[redacted]

Statistical Analysis Plan

Sponsor:	Avanir Pharmaceuticals, Inc.
Protocol No:	15-AVP-786-301
Protocol Version No./ Date	Amendment 3 16-May-2016
Title	A Phase 3, multicenter, randomized, double-blind, placebo-controlled study to assess the efficacy, safety, and tolerability of AVP-786 (deuterated [d6]-dextromethorphan bydrobromide [d6-DM]/quinidine sulfate [Q]) for the treatment of agitation in patients with dementia of the Alzheimer's type
CRF Version No./ Date	6.0 19-Aug-2016
[redacted]	
SAP Version No./ Date:	3.0 16-Nov-2018

Approvals

Prepared by

[redacted]
Biostatistician/ Title:
Signature/ Date:

[redacted]
Project Manager/ Title:
Signature/ Date:

Sponsor
Sponsor Name: Avanir Pharmaceuticals, Inc.
Biostatistician / Title: [redacted] Avanir Biostatistics, Clinical Development
Signature/ Date: [redacted]

CDL / Title: [redacted] Avanir Clinical Development
Signature/ Date: [redacted]

CSM / Title: [redacted] Avanir Clinical Research
Signature/ Date: [redacted]

Non-Clinical Dev Lead / Title: [redacted] Avanir Non-Clinical Development
Signature/ Date: [redacted]

NCT Number: NCT02442765
This NCT number has been applied to the document for purposes of posting on Clinicaltrials.gov

[redacted]

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 Avanir Pharmaceuticals, Inc. Protocol 15-AVP-786-301.

This SAP should be read in conjunction with the study protocol and case report form (CRF). This version of the plan has been developed using the Protocol Amendment 3 dated 16-May-2016 and CRF dated 19-Aug-2016. Any further changes to the protocol or CRF may necessitate updates to the SAP.

1.1 Changes from Protocol

Not applicable.

2.0 Study Objectives

The objectives of the study are to evaluate the efficacy, safety, and tolerability of AVP-786 compared to placebo for the treatment of agitation in patients with dementia of the Alzheimer's type.

3.0 Study Design

3.1 General Description

This is a Phase 3, multicenter, randomized, double-blind, placebo-controlled study with a duration of 12 weeks, consisting of 2 consecutive double-blind treatment stages (Stage 1 and Stage 2) of 6 weeks duration each. The study utilizes a sequential parallel comparison design (SPCD) with a re-randomization for Stage 2. Table 1 provides an overview of the study duration for each stage:

TABLE 1. OVERVIEW OF STUDY DURATION

Sequence	Screening	Stage 1	Stage 2
Duration in Days	Days -28 to -1	Days 1 to 42	Days 43 to 85

Approximately 380 patients with a diagnosis of probable Alzheimer's disease (AD) and clinically meaningful moderate/severe agitation secondary to AD will be enrolled at approximately 60 research centers in the United States. There will be 8 scheduled clinic visits including a screening visit and 2 safety follow-up phone visits in the study. Clinic visits include Screening, Baseline (Day 1), and Weeks 1, 2, 3, 6, 9, and 12. Safety follow-up phone calls will occur on Week 4 and Week 10. Visits at Weeks 6, 9 and 12 correspond to Stage 2 Baseline, Week 3 and Week 6. The schedule of events is provided in Table 2:

TABLE 2. SCHEDULE OF EVENTS AND VISITS

Procedure	Visit:	Screening[1]	Baseline	Visit 2[1]	Visit 2.1[1]	Visit 3[1]	Phone Call[1,2]	Visit 4[1]	Visit 5[1]	Phone Call[1,2]	Visit 6[1] /ET[3,4]
	Study Day:	Day -28 to -1	Day 1	Day 8	Day 15	Day 22	Day 29	Day 43	Day 64	Day 71	Day 85
	End of Study Week:	Week -4 to -1		Week 1	Week 2	Week 3	Week 4	Week 6	Week 9	Week 10	Week 12
Sign informed consent forms		X									
Medical history		X									
Review of eligibility[5]		X	X								
Randomization			X								
Physical and neurological examination		X						X			X
Vital signs and weight		X	X[6]	X	X	X		X	X		X[6]
ADCS-CGIC-Overall			X[7]					X			X
CGIS-Agitation		X	X					X			X
mADCS-CGIC-Agitation			X[8]					X			X
Risk assessment for falls (worksheet and TUG test)		X						X[9]			X[9]
ECG		X[10]	X[11]		X	X		X[11]	X		X
AEs			X	X	X	X	X	X	X	X	X
Prior and concomitant: medications, nondrug therapies, and nonpharmacological interventions for agitation		X	X	X	X	X	X	X	X	X	X
MMSE		X	X					X			X
GMHR		X									X
CMAI		X	X	X	X	X		X	X		X
NPI		X[12]	X	X[12]	X[12]	X		X	X		X

Procedure	Visit:	Screening[1]	Baseline	Visit 2[1]	Visit 2.1[1]	Visit 3[1]	Phone Call[1,2]	Visit 4[1]	Visit 5[1]	Phone Call[1,2]	Visit 6[1] /ET[3,4]
	Study Day:	Day -28 to -1	Day 1	Day 8	Day 15	Day 22	Day 29	Day 43	Day 64	Day 71	Day 85
	End of Study Week:	Week -4 to -1		Week 1	Week 2	Week 3	Week 4	Week 6	Week 9	Week 10	Week 12
CSDD		X						X			X
ZBI			X					X			X
DEMQOL[13]			X					X			X
ADAS-cog[14]			X					X			X
PGIC[15]								X			X
RUD			X					X			X
ESS			X					X			X
S-STS		X	X	X	X	X		X	X		X
Administer morning dose of study medication in clinic			X	X[16]	X[16]	X		X	X		X
Chemistry, hematology, and urinalysis		X[17]				X		X	X		X[17]
Urine pregnancy test[18]		X	X					X			X
PK blood sample								X			X
[redacted]			X								
Dispense study drug and diary card			X			X		X	X		
Review and return unused study medication and diary card				X[16]	X[16]	X		X	X		X

ADAS-cog = Alzheimer's Disease Assessment Scale-cognitive subscale; ADCS-CGIC-Overall = Alzheimer's Disease Cooperative Study-Clinical Global Impression of Change for Overall Clinical Status; AE = adverse event; CGIS-Agitation = Clinical Global Impression of Severity of Illness scale for Agitation; CMAI = Cohen-Mansfield Agitation Inventory; CSDD = The Cornell Scale for Depression in Dementia; DEMQOL = Dementia Quality of Life scale; ECG = electrocardiogram; ESS = Epworth Sleepiness Scale; ET = early termination; GMHR = General Medical Health Rating; mADCS-CGIC-Agitation = modified Clinical Global Impression of Change scale for Agitation; MMSE = Mini- Mental State Examination; NPI = Neuropsychiatric Inventory; PGIC = Patient Global Impression of Change rated by the caregiver; PK = pharmacokinetics; RUD = Resource Utilization in Dementia; S-STS = Sheehan Suicidality Tracking Scale; TUG = Timed Up and Go; ZBI = Zarit Burden Interview

Note: Whenever possible, each patient and caregiver should have the rating scales administered by the same raters throughout the study, for consistency of ratings. The following scales MUST be administered by the same rater at each visit: CMAI, NPI, mADCS-CGIC-Agitation, and CGIS-Agitation.

1 Study visits have a +/- 3-day window except Screening, Visit 2, and phone calls. Screening, Visit 2, and phone calls have a +3-day window. The screening period may be extended after discussion with and approval by the medical monitor.

2 Phone call should be made to patient/caregiver to collect adverse events and query on concomitant medication use

3 Early termination visit for patients who withdraw prior to study completion

4 Patients who terminate early from the study or who do not roll over to the extension study (Study 15-AVP-786-303) will receive a safety follow-up phone call 30 days after the last dose of study medication.

5 For each patient, a protocol eligibility form will be completed

6 Weight should be measured only at the Baseline Visit and Visit 6

7 The ADCS-CGIC-Overall baseline evaluation worksheet should be completed to record baseline information for assessing change at Visits 4 and 6

8 The mADCS-CGIC-Agitation baseline evaluation worksheet should be completed to record baseline information for assessing change at Visits 4 and 6

9 Only the TUG test should be performed for risk assessment of falls at Visits 4 and 6

10 ECG should be performed in triplicate at the Screening visit

11 ECG to be performed pre-dose and post-dose

12 Only the Agitation/Aggression domain of the NPI should be performed at the Screening Visit, Visit 2, and Visit 2.1

13 The proxy version is to be rated by the caregiver. The non-proxy version is to be rated only by patients with an MMSE score of ≥ 10 at baseline

14 ADAS-cog is to be rated only by patients with an MMSE score of ≥ 10 at baseline

15 PGIC is to be rated by the caregiver

16 The morning dose of study medication can be administered at home if the visit will occur within 2 hours of dosing; the time of dosing should be noted by the patient/caregiver. The blister card and diary card should be returned to the patient/caregiver after reviewing for compliance.

17 Thyroid function tests (TSH, and reflex T3 and T4 if TSH is abnormal) should be performed at the Screening visit. Glycosylated hemoglobin (HbA1c) test should be performed at the Screening visit and Visit 6.

18 Urine pregnancy test to be performed for females of child bearing potential only

3.2 Stage 1

Following screening procedures for assessment of inclusion and exclusion criteria, eligible patients will be randomized into the study at Baseline in a [redacted] (active:active:placebo) ratio to receive either AVP-786-28/4.9 capsules (referred to as AVP-786 28 mg from this point forward), AVP-786-18/4.9 capsules (referred to as AVP-786 18 mg from this point forward), or matching placebo capsules. The randomization will be stratified by the Neuropsychiatric Inventory (NPI) Agitation/Aggression (AA) domain score (≤ 6 vs. > 6), risk assessment for falls (normal/mild vs. moderate/severe), and concomitant use of antipsychotic medications (yes vs. no). Study medication (active or placebo) will be administered orally twice daily (BID) with approximately a 12 hour separation between doses throughout the treatment period. Dosing schedules for each treatment group are provided below:

- Patients randomized to receive AVP-786 28 mg will start with AVP-786-18 mg once a day in the morning and matching placebo in the evening for the first 7 days of the study. From Day 8, these patients will receive AVP-786 18 mg BID for 14 days. From Day 22, patients will receive AVP-786 28 mg BID for the remaining 3-week duration of Stage 1.
- Patients randomized to receive AVP-786 18 mg will start with AVP-786 18 mg once a day in the morning and matching placebo in the evening for the first 7 days of the study. From Day 8, these patients will receive AVP-786 18 mg BID for the remaining 5-week duration of Stage 1.
- Patients randomized to receive placebo will be dosed with placebo BID during Stage 1.

3.3 Stage 2

Patients who complete Stage 1 are eligible to participate in the 6-week Stage 2 of the study. In Stage 2, patients will be assigned to study treatment as follows:

- Patients who receive AVP-786 in Stage 1 will continue to receive the same dose of AVP-786 BID for the entire 6-week duration of Stage 2.
- Patients who receive placebo in Stage 1 will be re-randomized in a [redacted] (active:active:placebo) ratio in Stage 2 to receive AVP-786 28 mg capsules, AVP-786 18 mg capsules, or matching placebo stratified by two subgroups ("responders" and "non-responders") on Week 6 . A placebo responder is defined as a patient with [redacted] otherwise a placebo patient will be considered a non-responder. Dosing schedule for each treatment group are provided below:
    - 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 28 mg will receive AVP-786 18 mg once a day in the morning and matching placebo in the evening for the first 7 days of Stage 2 (Days 43-49). From Day 50, these patients will receive AVP-786 18 mg BID for 14 days (Days 50-63). From Day 64, these patients will receive AVP-786 28 mg BID for the remaining 3 weeks of Stage 2.
    - Patients who are re-randomized to AVP-786-18 mg will receive AVP-786-18 mg once a day in the morning and matching placebo in the evening for the first 7 days of Stage 2 (Days 43-49). From Day 50, patients will receive AVP-786 18 mg BID for the remaining 5 weeks of Stage 2.
- Patients who receive placebo and who drop out early in Stage 1 will also be assigned a re-randomization treatment in the same manner as the other placebo patients described above. Their "responders" and "non-responders" status will be based on their treatment responses using measurements at their early termination visit.

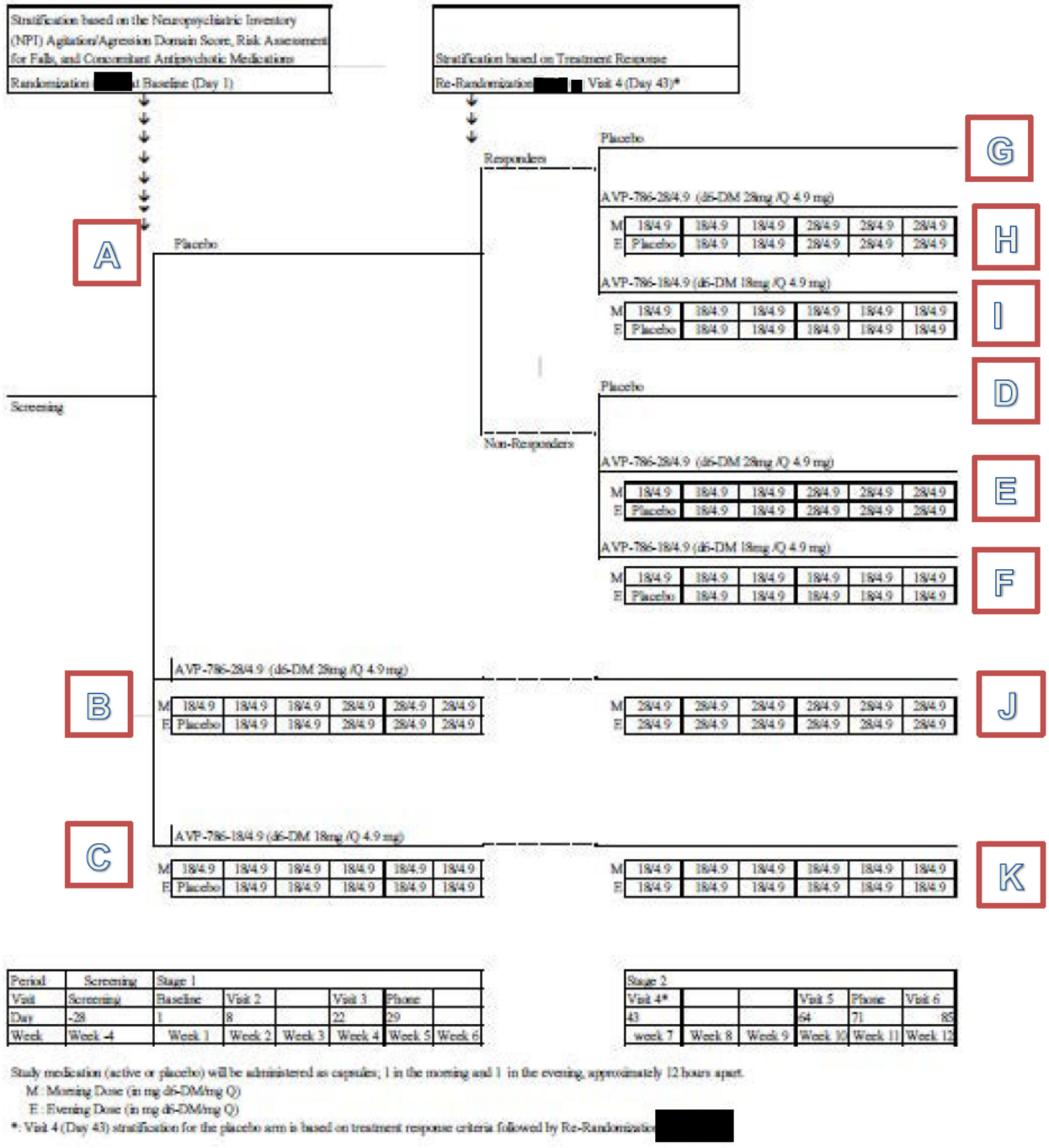
[redacted]

3.4 SPCD Schematic

A schematic of the study along with the treatment segments is shown in Figure 1. The letters for each treatment segment are described below and will be used throughout the SAP for clarification:

- A: Stage 1 Placebo data. A=A1 + A2, A1 and A2 are Stage 1 data for patients who were randomized to Placebo/Placebo and Placebo/AVP-786, respectively
- B: Stage 1 AVP-786 28 mg data
- C: Stage 1 AVP-786 18 mg data
- D: Stage 2 data for patients who were placebo non-responders in Stage 1 and re-randomized to placebo in Stage 2
- E: Stage 2 data for patients who were placebo non-responders in Stage 1 and re-randomized to AVP-786 28 mg in Stage 2
- F: Stage 2 data for patients who were placebo non-responders in Stage 1 and re-randomized to AVP-786 18 mg in Stage 2
- G: Stage 2 data for patients who were placebo responders in Stage 1 and re-randomized to placebo in Stage 2
- H: Stage 2 data for patients who were placebo responders in Stage 1 and re-randomized to AVP-786 28 mg in Stage 2
- I: Stage 2 data for patients who were placebo responders in Stage 1 and re-randomized to AVP-786 18 mg in Stage 2
- J: Stage 2 AVP-786 28 mg data for patients randomized to AVP-786 28 mg in Stage 1
- K: Stage 2 AVP-786 18 mg data for patients randomized to AVP-786 18 mg in Stage 1

FIGURE 1. STUDY SCHEMATIC



Study medication (active or placebo) will be administered as capsules; 1 in the morning and 1 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 for the placebo arm is based on treatment response criteria followed by Re-Randomization [redacted]

3.5 Sample Size Considerations

Power calculations were performed assuming a bivariate normal distribution for the primary efficacy endpoint with AVP-786 28 mg (high dose) versus placebo. [redacted] t is further assumed that 87% of patients will complete Stage 1, 70% of these patients assigned to placebo will be placebo non-responders and 92% of these patients will complete Stage 2. The total cumulative dropout rate is 20%. The planned enrollment of 380 patients in a [redacted] (active:active:placebo) randomization ratio yields approximately 93% power to reject the null hypothesis in the comparison of AVP-786 28 mg versus placebo with type I error rate at two-sided $\alpha = 0.05$ level. To control the overall type I effort rate at 0.05 level, the null hypothesis about the AVP-786 18 mg dose will be tested only if the null hypothesis about the AVP-786 28 mg dose is rejected.

3.6 Randomization

Stage 1

Eligible patients will be randomized to receive AVP-786 28 capsules, AVP-786 18 capsules, or matching placebo capsules in a [redacted] ratio on Day 1 (Baseline, Stage 1) in a double-blind manner according to a randomization scheme devised by Avanir or its representative and managed within an IWRS. The randomization will be stratified by NPI Agitation/Aggression domain score (≤ 6 vs. > 6), risk assessment for falls (normal/mild vs. moderate/severe), and concomitant use of antipsychotic medications (yes vs. no).

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 two sub-groups ("responders" and "non-responders") based on their treatment responses. Patients will be considered "responders" if [redacted] Patients who do not meet [redacted] will be considered "non-responders". Patients within each placebo sub-group will be re-randomized in a [redacted] (active:active:placebo) ratio to receive either AVP-786 28, AVP-786 18, , or matching placebo capsules.
- Patients who receive placebo and who drop out early in Stage 1 will also be assigned a re-randomization treatment in the same manner as the other placebo patients. Their "responders" and "non-responders" status will be based on their treatment responses 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 (either AVP-786 28 or AVP-786 18) for the entire 6-week duration of Stage 2.

4.0 Study Endpoints and Covariates

4.1 Scales and Questionnaires

4.1.1 Cohen-Mansfield Agitation Inventory (CMAI)

The CMAI will be used as the primary efficacy measure in this study. The CMAI is used to assess the frequency of manifestations of agitated behaviors in elderly persons. The responses are based on the 2 weeks preceding the assessment of the CMAI and will be assessed at Screening, Baseline, Week 1, Week 2, Week 3, Week 6, Week 9, and Week 12.

The CMAI consists of 29 agitated behaviors [redacted]

Each of the 29 items is rated on a 7-point scale of frequency consisting of the following:

CMAI Item Scores

- 1 = Never
- 2 = Less than once a week
- 3 = Once or twice a week
- 4 = Several times a week
- 5 = Once or twice a day
- 6 = Several times a day
- 7 = Several times an hour

The CMAI total score is calculated as the sum of ratings for all 29 items and range from 29 to 203. The CMAI total score will be unevaluable if less than 24 of the 29 items have recorded responses. If 24 to 28 items are recorded, then the total score will be the mean of the recorded items multiplied by 29 rounded to the first decimal place. [redacted]

4.1.2 Neuropsychiatric Inventory (NPI)

The NPI is a validated clinical instrument for evaluating psychopathology in a variety of disease settings, including dementia. The NPI will be administered to the patient's caregiver at Baseline, Week 1, Week 2,

Week 3, Week 6, Week 9, and Week 12. Additionally, the Agitation/Aggression domain of the NPI will be administered to the patient's caregiver at Screening.

The NPI is a retrospective caregiver-informant interview covering 12 neuropsychiatric symptom domains:

- Delusions
- Hallucinations
- Agitation/Aggression
- Depression/Dysphoria
- Anxiety
- Elation/Euphoria
- Apathy/Indifference
- Disinhibition
- Irritability/Lability
- Aberrant Motor Behavior
- Sleep and Nighttime Behavioral Disorders
- Appetite/Eating Disorders

The scripted NPI interview includes a compound screening question for each symptom domain, followed by a list of interrogatives about domain-specific behaviors that is administered when a positive response to a screening question is elicited. Neuropsychiatric manifestations within a domain are collectively rated by the caregiver in terms of frequency and severity, yielding a domain score. Domain scores range from 1 to 12 and are calculated using the following formula:

domain score = frequency x severity

where, frequency and severity are rated according to the following scales:

Frequency
- 1 = Rarely – less than once per week
- 2 = Sometimes – about once per week
- 3 = Often – several times per week but less than every day
- 4 = Very often – once or more per day

Severity
- 1 = Mild – produces little distress in the patient
- 2 = Moderate – more disturbing to the patient but can be redirected by the caregiver
- 3 = Severe – very disturbing to the patient and difficult to redirect

Frequency and severity rating scales have defined anchor points to enhance the reliability of caregiver responses. Caregiver distress is also rated for each positive neuropsychiatric symptom domain using the following anchored scores:

Distress
- 0 = Not at all
- 1 = Minimally (almost no change in work routine)
- 2 = Mildly (some change in work routine but little time rebudgeting required)
- 3 = Moderately (disrupts work routine, requires time rebudgeting)
- 4 = Severely (disruptive, upsetting to staff and other residents, major time infringement)
- 5 = Very Severely or Extremely (very disruptive, major source of distress for staff and other residents, requires time usually devoted to other residents or activities)

The NPI nursing-home (NPI-NH) version will be used for patients from in-patient or assisted living facilities. The questions in the NPI-NH are rephrased for professional caregivers who may not know the patient prior to the onset of illness however, the overall instrument domains and scoring is identical to the NPI except for the caregiver distress section which is replaced with occupational disruptiveness in the NPI-NH version.

A total NPI score is calculated by adding the scores of all 12 domain scores together. The possible total scores are from 0 to 144. The NPI Total Score will be unevaluable if less than 10 of the 12 items are recorded. If 10 or 11 of the 12 items are available, then the total score is the mean of the available scores times 12. All imputed scores are rounded to the first decimal place. The distress score is not included in the total NPI score.

[redacted]

The total caregiver distress score is generated by adding together the scores of all 12 domains of the NPI distress questions. Scores range from 0 to 60 with a higher score indicating more distress. Total score calculation with missing items follows the same algorithm as the NPI total score.

4.1.3 Clinical Global Impression of Severity of Illness-Agitation (CGIS-Agitation)

The CGIS-Agitation is an observer-rated scale that measures illness severity and will be assessed at Screening, Baseline, Week 6, and Week 12.

The clinician is asked to rate the patient relative to their past experience with other patients with the same diagnosis, with or without collateral information. The CGIS-Agitation is a 7-point scale and will be assessed for severity of agitation. A value of 0 is given to patients who are not assessed. The 7-point scale for CGIS-Agitation is the following:

Severity of Agitation Scores
- 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 patient

4.1.4 Alzheimer's Disease Cooperative Study – Clinical Global Impression of Change Rating (ADCS-CGIC-Overall)

The ADCS version of the Clinical Global CGIC, referred to as ADCS-CGIC-Overall, provides a reliable means to assess change from a baseline level of global function within the timeframe of a clinical trial. The ADCS-CGIC-Overall focuses on clinicians' observations of change in the patient's cognitive, functional, and behavioral performance since the beginning of the trial. Once the baseline level of severity is established, the change score at the follow-up visits is based on information gathered from the patient and caregiver interviews by an independent evaluator. The response for the ADCS-CGIC-Overall will consist of one of the following:

Overall Clinical Status Responses
- 1 = Marked Improvement
- 2 = Moderate Improvement
- 3 = Minimal Improvement
- 4 = No Change
- 5 = Minimal Worsening
- 6 = Moderate Worsening
- 7 = Marked Worsening

The baseline ADCS-CGIC-Overall evaluation will be conducted at the Baseline visit. The ADCS-CGIC-Overall will be assessed at Week 6 and Week 12 for change in Overall Clinical Status. At Week 6, change from the Baseline visit will be assessed. At Week 12, change from Week 6 and change from the Baseline visit will be assessed.

4.1.5 Modified Alzheimer's Disease Cooperative Study – Clinical Global Impression of Change Rating (mADCS-CGIC-Agitation)

The mADCS-CGIC-Agitation is a modified version of the ADCS-CGIC containing additional questions related to agitation and an assessment of the Clinician's Impression of Change focused specifically on agitation. The response for the mADCS-CGIC-Agitation will consist of one of the following:

Responses
- 1 = Marked Improvement
- 2 = Moderate Improvement
- 3 = Minimal Improvement
- 4 = No Change
- 5 = Minimal Worsening
- 6 = Moderate Worsening
- 7 = Marked Worsening

The baseline mADCS-CGIC-Agitation evaluation will be conducted at the Baseline visit. The mADCS-CGIC-Agitation will be assessed at Week 6 and Week 12 for change in agitation. At Week 6, change from the Baseline visit will be assessed. At Week 12, change from Week 6 and change from the Baseline visit will be assessed.

4.1.6 Zarit Burden Interview (ZBI)

The ZBI is a 22-item scale used to assess the impact of patient's disabilities on the caregiver's life and will be assessed at Baseline, Week 6, and Week 12. It is designed to reflect the burden experienced by caregivers of dementia patients and can either be completed by the caregiver or administered as an interview. Each question is rated on the following 5-point scale:

Zarit Burden Interview Scores for Individual Questions
- 0 = Never
- 1 = Rarely
- 2 = Sometimes
- 3 = Quite Frequently
- 4 = Nearly Always

The ZBI is scored by summing the responses of the individual questions and ranges from 0 to 88. Higher scores indicate greater caregiver distress. The ZBI score will be unevaluable if 4 or more of the 22 items are missing. If 19 to 21 of the 22 items are available, then the total score is the mean of the available scores times 22. All imputed scores are rounded to the first decimal place.

4.1.7 Patient Global Impression of Change (PGIC)

The PGIC is a 7-point scale used to assess treatment response and will be assessed and rated by the patient's caregiver at Week 6 and Week 12 and will focus on the patient's agitation. The response values for the PGIC are provided below:

Patient Global Impression of Change Scale
- 1 = Very much improved
- 2 = Much improved
- 3 = Minimally improved
- 4 = No change
- 5 = Minimally worse
- 6 = Much worse
- 7 = Very much worse

4.1.8 Dementia Quality of Life (DEMQOL)

The DEMQOL is a scale used to evaluate health related quality of life in patients with dementia and their caregivers. There are 2 versions of the DEMQOL, a 28-item version (rated by the patient) and a 31-item version (DEMQOL-proxy, rated by the caregiver). The following provides the response categories for the individual DEMQOL questions and the overall quality of life question:

DEMQOL Responses for Individual Questions
- 1 = A lot

Overall Quality of Life Question
- 1 = Very good

[redacted]

- 2 = Quite a bit
- 3 = A little
- 4 = Not at all

- 2 = Good
- 3 = Fair
- 4 = Poor

The DEMQOL-proxy (and DEMQOL for patients with a Mini Mental State Examination (MMSE) score ≥ 10 at the Baseline visit) will be assessed at Baseline, Week 6, and Week 12.

Scores for DEMQOL Item 1, 3, 5, 6, 10, 29, DEMQOL-proxy item 1, 4, 6, 8, 11, 32, need to be reversed for analysis, i.e., 4=A lot, 3=Quite a bit, 2=A little, 1=Not at all. Total score is derived by sum of all item scores with item 29 in DEMQOL and item 32 in DEMQOL-proxy being excluded. Missing item handling for DEMQOL, 1) count the number of missing item for question 1 to 28, 2) if the number < 15, set the missing item values to the mean of the non-missing item 1 to 28, then add them up; if the number of missing >= 15, then the total is missing. Missing item handling for DEMQOL-proxy, 1) count the number of missing item for question 1 to 31, 2) if the number < 16, set the missing item values to the mean of the non-missing item 1 to 31, then add them up; if the number of missing >= 16, then the total is missing.

4.1.9 Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-cog)

The ADAS was designed to evaluate the cognitive and non-cognitive behavioral dysfunction characteristics of patients with AD and will be assessed at Baseline, Week 6, and Week 12 for patients with an MMSE score ≥ 10 at the Baseline visit. The cognitive subscale (ADAS-cog) consists of 11 subsets related to memory, praxis, and language. Scoring of the individual subsets is specific to each particular subset. The ADAS-cog is scored by summing the 11 subset scores and ranges in scale from 0 to 70. Scoring for each component of the ADAS-cog is as follows:

Component	Measurement	Scoring
Word Recall Task	Mean number of words not recalled on 3 trials	Maximum score = 10
Naming Task	Number of objects named incorrectly	0 (0 – 2 items) to 5 (15-17 items)
Commands	Number of correct and incorrect commands	0 (all commands correct) to 5 (all commands incorrect)
Constructional Praxis	Number of drawings correct	0 (all drawings correct) to 5 (no figures drawn, no recognizable attempt at drawing any side/section of any figure)
Ideational Praxis	Number of components performed correctly	0 (all components performed correctly) to 5 (failure to perform 5 components)
Orientation	One point given for each incorrect response	Maximum score = 8
Word Recognition	Mean number of words not recognized on 3 trials	Maximum error score = 12
Remembering Test Instructions	Ability to remember the requirements of the Word Recognition Task	0 (patient never needs extra reminders) to 5 (severe – must be reminded 7 or more times)
Spoken Language Ability	Quality of speech	0 (no instances of difficulty understanding the patient) to 5 (severe – 1 or 2 word utterance; fluent but empty speech; mute)
Word-finding Difficulty	Ability in expressive language	0 (no evidence) to 5 (severe – nearly total loss of content words; speech sound empty; 1 or 2 word utterances)

[redacted]

Comprehension of Spoken Language	Ability to understand what is being said to them	0 (patient understands) to 5 (severe – patient rarely responds to questions appropriately; not due to po rty of speech)

4.1.10 Cornell Scale for Depression in Dementia (CSDD)

The CSDD was developed to assess signs and symptoms of major depression in patients with dementia and will be assessed at Screening, Week 6, and Week 12. Information is elicited through one interview with the caregiver and one interview with the patient. The CSDD has 19 items and each item is rated for severity on the following scale:

Severity Scale for CSDD Items
- 0 = Absent
- 1 = Mild or intermittent
- 2 = Severe

The CSDD score is calculated by summing the non-missing scores from each item score. Items which could not be evaluated are assigned value 0 (Psych Congress Network: https://www.psychcongress.com/saundras-corner/scales-screeners/depression/cornell-scale-depression-dementia-csdd). Scores above 10 indicate probable major depression. Scores above 18 indicate definite major depression. Scores below 6 as a rule are associated with absence of significant depressive symptoms.

4.1.11 Resource Utilization in Dementia (RUD)

The RUD is used to calculate healthcare costs associated with dementia and will be assessed at Baseline, Week 6, and Week 12.

The RUD evaluates dementia patients' utilization of formal and informal healthcare resources, including hospitalizations and doctor visits, living assistance, and time spent by nonprofessional caregivers. The RUD is administered as a semi-structured interview with the patient's primary caregiver. One section focuses on caregiver impact (loss of work and leisure time incurred by the caregiver) and the other section focuses on the patient's use of healthcare resources. The total healthcare costs associated with the patient's dementia is estimated by multiplying the number of units used (e.g., hours of caregiver time, visits to doctors, nights in accommodation) by the corresponding unit price vector.

4.1.12 General Medical Health Rating (GMHR)

The GMHR is a global clinical rating designed to quantify the severity of general co-morbidity in a patient with dementia and will be assessed at Screening and Week 12.

The GMHR is rated according to the following scale:

GMHR Rating Scale
- 1 = Poor
- 2 = Fair
- 3 = Good
- 4 = Excellent to very good

4.1.13 EuroQol 5-Dimension 5-Level (EQ-5D-5L)

The EQ-5D-5L is composed of a descriptive system and the EQ Visual Analog Scale (EQ VAS). The descriptive system comprises 5 dimensions. The 5 dimensions and the responses for the dimensions are provided below:

EQ-5D-5L Dimensions	Responses for Each Dimension

[redacted]

• Mobility • Self-care • Usual activities • Pain/Discomfort • Anxiety/Depression	• Slight Problems • Moderate Problems • Severe Problems • Extreme Problems

The EQ VAS records the respondent's self-rated health on a vertical VAS where the endpoints are labeled 'Best imaginable health state' and 'Worst imaginable health state'. This information can be used as a quantitative measure of health outcome as judged by the respondent. There are 2 versions of the EQ-5D-5L, a version rated by the patient and a version (EQ-5D-5L-proxy) rated by the caregiver.

The EQ-5D-5L-proxy (and EQ-5D-5L for patients with an MMSE ≥ 10) was to be assessed at Baseline, Week 6, and Week 12 however, this data collection tool was removed with Protocol Amendment 2. Data collected from patients enrolled prior to this amendment will only be provided in a data listing.

4.2 Efficacy Endpoints

4.2.1 Primary Efficacy Endpoints

The primary efficacy endpoint is the change from Baseline to Week 6 (Stage 1) and from Week 6 to Week 12 (Stage 2) in the CMAI total score.

4.2.2 Secondary Efficacy Endpoints

Secondary efficacy endpoints include change from Baseline to Week 6 (Stage 1) and from Week 6 to Week 12 (Stage 2) for the following efficacy measures:

- mADCS-CGIC-Agitation
- NPI - Agitation/Aggression domain score
- NPI – Agitation/Aggression domain Caregiver Distress score
- NPI – Aberrant Motor Behavior domain score
- ZBI
- NPI – Irritability/Lability domain score
- Total NPI
- CGIS-Agitation
- ADCS-CGIC-Overall
- PGIC
- DEMQOL
- CSDD
- ADAS-cog
- RUD
- GMHR (change from Baseline to Week 12)

At post Baseline visits, raw score for mADCS-CGIC-Agitation, ADCS-CGIC-Overall, and PGIC assess change from Baseline. At Week 6 visit, raw scores for mADCS-CGIC-Agitation, ADCS-CGIC-Overall, and PGIC assess change from Baseline (Day 1). At Week 12 visit, raw scores for mADCS-CGIC-Agitation and ADCS-CGIC-Overall assess both change from Baseline (Day 1) and change from Stage 2 Baseline (Week 6); raw score for PGIC assesses only change from Baseline (Day 1).

[redacted]

[redacted]

- CMAI agitation status improvement, and total score 30% and 50% improvement responders
- NPI other individual domain scores, NPI2 (Agitation/Aggression, Aberrant Motor Behaviors), NPI4D (Agitation/Aggression, Irritability/Lability, Disinhibition, Aberrant Motor Behaviors), NPI4A (Agitation/Aggression, Irritability/Lability, Anxiety, Aberrant Motor Behaviors), and all corresponding NPI Caregiver Distress Scores.
- NPI AA 30% and 50% improvement responders.
- mADCS-CGIC-Agitation and PGIC responders

4.3 Safety Endpoints

Safety and tolerability measurements of AVP-786 will include: adverse events (AEs), physical and neurological examinations, vital signs, clinical laboratory assessments, resting 12-lead electrocardiograms (ECGs), Sheehan Suicidality Tracking Scale (S-STS), MMSE, Time Up and Go (TUG) test, and the Epworth Sleepiness Scale (ESS).

4.4 Pharmacokinetic [redacted] Data

Blood samples will be collected between 0 to 3 hours after the morning dose of study medication on Week 6 and Week 12 for analysis of plasma levels of d6-dextromethorphan (DM), d6-DM metabolites, and quinidine (Q).

[redacted]

5.0 Analysis Populations

Three analysis populations will be used, modified intent-to-treat (mITT), intent-to-treat (ITT), and Safety. In addition, 12-Week parallel group population is also defined.

5.1 mITT Population

The mITT population will be used for all efficacy and health outcome analyses. 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. Patients will be included in the treatment group to which they were randomized regardless of treatment received. The mITT population is defined below for each stage:

- Stage 1: All patients randomized in Stage 1 who had at least one post-baseline efficacy assessment.
- Stage 2: All patients who were re-randomized into Stage 2 and had at least one efficacy assessment in Stage 2 (after Week 6).

5.2 ITT Population

The ITT population will be used for sensitivity analyses. Patients will be included in the treatment group to which they were randomized regardless of treatment received. The ITT population is defined below:

- Stage 1: All patients randomized in Stage 1
- Stage 2: All patients who were re-randomized in Stage 2

5.3 Safety Population

The Safety population includes all patients who received at least one dose of study medication. The Safety population will be used for all analyses of safety data. Patients will be included in the treatment group based on the actual treatment received.

5.4 12-Week Parallel Group Population

The 12-Week Parallel Group population is the patient cohort who are randomized to the same treatment in both stages. Since all Stage 1 placebo patients including those who drop out in Stage 1 are re-randomized and assigned a treatment group in Stage 2, this population is similar to a 12 week randomized parallel group design with a total sample size 254 (2/3 of the original sample size 380) and treatment ratio of [redacted] (active:active:placebo).

Derivation of 12-Week Parallel Group Population:

Let N=380, Stage 1 and 2 randomization ratio = [redacted], and [redacted] for AVP-786 18 mg: AVP-786 28 mg: Pbo, respectively.

12-Week Parallel Group sample size for AVP-786 18 mg = [redacted]*N
12-Week Parallel Group sample size for AVP-786 28 mg = [redacted]*N
12-Week Parallel Group sample size for Placebo = [redacted] N* [redacted] = [redacted] N
Randomization ratio in 12-Week Parallel Group = [redacted] =
12-Week Parallel Group total sample size = [redacted]* N + [redacted]* N + [redacted]* N = [redacted]*N ~ 254, or ~ [redacted]

This population is intended to be used to evaluate efficacy and safety over 12 weeks of treatment comparing AVP-786 28 mg, AVP-786 18 mg and placebo in a parallel group design setting. It includes patients from treatment segments (A, D, and G) in Placebo, (B and J) AVP-786 28 mg, and (C and K) AVP-786 18 mg (see Figure 1).

- mITT 12-Week Parallel Group Population: patients in the 12-Week Parallel Group population who have at least one post-baseline efficacy assessment.
- Safety 12-Week Parallel Group Population: patients in the 12-Week Parallel Group population who received at least one dose of study medication.

6.0 Definitions

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.

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. Specific considerations for Baseline are provided below:

- For primary and secondary efficacy analyses, Stage 1 Baseline is the last assessment prior to the first dose (typically Day 1). Stage 2 Baseline is the Visit 4/Week 6 (Day 43) assessment (re-randomization visit). The Stage 2 Baseline only applies to patients who were randomized to placebo in Stage 1 and re-randomized in Stage 2.
- Baseline for the analyses on the 12-Week Parallel Group population is derived from the last assessment prior to the first dose of study drug.
- Baseline for descriptive safety analyses are provided below:
    - For patients receiving placebo, AVP-786 28 mg, or AVP-786 18 mg for the entire study duration, Baseline is defined as the last non-missing assessment prior to the first dose of study drug.
    - For patients randomized to placebo in Stage 1 and then re-randomized to AVP-786 in Stage 2, the Stage 1 Baseline will be defined as the last non-missing assessment prior to

the first dose of study drug. The Stage 2 Baseline will be defined as the last non-missing assessment occurring after Day 1 and prior to re-randomization at Week 6.

Change from Baseline

Change from baseline (CFB) will be calculated as (post-baseline – baseline). CFB 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 CFB will not be calculated for that parameter. Patients with missing CFB values will be excluded from analyses in which CFB is the endpoint.

Percent CFB, when needed, will be calculated by dividing CFB by the baseline value multiplied by 100. Patients with a value of 0 at baseline will not have a percent CFB calculated.

Concomitant medications

Concomitant medications are defined as any medications taken on or after the date of first dose of study drug.

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.

End of Treatment (EOT)

EOT 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 analyses that are done by study stage.

Enrolled patient

An enrolled patient is one with a record in the database that is not a screen failure.

Prior medication

Prior medications are defined as any medications with start and stop dates prior to the date of first dose of study drug. Medications are defined as prior or concomitant, but not both (see also definition of concomitant medication in this section).

Protocol deviation

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.

Relationship to study treatment

AEs 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.

Placebo Responder

In this study, patients who are randomized to placebo in Stage 1 will be classified as responders or non-responders at the end of Stage 1. A placebo responder is defined as a patient with [redacted] otherwise a placebo patient will be considered a non-responder.

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.

Treatment-emergent adverse event
An AE will be considered to be a treatment-emergent AE (TEAE) if it begins or worsens on or after the first dose date and before the last dose date + 30 days.

7.0 Interim Analyses

There is no Interim analysis planned for this study.

8.0 Data Handling and Review

8.1 Visit Windows

Data at scheduled visits will be assigned to analysis visits as defined in the Visit Window tables below. Visit Windows will be used to classify observed data including unscheduled and early termination visits using study days. If 2 or more visits occur within the same analysis window, then the values closest to the target day will be used. If the assessment is the same distance from the target day, then the latest one will be used.

The tables are divided into patients who have advanced to Stage 2 and those who have not.

TABLE 3-1. VISIT WINDOWS FOR ASSESSMENTS COLLECTED ON-TREATMENT AT WEEKS 1, 2, 3, 6, 9, AND 12

Stage Week/Visit	Target Day	Study Days (Advancing to Stage 2)	Study Days (Not Advancing to Stage 2)
Stage 1			
Week 1/Visit 2	8	2, 11	2, 11
Week 2/Visit 2.1	15	12, 18	12, 18
Week 3/Visit 3	22	19, (re-randomization visit day -1)	19, 32
Week 6/Visit 4	43	re-randomization visit	33, day of last visit
Stage 2			
Baseline		re-randomization visit	N/A
Week 9/Visit 5	64	(re-randomization visit day +1), 74	N/A
Week 12/Visit 6	85	75, day of last visit	N/A

Note: for CMAI, NPI AA, and Vital Signs.

TABLE 3-2. VISIT WINDOWS FOR ASSESSMENTS COLLECTED ON-TREATMENT AT WEEKS 3, 6, 9, AND 12

Stage Week/Visit	Target Day	Study Days (Advancing to Stage 2)	Study Days (Not Advancing to Stage 2)
Stage 1			
Week 3/Visit 3	22	2, (re-randomization visit day -1)	2, 32
Week 6/Visit 4	43	re-randomization visit	33, day of last visit
Stage 2			
Baseline		re-randomization visit	N/A
Week 9/Visit 5	64	(re-randomization visit day +1), 74	N/A
Week 12/Visit 6	85	75, day of last visit	N/A

NOTE: FOR NPI OTHER THAN NPI-AA, AND LABS.

TABLE 3-3. VISIT WINDOWS FOR ASSESSMENTS COLLECTED ON-TREATMENT AT WEEKS 6 AND 12

Stage Week/Visit	Target Day	Study Days (Advancing to Stage 2)	Study Days (Not Advancing to Stage 2)
Stage 1			
Week 6/Visit 4	43	re-randomization visit	2, day of last visit
Stage 2			

Baseline		re-randomization visit	N/A
Week 12/Visit 6	85	(re-randomization visit day +1), day of last visit	N/A

Note: for mADCS-CGIC-Agitation, ZBI, CGIS-Agitation, ADCS-CGIC-Overall, PGIC, DEMQOL, CSDD, ADAS-cog, and RUD.

TABLE 3-4. VISIT WINDOWS FOR ASSESSMENTS COLLECTED ON-TREATMENT AT WEEK 12 ONLY

Week/Visit	Target Day	Study Days (Advancing to Stage 2)	Study Days (Not Advancing to Stage 2)
Week 12/Visit 6	85	(re-randomization visit day +1), day of last visit	2, day of last visit

Note: for GMHR.

TABLE 3-5A. VISIT WINDOWS FOR ASSESSMENTS USED IN THE ANALYSIS OF THE 12-WEEK PARALLEL GROUP POPULATION

Week/Visit	Target Day	Study Days
Week 1/Visit 2	8	2, 11
Week 2/Visit 2.1	15	12, 18
Week 3/Visit 3	22	19, 32
Week 6/Visit 4	43	33, 53
Week 9/Visit 5	64	54, 74
Week 12/Visit 6	85	75, day of last visit

Note: for CMAI, NPI AA, S-STS and Vital Signs.

TABLE 3-5B. VISIT WINDOWS FOR ASSESSMENTS USED IN THE ANALYSIS OF THE 12-WEEK PARALLEL GROUP POPULATION

Week/Visit	Target Day	Study Days
Week 3/Visit 3	22	19, 32
Week 6/Visit 4	43	33, 53
Week 9/Visit 5	64	54, 74
Week 12/Visit 6	85	75, day of last visit

Note: for NPI other than NPI-AA, and labs.

TABLE 3-5C. VISIT WINDOWS FOR ASSESSMENTS USED IN THE ANALYSIS OF THE 12-WEEK PARALLEL GROUP POPULATION

Week/Visit	Target Day	Study Days
Week 6/Visit 4	43	23, 63 (2, re-randomization visit)
Week 12/Visit 6	85	64 (re-randomization visit day +1), day of last visit

Note: for mADCS-CGIC-Agitation, ZBI, CGIS-Agitation, ADCS-CGIC-Overall, PGIC, DEMQOL, CSDD, ADAS-cog, RUD, TUG, MMSE, and ESS.

TABLE 3-6A VISIT WINDOWS FOR ASSESSMENTS COLLECTED ON-TREATMENT AT WEEKS 2, 3, 6, 9, AND 12

Stage Week/Visit	Target Day	Study Days (Advancing to Stage 2)	Study Days (Not Advancing to Stage 2)
Stage 1			
Week 2/Visit 2.1	15	2, 18	2, 18
Week 3/Visit 3	22	19, (re-randomization visit day -1)	19, 32
Week 6/Visit 4	43	re-randomization visit	33, day of last visit
Stage 2			
Baseline		re-randomization visit	N/A

Week 9/Visit 5	64	(re-randomization visit day +1), 74	N/A
Week 12/Visit 6	85	75, day of last visit	N/A

NOTE: FOR ECG

TABLE 3-6B. VISIT WINDOWS FOR ASSESSMENTS USED IN THE ANALYSIS OF THE 12-WEEK PARALLEL GROUP POPULATION

Week/Visit	Target Day	Study Days
Week 2/Visit 2.1	15	12, 18
Week 3/Visit 3	22	19, 32
Week 6/Visit 4	43	33, 53
Week 9/Visit 5	64	54, 74
Week 12/Visit 6	85	75, day of last visit

NOTE: FOR ECG

8.2 Missing Data Conventions

Missing data will be handled differently depending on the parameter and analysis. Analyses that are done on 'observed cases' will not follow any imputation rules below. Imputation rules are as follows:

- Missing baseline values will not be imputed in any situation.
- 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) and worst observation carried forward (WOCF) + LOCF for SPCD.

Missing Imputation for LOCF and WOCF for SPCD		
Stage	LOCF	WOCF + LOCF [1]
Stage 1	Impute with last non-missing value within Stage 1 including Baseline.	Missing due to lack of efficacy – Impute with worst non-missing value within Stage 1 including Baseline. Missing due to other reasons - Impute with last non-missing value within Stage 1 including Baseline.
Stage 2	Impute with last non-missing value within Stage 2 including Stage 2 Baseline.	Missing due to lack of efficacy – Impute with worst non-missing value within Stage 2 including Stage 2 Baseline. Missing due to other reasons - Impute with last non-missing value within Stage 2 including Stage 2 Baseline.
Stage 2 – Missing data for all timepoints	No imputation	No imputation
[1] Only applicable for the primary efficacy endpoints.		

- Any safety assessments for patients without a Week 12 value will use the last non-missing post-baseline observation as EOT.

[redacted]

- Missing post-baseline values for by-visit safety data will be summarized using the Visit Windows from Section 8.1. If a value is not available within a given window, then no imputation will be done.
- Missing data for AE relationship will be imputed as "Related".
- Rules for partial dates are provided in Appendices 2 and 3. These rules will apply to adverse events and medications.

8.3 Treatment Misallocations

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

Safety data will be summarized, "as treated", according to the treatment the patient actually received.

8.4 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 v1.3). 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. AEs severity will be graded on a 3-point scale and reported in detail as indicated on the eCRF.

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

8.5 Data Screening

The programming of analysis datasets, tables, figures, and listings (TFL) 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 (Dry Run #1) on clean patients and a post-freeze TFL run (Dry Run #2) on the frozen database allow for further data screening prior to database lock (DBL). The post-freeze TFLs will be discussed with AVANIR in a data review meeting to identify any final data issues and seek corrections prior to DBL. The [redacted] statistician and AVANIR must approve database lock.

8.6 Data Safety Monitoring Board (DSMB)

This study is monitored on an ongoing basis by an external data safety monitoring board (DSMB). Safety analyses will be provided to the DSMB approximately every 3 months. Select efficacy results may be included, although no p-values will be provided. Details of the DSMB can be found in the DSMB charter and DSMB SAP.

9.0 Overall Statistical Considerations

9.1 Summary Statistics

All analyses will use SAS® version 9.4 or higher. Summary tables will be organized by treatment group. All available data will be used in the analyses, and important data will be included in data listings, sorted by Stage 1 treatment group, patient, and by visit within patient.

Unless otherwise noted, categorical data will be presented using counts and percentages, with the number of patients in the analysis population by treatment group as the denominator for percentages.

Percentages are rounded to one decimal place except 100% will be displayed without any decimal places and percentages will not be displayed for zero counts. Continuous data, unless otherwise noted, will be summarized using the number of observations (n), mean, standard deviation (SD), median, minimum, and maximum. The mean and median will be rounded to 1 decimal place greater than the precision of the original value, up to a maximum of 3 decimal places. The SD will be rounded to 2 decimal places greater than the precision of the original value, up to a maximum of 3 decimal places.

Supporting figures will be provided for some efficacy or safety analyses in addition to the summary tables. Efficacy listings will include analysis flags and/or imputed values to show which values were used in the summary tables. In addition, analysis visits will be displayed in the listing if it is important for the corresponding summary.

9.2 Hypothesis Testing and Multiplicity of Endpoints

The primary efficacy endpoint is change from Baseline to Week 6 [Stage 1] and from Week 6 to Week 12 [Stage 2] in the CMAI total score. For the primary efficacy endpoint, the null hypothesis is that there is no treatment effect in Stage 1 and Stage 2 and it will be tested against the alternative that there is a treatment effect in at least one of the 2 stages.

A gate-keeping procedure will be used to control the overall type 1 error at 2-sided $\alpha = 0.05$ for the primary and one key secondary (mADCS-CGIC-Agitation) efficacy comparisons.

Specifically, the treatment comparison testing sequence is as below:

1. CMAI total score – high dose vs. placebo
2. mADCS-CGIC-Agitation – high dose vs. placebo
3. CMAI total score – low dose vs. placebo
4. mADCS-CGIC-Agitation – low dose vs. placebo

For the remaining secondary efficacy endpoints, treatment comparisons will be performed at nominal significance level of 0.05.

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 Consideration of Primary Estimand and Sensitivity Analysis

The purpose of this study is to assess the expected drug effect in a future population that results from patients initiating AVP-786 18 mg, AVP-786 28 mg versus Placebo. The estimand of primary interest is defined as follows:

- Population: All patients who meet the inclusion/exclusion criteria, and who have at least 1 post baseline efficacy assessment (mITT).
- Variable: Change from baseline to Week 6, Week 6 to 12 in the CMAI total score.
- Intercurrent event(s): The “treatment policy” strategy will be followed, whereby the value for the variable of interest will be used regardless of adherence to randomized treatment and/or initiation of any concomitant medications which may include rescue medication (Lorazepam) and disallowed medications.
- Population-level summary: The between-treatment difference in the mean changes from baseline to Week 6 (Stage 1) and change from Stage 2 baseline to Week 12 (Stage 2) in the CMAI total score.

All data collected during the study treatment period will be used for statistical analysis. For the primary efficacy analysis, the treatment effect will be estimated using a likelihood-based mixed model repeated

measures (MMRM) on observed data at each Stage separately. Under the MAR assumption, MMRM provides an unbiased estimate of treatment effect for the treatment period. Sensitivity analyses of the primary efficacy endpoint will be performed with missing values imputed by multiple imputation under MNAR (Section 10.4.3). In addition, missing value imputed by LOCF and WOCF+LOCF will also be performed. These sensitivity results in totality will be used to substantiate the credibility of the results.

10.2 Derivation of Primary Efficacy Endpoints

CMAI was assessed at Stage 1 - Baseline, Week 1, 2, 3, 6, and Stage 2 - Baseline (same as Stage 1 Week 6), Week 3 (study Week 9) and 6 (study Week 12).

Observed CMAI total scores including those from unscheduled visits and early discontinuation, are assigned analysis visit based on visit window Table 3-1 using assessment's study days. If 2 or more visits occur within the same analysis window, then the values closest to the target day will be used. If the assessment is the same distance from the target day, then the latest one will be used. After applying visit window, there is only one non-missing or missing value for each analysis visit. These data are called observed data and change from baseline in CMAI total score (observed data) for SPCD analysis are derived by subtracting corresponding baseline value at each stage and visit.

Change from baseline in CMAI total score (observed data) for 12-Week Parallel Group analysis can be similarly derived by applying visit window in Table 3-5A and other efficacy endpoint derivation follow the same algorithm.

10.3 Primary Efficacy: SPCD Mixed Model Repeated Measures (MMRM)

The primary efficacy endpoints (change from Baseline to Week 6 [Stage 1] and from Week 6 to Week 12 [Stage 2] in the CMAI total score) will be analyzed using a weighted test statistic with the treatment effects in each Stage estimated by a likelihood-based mixed model repeated measures (MMRM) analysis (Chen et al., 2011). This analysis will include observed data from all Stage 1 mITT patients and from Stage 2 data for Stage 1 Placebo Non-responder Subset. 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 change in CMAI total score between AVP-786 and placebo in Stage 1 and Stage 2 for each dose as specified in Section 9.2.

Separate MMRMs will be used for Stage 1 and Stage 2. Stage 1 treatment effect will be estimated by the 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 CMAI total score, baseline-by-visit interaction, baseline NPI AA (≤ 6 vs. > 6), risk assessment for falls (normal/mild vs. moderate/severe), and concomitant use of antipsychotic medications (yes vs. no). Stage 2 baseline will be used for Stage 2 model. Unstructured covariance matrix will be used. If there are convergence issues, then the following covariance structures other than the unstructured will be used in the order of 1) autoregressive of order 1 (AR1), 2) compound symmetry (CS) and the covariance structure converging to the best fit will be used as the primary analysis.

The treatment effect estimates in each stage will be used to construct a combined weighted test statistic, Z_{MMRM}. The formula is given below:

$$Z_{MMRM} = \frac{weight \times (stage\ 1\ treatment\ effect) + (1-weight) \times (stage\ 2\ treatment\ effect)}{\sqrt{(weight)^2 \times (stage\ 1\ standard\ error)^2 + (1-weight)^2 \times (stage\ 2\ standard\ error)^2}}$$

A weight of 0.6 will be used for Stage 1 and a weight of 0.4 for Stage 2. The treatment effects and standard errors will be obtained directly from the model output. The test statistic will be used to generate a 2-sided p-value for the hypothesis test.

Model estimates (treatment difference and its 95% CI) will be reported for each stage however, only the overall p-value from the combined weighted test statistic will be provided.

Treatment benefit is considered achieved if the CMAI total score shows statistically significant improvement.

10.4 Sensitivity Analyses for the Primary Efficacy Endpoints

10.4.1 Seemingly Unrelated Regression (SUR) Method

As an additional sensitivity analysis, a SUR analysis as described in Tamura and Huang (2007), will be run on the primary efficacy endpoint using the mITT population. The SUR method is similar to the weighted test statistics method in that it will utilize data from all mITT patients in Stage 1 and the Placebo Non-responder Subset from the mITT in Stage 2. Also, as with the primary analysis the models from Stage 1 and Stage 2 will be run separately and combined for an overall p-value. LOCF within a stage will be used for missing values. The SUR method takes into account the fact that random error from the two stages of the study may be correlated for patients with data in both stages. SAS PROC MODEL will be implemented for this analysis.

Stage 1 and Stage 2 models will be included within the PROC MODEL statement using appropriate baseline and CFB values. For patients who received AVP-786 in Stage 1 or who were placebo responders will have Stage 2 data set to missing. The test statistic, Z_{sur}, is defined below:

$$\frac{weight \times (stage\ 1\ treatment\ effect) + (1 - weight) \times (stage\ 2\ treatment\ effect)}{\sqrt{(weight)^2 \times (stage\ 1\ standard\ error)^2 + 2 \times (weight) \times (1 - weight) \times Cov(stage\ 1\ and\ stage\ 2\ treatment\ effect) + (1 - weight)^2 \times (stage\ 2\ standard\ error)^2}}$$

The weights will be the same as other models (0.6 in Stage 1 and 0.4 in Stage 2). The test statistic will be used to generate a 2-sided p-value for the hypothesis test.

10.4.2 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 will 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.

10.4.3 SPCD Ordinary Least Squares (OLS) ANCOVA

As a sensitivity analysis, the change in CMAI total score will be analyzed using a weighted ordinary least squares (OLS) test statistics with treatment effect estimated by analysis of covariance (ANCOVA). This analysis will include all mITT patients in Stage 1 and Placebo Non-responder Subset in Stage 2.

As with the MMRM, separate ANCOVA models will be used for Stage 1 and Stage 2. The model will include treatment, baseline NPI AA (≤ 6 vs. > 6), risk assessment for falls (normal/mild vs. moderate/severe), and concomitant use of antipsychotic medications (yes vs. no) as factors, and baseline CMAI total score as covariates. The baseline CMAI total score for each stage is defined in Section 6.

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

Summary statistics including the change and percent change from baseline, model estimates (LS mean difference, 95% CI and p-value), and the standard effect size (SES) will be reported for each stage. An overall p-value based on the weighted OLS z-statistic will be provided.

The SES will be calculated using the following formula:

$$\frac{Mean\ CFB\ AVP - Mean\ CFB\ Placebo}{change\ from\ baseline\ pooled\ SD}$$

Note that this analysis will also contain the result of the Stage 1 (CFB to Week 6) ANCOVA for the Stage 1 mITT population. In addition, comparisons between placebo and each AVP-786 dose for the placebo non-responders only in Stage 2 (change from Week 6 to Week 12) will be available from this analysis.

The above analysis will be carried out using below 2 missing value imputation methods (within stage):

- LOCF – Missing values will be imputed by the last observation in each stage as described in Section 8.2
- WOCF + LOCF – Missing values due to lack of efficacy will be imputed using the worst case value in each stage and missing values due to other reasons will be imputed by LOCF as described in Section 8.2

10.4.4 SPCD Analysis on the ITT Population

The primary endpoint will be analyzed using the ITT population, by the same method described in Section 10.3.

10.5 Secondary Efficacy Endpoint Analyses

Continuous secondary efficacy endpoints described in Section 4.2.2 will be analyzed using the statistical methodologies described in the following sections.

10.5.1 Secondary Efficacy Endpoint SPCD MMRM Analysis

The SPCD MMRM described in Section 10.3 will be used for analysis of CFB for the following secondary efficacy endpoints (which are assessed at interim visits between Baseline and Week 6, Week 6 and 12) for the mITT population (Stage 1 plus the placebo non-responder subset for Stage 2):

- [redacted]
- NPI – domain score, total score, and caregiver distress scores

10.5.2 Secondary Efficacy Endpoint SPCD OLS ANCOVA Analysis

The SPCD OLS ANCOVA described in Section 10.4.1 will be used for analysis of CFB for the following secondary efficacy endpoints (which are not assessed at interim visits between Baseline and Week 6, Week 6 and 12) for the mITT population (Stage 1 plus the placebo non-responder subset for Stage 2). LOCF within stage will be used for missing value imputation.

- mADCS-CGIC – Agitation
- Zarit Burden Interview Scale
- CGIS-Agitation
- ADCS-CGIC-Overall
- PGIC
- DEMQOL
- ADAS-cog
- CSDD

Note that the raw score at post baseline visits for mADCS-CGIC-Agitation, ADCS-CGIC-Overall, and PGIC measure change from corresponding baseline.

The ANCOVA model will include treatment, baseline NPI AA (≤ 6 vs. > 6), risk assessment for falls (normal/mild vs. moderate/severe), concomitant use of antipsychotic medications (yes vs. no) as factors, and baseline value as a covariate. For mADCS-CGIC-Agitation, ADCS-CGIC-Overall, and PGIC, baseline CMAI total score will be used as the covariate. Baseline values for each stage are defined in Section 6.

10.5.3 Response Analysis

The number and percentage of patients who have favorable treatment response according to the CMAI total score and the NPI AA domain score will be summarized using the mITT population (Stage 1 plus the Stage 1 placebo non-responder subset in Stage 2). The following categories will be used to classify response patients:

- [redacted]
- Response: Patients with a 30% reduction in the CMAI total score
- Response: Patients with a 50% reduction in the CMAI total score
- Response: Patients with a 30% reduction in the NPI AA domain score
- Response: Patients with a 50% reduction in the NPI AA domain score
- Response: Patients with score of 1 or 2 (marked improvement or moderate improvement) in mADCS-CGIC-Agitation.
- Response: Patients with score of 1 or 2 (very much improved or much improved) in PGI-C.

[redacted]

The number and percentage of response will be provided by stage and treatment group. LOCF will be used for patients with missing data. 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

[redacted]

(Anastasia 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).

10.5.4 Resource Utilization in Dementia (RUD) and General Medical Health Rating (GMHR) Analysis

Descriptive analyses of the following RUD variables will be provided at baseline, week 6 and 12:

- Primary caregiver demographic and other characteristics: age, sex, relationship to subject, number of children, cohabitation (Yes/No), number of additional caregivers, level of contribution,
- Subject living accommodation characteristics: living accommodation; living arrangements, number of subjects with temporary arrangements and type, average nights in temporary accommodations
- Caregiving time
    - Average time spent assisting the subject each day for items 2a) 3a) 4a)
    - Average number of days per month for items 2b) 3b) 4b)
- Aspects of caregiver burden
    - Sleeping hours of the caregiver: average daily sleep time, and the change from baseline on this outcome
    - Caregiver responsibilities affect their work: work (Yes/No) and reasons for No; average paid hours, average hours paid to care for subject, average hours cut down due to caring for subject, average number of days missed work due to caregiving, average number of part of a day missed due to caregiving
    - Among the subset of subjects whose caregiver has reported they work for pay at baseline, the following will be analyzed:
        - Number of days of work missed completely
        - Number of days of work missed partially or completely
- Subject temporary institutionalization; institutionalisation (identified as the following living accommodations: "dementia-specific residential accommodation" and "long-term institutional care"):
    - Proportion of subjects reporting at least one institutionalisation
- Subject healthcare resource utilization
    - Number of hospitalizations and type of ward, number of emergency visits, number of health services and type of healthcare professional, and type and duration of services
    - Overall amount of utilization expressed in the appropriate unit (visits/contacts, days) including number of days receiving care
    - Percentage of subjects with at least one utilisation
    - Amount of utilization expressed in the appropriate unit (visits/contacts, days) within subjects with at least one utilisation

Descriptive analyses of the GMHR variable will be provided at baseline (Screening visit) and Week 12.

10.5.5 12-Week Parallel Group Analysis

To evaluate the treatment effect under 12 weeks of exposure to the same treatment, a repeated measures analysis using observed data from all scheduled visits will be performed on the mITT 12-Week Parallel Group Population for all primary and secondary efficacy variables.

This analysis will compare treatment groups over time using a linear mixed effects repeated measure model which includes fixed effects for treatment, visit, treatment-by-visit interaction, baseline value, baseline-by-visit interaction, baseline NPI AA (≤ 6 vs. > 6), risk assessment for falls (normal/mild vs. moderate/severe), and baseline concomitant use of antipsychotic medications (yes vs. no). For mADCS-CGIC-Agitation, ADCS-CGIC-Overall, and PGIC, baseline CMAI total score will be used as the baseline value covariate. For mADCS-CGIC-Agitation and ADCS-CGIC-Overall, change relative to Baseline (Day 1) will be used. The same selection of covariance matrix as the primary MMRM Section 10.0 (i.e., unstructured as first preference) will be employed.

For the response (Section 10.5.3) analyses, a GEE model will also be performed with the same effects used in the MMRM model.

10.6 Subgroup Analyses

The primary efficacy endpoint will be analyzed for the below subgroups to evaluate potential differential treatment effect under 6 and 12 weeks treatment exposure.

[redacted]

[redacted]

10.6.3 Subgroup Patients Who Did Not Use Antipsychotics at Baseline

This subgroup patient must meet protocol inclusion requirement for CGIS-Agitation score of ≥ 4 and did not use antipsychotics at baseline.

The primary efficacy analysis will be performed using this subgroup. Analyses will also be done on the mITT 12-Week Parallel Group populations.

10.6.4 Additional Subgroups

Additional analyses of subgroups, such as age group, sex, and baseline stratification factors, may be performed for the primary efficacy endpoint if deemed important and sample size permits.

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, MMSE, TUG test, and ESS. All safety analyses will be completed on the Safety population.

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

1. Placebo/Placebo: patients who were randomized to Placebo/Placebo, receiving placebo for the entire duration of the study (treatment segments D and G from the study schematic, including data from treatment segment A1 during Stage 1). A=A1 + A2, which are Stage 1 data for patients who were randomized to Placebo/Placebo and Placebo/AVP-786, respectively. Note that patients randomized to Placebo/AVP-786 (A2) but dropped out in Stage 1 (part of A2) are not included in this population. Instead, their data will be summarized under the 'All Placebo' treatment group.
2. AVP-786 28/28 mg: patients receiving AVP-786 28 mg for the entire duration of the study (B and J).
3. AVP-786 18/18 mg: patients receiving AVP-786 18 mg for the entire duration of the study (C and K).
4. Placebo/AVP-786 28 mg: patients who switched from placebo to AVP-786 28 mg including those who received placebo and dropped out in Stage 1. This group will be further divided into data that occurred on placebo (A) and data that occurred on AVP-786 28 mg (E and H).
5. Placebo/AVP-786 18 mg: patients who switched from placebo to AVP-786 18 mg including those who received placebo and dropped out in Stage 1. This group will be further divided into data that occurred on placebo (A) and data that occurred on AVP-786 18 mg (F and I).

[redacted]

6. All Placebo: This includes data from the Stages when patients who received placebo (A, D, and G).
7. All AVP-786 28 mg: This includes data from the Stages when patients who received AVP-786 28 mg (B, J, E, and H).
8. All AVP-786 18 mg: This includes data from the Stages when patients who received AVP-786 18 mg (C, K, F, and I).

Placebo/Placebo, AVP-786 18/18 mg and AVP-786 28/28 mg treatment groups summarize the safety information for the Safety 12-Week Parallel Group population, which receive 12 week treatment exposure. It is what would be summarized if the study had been a 12 week parallel group design, and a conventional statistical test to compare these two groups can be performed if needed.

All Placebo, All AVP-786 18 mg, and All AVP-786 28 mg 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.

11.1 Adverse Events

AE 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 ≤ AE start date ≤ last dose date + 30 days).

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

- Number of total AEs, TEAEs and deaths
- Incidence of patients with at least one TEAE, drug-related TEAE, non-serious, serious adverse events (SAEs), drug-related serious adverse event, 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 will be summarized by system organ class (SOC) and preferred term (PT), descending frequency of 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.

AEs 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 n (%) of patients with recurrences will also be given.

SAEs 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 28/28 mg, AVP-786 18/18 mg 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 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 or more 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 with 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 date 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 date of study drug.

    - If the same AE occurs more than once while the patient is on the same treatment, total 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) x 100.

- Recurrence is defined as a new report of the same TEAE with a new AE start date within a given treatment (e.g. 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 3, Week 6, Week 9, and Week 12. The following parameters will be summarized descriptively through change from baseline and percent change from baseline by treatment, visit for both Safety Population and Safety 12-Week Parallel Group Population.

- 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])

- 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, blood, leucocyte esterase, nitrates, and microscopic appearance)

- Thyroid function tests: (thyroid stimulating hormone [TSH], and reflex T3 and T4 if TSH is abnormal). Thyroid function tests are collected at the Screening visit only and will only be provided in a data listing.

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 28 mg, and AVP-786 18 mg; Stage 2 Placebo (Stage 1 Placebo re-randomized to Placebo), Stage 2 AVP-786 28 mg (Stage 1 Placebo re-randomized to AVP-786 28 mg), and Stage 2 AVP-786 18 mg (Stage 1 Placebo re-randomized to AVP-786 18 mg). 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. Shift tables will also be created for Placebo/Placebo, AVP-786 28/28 mg and AVP-786 18/18 mg for the Safety 12-Week Parallel Group population.

Potentially clinically significant (PCS) tables will also be used to summarize out-of-range values. PCS values are found in Table 4. The number and percentage 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 4. 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	≥3X ULN		Glucose	mmol/L	≤2.775	≥11.1
ALT (SGPT)	U/L	None	≥3X ULN		LDH	U/L	None	≥3X ULN
AST (SGOT)	U/L	None	≥3X ULN		Magnesium	mmol/L	<0.37	>1.23
Bilirubin	umol/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

Laboratory Parameter	Unit	Low PCS Criteria	High PCS Criteria		Laboratory Parameter	Unit	Low PCS Criteria	High PCS Criteria
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		Urate (Male)	umol/L	None	≥624.54
Creatine Kinase	U/L	None	≥3 ULN		Urate (Female)	umol/L	None	≥505.58
Creatinine	umol/L	None	>132.6					
Hematology								
Hemoglobin	g/L	<100	>180		Monocytes	x10^9/L	None	>1
Hematocrit	proportion of 1.0	<0.3	>0.5		Monocytes/ Leukocytes	%	None	≥15
Basophils	x10^9/L	None	>0.3		Neutrophils/ Leukocytes	%	≤15	None
Eosinophils/Leukocytes	%	None	≥10		Leukocytes	x10^9/L	≤2.8	≥16
Lymphocytes	x10^9/L	≤0.5	>4		Erythrocytes	x10^12/L	≤2.5	≥7.0
Lymphocytes/ Leukocytes	%	≤10	≥60		Platelet Count	x10^9/L	≤100	≥700

Over the course of the study, there may be some lab tests performed that are 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

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

- Screening (performed in triplicate)
- Baseline (pre-dose)
- Baseline (2-3 hours post-dose)
- Week 2
- Week 3
- Week 6 (pre-dose)
- Week 6 (2-3 hours post-dose)
- Week 9
- Week 12

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 11. Note that for patients who switched from placebo to one of the AVP-786 doses, baseline is defined as described in Section 6.

In addition, since ECGs are recorded pre- and post-dose at Baseline and Week 6, 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 5 below. The number and percentage 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 5. 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	>470 to ≤485, >485 to ≤500, >500
	Both	Change from baseline (increase)	>30, >60

ECG overall interpretations will be summarized by the number and percentage that were normal or abnormal. The interpretations by the cardiologist (i.e., central ECG) will be used for these summaries. The listings will provide all interpretations and corresponding details.

11.4 Vital Signs

Vital signs will be assessed at all visits. Starting with Protocol Amendment #3, orthostatic blood pressure and heart rate measurements will be performed at all clinic visits. Supine blood pressure and heart rate will be measured twice. The average of the measurements will be presented. Additionally, a single measurement of standing will be collected. The following parameters will be summarized: systolic blood pressure (SBP), diastolic blood pressure (DBP), heart rate, respiratory rate, and temperature. These parameters will be summarized through change from baseline and percent change from baseline in similar fashion as the ECG parameters.

Vital signs will also be assessed through PCS criteria, which are given in Table 6. 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 6. 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
Heart rate (bpm)	≥120 AND ≥15 increase from baseline	≤50 AND ≥15 decrease from baseline
SBP and heart rate	SBP ≥10 increase from baseline AND heart rate ≥5 increase from baseline	
DBP and heart rate	DBP ≥5 increase from baseline AND heart rate ≥5 increase from baseline	

Orthostatic changes in blood pressure and heart rate from supine to standing will also be summarized. Additionally, patients meeting orthostatic hypotension or postural tachycardia PCS criteria at any time post-baseline will be summarized according to the criteria in Table 7.

TABLE 7. ORTHOSTATIC HYPOTENSION AND POSTURAL TACHYCARDIA PCS CRITERIA

Category	PCS Criterion
Orthostatic hypotension	≥ 20 mmHg decrease in SBP or ≥ 10 mmHg decrease in DBP from supine to standing
Postural tachycardia	≥ 30 bpm increase in heart rate from supine to standing or a standing heart rate ≥ 120 bpm

11.5 Physical and Neurological Examinations

Physical and neurological examinations will be performed at Screening, Week 6, and Week 12 and include assessments of the head, eyes, ears, nose, throat, lymph nodes, skin extremities, respiratory, gastrointestinal, musculoskeletal, cardiovascular, and nervous systems. Physical and neurological examination data will be provided in a data listing.

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 can be analyzed as individual item scores, suicidal ideation subscale score, suicidal behavior subscale score, and a total score. For the screening visit, the timeframe for the items 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. Scores, CFB in scores, and percent CFB 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	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:	if Questions 12 or 14 are missing, then score will be missing

	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: 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.	max(SSTS112 or SSTSQ16C or SSTSQ16D), max(SSTS114 or SSTSQ15C or SSTSQ15D), SSTS17=Yes add 100, SSTS20=Yes add 4	
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

11.7 Mini Mental State Examination (MMSE)

The MMSE is a brief 30-point questionnaire test that is used to screen for cognitive impairment and severity of cognitive impairment. The MMSE scale comprises 11 questions or simple tasks concerning orientation, memory, attention, and language to evaluate a patient's cognitive state and are scored according to the following ranges:

Item	Score Range
• Orientation to Time	• 0 to 5
• Orientation to Place	• 0 to 5
• Registration	• 0 to 3
• Attention and Calculation	• 0 to 5
• Recall	• 0 to 3

[redacted]

• Naming	• 0 to 2
• Repetition	• 0 to 1
• Comprehension	• 0 to 3
• Reading	• 0 to 1
• Writing	• 0 to 1
• Drawing	• 0 to 1

The total score is calculated by summing all of the item scores and ranges from 0 to 30. Higher scores indicate milder cognitive impairment.

The MMSE will be assessed at Screening, Baseline, Week 6, and Week 12. The MMSE total score, CFB in total score, and percent CFB in total score will be summarized descriptively by visit and treatment group.

11.8 Timed Up and Go (TUG) Test

The TUG test measures the time (in seconds) taken for an individual to stand up from a standard armchair, walk 3 meters, turn, walk back to the chair and sit down. It is a commonly used scale for measuring functional mobility and risk of falls.

The TUG test will be performed at Screening, Week 6, and Week 12. The TUG time, CFB in TUG time, and percent CFB in TUG time will be summarized descriptively by visit and treatment group.

11.9 Epworth Sleepiness Scale (ESS)

The ESS is an 8-item questionnaire that is used to measure sleepiness by rating the probability of falling asleep on 8 different situations that most people engage in during the day. The questions are rated on a 4 point scale (0 to 3) where 0 = would never doze, 1 = slight chance of dozing, 2 = moderate chance of dozing, and 3 = high chance of dozing. A total score of 0 to 9 is considered to be normal.

The ESS will be assessed at Baseline, Week 6, and Week 12. The total score, CFB in total score, and percent CFB in total will be summarized descriptively by visit and treatment group.

12.0 Additional Summaries

12.1 Patient Disposition

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

12.2 Protocol Deviations

Protocol deviations (see definition in Section 6) for patients in the Safety population will be reported by category for each treatment group. Protocol deviations will be listed and summarized.

12.3 Demographic and Baseline Characteristics

Demographics will be summarized by Stage 1 randomized treatment group and overall for the mITT, Safety populations and mITT 12-Week Parallel Group Population. 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 (<65, ≥65)
- Weight

In addition, below information will be summarized by Stage 1 randomized treatment group unless specified otherwise.

- Baseline disease characteristics: Safety Population, caregiver relationship to patient, patient living arrangement,
- Randomization Stratification factors: mITT population, Baseline NPI AA domain score (≤ 6 and > 6), Risk for falls (normal/mild and moderate/severe), Use of concomitant antipsychotic medications (yes/no).
- Percent of Responders and Non-Responders at Week 6: mITT Population
- Stage 1 baseline efficacy assessment: mITT population, for CMAI, NPI, CGIS-Agitation
- Stage 2 Baseline Efficacy Assessments - Stage 1 Placebo Non-Responders: Stage 2 mITT Population and treatment groups, for CMAI, NPI, CGIS-Agitation

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 for the Safety populations by SOC and PT based on the Stage 1 randomization and will be provided in data listings.

12.5 Exposure

Duration of exposure will be summarized quantitatively for the Safety population using the number of days on study medication for each patient, displayed by treatment group. Summaries will be done for Placebo/Placebo, AVP-786 28/28 mg, AVP-786-18/18 mg, and each arm of the Placebo/AVP-786 18 mg, Placebo/AVP-786 28 mg groups. Duration for the Placebo/Placebo and AVP-786 28/28 mg, AVP-786 18/18 mg groups will be calculated as (last dose date – first dose date + 1). For Stage 1 Placebo/AVP-786 28 mg, Placebo/AVP-786 18 mg (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 groups. Since the dosing regimen increases after one week on treatment but a visit does not occur for two 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 18 mg QD, 18 mg BID, and 28 mg BID. Dosing regimen can be found in Figure 1.

12.6 Compliance

Overall treatment compliance will be calculated as a percentage using the total number of capsules that were taken. 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 for the Safety population as described above and using descriptive statistics for Placebo/Placebo, AVP-786 28/28 mg, AVP-786 18/18 mg, and each arm of the Placebo/AVP-786 groups. 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 subjects 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 = actual amount taken
2. Doses should have taken:

Stage	Patients re-randomized	Patients not re-randomized
Stage 1	2 * (last dose date Stage 1 – first dose date Stage 1) + 1	2 * (last dose date Stage 1 – first dose date Stage 1) + 1
Stage 2	2 * (last dose date Stage 2 – first dose date Stage 2) +1	2 * (last dose date Stage 2 – first dose date Stage 2) +1
Overall	2 * (last dose date Stage 2 – first dose date Stage 1) + 1	2 * (last dose date Stage 2 – first dose date Stage 1) + 1

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

For the calculations in the above table, 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.

12.7 Prior and Concomitant Medications, Non-drug Therapies and Interventions

The number and percentage of prior and concomitant medications will be provided by the treatment groups used for the safety analyses. Summaries will be provided by anatomical therapeutic chemical (ATC) and preferred term for the Safety population. Non-drug therapies and nonpharmacological interventions will be only provided in data listings.

In addition, below information will be summarized;

- At baseline, number and % patients who used medications for AD (donepezil, rivastigmine, galantamine, memantine), medications/medication classification for agitation secondary to AD (atypical antipsychotics, antidepressants, butyrophenones, buspirone).
- Cumulative number and % patients who used short term rescue medication lorazepam by visit.

12.8 Pharmacokinetics and Pharmacodynamics

Plasma concentrations for d6-DM, d6-DM metabolites, and Q will be collected at Week 6 and Week 12. Plasma concentration will be summarized by visit and treatment group for the Safety population overall [redacted]. Additionally, PK/PD correlations will be provided.

13.0 Validation

[redacted]

14.0 References

[redacted]

[redacted]

Appendix 1 Glossary of Abbreviations

Glossary of Abbreviations:	
AA	Agitation/Aggression
AD	Alzheimer's disease
ADAS-cog	Alzheimer's Disease Assessment Scale – Cognitive Subscale
ADCS	Alzheimer's Disease Cooperative Study
ADCS-CGIC-Overall	Alzheimer's Disease Cooperative Study – Clinical Global Impression of Change Rating for Overall Clinical Status
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	Anatomic therapeutic classification
BID	Twice daily
BP	Blood pressure
BUN	Blood urea nitrogen
CDISC	Clinical Data Interchange Standards Consortium
CFB	Change from baseline
CGIC	Clinical Global Impression of Change
CGIS-Agitation	Clinical Global Impression of Severity of Illness-Agitation
CK	Creatinine kinase
CRF	Case report form
CS	Compound symmetry
CSDD	Cornell Scale for Depression in Dementia
CSR	Clinical study report
CTCAE	Common Terminology Criteria for Adverse Events
[redacted]	
D6-DM	Deuterated (d6)-dextromethorphan hydrobromide
DBP	Diastolic blood pressure
DEMQOL	Dementia Quality of Life
DM	Dextromethorphan hydrobromide
DSMB	Data Safety Monitoring Board
ECG	Electrocardiogram

EQ-5D-5L	EuroQol 5-Dimension 5-Level
EOT	End of treatment
ESS	Epworth Sleepiness Scale
GGT	Gamma-glutamyl transferase
GMHR	General Medical Health Rating
ITT	Intent-to-Treat
IWRS	Interactive Web Response System
LDH	Lactate dehydrogenase
LOCF	Last observation carried forward
LS	Least squares
mADCS-CGIC-Agitation	Modified Alzheimer's Disease Cooperative Study-Clinical Global Impression of Change Rating for Agitation
MAR	Missing at random
MCMC	Markov Chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple imputation
mITT	Modified Intent-to-Treat
MMRM	Mixed effects model repeated measures
MMSE	Mini Mental State Examination
MNAR	Missing not at random
NPI	Neuropsychiatric Inventory
NPI AA	Neuropsychiatric Inventory (NPI) Agitation/Aggression (AA) domain score
NPI-NH	Neuropsychiatric Inventory Nursing-Home version
OLS	Ordinary least squares
PCS	Potentially clinically significant
PGIC	Patient Global Impression of Change
pH	Potential hydrogen
PR	The P-R interval from an ECG tracing
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

QTcB	QT interval corrected for heart rate using the Bazett's formula
QTcF	QT interval corrected for heart rate using the Fridericia's formula
RBC	Red blood cell
RUD	Resource Utilization in Dementia
SAE	Serious adverse event
SAP	Statistical analysis plan
SBP	Systolic blood pressure
SD	Standard deviation
SNRI	Serotonin-norepinephrine reuptake inhibitor
SSRI	Selective serotonin reuptake inhibitor
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
TUG	Timed Up and Go
VAS	Visual analog scale
WBC	White blood cells
WHODRUG DDE	World Health Organization Drug Dictionary Enhanced
WOCF	Worst observation carried forward
ZBI	Zarit Burden Interview

[redacted]

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.

[redacted]

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.