -25\%.$$

Patients who do not meet either of these criteria are considered non-responders.

2.4 Study Drug Administration

The investigational product (IP) is AVP-786 (d6-DM/Q), which is titrated over a 2-week period to d6-DM 42.63 mg/Q 4.9 mg (AVP-786-42.63/4.9) twice daily (BID). Placebo capsules appearing identical to study medication will be used as control. Eligible patients will be randomized into the study in a 1:1 ratio to receive AVP-786 or placebo in Stage 1. Study medication will be administered orally twice daily (BID), 1 capsule in the morning and 1 capsule in the evening approximately 12 hours apart) throughout the study. The morning dose of study medication will be administered at the site during scheduled visits.

For patients randomized to receive AVP-786, their dosage regimen will begin with d6-DM 28 mg/Q 4.9 mg (AVP-786-28/4.9) once a day in the morning and placebo in the evening for the first 7 days followed by AVP-786-28/4.9 BID for the next 7 days. Patients will then receive AVP-786-42.63/4.9 BID for the remaining duration of the study.

Patients who are assigned to placebo in Stage 1 will be re-randomized in a 1:1 ratio to AVP-786 or placebo for their Stage 2 treatment, stratified by their Stage 1 efficacy response status (responder or non-responder). Patients who receive placebo in Stage 1 and are re-randomized to AVP-786 will have their dosage titrated similarly as described above, over a 2-week period to d6-DM 42.63 mg/Q 4.9 mg BID in Stage 2.

Table 3 Schedule of Events and Visits

Procedure	Visit:	Screening[1]	Visit 1 Baseline[1]	Visit 2 Phone Call[1]	Visit 3[1]	Visit 4[1]	Visit 5[1]	Visit 6/ET[1,2,3]	Follow-up Phone Call[1]
	Study Day:	**Day -28 to -1**	**Day 1**	**Day 8**	**Day 22**	**Day 43**	**Day 64**	**Day 85**	**Day 92**
	End of Study Week:	**Week -4 to -1**		**Week**	**Week**	**Week**	**Week 9**	**Week 12**	**Week 13**
Sign informed consent forms		X							
Medical history (including TBI severity classification)		X							
Review of eligibility		X[4]	X						
Randomization			X						
Physical and neurological examination		X						X	
Risk assessment of falls		X							
Vital signs, height and weight		X	X[5]		X	X	X	X[5]	
Electrocardiogram		X[6]	X[7]		X[7]	X[7]	X[7]	X[7]	
Adverse events		X	X	X	X	X	X	X	X
Prior and concomitant medications use assessments		X	X[8]	X	X[8]	X[8]	X[8]	X[8]	X
NPI-C subscales: Agitation, Aggression, Irritability/Lability, and Disinhibition[9]		X	X		X	X	X	X	
mCGI-S		X	X			X		X	
mCGI-C						X		X	
S-STS		X	X		X	X	X	X	
PHQ-9		X							
PGI-S			X			X		X	
PGI-C						X		X	
Administer morning dose of study medication in clinic			X		X	X	X	X	
Chemistry, hematology, and urinalysis		X[10]				X		X[10]	
Pregnancy test[11]		X	X		X	X	X	X	

Blood sample for pharmacokinetic analysis					X[12]	X[13]	X[12]	
Blood sample for CYP2D6		X						
Urine sample for drug screening/toxicology	X			X		X	X	
Dispense study medication blister card and diary card		X		X	X	X		
Collect study medication blister card and review diary card[14]				X	X	X	X	

CYP2D6 = cytochrome P450 isoenzyme 2D6; ET = early termination; mCGI-C = modified Clinical Global Impression of Change (assessment of change in aggression, agitation, and irritability symptoms); mCGI-S = modified Clinical Global Impression of Severity (assessment of severity of aggression, agitation, and irritability symptoms); NPI-C = Neuropsychiatric Inventory Clinician; PGI-C = Patient Global Impression of Change; PGI-S = Patient Global Impression of Severity; PHQ-9 = Patient Health uestionnaire; PK = pharmacokinetic; S-STS = Sheehan Suicidality Tracking Scale; TBI = Traumatic Brain Injury.

Note: Whenever possible, for consistency of ratings, the same raters should rate a patient (and the informant /caregiver) at each of the patient's visits for all 3 of the following scales: the NPI-C, the mCGI-S, and the mCGI-C. The mCGI-S and mCGI-C must be administered after the NPI-C subscales. If it is not possible to have the same rater for the NPI-C and the mCGI subscales (mCGI-S and mCGI-C); it is critical that 1 rater assesses both of the mCGI subscales for a patient. In addition, if the NPI-C rater differs from the mCGI rater, prior to making the mCGI assessment, the mCGI rater must review the NPI-C results, discuss the results with the NPI-C rater, and review any additional pertinent information.

1.Study visits after Baseline have a ± 3-day window except Visit 2 and the Follow-up Phone Call, which have a + 3-day window. The screening period may be extended after discussion with and approval by the Medical Monitor.
2.Early termination visit for patients who withdraw prior to study completion.
3.A safety Follow-up Phone Call should be made to the patient or informant/caregiver 7 (+3) days after last dose of study medication.
4.For each patient, a protocol eligibility form will be completed.
5.Weight should be measured at the Baseline visit and Visit 6 (Day 85); height is only measured at Baseline.
6.Electrocardiogram should be performed in triplicate at the Screening visit.
7.Electrocardiogram to be performed pre-dose and post-dose at the Baseline Visit (Day 1) and Visit 4 (Day 43). To be performed post-dose at Visit 3 (Day 22), Visit 5(Day 64), and Visit 6 (Day 85).
8.On the days of study visits (i.e., Baseline [Day 1], Visit 3 [Day 22], Visit 4 [Day 43], Visit 5 [Day 64], and Visit 6 [Day 85]), patients are instructed to refrain from the use of benzodiazepine medications, medicinal/recreational marijuana, and alcohol for 12 hours prior to the visit and until all study assessments have been completed.
9.For eligibility, the responses to the Agitation/Aggression and Irritability/Lability subscales of the NPI will be derived directly from the NPI-C at Screening and Baseline.
10.Thyroid function tests (thyroid-stimulation hormone [TSH] and, if TSH is abnormal, reflex triiodothyronine [T3], reflex total thyroxine [T4], and reflex free T4) and glycosylated hemoglobin (HbA1c) test should be performed at the Screening visit and Visit 6 (Day 85).
11.Serum pregnancy test to be performed for patients of childbearing potential at the Screening visit. Urine pregnancy tests to be performed for patients of childbearing potential at all other visits prior to dosing; if positive, the patient must be discontinued from the study.
12.Blood sample for pharmacokinetic analysis should be collected 1 to 3 hours after administering the morning dose.

13.Blood sample for pharmacokinetic analysis should be collected before administering the morning dose.
14.The study medication blister card and diary card should be collected and reviewed at each visit for compliance.

2.5 Sample Size Considerations

Sample size calculations are performed assuming a normal distribution for the primary efficacy endpoint of change from baseline in NPI-C-3 composite score. It is assumed that the standard effect size is -0.50. It is further assumed that 80% of patients will complete Stage 1; additionally, 80% of the Stage 1 completers are expected to complete Stage 2. A sample size of 150 patients randomized in a 1:1 ratio (AVP-786:Placebo) in Stage 1 will have an approximate 80% power for this SPCD study with weight = 0.7 for the Stage 1 data, and 2-sided type I error α=0.05. With this assumption, the primary efficacy endpoint point will have about 86% power at Stage I.

3.0 Study Endpoints and Covariates

3.1 Scales and Questionnaires

The scales and questionnaires will be listed in addition to any efficacy and safety analyses described in the subsequent sections.

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

The NPI-C is completed via a set of comprehensive interviews by the clinician with the informant/caregiver (first) and patient (after informant), covering 14 neuropsychiatric symptom domains. Behaviors within a domain are rated by the informant/caregiver in terms of frequency (0 to 4), severity (0 to 3) and level of distress to the informant/caregiver (0 to 5), then rated by the patient in terms of frequency (0 to 4), and finally adjudicated by the clinician with a clinical impression severity rating (0 to 3) that is based on all available information. The NPI-C allows for use of each domain independently as stand-alone based on the objectives of the research question. In this study, scores for the NPI Aggression/Agitation and Irritability/Lability subscales will be derived from the NPI-C at the Screening and Baseline visits, to determine if the patient's behavior meets a clinically relevant threshold (score of ≥ 4 on any of the three subscales) for entry into the study. Ratings on the clinician impression severity scores of the Agitation, Aggression, and Irritability/Lability domains of the NPI-C will be combined into a composite score, referred to as NPI-C-3, that serves as the primary efficacy endpoint of the study.

NPI-C-3 composite score will be the sum of below 3 domain scores based on Clinical Impression Severity:

- NPI-C Agitation domain score
    - Sum of Clinical Impression Severity scores for Agitation questions 1-13, score ranges from 0-39
- NPI-C Aggression domain score
    - Sum of Clinical Impression Severity scores for Aggression questions 1-8, score ranges from 0-24
- NPI-C Irritability/Lability domain score

  - Sum of Clinical Impression Severity scores for Irritability/Lability questions 1-12, score ranges from 0-36

- NPI-C-3 composite score
  - NPI-C Agitation + NPI-C Aggression + NPI-C Irritability/Lability, socre ranges from 0 – 99.

In addition, NPI-C Disinhibition domain score will be calculated similarly as: Sum of Clinical Impression Severity scores for Disinhibition questions 1-16, score ranges from 0-48.

The NPI-C will be administered to the patient and caregiver at Screening (Day -28 to Day -1) Baseline (Day 1), Visit 3 (Week 3), Visit 4 (Week 6), Visit 5 (Week 9), and Visit 6 (Week 12). The recall period for the behaviors in each domain will be 3 weeks, for all visits.

3.1.2 Clinical Global Impression (CGI) Scales

In patients with TBI who present with marked impairments in motor function, cognition, or other areas, global assessments may not be sensitive to specific changes in their target neuropsychiatric symptoms of aggression, agitation and irritability. The CGI scales are modified in this study, and the clinician is explicitly instructed to rate the severity (mCGI-S) or change (mCGI-C) of the patient's aggression, agitation and irritability symptoms, to ensure that the global ratings specifically evaluate the target symptoms. At each visit, the NPI-C is administered before the mCGI-C and mCGI-S scales are rated, to ensure the mCGI-C and mCGI-S scale ratings reflect a thorough assessment of the patient's neurobehavioral symptoms.

The modified scales will be used as part of the secondary efficacy endpoints.

3.1.2.1 Modified Clinical Global Impression of Severity (mCGI-S)

The mCGI-S is a 7-point (1-7) modified version of the CGI-S scale, and requires that the clinician rate the severity of the patient's neurobehavioral disinhibition including aggression, agitation, and irritability after the NPI-C interview, at the time of assessment relative to the clinician's past experience with patients who have the same diagnosis. Considering total clinical experience, a patient is assessed on severity of illness at the time of rating as:

0, not assessed;
1, normal, not at all ill;
2, borderline ill;
3, mildly ill;
4, moderately ill;
5, markedly ill;
6, severely ill; or
7, among the most extremely ill patients.

The mCGI-S will be assessed at Screening (Day -28 to Day -1), Baseline (Day 1), Visit 4 (Week 6), and Visit 6 (Week 12). A score of 0 will be treated as missing data.

3.1.2.2 Modified Clinical Global Impression of Change (mCGI-C)

The mCGI-C is a 7-point modified version of the CGI-C scale, that requires the clinician to rate the change of the patient's neurobehavioral disinhibition including aggression, agitation, and

irritability at the time of assessment after the NPI-C interview, relative to the clinician's past experience with the patient's neurobehavioral disinhibition at admission. Considering total clinical experience, a patient is assessed for change of mental illness as:

0, Not assessed;
1, Very much improved;
2, Much improved;
3, Minimally improved;
4, No change;
5, Minimally worse;
6, Much worse; or
7, Very much worse.

The mCGI-C will be assessed at Visit 4 (Week 6) and Visit 6 (Week 12) for change in overall clinical status. At Visit 4 (Week 6), change from the Baseline visit (Day 1) will be assessed. At Visit 6 (Week 12), change from Visit 4 (Week 6) and change from the Baseline visit (Day 1) will also be assessed. A score of 0 will be treated as missing data.

3.1.3 Patient Health Questionnaire (PHQ-9)

The PHQ-9 is a multiple-choice self-report inventory that is used in this studyfor the evaluation of depression symptoms. The choices for all questions aim to capture the severity of depression symptoms:

0 = Not at all;
1 = Several days;
2 = More than half the days;
3 = Nearly everyday.

The PHQ-9 is scored by summing the responses of the first 9 individual questions ranging from 0 to 27, while the 10th question is not included in the score. The PHQ-9 will be assessed at Screening only.

3.1.4 Patient Global Impression of Severity (PGI-S)

The PGI-S is a single-question scale that will specifically assess the severity of symptoms of neurobehavioral disinhibition including aggression, agitation, and irritability, on a 7-point scale, as 1, normal, not at all ill; 2, borderline ill; 3, mildly ill; 4, moderately ill; 5, markedly ill; 6, severely ill; or 7, extremely ill. The PGI-S will be completed by the patient at Baseline (Day 1), Visit 4 (Week 6), and Visit 6 (Week 12).

3.1.5 Patient Global Impression of Change (PGI-C)

The PGI-C is a 7-point (1-7) scale used to assess treatment response, and it is rated as: 1 = very much improved, 2 = much improved, 3 = minimally improved, 4 = no change, 5 = minimally worse, 6 = much worse, or 7 = very much worse. The PGI-C will be assessed and rated by the patient at Visit 4 (Week 6), and Visit 6 (Week 12).

3.1.6 Sheehan Suicidality Tracking Scale (S-STS)

Different from questionnaires listed above, the S-STS is a safety assessment. Each item of the S-STS is scored on a 5-point Likert scale (0 = not at all; 1 = a little; 2 = moderate; 3 = very; and

4 = extremely). The S-STS will be analyzed as suicidal ideation subscale score, suicidal behavior subscale score and total score. For the screening visit, the timeframe for the items on the scale will be 'in the past 6 months' and for all other visits it will be 'since last visit'. The S-STS will be assessed at all clinic visits.

3.2 Efficacy Endpoints

3.2.1 Primary Endpoints:

Change from Baseline to Week 6 (Stage 1) and from Week 6 to Week 12 (Stage 2) in the NPI-C-3 composite score.

3.2.2 Secondary Endpoints:

Change from Baseline to Week 6 (Stage 1) and from Week 6 to Week 12 (Stage 2) in the following:

- NPI-C subscale scores (Agitation, Agreesion, Irritability/Liability, and Disinhibition)
- Modified Clinical Global Impression of Severity (mCGI-S) score
- Patient Global Impression of Severity(PGI-S)

Week 6 and Week 12 for the following measures,

- mCGI-C Raw score from baseline at first randomization and re-randomization respectively
- PGI-C Raw score

3.2.3 Safety and Tolerability

Safety and tolerability of AVP-786 will be assessed by analysis of the adverse events (AEs), physical and neurological examinations, vital signs, clinical laboratory assessments, resting 12-lead electrocardiograms (ECGs), and the S-STS.

Pregnancy tests will be conducted for females of childbearing potential.

3.2.4 Pharmacokinetic, Genotype and Biomarker Parameters

Information on the cytochrome P450 (CYP) 2D6 enzyme phenotype will be used to classify patients as either poor, intermediate, extensive, or ultra-rapid metabolizers of d6-DM. Plasma concentrations of d6-DM, its metabolites, Q, and second generation antipsychotic medications will be assessed at Visit 4 (Week 6) pre-dose, Visit 5 (Week 9) and Visit 6/Early Termination (Week 12), and summarized by metabolizer groups. Safety population will be used for this summary.

4.0 Analysis Populations

Analysis populations are defined below. Number and percent of patients meeting definition of each analysis population will be summarized by treatment group

4.1 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 took at least one dose of study medication and had at least one post-baseline NPI-C-3 composite 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 NPI-C-3 composite score efficacy assessment in Stage 2 (after Week 6/Visit 4).

4.2 Per-Protocol Population

The PP population is defined as those mITT patients who have no major protocol violations which may substantially impact the primary efficacy assessment. The following criteria will be used as a guide to exclude patients from the PP analysis:

- Violation of inclusion and exclusion criteria which may substantially impact the primary efficacy assessment.
- Study medication compliance < 80%.
- Receiving incorrect treatment during the study. Specifically, active treatment group patient received placebo or placebo treatment group patient received active re treatment.
- Taking disallowed medication(s) post-baseline which may substantially impact the primary efficacy assessment.
- Below the limit of quantification d6-dM or Q concentrations at Week 6 or 12 when receiving active treatment.

Patients excluded from this population, as well as any medications taken in violation of the study protocol, will be provided in data listings.

4.3 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.

4.4 12-Week Parallel Group Populations

The 12-Week Parallel Group population includes those patients who are randomized into Placebo/Placebo or AVP-786/AVP-786 (randomized to AVP-786 in Stage 1). It is intended to assess efficacy or safety under a 12-week study duration, and what would be done if the study had been a parallel group design.

4.4.1 mITT 12-Week Parallel Group Population

The mITT 12-Week Parallel Group population are those patients in the 12-Week Parallel Group population who take at least one dose of IP and have at least one post-baseline NPI-C-3 composite score.

Note that patients who are randomized to Placebo/AVP-786 and patients who are randomized to Placebo in Stage 1 but terminated prior to rerandomization, are not part of this population (See Study schema).

4.4.2 Safety 12-Week Parallel Group Population

The Safety 12-Week Parallel Group population are those patients in the 12-Week Parallel Group population who received at least one dose of study medication.

4.4.3 Per-Protocol 12-Week Parallel Group Population

The PP 12-Week Parallel Group population is defined as those patients in the mITT 12-Week Parallel Group population who are also in the PP population.

5.0 Overall Statistical Considerations

5.1 Definition of Baseline Value and Post- Baseline Value

Baseline value 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 specifics:

- For the primary efficacy analysis and secondary efficacy analyses on the mITT or PP populations, Stage 1 Baseline value 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 value 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.
- Baseline values used for supplementary analyses on the mITT 12-Week Parallel Group population are derived from the last assessment prior to first dose of study drug.
- 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.

5.2 Definition of End of Treatment

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

5.3 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.

5.4 Discontinuation of Study

A patient will be considered discontinued from the study when a Study Exit CRF page is completed indicating primary reason for discontinuation. Discontinuation will be classified into Stage 1 or Stage 2 according to when a patient discontinues.

5.5 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 more than one observation falls within a particular study day interval, then the last observation within that interval is used. Evaluations occurring more than 3 days after the last double-blind dosing date will not be mapped into study visit windows, and will be excluded from the analysis.

Study visit windows will be used to map visits using study day intervals. This visit window convention applies to tables and listings for all efficacy and safety scales (NPI-C-3, NPI-C, mCGI-S, mCGI-C, PGI-C, PGI-S, and S-STS). This derived study window variable will be named as WEEK and will be footnoted. In listings it will be listed along with the CRF study visit.

Separate visit window tables are provided based on how frequently the assessments were done. Table 4 and Table 5 are divided into patients who have advanced to Stage 2 and those that have not. Table 6 is for the patients who are in 12-week parallel group and patients who were randomized into AVP-786 group in Stage 1 randomization (Study Schematic B and C).

Table 4 shows classifications for study day intervals in Stage 1. The variable “Target Day” is defined using the number of days since the start of double-blind dosing in Stage 1. The first day of double-blind dosing in Stage 1 is defined as “Day 1”. Baseline observations are the last available observations that are less than or equal to “Day 1” first dose of double-blind dosing in Stage 1.

Table 4 Study Day and Visit Window Stage 1

Week	Target Day [a]	Study Day Interval [a]
1	7	2-10
2	14	11-17
3	21	18-24
4	28	25-31
5	35	32-38
6	42	39-45 [b]

[a] Relative to the first day of double-blind study medication in the double-blind treatment period of Stage 1.
[b] Evaluations occurring more than 3 days after the last dosing date of double-blind study medication in the double-blind treatment period will be excluded from the efficacy analyses.

Table 5 shows classifications for study day intervals in Stage 2. The variable “target day” is defined using the number of days since the start of double-blind dosing in Stage 2. The first day

of double-blind dosing in Stage 2 is defined as “Day 1”. An end of Stage 1 observation is the last available observation in Stage 1 on or before “Day 1”. If that measurement took place prior to the lower bound of the Week 6 study visit window, i.e. day 39, see Table 3, it does not qualify as an End of Stage 1 measurement.

Table 5 Study Day and Visit Window Stage 2

Week	Target Day [a]	Study Day Interval [a]
Baseline		re-randomization visit
7	49	46-52
8	56	53-59
9	63	60-66
10	70	67-73
11	77	74-80
12	84	81-87 [b]

[a] Relative to the first day of double-blind study medication in the double-blind treatment period of Stage 2.
[b] Evaluations occurring more than 3 days after the last dosing date of double-blind study medication in the double-blind treatment period of Stage 2 will be excluded from the efficacy analyses.

Table 6 shows classifications for study day intervals in the double-blind period for 12-Week treatment group. The variable “target day” is defined using the number of days since the start of double-blind dosing. The first day of double-blind dosing is defined as “Day 1”. If more than one observation falls within a particular study day interval, then the last observation within that interval is used. Evaluations occurring more than 3 days after the last double-blind dosing date will not be mapped into study visit windows and will be excluded from the analysis.

Table 6 Study Day and Visit Window 12-Week Group

Week	Target Day [a]	Study Day Interval [a]
1	7	2-10
2	14	11-17
3	21	18-24
4	28	25-31
5	35	32-38
6	42	39-45
7	49	46-52
8	56	53-59
9	63	60-66
10	70	67-73
11	77	74-80
12	84	81-87 [b]

[a] Relative to the first day of double-blind study medication in the double-blind treatment period.
[b] Evaluations occurring more than 3 days after the last dosing date of double-blind study medication in the double-blind treatment period will be excluded from the efficacy analyses.

5.6 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 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 5.5. 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 Appendices 2 and 3. These will apply to AE and concomitant medications when applicable.
- The primary endpoint NPI-C-3 composite score is set as missing if any of the 3 subscales is missing.
- For SPCD and by-stage and visit efficacy analyses (excluding observed case analyses), missing data will be imputed within a study stage. The table below provides imputation rules for last observation carried forward (LOCF)

Stage	LOCF
Stage 1	Impute with last non-missing value within Stage 1 excluding Baseline.
Stage 2	Impute with last non-missing value within Stage 2 excluding Stage 2 Baseline.

5.7 Treatment Misallocations

When mITT or PP (including parallel groups) populations are used, then 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.

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

5.8 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.

5.9 Data Listings

Unless otherwise specified, data listings will be provided on observed values. Efficacy listings for the primary and secondary endpoints will be flagged for imputations. In addition, the analysis visit will be displayed if it is important for the corresponding summary. Listings will be provided to serve as support for all summary tables or figures.

5.10 Graphical Displays

Supporting figures may be used for some efficacy 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.

5.11 Hypothesis Testing

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

5.12 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.

5.13 Definition of Age

The following SAS® code will be used to calculate patient age (years):

```
Age = floor ([intck{'month', birth date, screen date} - {day(informed consent date)
< day(birth date)}] / 12),
```

where intck is a SAS® function counting integer days.

5.14 Study Day

Study day is defined relative to the date of the first dose of study drug. For assessments that occur after this visit date, study day is calculated as (assessment date – study drug first dose date +1). For assessments that occur prior to study drug first dose date, study day will be calculated as (assessment date – study drug first dose date); there is no Study Day 0.

6.0 Interim Analyses

No Interim analysis is planned for this trial.

7.0 Data Handling and Transfer

Data will be entered by investigational sites into a clinical database built with Bioclinica and exported as SAS® version 9.4 or higher datasets (SAS Institute, Inc., Cary, NC). Converted datasets are created using SAS® and following Clinical Data Interchange Standards Consortium (CDISC) Standard Data Tabulation Model conventions (v3.1.3 implementation guide v1.2). Derived analysis datasets are generated using SAS® and following standard CDISC Analysis Dataset Model conventions (implementation guide v1.0). Data analyses including summary tables, figures, and listings (TFLs) are produced using SAS®.

Medical history and AEs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) version 18.1 to assign a system organ class (SOC) and preferred term (PT) to each AE. AE severity will be graded using National Cancer Institute (NCI; US) Common Toxicity Crteria for Adverse Events (CTCAE) version 4.03.

Prior and concomitant medications are coded to preferred drug names using the World Health Organization Drug Dictionary Enhanced (WHODRUG DDE, 2015SEP01).

8.0 Data Screening

The programming of analysis datasets and TFLs provides additional data screening. Presumed data issues will be output into SAS logs identified by the word "Problem" and extracted from the logs by a SAS macro and sent to Data Management.

Review of a pre-freeze TFL run on clean patients and a post-freeze TFL run on the frozen database allow for further data screening prior to lock.

9.0 Data Safety Monitoring Board (DSMB)

Not applicable.

10.0 Efficacy Statistical Analysis Methods

The study schematic in Figure 1 should 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 NPI-C-3 composite score) for the SPCD, 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 NPI-C-3 composite 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 group and Stage 2 mITT population. Study segments E and F will be compared. Stage 1 treatment effect will be estimated by Week 6 CFBL and the Stage 2 treatment effect will be estimated by the Week 12 CFBL (baseline defined as stage 2 baseline). The model will include terms for treatment, visit, treatment-by-visit interaction, baseline NPI-C-3 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:

```
proc mixed data=dset method=reml covtest;
```

```
class subject trt visit site;
model change = trt visit baseline trt*visit baseline*visit site/ ddfm = kr;
repeated visit / type=un subject=subject;
lsmeans trt*visit/cl pdiff;
```

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 $w = 0.7$ is used for Stage 1, which means a weight of $1 - w = 0.3$ 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)));
```

The primary efficacy endpoint will be tested at 0.05 significance level. A hierarchical testing procedure will be used in the order of NPI-C-3 SPCD score calculated from Stage 1 and Stage 2, Stage 1 NPI-C-3 score, and Stage 2 NPI-C-3 Score. If the test of NPI-C-3 SPCD score is significant, then NPI-C-3 score from Stage 1 will be tested at 0.05 significance level; the NPI-C-3 score from Stage 2 will be tested last upon proven significance of the NPI-C-3 at Stage 1. The NPI-C-3 change from baseline as well as the component score will be plotted over time by treatment.

10.2 Sensitivity Analyses for the Primary Efficacy Endpoints

10.2.1 Sensitivity Analysis with MNAR Using Multiple Imputation (MI)

MMRM assumes data are MAR, which is a reasonable assumption in longitudinal clinical trials. However, the possibility of missing not at random (MNAR) data can never be ruled out. In order to further evaluate robustness of the primary results to deviations from MAR assumptions, additional sensitivity analysis may be conducted. Sensitivity analyses based on selection model, pattern-mixture model, and/or shared parameter model will be performed in order to explore data missing mechanisms of MNAR and investigate the response profile of dropout reason.

Pattern Mixture Models (PMM) based on MI with mixed missing data mechanisms will be used to investigate the response profile of dropout subjects by last dropout reason under MNAR mechanism for the following two scenarios:

1) Dropout reasons due to AE as MNAR

2) All dropouts as MNAR

Delta Adjustment Imputation Methods

This MNAR sensitivity analysis is to departure from MAR assumption by progressively increasing the delta until conclusion from the primary analysis is overturned. The delta is 0%,

5%, 10%, 15%, .., 100% of the observed treatment difference in both stages between AVP-786 and Placebo from the primary analysis of MMRM model until conclusion of the primary analysis is overturned. When delta = 0 the missing data are assumed to be MAR. When delta > 0, the missing data are assumed to be MNAR.

1) Using Monte Carlo Markov Chain (MCMC) methodology from PROC MI to impute the intermittent missing data to a monotone missing pattern;
2) Using a standard MAR-based multiple imputation approach from PROC MI to impute the monotone missingness data
3) For subjects in the group treated with AVP-786 and with a dropout reason of AE, a delta will be added for all the values after the dropout time.
4) Using MMRM model in the primary analysis to analyze the completed data using PROC MIXED on the multiple imputed data
5) Obtaining the overall results using PROC MIANALYZE.

Placebo Based Imputation Methods

Similar to "Standard" multiple imputations, except parameters for imputation model obtained from only the placebo (control) group. Missing data for both placebo and drug group are imputed based on the imputation model derived from placebo data within each stage. If drug improved outcomes prior to dropout, this benefit is carried into subsequent imputed values, but will diminish over time in accordance with the correlation structure.

10.2.2 SPCD MMRM on the Per-Protocol Population for the NPI-C-3

The analysis described in Section 10.1 will be repeated for the PP population.

10.2.3 Effect of COVID-19 on the Primary Efficacy Endpoint

In earlier 2020, an outbreak of respiratory disease caused by a novel coronavirus named "Coronavirus Disease 2019" (COVID-19) had widely spread all over the world. On March 13th, 2020, the President of the United States declared a national emergency in response to COVID-19. It is after this date, centers enrolled in this study had implemented remote visits in response to this pandemic. In order to assess the effects of this change on the primary efficacy endpoint, subgroup analysis is planned to assess the effect of COVID-19 on the primary analysis.

The COVID subgroup is defined for all randomized subjects: subjects randomized after March 13th 2020 will start the remote access or remote changes for their visits and will be classified as COVID group = COVID; subjects randomized on or before March 13th 2020 will have COVID group = PRE-COVID. The COVID-19 related changes will include but not limit to remote visit, missed visit (especially for efficacy assessments), missed assessment, discontinuation due to COVID-19.

The analysis described in Section 10.1 will be repeated for the subgroup analysis by COVID group. For subjects who are in pre-COVID group, if the are still ongoing after March 13th 2020, their NPI-C-3 measurments will be set to missing in this analysis.

10.3 Secondary Efficacy Analyses

The following secondary endpoints will be analyzed using the methods described in the sections below.

Mean score at Week 6 (Stage 1) and Week 12 (Stage 2) in the following:

- Modified Clinical Global Impression of Change (mCGI-C from baseline at first randomization and re-randomization respectively)
- Patient Global Impression of Change (PGI-C Raw score)

Mean Change from Baseline to Week 6 (Stage 1) and from Week 6 to Week 12 (Stage 2) in the following:

- NPI-C subscale scores (Agitation, Agreesion, Irritability/Liability and Disinhibition)
- Modified Clinical Global Impression of Severity (mCGI-S) score
- Patient Global Impression of Severity (PGI-S) score

10.3.1 Secondary SPCD MMRM and SPCD OLS ANCOVA

All quantitative secondary endpoints listed above will be analyzed using the mITT population.

The NPI-C subscales measured at baseline, Week 3, 6, 9 and 12, will be analyzed using the SPCD MMRM method as described in Section 10.1.

For the mCGI-C, PGI-C, mCGI-S and PGI-S, an ANCOVA model will be run using LOCF data. Note that baseline NPI-C-3 composite score depends on the study stage and is defined in Section 5.1.

SPCD OLS ANCOVA models will be run using data from Stage 1 and Stage 2. Study segments A and B will be compared in the Stage 1 model and study segments E and F (i.e., Stage 1 Placebo Non-Responders) will be used to compare Stage 2 data. The model will include treatment as a factor and baseline NPI-C-3 composite score, site as covariates.. The SAS® code for the models is shown below:

```
proc glm data = dset;
 class treatment site;
 model change = treatment baseline site/ solution;
 lsmeans treatment / pdiff stderr cl cov out = dset;
```

The parameter estimates will be combined in each stage in the same way described for the primary efficacy analysis.

Summary statistics will include the change and percent CFBL for each stage. Also included in the ANCOVA tables will be the standard effect size, which will be calculated using the following formula:

$$\text{Effect size} = \frac{\text{Mean}(\text{CFBL}_{\text{AVP-786}}) - \text{Mean}(\text{CFBL}_{\text{PBO}})}{\text{SD}(\text{CFBL}_{\text{POOL}})}$$

where the pooled SD of the CFBL is calculated from the within-treatment standard deviations of the CFBL as follows:

$$\mathrm{SD}(\mathrm{CFBL}_{\mathrm{POOL}}) = \sqrt{\frac{(n_{\mathrm{PBO}} - 1)s^2_{\mathrm{PBO}} + (n_{\text{AVP-786}} - 1)s^2_{\text{AVP-786}}}{n_{\mathrm{PBO}} + n_{\text{AVP-786}} - 2}}$$

Model estimates (LS mean treatment difference, confidence interval and p-value) will be reported for each stage, and an overall p-value based on the OLS Z-statistic will also be provided, which uses the similar formula as Z_{MMRM} described in Section 10.1.

Note that this analysis will also contain the result of the Stage 1 ANCOVA for the Stage 1 mITT Population. In addition, the comparison between placebo and AVP-786 for the placebo non-responders only in Stage 2 will be available from this analysis.

10.4 Exploratory Analyses

10.4.1 NPI-C-3 Composite Score Response Analysis

The number and percentage of patients who have favorable treatment response according to the NPI-C-3 composite score will be summarized using the mITT population (Stage 1 plus the Stage 1 Placebo Non-Responder Subset in Stage 2), LOCF data. The following categories will be used to classify response patients:

- Response: Patients with a ≥ 30% reduction in the NPI-C-3 composite score
- Response: Patients with a ≥ 40% reduction in the NPI-C-3 composite score
- Response: Patients with a ≥ 50% reduction in the NPI-C-3 composite score

The number and percentage of response will be provided by stage and treatment group. Overall Stage 1 and Stage 2 treatment differences will be tested via SPCD one degree of freedom score test assuming Stage 2 and Stage 1 treatment effect ratio ρ=1 (Ivanova et al., 2011). In addition, the treatment effect will also be tested at each visit using Chi-square or Fisher's Exact tests (if expected cell counts are < 5). Ralative risk and p-values based on Cochran-Mantel-Haenszel (CMH) general association test controlling for study center will be explored and provided.

10.4.2 12-Week Parallel Group Efficacy Analysis

To evaluate the treatment effect under 12 weeks of exposure to the same treatment, a repeated measures analysis of the NPI-C-3 composite score using observed data from all scheduled visits will be performed on the mITT 12-Week Parallel Group Population and PP 12-Week Parallel Group Population. This analysis will compare treatment groups over time using MMRM, observed data. The model will include fixed effects for treatment, visit, site, treatment-by-visit interaction, baseline NPI-C-3 composite score, and baseline-by-visit interaction. The same selection of covariance matrix as the primary MMRM (i.e., unstructured as first preference) will be employed. The model statement is shown below:

```
proc mixed data=dset method=reml;
 class subject trt visit site;
 model change = trt visit baseline trt*visit baseline*visit site;
 repeated visit / type=un subject=subjid;
 lsmeans trt*visit/cl pdiff;
run;
```

The analysis above will be performed on the NPI-C-3 composite score in addition to all secondary endpoints for the mITT 12-Week parallel group population, observed data.

In addition to evaluate the efficacy of one dose of AVP-786 42.63/4.9 mg twice daily compared with placebo on NPI-C-3 composite score at Week 6, exploratory analysis will be performed using the above MMRM model on:

- mean change from Baseline to Week 6 (Stage 1) in NPI-C3 (mITT population)

10.4.3 mCGI-C and PGI-C Proportional Odds

For the mCGI-C and PGI-C, odds ratios (ORs) will be calculated at week 6 and 12 for the odds of moderate improvement or better response (defined as responder) on AVP-786 vs. placebo. This will be performed on the mITT population and mITT 12-Week parallel group population using LOCF data, only Stage 1 (A and B) and the Stage 1 Placebo Non-responders in Stage 2 (E and F) will be included in the analyses. For stage 2, the CGI-C and PGI-C relative to Week 6 will be used.

The ORs will be obtained through a proportional odds regression model. The following model statements can be used:

```
proc logistic data = dset;
class treatment site(param=ref ref='0');
model var (order=internal) = treatment site/ link=logit aggregate =
      (treatment) scale = none;
ods output type3 = dset oddsratios = dset;
run;
```

ORs greater than 1 indicate an overall increased likelihood of a favorable response in the AVP-786 group.

The same analysis will be repeated for the mITT 12 Week Parallel Group population. In this case, compares A (for segments C and E) versus B

10.4.4 SPCD Analysis and 12-week parallel group analysis by Subgroup

The primary endpoint will be analyzed using the subgroups defined by each category below for the mITT population and mITT 12-Week parallel group population, using MMRM, observed data, the same model as described in Section 10.1 and Section 10.4.2, but not controlling for study site.

- Age group (< 45, ≥ 45)
- Gender (M, F)

11.0 Safety Statistical Analysis Methods

Safety will be assessed through the analysis of AEs, clinical laboratory assessments, ECGs, vital signs, physical and neurological examinations, S-STS. All safety analyses will be completed on the Safety Population. Safety measures will be summarized by Stage 1, Stage 2, and the 2 stages combined.

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).
- AVP-786/AVP-786: patients receiving AVP-786 for the entire duration of the study (B).
- 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 any placebo (A, C, and E).
- All AVP-786: This includes data from the stages when patients received any AVP-786 (B, D, F).

Placebo/Placebo and AVP-786/AVP-786 treatment groups summarize the safety information for the Safety 12-Week Parallel Group population, which is under 12 week treatment exposure. It is what would be summarized if the study had been a 12 week parallel group design and conventional statistical test to compare these 2 groups can be performed if needed.

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 6.1, 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 treatment-emergent adverse events (TEAEs). TEAEs 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 ≤ AE start date ≤ last dose date + 30 days).

TEAEs related to study treatment will be defined as those entered as possibly related or related. Unlikely related or not related entries will be considered not related to study treatment. If relationship is missing, then it will be considered as related to study treatment.

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

TEAEs, drug-related TEAEs, SAEs, drug-related SAEs, TEAEs leading to discontinuation, and non-serious TEAEs will be summarized by SOC and PT. TEAEs and drug-related TEAEs will also be summarized by PT ordered by descending frequency of PT. All TEAEs will be summarized by SOC, PT, and by maximum severity. Summaries of PTs will also be done for those occurring in at least 5% of patients in any treatment or study stage.

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. The number of patients with at least one recurrence of common TEAEs will be summarized by PT.

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 multiple AEs of the same SOC and PT start in both study stages, it should be counted under both placebo and AVP-786.
- For patients who switched from placebo to AVP-786: If one AE starts in Stage 1 and ends in Stage 2 or has an unknown end date, it should be counted under Stage 1 (i.e. placebo).
- 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.
- A common TEAE is defined as a TEAE an incidence of ≥ 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 date of follow up + 14 days.

  - 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 24.1.

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 6 / Early Termination)
- 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). Safety analysis will be performed by treatment group and stage. 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. Shift tables will also be created for Placebo/Placebo and AVP-786/AVP-786 for the Safety 12-Week Parallel Group population.

Potentially clinically significant (PCS) tables will also be used to summarize out-of-range values. The PCS values are found in Table 7. 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 7 Laboratory 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) (two ECGs pre-dose)

- Baseline/Visit 1 (Day 1) (2-3 hours post-dose)
- Visit 2/Week 1 (Day 8)
- Visit 3/Week 3 (Day 22)
- Visit 4/Week 6 (Day 43) (two ECG, pre-dose)
- Visit 4/Week 6 (Day 43) (2-3 hours post-dose)
- Visit 5/Week 9 (Day 64)
- Visit 6/Week 12/Early Termination (Day 85) (pre-dose)

The following quantitative parameters will be reported by the central reader: heart rate, PR interval, QRS duration, QT interval (uncorrected), and QT interval due to 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 12.0, except All Placebo and All AVP-786 groups. Note that for patients who switched from placebo to AVP-786, baseline is defined as described in Section 5.1. Screening results will not be included in this summary.

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 8 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.

Early termination and unscheduled visits, as well as multiple visit per visit window, will be included based on Section 5.5 Visit Windows.

Aside from summaries described above, ECG results and changes from based, as well as central and local lab interpretations will be listed.

Table 8 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, measured only at Baseline and Visit 6 (Week 12), will also be summarized. These parameters will be summarized through CFBL and percent CFBL in similar fashion as the ECG parameters.

Vital signs will also be assessed through PCS criteria, which are given in Table 9. 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 9 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

Vital sign results and changes from baseline will be listed. Post-Baseline vital sign PCS results and change from baseline will be listed separately.

11.5 Physical and Neurological Exams

Physical and neurological exams are assessed at Screening, Week 6 and Week 12. Findings will only be provided in listings. Physical exam clinically significant abnormal findings will be listed separately.

11.6 Sheehan Suicidality Tracking Scale (S-STS)

The S-STS is a prospective scale that assesses treatment-emergent suicidal thoughts and behaviors. Each item of the S-STS is scored on a 5-point Likert scale as shown below:

S-STS Item Scale:

0 = Not at all
1 = A little
2 = Moderate
3 = Very
4 = Extremely

The S-STS will be assessed at all clinic visits. Scores, CFB in scores, and percent CFBL in scores will be summarized descriptively by visit and treatment group for each subscale and total score.

S-STS suicidal ideation subscale score, suicidal behavior subscale score, and a total score will be calculated as follow:

Parameter	Description of Derivation	Programming Algorithm	Missing Data Handling
Suicidal Ideation	Sum of: Questions 2 - 11; Questions are on a 0-4 scale	Sum of: SSTS102, SSTS103, SSTS104, SSTS105, SSTS106, SSTS107, SSTS108, SSTS109, SSTS110, SSTS111	if any missing for Questions 2 - 11, then score will be missing
Suicidal Behavior	If Question 1b=Yes, then use the following algorithm: Sum of: Questions 1a, highest score of (Q12 or any row of Q16), highest score of (Q14 or any row of Q15), 17, and 20 If Question 1a is not present, then use the following algorithm: Sum of: highest score of (Q12 or any row of Q16), highest score of (Q14 or any row of Q15), 17, and 20 Notes:	If SSTS101B=Yes (looks like this is only present in data if =Yes, then do: Sum of: SSTS101A, max(SSTS112 or SSTSQ16C or SSTSQ16D), max(SSTS114 or SSTSQ15C or SSTSQ15D), SSTS17=Yes add 100, SSTS20=Yes add 4 If SSTS101B not present or No, then do: Sum of: max(SSTS112 or SSTSQ16C or SSTSQ16D), max(SSTS114 or SSTSQ15C or SSTSQ15D), SSTS17=Yes add 100, SSTS20=Yes add 4	if Questions 12 or 14 are missing, then score will be missing

	Questions 15, 16, 17, and 20 are optional and only included in the QS datasets if populated with data. Scores should still be calculated if these Q15, Q16, Q17, Q20 are missing.		
Total Scale Score	If Question 1b=Yes, then use the following algorithm: Sum of: Questions 2 - 11, Questions 1a, highest score of (Q12 or any row of Q16), highest score of (Q14 or any row of Q15), 17, and 20 If Question 1a is not present, then use the following algorithm: Sum of: Questions 2 - 11, highest score of (Q12 or any row of Q16), highest score of (Q14 or any row of Q15), 17, and 20 Notes: Questions 15, 16, 17, and 20 are optional and only included in the QS datasets if populated with data. Scores should still be calculated if these Q15, Q16, Q17, Q20 are missing.	If SSTS101B=Yes (looks like this is only present in data if =Yes, then do: Sum of: SSTS102, SSTS103, SSTS104, SSTS105, SSTS106, SSTS107, SSTS108, SSTS109, SSTS110, SSTS111, SSTS101A, max(SSTS112 or SSTSQ16C or SSTSQ16D), max(SSTS114 or SSTSQ15C or SSTSQ15D), SSTS17=Yes add 100, SSTS20=Yes add 4 If SSTS101B not present or No, then do: Sum of: SSTS102, SSTS103, SSTS104, SSTS105, SSTS106, SSTS107, SSTS108, SSTS109, SSTS110, SSTS111, max(SSTS112 or SSTSQ16C or SSTSQ16D), max(SSTS114 or SSTSQ15C or SSTSQ15D), SSTS17=Yes add 100, SSTS20=Yes add 4	if any missing for Questions 2 - 12, 14, then score will be missing

12.0 Additional Summaries

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

The columns will be treatment as randomized as opposed to actual medication received.

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, mITT 12-Week Parallel Group and Safety populations will be provided.

12.2 Protocol Deviations

Potential planned or unplanned protocol deviations noted during clinical monitoring will be documented by category (i.e., inclusion criteria, exclusion criteria, study drug, safety assessment, efficacy assessment, visit window, informed consent, prohibited medication, and other). All deviations will be reviewed, categorized, and finalized prior to database lock. All protocol deviations will be listed.

12.3 Demographic and Baseline Characteristics

Demographics will be summarized by randomized treatment group and overall for the mITT, mITT 12-week parallel group, 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 (< 45, $45 \leq$ and < 60, ≥ 60)
- Weight
- Time since TBI (months)
    - $<$ 1 year
    - 1-5 years
    - 5-10 years
    - $>$ 10 years
- TBI Severity
    - Mild
    - Moderate
    - Severe
- Time since most recent TBI (months)
- Most recent TBI Severity
    - Mild
    - Moderate
    - Severe

- Time since TBI used for eligibilitiy (months)
- TBI used for eligibilitiy severity
    - Mild
    - Moderate
    - Severe

- Concomitant medications

For subjects with mssing TBI dates, if the day is missing the first day of the month will be used; if the month is missing then July will be used to impute the month; if year is missing no immputation will be performed. A summary of baseline characteristics will include baseline values of all primary and secondary endpoints except mCGI-C, and PGI-C, 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.

12.4 Medical History

Medical history will be summarized by SOC and PT by Stage 1 randomization and will be presented in the listings. The Safety Population will be used. Rows will be sorted by decreasing frequency of "All Patients" column for both SOC and PT.

12.5 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 24 mg QD, 24 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 Terminated After Re-randomization	Patients Terminated Prior to Re-randomizations
Stage 1	2 × (re-randomization day – first dose day) + 1	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 added 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.

12.6 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. If the stop date is missing then it is assumed to be after first dose date. Note that medication that is ongoing at the time of first dose is also considered concomitant. Partial dates will be imputed using rules in Appendix 3. Medications are defined as prior or concomitant, but not both.

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 PT.

13.0 Validation

ICON/PRA's goal is to ensure that each TFL delivery is submitted to the highest level of quality. Our quality control procedures will be documented separately in the study specific quality control plan.

14.0 References

1. Yeh-Fong Chen, Yang Y., James Hung HM, Wang S-J. Evaluation of performance of some enrichment designs dealing with high placebo response in psychiatric clinical trials. Contemporary Clinical Trials 2011;32:592-604.

2. Roy N Tamura, Huang X. An examination of the efficacy of the sequential parallel design in psychiatric clinical trials. Clinical Trials 2007;4:309-17

3. Ivanova Anastasia, Qaqish, B, Schoenfeld, DA. Optimality, sample size, and power calculations for the sequential parallel comparison design. Statistics in Medicine 2011;30:2793-2803.

Appendix 1 Glossary of Abbreviations

Abbreviation	Term
AE	Adverse event
ALT/SGPT	Alanine aminotransferase/serum glutamic pyruvic transaminase
ANCOVA	Analysis of covariance
AR1	First-order autoregressive
AST/SGOT	Aspartate aminotransferase/serum glutamic oxaloacetic transaminase
ATC	Anatomical therapeutic chemical (classification)
BID	Twice daily
BUN	Blood urea nitrogen
CDISC	Clinical Data Interchange Standards Consortium
CFBL	Change from baseline
CGI-C	Clinical Global Impression of Change
CGI-S	Clinical Global Impression of Severity
CI	Confidence interval
CK	Creatinine kinase
CMH	Cochran-Mantel-Haenszel
CRF	Case report form
CTCAE	Common Terminology Criteria for Adverse Events
CYP	Cytochrome P450
D6-DM	Deuterated (d6)-dextromethorphan hydrobromide
DBP	Diastolic blood pressure
DSMB	Data Safety Monitoring Board
ECG	Electrocardiogram
GGT	Gamma-glutamyl transferase
HbA1c	Glycosylated hemoglobin
LDH	Lactate dehydrogenase
LS	Least squares
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified Intent-to-Treat
MMRM	Mixed effects model repeated measures
MI	Multiple Imputation
NCI	National Cancer Institute (US)
NPI	Neuropsychiatric Inventory
NPI-C	Neuropsychiatric Inventory- Clinician Rating Scale

Abbreviation	Term
NPI-C-3	Neuropsychiatric Inventory- Clinician Rating Scale composite score
OLS	Ordinary least squares
OR	Odds ratio
PCS	Potentially clinically significant
PGI-C	Patient Global Impression of Change
PGIS	Patient Global Impression of Severity
PHQ-9	Patient Health Questionnaire
PK	Pharmacokinetics
PP	Per-Protocol
PT	Preferred term
Q	Quinidine sulfate
QD	Once daily
QTcF	QT interval corrected for heart rate using the Fridericia's formula
RBC	Red blood cell
SAE	Serious adverse event
SAP	Statistical analysis plan
SBP	Systolic blood pressure
SD	Standard deviation
SOC	System organ class
SPCD	Sequential parallel comparison design
S-STS	Sheehan Suicidality Tracking Scale
SUR	Seemingly unrelated regression
T3	Triiodothyronine
T4	Thyroxine
TEAE	Treatment-emergent adverse event
TFLs	Tables, figures, and listings
TSH	Thyroid stimulating hormone
UN	Unstructured covariance matrix
VAS	Visual analogue scale
WBC	White blood cells
WHODRUG DDE	World Health Organization Drug Dictionary Enhanced

Appendix 2 Adverse Event Start/Stop Date Imputation

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

Parameter	Missing	Additional Conditions	Imputation
Start date for AEs	D	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
	D and M	Y same as Y of first dose of study drug	Date of first dose of study drug
		Y prior to Y of first dose of study drug but same as Y of screening date	Date of screening date
	D, M, Y	None - date completely missing	Date of first dose of study drug
Stop date for AEs	D	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	Use last day of month
	D and M	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
	D, M, Y	None - date completely missing	No imputation, but assume ongoing

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.

Appendix 3 Prior and Concomitant Medication Start/Stop 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.

Appendix 4 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

This page is a manifestation of an electronically captured signature

SIGNATURE PAGE

Document Name: AVP-786-205 Statistical Analysis Plan_Final

Document Number: 1000205368

Document Version: 2.0

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy hh:min) - UTC timezone
PPD	Biostatistics Approval	23-Aug-2022 11:32:01
PPD	Clinical Approval	23-Aug-2022 17:59:16