



---

21Aug2014

# **Statistical Analysis Plan**


## **Protocol 12-AVR-131**

**Version 1.1 (supercedes Version 1.0 dated 07Aug2014 )**

This document contains confidential and proprietary information, and is not to be distributed to any third party.


SIGNATURE PAGE

Written by:

  
Harry Haber, MPH  
Senior Statistician  
MMS Holdings Inc.

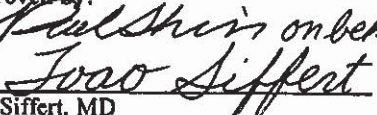
21 Aug 2014  
Date

Reviewed by:

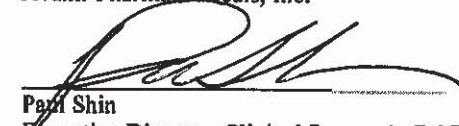
  
Linda LaMoreaux, MPH  
Principal Biostatistician  
MMS Holdings Inc.

21 Aug 2014  
Date

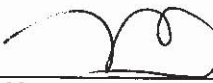
Approved by:

*Paul Shin on behalf of*  
  
Joao Siffert, MD  
Senior Vice President of R&D and Chief Medical Officer  
Avanir Pharmaceuticals, Inc.

21 Aug 2014  
Date

  
Paul Shin  
Executive Director, Clinical Research, R&D  
Avanir Pharmaceuticals, Inc.

21 Aug 2014  
Date

  
Uyen Nguyen  
Sr. Clinical Study Manager  
Avanir Pharmaceuticals, Inc.

21 Aug 2014  
Date

## TABLE OF CONTENTS

TITLE PAGE .....	1
SIGNATURE PAGE .....	2
1.0 LIST OF ABBREVIATIONS.....	6
2.0 INTRODUCTION .....	8
2.1 Purpose of Statistical Analysis Plan .....	8
2.2 Study Objectives .....	8
2.2.1 Primary Objectives.....	8
2.2.2 Secondary Objectives.....	8
2.3 Summary of Study Design .....	8
3.0 DEFINITION OF STUDY POPULATIONS .....	11
3.1 Modified Intent-to-Treat Population.....	11
3.2 Intent-to-Treat Population.....	11
3.3 Week 4 Evaluable Population.....	11
3.4 Safety Population .....	11
3.5 Treatment Misallocations.....	11
3.6 Visit Windows .....	11
3.7 Change from Baseline Calculation and Missing Data .....	12
4.0 SCALES AND QUESTIONNAIRE ASSESSMENTS .....	13
4.1 Clinical Global Impression of Severity of Illness (CGI-S).....	13
4.2 Neuropsychiatric Inventory (NPI) .....	13
4.3 Mini-Mental State Examination (MMSE) .....	14
4.4 Alzheimer’s Disease Cooperative Study – Activities of Daily Living Inventory (ADCS-ADL).....	15
4.5 Cornell Scale for Depression in Dementia (CSDD) .....	15
4.6 Caregiver Strain Index (CSI) .....	15
4.7 Alzheimer’s Disease Assessment Scale – Cognitive Subscale (ADAS-cog) .....	15
4.8 Patient Global Impression of Change (PGI-C) .....	16
4.9 Alzheimer’s Disease Cooperative Study – Clinical Global Impression of Change Rating (ADCS-CGIC).....	16
4.10 Quality of Life-Alzheimer’s Disease Measure (QoL-AD) .....	17
5.0 STATISTICAL ANALYSIS METHODS .....	17
5.1 Patient Disposition and Patient Status .....	17
5.2 Demographic Characteristics .....	17
5.3 Treatment Compliance.....	18
5.4 Primary Efficacy Analysis .....	19
5.4.1 Primary Efficacy Endpoint .....	19
5.4.2 Primary Efficacy Statistical Methods .....	19

5.5	Secondary Efficacy Analyses .....	20
5.5.1	Secondary Efficacy Endpoints .....	20
5.5.2	Other Secondary Endpoints .....	21
5.6	Exploratory Analyses.....	22
5.6.1	Analysis of ADAS-cog .....	22
5.6.2	Sensitivity Analysis of the Primary Efficacy Endpoint .....	22
5.6.3	Additional Analyses of Efficacy Endpoints.....	23
5.6.4	Comparisons Between Additional Patient Groupings .....	24
5.6.5	Subset Analyses Based on Stage 1 Baseline Responses .....	25
5.7	Safety Analysis .....	26
5.7.1	Safety Endpoints .....	26
5.7.2	Statistical Methods for Safety Analysis.....	26
5.7.2.1	Adverse Events (AEs).....	26
5.7.2.1.1	Overview of Adverse Events .....	28
5.7.2.1.2	All Treatment Emergent Adverse Events .....	28
5.7.2.1.3	Serious Treatment Emergent Adverse Events .....	28
5.7.2.1.4	Related Treatment Emergent Adverse Events .....	29
5.7.2.1.5	Deaths .....	29
5.7.3	Physical and Neurological Examination.....	29
5.7.4	Vital Signs.....	30
5.7.5	Clinical Laboratory Parameters .....	30
5.7.6	ECG.....	32
5.7.7	Study Discontinuation due to AE/Lack of Efficacy .....	32
5.7.8	Extent of Exposure.....	33
5.7.9	Prior and Concomitant Medications .....	33
5.8	Pharmacokinetic/Pharmacodynamic Analysis.....	33
6.0	GENERAL CONSIDERATIONS .....	33
6.1	Interim Analysis.....	33
6.2	Changes from Protocol .....	34
6.3	Pooling of Centers.....	34
6.4	Multiple Comparisons/Multiplicity .....	34
6.5	Statistical Software Used for Analysis .....	34
6.6	Treatment Group Labels .....	34
7.0	CHANGES FROM THE PREVIOUS VERSION.....	34
8.0	REFERENCE(S).....	36
9.0	APPENDICES .....	37

## LIST OF IN-TEXT TABLES

Table 1: Visit Windows .....	12
Table 2: NPI Domains .....	13
Table 3: Mini-Mental State Examination (MMSE) .....	14
Table 4: Demographic Characteristics .....	18
Table 5: Vital Signs PCS Criteria .....	30
Table 6: Potentially Clinically Significant Abnormal Laboratory Values .....	31
Table 7: ECG Criteria .....	32

## LIST OF IN-TEXT FIGURES

Figure 1: Study Schematic .....	9
---------------------------------	---

## LIST OF APPENDICES

Appendix 1: Study Schedule .....	37
Appendix 2: Algorithms for Study Instruments .....	38
Appendix 3: Treatment Group Rationale by Table Type .....	40
Appendix 4: Efficacy Analyses .....	43
Appendix 5: Adverse Event Start/Stop Date Imputation .....	54
Appendix 6: Prior and Concomitant Medication Start Date Imputation .....	55
Appendix 7: Medical History Start Date Imputation .....	56

## 1.0 LIST OF ABBREVIATIONS

Abbreviation	Definition
AE	Adverse Event
AD	Alzheimer's Disease
ADAS-cog	Alzheimer's Disease Assessment Scale – cognitive subscale
ADCS-ADL	Alzheimer's Disease Cooperative Study – Activities of Daily Living Inventory
ADCS-CGIC	Alzheimer's Disease Cooperative Study – Clinical Global Impression of Change
ALT/SGPT	Alanine aminotransferase/serum glutamic-pyruvic transaminase
ANCOVA	Analysis of covariance
AR1	Auto-regressive (1)
ALT(SGPT)	Alanine aminotransferase (serum glutamic-pyruvate transaminase)
AST(SGOT)	Aspartate aminotransferase (serum glutamic-oxaloacetic transaminase)
AUC0-12	Area under the curve 0-12 hours
BP	Blood Pressure
CGIS	Clinical Global Impression of Severity of Illness
CiTAD	Citalopram for agitation in Alzheimer's Disease
CMAI	Cohen Mansfield Agitation Inventory
CS	Compound symmetry
CSDD	Cornell Scale for Depression in Dementia
CSI	Caregiver Strain Index
Cmax	Maximum concentration
DBP	Diastolic Blood Pressure
DM	Dextromethorphan
DX	Dextromethorphan
ECG	Electrocardiogram
GGT	Gamma-glutamyltransferase
ITT	Intent-to-treat
LOCF	Last observation carried forward
MITT	Modified intent-to-treat
MedDRA	Medical Dictionary for Regulatory Activities
Msec	Millisecond
MMSE	Min-Mental State Examination
NBRSA	Neurobehavioral Rating Scale-agitation subscale
NPI	Neuropsychiatric Inventory
OLS	Ordinary Least Squares
PCS	Potentially clinically significant
PGI-C	Patient Global Impression of Change
PK	Pharmacokinetics
PR	The P-R interval from an ECG tracing
PT	Preferred Term
Q	Quinidine
QoL-AD	Quality of Life – Alzheimer's Disease Measure

<b>Abbreviation</b>	<b>Definition</b>
QRS	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
RR	RR interval from an ECG tracing
SAF	Safety population
SAP	Statistical Analysis Plan
SBP	Systolic Blood Pressure
SD	Standard Deviation
SI units	International System of Units
SOC	System organ class
SPCD	Sequential Parallel Comparison Design
TEAE	Treatment-emergent Adverse Event

## **2.0 INTRODUCTION**

### **2.1 Purpose of Statistical Analysis Plan**

The purpose of this statistical analysis plan is to describe in detail the procedures and statistical methods required for completing the statistical analysis for Study 12-AVR-131 (Protocol Amendment 3, Version 4.0, dated 28 January 2014).

### **2.2 Study Objectives**

#### **2.2.1 Primary Objectives**

The primary objective of the study is to evaluate the efficacy of AVP-923 compared to Placebo, for the treatment of symptoms of agitation in patients with Alzheimer's disease (AD). AVP-923 is a combination of dextromethorphan (DM) and quinidine (Q).

#### **2.2.2 Secondary Objectives**

The secondary objectives of the study are to evaluate the safety, tolerability, and pharmacokinetics (PK) of AVP-923 (AVP-923-20 and AVP-923-30) in patients with AD.

### **2.3 Summary of Study Design**

This is a multicenter, randomized, double-dummy, double-blind, placebo-controlled, 2-stage, sequential parallel study using the Sequential Parallel Comparison Design (SPCD)<sup>1</sup>. The study consists of 2 consecutive double-blind treatment stages (Stage 1 and Stage 2). Each stage is of 5-week duration. (See [Figure 1](#) Study Schematic)

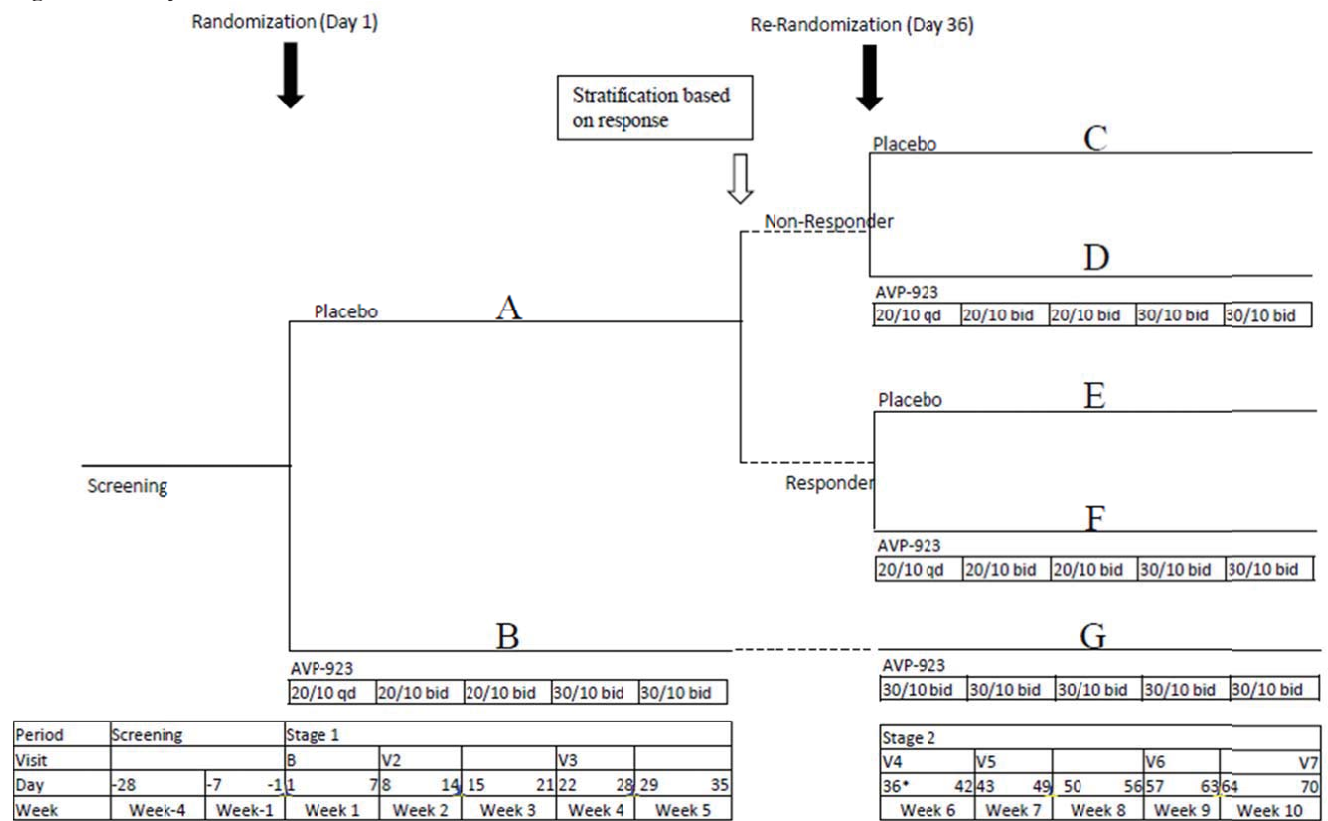
Up to approximately 200 patients will be enrolled at approximately 30 to 40 centers in the US.

Eligible patients will be randomly assigned at the Baseline visit to receive AVP-923 or matching Placebo. Study medication will be administered orally twice-daily from Day 1 through Day 70.

Screening procedures must occur within 4 weeks prior to randomization. Following screening procedures for assessment of inclusion and exclusion criteria, eligible patients will be randomized into Stage 1 of the study.



**Figure 1: Study Schematic**



\*Day 36 (V4) assessments are part of Stage 1 and also serve as baseline assessments for Stage 2

## Stage 1

Eligible patients will be randomized into Stage 1 of the study in a 3:4 (active:placebo) ratio to receive either AVP-923 capsules or matching Placebo capsules administered orally for 5 consecutive weeks. All patients receiving AVP-923 will start at AVP-923-20 (20 mg of DM and 10 mg of Q) once a day and be escalated up to AVP-923-30 (30 mg of DM and 10 mg of Q) twice a day (BID). For the initial 7 days of the study, randomized patients will receive AVP-923-20 in the morning and Placebo in the evening, or Placebo twice-a-day (Stage 1, Days 1-7). Starting on Day 8, patients will receive AVP-923-20 BID or Placebo BID for 2 consecutive weeks (Stage 1, Days 8-21), taking one capsule in the morning and one capsule in the evening, approximately 12 hours apart. On Day 22 of the study, the dose of study medication will be escalated in a double-blind manner. Patients receiving AVP-923-20 BID will increase to AVP-923-30 BID, and patients receiving Placebo BID will continue receiving Placebo for the remaining 2 weeks (Stage 1, Days 22-35) of the study. All study medication including AVP-923-20 capsules, AVP-923-30 capsules, and Placebo capsules are of identical appearance in order to maintain the integrity of the blind.

## Stage 2

Patients who have completed Stage 1 are eligible to participate in the 5-week Stage 2 of the study. Study medication will be administered orally twice daily throughout Stage 2. Patients will be assigned to a double-blind treatment for additional 5 weeks as follows:

- Patients who received AVP-923 in Stage 1 (Days 1-35), will receive AVP-923-30 BID for the entire 5-week duration of Stage 2 (Days 36-70).
- Patients who received Placebo in Stage 1 will be stratified into two sub-groups (“responders” and “non-responders”) depending on their clinical response at Visit 4. Each Placebo sub-group will then be re-randomized to receive either AVP-923 or matching Placebo in a 1:1 ratio.

Patients on Placebo will be considered “responders” if their CGI-S score is between 1 and 3 (inclusive) and their score in the Agitation/Aggression domain in the Neuropsychiatric Inventory (NPI) has decreased by 25% or greater compared to Baseline, at Visit 4 (end of Stage 1). Patients who do not meet these criteria will be considered “non-responders.” Assessment of CGI-S and NPI at Visit 4 should be performed, whenever possible, by the same rater who has assessed CGI-S and NPI prior to randomization into Stage 1 of the study.

Patients who received Placebo during Stage 1 and are re-randomized to AVP-923 in Stage 2 will receive AVP-923-20 in the morning and matching Placebo in the evening for the initial 7 days (Stage 2, Days 36-42) of the study. Starting on Day 43, patients will receive AVP-923-20 twice a day for 2 consecutive weeks (Stage 2, Days 43-56) and starting on Day 57, patients will receive AVP-923-30 BID for the remaining 2 weeks (Stage 2, Days 57-70) until study completion.

Those who received Placebo during Stage 1 and are re-randomized to Placebo in Stage 2 will receive Placebo BID throughout Stage 2. Patients who received AVP-923 during Stage 1 are not re-randomized and will continue to take AVP-923-30 for the remaining 5 weeks of the study. All study medication including AVP-923-20 capsules, AVP-923-30 capsules, and Placebo capsules are of identical appearance in order to maintain the integrity of the blind.

### **3.0 DEFINITION OF STUDY POPULATIONS**

#### **3.1 Modified Intent-to-Treat Population**

The Modified Intent-to-Treat Population (MITT) will be used for all primary and secondary analyses of efficacy. The MITT population will also be used for some exploratory analyses, to enable meaningful comparisons between the exploratory results and the primary results. Patients will be included in the treatment group to which they were randomized, regardless of treatment received. The MITT population is defined as follows:

- All patients randomized in Stage 1 who had at least one post-Baseline NPI Agitation/Aggression score in Stage 1
- All Placebo non-responders from Stage 1 who are re-randomized in Stage 2 and had at least one post-Week 4 NPI Agitation/Aggression score in Stage 2

#### **3.2 Intent-to-Treat Population**

The Intent-to-Treat population (ITT) will include all randomized patients in Stage 1 and Stage 2. The ITT population will be used for some exploratory efficacy analyses.

#### **3.3 Week 4 Evaluable Population**

The Week 4 Evaluable Population will be used for exploratory analyses of the primary efficacy parameter and ADCS-CGIC. Patients will be included in the treatment group to which they were randomized, regardless of treatment received. The Week 4 Evaluable Population is defined as follows:

- All patients randomized in Stage 1 who had at least one post-Baseline NPI Agitation/Aggression score in Stage 1 at or following Visit 3
- All Placebo non-responders from Stage 1 who are re-randomized in Stage 2 and had at least one post-Week 4 NPI Agitation/Aggression score in Stage 2 at or following Visit 6

#### **3.4 Safety Population**

The safety population (SAF) will be used in the statistical analysis of safety. The SAF population includes all randomized patients who received at least one dose of study treatment. The SAF population will be used for all analyses of safety, and patients will be included in the treatment group based on treatment received.

#### **3.5 Treatment Misallocations**

In instances where treatment is improperly allocated to a patient for the entire study, data will be summarized and analyzed as randomized, with the exception of safety data, which will be summarized by actual treatment received.

#### **3.6 Visit Windows**

Data at scheduled visits will be assigned to the nominal visit at which the data were collected. The following visit windows in [Table 1](#) are defined to provide derived visits corresponding to

post-Baseline time points for data captured at early termination and unscheduled visits. If two or more visits (nominal and derived) occur within the same analysis window, the latest non-missing assessment will be used. Data from all visits will be presented in the listings.

**Table 1: Visit Windows**

Stage/Visit	Target Day	Day Range
Stage 1		
Visit 2	8	7-15
Visit 3	22	16-32 <sup>1</sup> /day prior to date of Visit 4
Visit 4	36 <sup>1</sup> /Date of Stage 2 Study Medication	33 <sup>1</sup> -36 <sup>1</sup> /Date of Stage 2 Study Medication
Stage 2		
Visit 5	43	Day after date of Stage 2 Study Medication - 53
Visit 6	57	54-67
Visit 7	70	≥ 68

<sup>1</sup> For subjects who did not receive Stage 2 study medication

### 3.7 Change from Baseline Calculation and Missing Data

All analyses will be presented as observed at each scheduled windowed visit (visit windows are defined in [Section 3.6](#)). Change from Baseline per stage (as defined for the relevant parameters and visits) at each visit will be calculated for patients with both a Baseline and post-Baseline value at the respective visit as follows:

$$C_{\text{visit } v} = M_{\text{visit } v} - M_{\text{Baseline}}$$

where  $C_{\text{visit } v}$  = change in parameter value at post-Baseline visit  $v$ ,  
 $M_{\text{visit } v}$  = the parameter value at post-Baseline visit  $v$ , and  
 $M_{\text{Baseline}}$  = the parameter value at Baseline.

If a Baseline value has not been recorded for a parameter, then change from Baseline cannot be calculated for that parameter, and the patient will be excluded from change from Baseline analysis. If a patient discontinues study before the end of a stage the last non-missing post-Baseline observation will be used as the final observation for that stage (LOCF).

If the parameter is related to secondary efficacy (Total NPI, ADCS-ADL, PGI-C, ADCS-CGIC, QoL-AD, CSDD, and CSI [refer to [Section 4](#) and its sub-appendices for details]), the same considerations as stated above hold.

For missing items within questionnaires and scales, please refer to [Appendix 2](#) for imputation rules.

For laboratory, electrocardiogram (ECG), or vital signs parameters, if there are missing values at the target days but there are non-missing values at unscheduled times within the visit windows of the target days, then the unscheduled results will be used as the visit values. For patients who withdraw early, the last non-missing values in either Stage 1 or Stage 2 prior to premature termination will be used as the patient's end-of-study results.

## 4.0 SCALES AND QUESTIONNAIRE ASSESSMENTS

Refer to [Appendix 2](#) for details on data handling and derivation of scale and questionnaire outcomes.

### 4.1 Clinical Global Impression of Severity of Illness (CGI-S)

CGI-S is a validated tool used to assess a patient's severity of illness. The assessment is based on a 7-point scale (1-7) where the anchor values are 1 = normal and 7 = among the most extremely ill patients. A value of 0 is given to patients that are not assessed. Patients are assessed at Screening (Day -28 to Day -1), Baseline, Visit 4, Visit 5, and at Visit 7.

The CGI-S Agitation assessment is included as an inclusion criteria requirement. Patients must have a CGI-S Agitation score  $\geq 4$  (moderately ill) at screening and Baseline to participate in the study.

### 4.2 Neuropsychiatric Inventory (NPI)

The NPI is a validated clinical instrument for evaluating psychopathology in a variety of disease settings, including dementia. When NPI is not assessed at the center, a nursing home version (NPI-NH) will be used for patients from in-patient or assisted living facilities. The questions in the NPI-NH were rephrased for professional caregivers who may not know the patients prior to the onset of illness; however, the overall instrument domains and scoring are identical to the NPI.

The NPI is a retrospective caregiver-informant interview covering 12 neuropsychiatric symptom domains displayed in [Table 2](#).

**Table 2: NPI Domains**

A: Delusions	G: Apathy/indifference
B: Hallucinations	H: Disinhibition
C: Agitation/Aggression	I: Irritability/Lability
D: Depression/Dysphoria	J: Aberrant motor behavior
E: Anxiety	K: Sleep/Nighttime behavior disorders
F: Euphoria/elation	L: Appetite/Eating changes

Neuropsychiatric manifestations within a domain are collectively rated by the caregiver in terms of both frequency (1 = Occasionally, less than once per week, to 4 = Very frequently, once or more per day) and severity (1 = Mild to 3 = Marked), yielding a composite symptom domain score.

$$\text{Domain Score} = \text{Frequency} \times \text{Severity}$$

For a given NPI domain, a potential assessed domain score has a range from 1 to 12. A total overall NPI score is calculated by summing the NPI domain scores for all 12 domains (A through L) together.

Caregiver distress is rated for each positive neuropsychiatric symptom domain on a scale ranging from 0 (Not at all) to 5 (Very severely or extremely). 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 the higher score indicating extreme distress.

The NPI is administered to the patient’s caregiver at Baseline, Visit 2, Visit 3, Visit 4, Visit 5, Visit 6, and Visit 7. At Baseline, Visit 4, and Visit 7, the NPI will be used to evaluate behaviors within the past 4 weeks. At other visits, behaviors will be evaluated from last visit.

The Agitation/Aggression domain in the NPI is the primary efficacy endpoint – see [Section 5.4.1](#).

Note that responders are defined as having a 25% or greater reduction from Baseline to Visit 4 in the Agitation/Aggression domain, in combination with a Visit 4 CGI-S Agitation score of 1, 2, or 3. This will be used to determine response status (“Responder” or “Non-Responder”) for Stage 1 treatment (patients on Placebo).

### 4.3 Mini-Mental State Examination (MMSE)

The MMSE is a brief test that is used to screen for cognitive impairment. The MMSE scale comprises 11 questions or simple tasks concerning orientation, memory, attention, and language to evaluate the patient’s cognitive state. The anchor values are not consistent for each task (refer to [Table 3](#)). The MMSE total score is calculated by summing the item scores across all 11 tasks. A patient’s total possible MMSE score ranges from 0 to 30 points. Higher scores indicate milder cognitive impairment.

At Stage 1 randomization, both CGI-S (severity of agitation – moderate vs. severe) and screening MMSE assessments (cognitive function – high vs. low) are used as stratification factors.

The MMSE assessment is included as an inclusion criteria requirement. Patients must have a MMSE score between 8 and 28 (inclusive) for study entry. The MMSE is assessed at Screening (Day -28 to Day -1), Visit 4, and Visit 7.

**Table 3: Mini-Mental State Examination (MMSE)**

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

#### **4.4 Alzheimer's Disease Cooperative Study – Activities of Daily Living Inventory (ADCS-ADL)**

The ADCS-ADL inventory measures basic activities of daily living such as dressing, conversation, eating, bathing, and grooming. The 19-item version, covering mainly basic ADL, is used for the assessment of patients with more severe disabilities. ADCS-ADL uses a scale from 0 to 54, with lower scores indicating declining ability. The ADCS-ADL is assessed at Baseline, Visit 4, and Visit 7.

#### **4.5 Cornell Scale for Depression in Dementia (CSDD)**

The CSDD was specifically developed to assess signs and symptoms of major depression in patients with dementia. Because some of these patients may give unreliable reports, the CSDD uses a comprehensive interviewing approach that derives information from the patient and the informant. Information is elicited through two semi-structured interviews: an interview with an informant and an interview with the patient. The interviews focus on depressive symptoms and signs occurring during the week preceding the assessment.

The CSDD has 19 items, and each item is rated for severity on a scale of 0-2 (0 = absent, 1 = mild or intermittent, 2 = severe). The item scores are added. Scores above 10 indicate a probable major depression. Scores above 18 indicate a definite major depression. Scores below 6 as a rule are associated with absence of significant depressive symptoms.

The CSDD is assessed at Screening (Day -28 to Day -1), Visit 4, and Visit 7.

#### **4.6 Caregiver Strain Index (CSI)**

The CSI is a tool that can be used to quickly identify families with potential caregiving concerns. It is a 13-question tool that measures strain related to care provision. There is at least one item for each of the following major domains: Employment, Financial, Physical, Social, and Time. Positive responses to 7 or more items on the index indicate a greater level of strain. A 0 (No) to 1 (Yes) scale is used for each of the 13 questions. The CSI is assessed at Baseline (Day 1), Visit 2, Visit 3, Visit 4, Visit 5, Visit 6, and Visit 7.

#### **4.7 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. The cognitive sub-scale (ADAS-cog) consists of 10 tasks related to memory, praxis, and language. Scoring for each component of the ADAS-cog is as follows:

- Word Recall Task: Mean number of words not recalled on three trials (maximum score = 10)
- Naming Task: Number of objects named incorrectly, ranging from 0 (0-2 items) to 5 (15-17 items)
- Commands: Number of correct and incorrect commands, ranging from 0 (all commands correct) to 5 (all 5 commands incorrect)
- Constructional Praxis: Number of drawings correct, ranging from 0 (all 4 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, ranging from 0 (all components performed correctly) to 5 (failure to perform 5 components)
- Orientation: One point given for each incorrect response (maximum = 8)
- Word Recognition: Mean number of words not recognized on 3 trials (maximum error score = 12)
- Remembering Test Instructions: Subject's ability to remember the requirements of the Word Recognition Task, ranging from 0 (subject never needs extra reminders of instructions) to 5 (severe – must be reminded 7 or more times)
- Spoken Language Ability: Subject's quality of speech, ranging from 0 (no instances when it is difficult to understand the subject) to 5 (severe – 1- or 2-word utterance; fluent but empty speech; mute)
- Word-finding Difficulty: Subject's ability in expressive language, ranging from 0 (no evidence) to 5 (severe – nearly total loss of content words; speech sound empty; 1- to 2-word utterances)
- Comprehension of spoken language: Subject's ability to understand what is being said to them, ranging from 0 (patient understands) to 5 (severe – patient rarely responds to questions appropriately; not due to poverty of speech)

The ADAS-cog is assessed at Baseline, Visit 4, and Visit 7. The total ADAS-cog score ranges from 0 to 70 where higher scores indicate greater cognitive impairment.

#### **4.8 Patient Global Impression of Change (PGI-C)**

The PGI-C is single judgment based on a 7-point (1-7) scale used to assess treatment response, and it is rated as: 1 (very much improved), 2 (much improved), 3 (minimally improved), 4 (no change), 5 (minimally worse), 6 (much worse), or 7 (very much worse).

In this study, PGI-C will be assessed and rated by the patient's caregiver at Visit 4 and Visit 7 and will focus on the patient's agitation. At both visits, the PGI-C will assess change from the Baseline (Day 1) visit.

#### **4.9 Alzheimer's Disease Cooperative Study – Clinical Global Impression of Change Rating (ADCS-CGIC)**

The intent of the ADCS version of the Clinical Global CGIC is to provide a means to reliably assess global change from Baseline in a clinical trial. It provides a semi-structured format to allow clinicians to gather necessary clinical information from both the patient and caregiver, in order to make a global impression of clinical change. ADCS-CGIC is rated as: marked improvement, moderate improvement, minimal improvement, no change, minimal worsening, moderate worsening, or marked worsening.

The standard ADCS-CGIC instrument was modified to better assess aspects relevant to studying agitation in AD. The modified version contains additional questions related to agitation and an additional assessment of the Clinician's Impression of Change focused specifically on agitation.



This modified version of the ADCS-CGIC was originally designed for the Citalopram study for Agitation in Alzheimer's disease (CitAD<sup>9</sup>) and is used with permission from the study group.

In this study, a Baseline ADCS-CGIC evaluation will be conducted, and then the modified ADCS-CGIC will be assessed at Visit 4 and Visit 7 for both Agitation and Overall Clinical Status. At Visit 4, the ADCS-CGIC will be completed to assess change from the Baseline (Day 1) visit. At Visit 7, the ADCS-CGIC will be completed to assess change from Visit 4 and change from the Baseline (Day 1) visit.

The ADCS-CGIC from Visit 4 to Visit 7 will be performed retrospectively for all patients who completed Visit 7 prior to Amendment 3, based on the existing ADCS-CGIC evaluation worksheets that allow the clinician to record assessments of clinical severity and of change over time.

#### **4.10 Quality of Life-Alzheimer's Disease Measure (QoL-AD)**

The QoL-AD is a brief, 13-item measure designed specifically to obtain a rating of the patient's quality of life from both the patient and the caregiver. It was developed for individuals with dementia, based on patient, caregiver, and expert input, to maximize construct validity, and to ensure that the measure focuses on QoL domains thought to be important in cognitively impaired older adults. It uses simple and straightforward language and responses and includes assessments of the individual's relationships with friends and family, concerns about finances, physical condition, mood, and an overall assessment of life quality. Caregivers complete the measure as a questionnaire about their patients' QoL, while patients complete it in interview format about their own QoL. The measure consists of 13 items, rated on a 4-point scale, with 1 being poor and 4 being excellent. Total scores range from 13 to 52, with higher score indicating better QoL. The QoL-AD is assessed at Baseline, Visit 4, and Visit 7.

### **5.0 STATISTICAL ANALYSIS METHODS**

#### **5.1 Patient Disposition and Patient Status**

Patient status will provide a global summary of the numbers of patients randomized to Placebo or AVP-923 and treated during Stage 1, the numbers in each treatment group who discontinued during Stage 1, the numbers re-randomized into Placebo or AVP-923 (either new to AVP-923 or continuing) and treated during Stage 2 by responder status, the numbers who discontinued during Stage 2 by treatment and responder status, and the numbers completing both stages of the study.

Final patient disposition will be summarized by the number of patients who completed a study stage, the primary reason for discontinuation, and the stage at which the patient discontinued. Patient Disposition and Patient Status will be tabulated for the MITT, ITT, and SAF populations using counts and percentages. Percentages will be determined using the counts of patients available for each population appropriate to Stage and treatment.

A supporting listing of patient disposition at the end of the study will also be provided.

#### **5.2 Demographic Characteristics**

Demographic characteristics will be summarized using number and percent for categorical variables and descriptive statistics (n, mean, standard deviation (SD), minimum, median,

maximum) for quantifiable variables. Since these summaries display data collected once, at the start of the study, the tables will display the groups to which the patients were originally randomized, as well as an overall column, for both the ITT and MITT populations. Demographic characteristics will also be summarized for all patients entering Stage 2 (by Stage 2 treatment group) and for Stage 1 Placebo non-responders re-randomized in Stage 2 (by Stage 2 treatment group). The variables to be summarized can be found in Table 4 below. For each categorical parameter, the denominator for the percentages will be the number of patients who had that parameter assessed. A supporting listing of patient demographics will also be provided.

**Table 4: Demographic Characteristics**

Characteristic	Summarized as:	Categories
Gender	Categorical	<ul style="list-style-type: none"> <li>Female</li> <li>Male</li> </ul>
Race	Categorical	<ul style="list-style-type: none"> <li>White</li> <li>Black or African American</li> <li>Asian</li> <li>American Indian or Alaska Native</li> <li>Native Hawaiian or Other Pacific Islander</li> <li>Other</li> </ul>
Ethnicity	Categorical	<ul style="list-style-type: none"> <li>Hispanic or Latino</li> <li>Not Hispanic or Latino</li> </ul>
Age/Age group (years)	Continuous and Categorical	Age Group: <ul style="list-style-type: none"> <li>&lt; 75</li> <li>≥ 75</li> </ul>
Patient Living Arrangements	Categorical	<ul style="list-style-type: none"> <li>Outpatient</li> <li>Assisted living</li> <li>Nursing home</li> </ul>
CGI-S Agitation Score	Continuous	
Metabolizer Status based on CYP2D6 genotype	Categorical	<ul style="list-style-type: none"> <li>Poor metabolizers</li> <li>Intermediate metabolizers</li> <li>Extensive metabolizers</li> <li>Ultra-rapid metabolizers</li> </ul>

### 5.3 Treatment Compliance

Study medication will be dispensed on Days 1, 8, 22, 36, 43, and 57 (Baseline to Visit 6). Patients will be considered compliant if at least 80% of their scheduled doses are taken. Overall compliance with the dosing regimen will be defined as the ratio of study medication taken to the expected amount of study medication taken.

$$\textit{Treatment Compliance} = 100 \times \frac{\textit{Amount of Study Medication Taken}}{\textit{Expected Amount of Study Medication Taken}}$$

When calculating overall treatment compliance, the overall number of days on study drug (last dose day – first dose day + 1) will be considered in determining the expected amount of study

medication taken. If a patient does not return any of their medication, the site will be queried to confirm the number of doses taken by the patient, based upon the returned source diary.

## 5.4 Primary Efficacy Analysis

All efficacy analyses are described in [Appendix 4](#).

### 5.4.1 Primary Efficacy Endpoint

The primary efficacy endpoint is based on both stages of the study; in Stage 1, it will assess the change in score of the Agitation/Aggression domain of the NPI from Baseline to Visit 4 and, in Stage 2, the change from Visit 4 (Stage 2 Baseline) to Visit 7.

Baseline for Stage 1 is the NPI value at Day 1. Placebo patients who are assessed as non-responders at Visit 4 and re-randomized to either continued Placebo or AVP-923 will have their Visit 4 value used as the Baseline for Stage 2.

### 5.4.2 Primary Efficacy Statistical Methods

The primary efficacy analysis will be based on a Sequential Parallel Comparison Design (SPCD) in which data from the 2 Stages are analyzed together using the MITT population. The method is described in detail by Chen et al<sup>1</sup>. The Ordinary Least Squares (OLS) analysis method will be used.

At the end of each Stage, an Analysis of Covariance (ANCOVA) model will be applied using data from that Stage to test the null hypothesis that the change due to Placebo is equal to that due to AVP-923 versus the alternative that there is a difference between treatments. Stage 1 will include all patients randomized to Placebo or AVP-923 who had data at Baseline and at least one post-Baseline visit during Stage 1. Stage 2 will consist only of data from patients randomized to Placebo at Stage 1 who completed the stage, were evaluated as “non-responders” according to pre-determined criteria, and were re-randomized to either Placebo or AVP-923.

The model will be:

$$C_{\text{Stage}i} = \alpha^{(i)} + \beta^{(i)} * \text{Baseline}_i + \theta^{(i)} * \text{Treatment} + \text{error},$$

where  $C_{\text{Stage}i}$  = change from Baseline to end of Stage  $i$ , where  $i = 1, 2$ ,

$\text{Baseline}_i$  = Stage  $i$  Baseline value at Stage  $i$ ,  $i = 1, 2$ , and

$\text{Treatment}$  = treatment AVP-923 indicator during Stage  $i$ , where  $i = 1, 2$  (value = 1 if subject assigned to AVP-923 during Stage  $i$  and 0 if patient assigned to Placebo)

From the individual Stage analyses, the treatment effect estimates  $\theta^{(i)}$  (change due to AVP-923 - change due to Placebo) at Stage  $i$  and variances of the treatment effects estimates  $\text{Var}(\theta^{(i)})$  at Stage  $i$  are included as part of the output. Then, using predetermined weights for each Stage (in this study the weights are  $w = w_i = 0.5$ , for  $i = 1, 2$ ), the values are used to construct a weighted OLS statistic to test the null hypothesis that the treatment effect among patients in Stage 1 and the treatment effect among Placebo non-responders in Stage 2 are equal to 0.

The test statistic is calculated as:

$$Z_{OLS} = (w\theta^{(1)} + (1-w)\theta^{(2)}) / \text{SQRT} (w^2\text{Var}(\theta^{(1)}) + (1-w)^2 \text{Var}(\theta^{(2)})),$$

Under the null hypothesis, the observed value is compared against a normal critical value of 1.96 (corresponding to a two-sided test at the alpha = 0.05 level of significance). Chen et al.<sup>1</sup> showed that such a procedure preserves the type I error.

## 5.5 Secondary Efficacy Analyses

The following secondary efficacy analyses will be carried out using the MITT population.

### 5.5.1 Secondary Efficacy Endpoints

The secondary efficacy endpoints are as follows:

- Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7 in the total NPI
- Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7 in all other domains of the NPI, separately
- Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7 in the sum of the following NPI domains: Agitation/Aggression (C), irritability/lability (I), disinhibition (H), and aberrant motor behaviors (J) – herein called ‘NPI4D’ – a cluster of neuropsychiatric manifestations common in patients with agitation which also approximates the scope of behaviors assessed by the Cohen Mansfield Agitation Inventory (CMAI) and Neurobehavioral Rating Scale-agitation subscale (NBRSA) scales<sup>7, 8</sup>
- Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7 in the sum of the following NPI domains: Agitation/Aggression (C), irritability/lability (I), anxiety (E), and aberrant motor behaviors (J) – herein called ‘NPI4A’
- Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7 in the NPI Caregiver Distress Score for the NPI Agitation/Aggression Domain
- Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7 in the NPI Caregiver Distress Score for Total NPI
- Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7 in the NPI Caregiver Distress Score for NPI4D
- Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7 in the NPI Caregiver Distress Score for NPI4A
- Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7 in the ADCS-ADL
- Change from Stage 1 Baseline to Visits 2 and 3, and change from Stage 2 Baseline to Visits 5 and 6 in the Agitation/Aggression domain of the NPI
- Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7 in the QoL-AD
- Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7 in the CSDD

- Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7 in the CSI
- Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7 in the MMSE
- ADCS-CGIC Agitation at Visit 4 and Visit 7 relative to change from Stage 1 Baseline
- ADCS-CGIC Agitation at Visit 7 relative to change from Visit 4
- ADCS-CGIC Overall Clinical Status at Visit 4 and Visit 7 relative to change from Stage 1 Baseline
- ADCS-CGIC Overall Clinical Status at Visit 7 relative to change from Visit 4
- PGI-C (rated by a caregiver) at Visit 4 and Visit 7

Per protocol, the same methodology used for the analysis of the primary endpoint will be used for these secondary endpoints, making adjustment to study weeks being compared as appropriate. For the analyses of ADCS-CGIC and PGI-C, ANOVA will be used instead of ANCOVA, as there is no Baseline assessment.

- Change from Stage 2 Baseline to Visits 5 and 6 in the Agitation/Aggression domain of the NPI

This analysis will be completed using an ANCOVA model with treatment as a factor and the Baseline value as a covariate.

### **5.5.2 Other Secondary Endpoints**

Changes in the use of concomitant psychotropic drugs - the concomitant use of psychotropic drugs is allowed. These drugs include:

donepezil, rivastigmine, galantamine, memantine, citalopram, escitalopram, fluvoxamine, fluoxetine, sertraline, paroxetine, venlafaxine, desvenlafaxine, duloxetine, milnacipram, midazolam, oxazepam, bromazolam, triazolam, alprazolam, trazodone, eszopiclone, zopiclone, zolpidem, zaleplon, apiprazole, asenapine maleate, clozapine, iloperidone, lurasidone, olanzapine, olanzapine/fluoxetine, paliperidone, quetiapine, risperidone, ziprasidone, butyrophenones, buspirone, lorazepam

Changes in the use of these drugs from Baseline use, eg, increases/decreases in dose, changes in frequency, addition/discontinuation, will be summarized. Patients may be counted in multiple categories of change.

Patients are allowed to receive oral lorazepam as rescue medication for the short-term treatment of symptoms of agitation if deemed necessary by the investigator. Lorazepam can be administered in a dose up to 1.5 mg/day and should not exceed 3 days in a 7-day period. A summary of the use of lorazepam as rescue medication, by visit, will be provided, along with a patient listing.

## 5.6 Exploratory Analyses

Exploratory efficacy analyses will be carried out using the ITT population, MITT population, or Week 4 Evaluable population, as appropriate. For each exploratory analysis below, the population to be used will be indicated. For patients receiving Placebo at Stage 1 and AVP-923 at Stage 2, Stage 1 Baseline is defined as the last non-missing assessment prior to Stage 1 randomization for assessments during Stage 1 and Stage 2 Baseline is defined as the last non-missing assessment prior to Stage 2 re-randomization for assessments during Stage 2.

### 5.6.1 Analysis of ADAS-cog

This exploratory analysis will be carried out using the ITT population.

Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7 in the total ADAS-cog will be performed using the same statistical methodology as described in [Section 5.4.2](#) for the primary analysis.

### 5.6.2 Sensitivity Analysis of the Primary Efficacy Endpoint

This exploratory analysis will be carried out using the ITT population and the MITT population.

An exploratory sensitivity analysis of the primary efficacy endpoint will be conducted using the repeated measures model described by Doros et al. (2013).<sup>6</sup> This model uses all available data for the NPI Agitation/Aggression domain, ie, Baseline plus the 3 visits from each of the 2 stages. Three separate models are used to estimate the treatment effect using data collected at Baseline, end of Stage 1, and end of Stage 2, with a general model that allows the data from intermediate visits to be part of the analysis (see Section 2.4 of the Doros et al. paper).

In this analysis, the unstructured covariance model will be used. However, if there are convergence problems, the use of the first-order autoregressive (AR1) and/or the compound symmetry (CS) covariance structures will be considered.

In addition, the following comparisons will be carried out for the primary endpoint and for Total NPI, NPI4A, and NPI4D, using the MITT population:

- exclude patients who did not use the same version of the NPI throughout the study (patients who used the NPI-NH at one or more visits and the NPI version at other visits), and
- exclude patients who used the NPI-NH version (rather than the NPI version) throughout the study.

These exploratory analyses will be carried out using the primary statistical analysis described in [Section 5.4.2](#).

### 5.6.3 Additional Analyses of Efficacy Endpoints

The following exploratory analyses will be carried out using the MITT population.

In this exploratory analysis, the following alternative definition of a “Responder” will be considered:

- 25% or greater reduction from Baseline to Day 36 (Visit 4) in the NPI Agitation/Aggression domain

For the above definition, the primary statistical analysis described in [Section 5.4.2](#), and the secondary analyses described in [Section 5.5.1](#) will be repeated. This alternative definition will likely result in a different group of patients included in the analysis, since Stage 2 data are only incorporated for Placebo non-responders at Stage 1.

In addition, the primary statistical analysis described in [Section 5.4.2](#) and the secondary analyses described in [Section 5.5.1](#) will be applied to a comparison including Placebo responders in Stage 1 as well as non-responders.

The following exploratory analyses will be carried out using the Week 4 Evaluable population:

- Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7 in the Agitation/Aggression domain of the NPI
- ADCS-CGIC Agitation at Visit 4 and Visit 7 related to change from Stage 1 Baseline
- ADCS-CGIC Agitation at Visit 7 related to change from Visit 4

The primary statistical analysis described in [Section 5.4.2](#) will be used for the analysis of NPI Agitation/Aggression change from Baseline and the same methodology as defined in [Section 5.5.1](#) will be used for the analysis of ADCS-CGIC.

The following exploratory analysis will be carried out using the MITT population:

- Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7 in the NPI Caregiver Distress Score

In addition, for ADCS-CGIC Agitation and PGI-C, to capitalize on the ordered categories of each rating scale, proportional odds regression will be used to compare responses between treatments at each visit using the MITT population. A treatment effect estimate will be generated as the Odds Ratio of being at or better than a given category for AVP-923 versus Placebo. A treatment effect estimate greater than 1.0 favors AVP-923. The SPCD design will be maintained as the results at Visit 7 will include only the Stage 1 Placebo non-responders re-randomized in Stage 2.

#### 5.6.4 Comparisons Between Additional Patient Groupings

Additional data groupings will be defined as described below and in [Figure 1](#):

- A: Stage 1 data for patients randomized to Placebo in Stage 1
- B: Stage 1 data for patients randomized to AVP-923 in Stage 1
- C: Stage 2 data for Placebo non-responders in Stage 1 who are randomized to Placebo in Stage 2
- D: Stage 2 data for Placebo non-responders in Stage 1 who are randomized to AVP-923 in Stage 2
- E: Stage 2 data for Placebo responders in Stage 1 who are randomized to Placebo in Stage 2
- F: Stage 2 data for Placebo responders in Stage 1 who are randomized to AVP-923 in Stage 2
- G: Stage 2 data for AVP-923 patients in Stage 1 who continue on AVP-923 in Stage 2

The following comparison will be tested for the primary endpoint and for NPI4A and NPI4D using the MITT population:

- Change from Stage 1 Baseline to Visit 4 comparing data for patients randomized to AVP-923 in Stage 1 (A from Figure 1) to data for patients randomized to Placebo in Stage 1 (B from Figure 1)



This analysis will be completed using an ANCOVA model with treatment as a factor and the Stage 1 Baseline value as a covariate.

The following comparisons will be carried for the primary endpoint and for Total NPI, NPI4A, and NPI4D using the MITT population:

- Number (%) of patients with  $\geq 25\%$  decrease at Visit 4 from Stage 1 Baseline,  $\geq 50\%$  decrease at Visit 4 from Stage 1 Baseline, and  $\geq 75\%$  decrease at Visit 4 from Stage 1 Baseline for patients in the Only Placebo group (A plus C plus E from Figure 1) vs patients in the Only AVP-923 group (B plus G from Figure 1)
- Number (%) of patients with  $\geq 25\%$  decrease at Visit 7 from Stage 2 Baseline,  $\geq 50\%$  decrease at Visit 7 from Stage 2 Baseline, and  $\geq 75\%$  decrease at Visit 7 from Stage 2 Baseline for Stage 1 Placebo patients re-randomized to Placebo in Stage 2 (C plus E from Figure 1) vs Stage 1 Placebo patients re-randomized to AVP-923 in Stage 2 (D plus F from Figure 1)

The following exploratory analyses will be carried out using the ITT population.

The following comparison will be tested for the primary and secondary efficacy endpoints, making adjustment to study weeks being compared as appropriate:

- Change from Baseline to Visit 7/LOCF comparing data for patients randomized to AVP-923 in Stage 1 (B plus G from Figure 1) to data for patients randomized to Placebo in Stage 1 who also received Placebo in Stage 2 (A plus C plus E from Figure 1)

This analysis will be completed using an ANCOVA model with treatment as a factor and the Stage 1 Baseline value as a covariate.

A longitudinal analysis of the NPI Agitation/Aggression domain scores using data from all visits (Baseline, Visits 1 through 7) will also be performed. This analysis will compare treatment groups over time using a linear mixed effects model with random intercept and slope for each patient. This analysis will be done comparing patients who received only Placebo (A plus C plus E from Figure 1) versus patients who received only AVP-923 (B plus G from Figure 1). The same analysis will be performed for the key secondary efficacy parameters that are collected at multiple visits (beyond just Baseline, Visit 4, and Visit 7) - total NPI, NPI4D, NPI4A, all other domains of the NPI separately, and CSI.

### **5.6.5 Subset Analyses Based on Stage 1 Baseline Responses**

The following patient subset of the ITT population, based on the results of the CiTAD study<sup>9</sup>, will be constructed based on Stage 1 Baseline:

- NPI Agitation/Aggression domain frequency score = 4 or (NPI Agitation/Aggression domain frequency score  $\geq 3$  and NPI Agitation/Aggression domain severity score  $\geq 2$ )

The primary statistical analysis described in [Section 5.4.2](#) will be applied for the primary and secondary efficacy endpoints for the ITT patient subset defined above.

- Descriptive statistics (n, mean, standard deviation (SD), minimum, median, maximum) will be calculated for the following efficacy parameters using the MITT population, for patients with CGI-S Agitation scores of 4 vs. >4 at Stage 1 Baseline: Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7 in the Agitation/Aggression domain of the NPI
- Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7 in the NPI4D
- Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7 in the NPI4A

## **5.7 Safety Analysis**

### **5.7.1 Safety Endpoints**

Safety will be assessed by the following:

- Adverse Events (AEs)
- Physical and Neurological examination
- Vital Signs
- Clinical laboratory values such as serum biochemistry, hematology, and urinalysis
- Resting 12-lead ECG
- Study discontinuation due to an adverse event or lack of efficacy

For patients only receiving AVP-923, Baseline is defined as the last non-missing assessment prior to Stage 1 randomization. Otherwise, for non-responders receiving Placebo at Stage 1 and re-randomized into Stage 2, Baseline is defined as the last non-missing assessment prior to Stage 1 randomization for assessments during Stage 1 and the last non-missing assessment prior to Stage 2 re-randomization for assessments during Stage 2.

In general, safety will be summarized by Only Placebo, Only AVP-923 and Placebo/AVP-923. Patients who received only Placebo or AVP-923 will be summarized as Only Placebo or Only AVP-923, respectively. Patients randomized to Placebo at Stage 1 and then re-randomized to AVP-923 at Stage 2 are summarized as Placebo/AVP-923. Safety data will also be summarized by All Placebo or All AVP-923. All Placebo and All AVP-923 are defined as patients receiving study medication (Placebo or AVP-923, respectively) at either Stage.

### **5.7.2 Statistical Methods for Safety Analysis**

#### **5.7.2.1 Adverse Events (AEs)**

AE summaries will be reported on the Safety Population. Medical Dictionary for Regulatory Activities (MedDRA) version 16.0 will be used to code AEs.

Treatment-emergent adverse events (TEAEs) are defined as adverse events which first occur, or worsen, after the first dose of study medication and within 30 days after the permanent discontinuation of the study medication (first dose date  $\leq$  AE start date  $\leq$  date of last dose + 30 days).

Key guidelines for counting the number and percent of patients with adverse events are as follows.

- When a patient has the same AE reported multiple times during an analysis period based on preferred terminology [System Organ Class (SOC), Preferred Term (PT)], the patient will only be counted once within a level of MedDRA. If a patient in the Placebo/AVP-923 group has an AE with onset during Stage 1 (while on Placebo) that continues into Stage 2 (while on AVP-923), as long as there is no increase in severity, the AE will be counted only for Placebo. If the AE increases in severity or has a new occurrence in Stage 2, it will also be counted for AVP-923.
- When assessing investigator reported relationship to study drug of the AEs, if an AE changes in causal relationship for a patient, the most related event will be chosen. Related AEs will include those reported as probably or possibly related by the investigator and those with a missing relationship.
- When assessing severity of the AEs, if an AE changes in severity (mild < moderate < severe) during an analysis period for a patient, the AE with the worst severity will be chosen. An AE with a missing severity will be excluded from summaries of intensity.
- When assessing severity for drug-related AEs as reported by the investigator, only drug-related AEs (possible, probably, and missing relationship) will be used in the analysis. If a patient has the same AE reported multiple times during an analysis period for a drug-related AE, the AE with the worst severity will be chosen. An AE with a missing severity will be excluded from summaries of intensity.
- Patients who prematurely discontinue at either stage and experience an AE within 30 days of final dose will have that AE counted for the treatment being taken during the stage at which discontinuation occurs.

All AE data will be presented in data listings which include both the stage/treatment group in which the AE occurred and the randomized treatment received.

In general, TEAE tables will be summarized by each treatment group (Only Placebo, Only AVP-923 and Placebo/AVP-923, All Placebo, All AVP-923) defined in [Appendix 3](#). For summaries by dose and age group, TEAEs will be summarized by All Placebo and All AVP-923; for summaries of duration and recurrence, and time to onset, TEAEs will be summarized by Only Placebo and Only AVP-923.

In summaries by SOC and PT, the SOCs will be presented in alphabetical order, and the PTs will be presented by descending frequency in the All AVP-923 group and then alphabetically within each SOC. In summaries by PT, the PTs will be presented in decreasing frequency based on the overall AVP-923 group.

Refer to [Appendix 5](#) for details of imputing partial Start Dates of Adverse Events.

#### 5.7.2.1.1 Overview of Adverse Events

The number and percent of patients experiencing AEs will be summarized for the following AE categories:

- Overall AEs (All AEs and TEAEs)
- Patients with at least one TEAE, drug-related TEAE, Serious TEAE, and drug-related Serious TEAE
- Patients who discontinued treatment due to TEAE, drug-related TEAE, Serious TEAE, and drug-related Serious TEAE
- Overall deaths (for any reason)
- Deaths due to drug-related TEAE

#### 5.7.2.1.2 All Treatment Emergent Adverse Events

The following summaries for all TEAEs will display the number and percent of patients for each group (Only Placebo, Only AVP-923, Placebo/AVP-923, All Placebo, All AVP-923) defined above:

- TEAEs by SOC and PT
- TEAEs by SOC and PT by maximum severity
- TEAEs by PT

The following will be summarized by All Placebo and All AVP-923 treatment groups:

- TEAEs by SOC, PT, and dose
- TEAEs by SOC, PT, and age group

#### 5.7.2.1.3 Serious Treatment Emergent Adverse Events

The following summaries will display the number and percent of patients with serious TEAEs by SOC and PT for each group (Only Placebo, Only AVP-923, Placebo/AVP-923, All Placebo, All AVP-923) defined above:

- All serious TEAEs
- Drug-related serious TEAEs
- Serious TEAEs leading to discontinuation of study drug

All serious AEs (treatment-emergent and non-treatment-emergent) will be listed.

#### 5.7.2.1.4. Related Treatment Emergent Adverse Events

Related TEAEs will include those reported as probably or possibly related to study drug as assessed by the investigator and those with a missing relationship. The number and percent of patients with related TEAEs will be summarized by SOC and PT, and by SOC, PT, and maximum severity for each group (Only Placebo, Only AVP-923, Placebo/AVP-923, All Placebo, All AVP-923) defined above. In addition, the number and percent of patients with serious related TEAEs will be summarized by PT for each treatment group.

The following summaries will also be done for the Only Placebo and Only AVP-923 treatment groups:

- Time (days) to first onset of common treatment-related TEAEs
- Duration (days), percentage of total study days, and recurrence of common treatment-related TEAEs
- Time to discontinuation due to treatment-related TEAEs

Common TEAEs are defined as TEAEs with PT = nausea, dizziness, somnolence, agitation, aggression, diarrhea, fall, fatigue, headache, or any other TEAE with an incidence of  $\geq 3\%$  in the Only AVP-923 treatment group and  $\geq 2$  times the incidence in the Only Placebo treatment group.

For the time to first onset summary, a Kaplan-Meier approach will be used where patients who do not have the event of interest will be right-censored at the time of their final visit.

Duration is defined as the number of days for the longest occurrence of the event (if experienced more than once); percentage of total study days is defined as total duration of event (sum of durations of same related TEAE) divided by the total study days completed by the patient  $\times 100$ ; recurrence is defined as a new report of the same related TEAE with a new start and stop date.

#### 5.7.2.1.5. Deaths

Any patient deaths during the course of this study will be included in the Overview of Adverse Events, Reasons for Discontinuation in the Study Disposition table and in by-patient data listings.

### **5.7.3 Physical and Neurological Examination**

Physical and Neurological examinations are assessed at Screening (Day -28 to Day -1), Visit 4, and Visit 7.

A shift in assessment tabulation from Baseline to Final Visit will be presented. Baseline will be defined as the last assessment prior to receiving study medication. For patients receiving Placebo during Stage 1 and re-randomized to AVP-923 at Stage 2, Baseline will be defined as last assessment prior to re-randomization and summarized as AVP-923. The counts and percentages of each Baseline and Final Visit tabulation will be based on the number of patients with both a Baseline and Final Visit value. The denominator used to calculate the percentages will be based on the Baseline assessment (Normal, Abnormal) totals.

### 5.7.4 Vital Signs

Systolic blood pressure, diastolic blood pressure, pulse rate, respiratory rate, and body temperature will be summarized by treatment (Only Placebo, Only AVP-923, Placebo/AVP-923, All Placebo, All AVP-923) with descriptive statistics (mean, standard deviation, minimum, median, and maximum) for the average Baseline value, average actual value, average change from Baseline value, and average percent change from Baseline value to each post-Baseline visit and Final Visit. The change from Baseline tabulations will only be performed on patients with both a Baseline and post-Baseline measurement. Refer to [Appendix 3 Treatment Group Rationale](#) by Table Type, for guidance on how the summaries will be displayed.

The number and percentage of patients who meet potentially clinically significant (PCS) criteria for Systolic BP, Diastolic BP, and Pulse Rate will be presented by Stage for each treatment group based on Any Visit during the Stage. The PCS criteria are presented in the Table 5 below:

**Table 5: Vital Signs PCS Criteria**

Vital Signs Parameter	Increases	Decreases
SBP	> 180 mmHg AND $\geq$ 20 mmHg increase from Baseline	$\leq$ 90 mmHg AND $\geq$ 20 mmHg decrease from Baseline
DBP	$\geq$ 105 mmHg AND $\geq$ 15 mmHg increase from Baseline	$\leq$ 50 mmHg AND $\geq$ 15 mmHg decrease from Baseline
Pulse Rate	$\geq$ 120 bpm AND $\geq$ 15 bpm increase from Baseline	$\leq$ 50 bpm AND $\geq$ 15 bpm decrease from Baseline
SBP and Pulse Rate	Increase from Baseline of $\geq$ 10 mmHg for SBP and $\geq$ 5 bpm for Pulse Rate	N/A
DBP and Pulse Rate	Increase from Baseline of $\geq$ 5 mmHg for DBP and $\geq$ 5 bpm for Pulse Rate	N/A

### 5.7.5 Clinical Laboratory Parameters

Hematology, Blood Chemistry, and Urinalysis laboratory safety parameters will be assessed at Screening (Days -28 through -1) and at Visits 4 and 7. The complete list of tests in each category can be found in Section 6.3.4 of the study protocol.

At each time point, for each treatment group, quantitative variables will be summarized by n, mean, standard deviation, minimum, median, and maximum. The International System of Units (SI units) will be used for these laboratory summaries.

Changes from Baseline will be calculated at Visits 4 and 7 and summarized using the same summary statistics. Refer to [Appendix 3 Treatment Group Rationale](#) by Table Type for details on treatment-group displays of the summary values.

Shift tables, relating Baseline to post-Baseline results with respect to normal limits for each test, will be generated by treatment. A shift in assessment tabulation from Baseline to post-Baseline will be presented. Baseline will be defined as the last assessment prior to receiving study medication. For patients receiving Placebo during Stage 1 and re-randomized to AVP-923 at Stage 2, Baseline will be defined as last assessment prior to re-randomization and summarized as AVP-923. The counts and percentages of each Baseline and post-Baseline tabulation will be based on the number of patients with both a Baseline and post-Baseline value. The denominator

used to calculate the percentages will be based on the Baseline assessment totals. Laboratory tests without normal limits will not have shift tables produced.

The number and percentage of patients who meet potentially clinically significant (PCS) criteria for selected laboratory tests will be presented by Stage for each treatment group based on any visit during the stage. The PCS criteria are presented in the Table 6 below:

**Table 6: Potentially Clinically Significant Abnormal Laboratory Values**

<b>Potentially Clinically Significant (PCS) Laboratory Criteria</b>			
		<b>PCS Criteria</b>	
<b>Laboratory Parameter</b>	<b>Unit</b>	<b>Low PCS Criteria</b>	<b>High PCS Criteria</b>
<b>Hematology</b>			
Hemoglobin	g/dL	< 10	> 18
Hematocrit	%	< 30	> 50
Platelet Count	x10 <sup>3</sup> /uL	≤ 75	≥ 700
Leukocyte (White Blood Cell Count)	x10 <sup>3</sup> /uL	≤ 2.8	≥ 16
Eosinophils	%	No lower limit	≥ 10
Neutrophils	%	≤ 15	No upper limit
Erythrocyte (Red Blood Cell Count)	x10 <sup>6</sup> /uL	≤ 2.5	≥ 7.0
Basophils		NA	> 300/3mm
Lymphocytes		≤ 500/3mm	> 4000/3mm
Monocytes		NA	> 1000/3mm
Bands		NA	NA
Morphology		NA	NA
<b>Chemistry</b>			
ALT (SGPT)	U/L	No lower limit	≥ 3 X ULN
AST (SGOT)	U/L	No lower limit	≥ 3X ULN
Total Bilirubin	mg/dL	No lower limit	≥ 1.5 ULN
BUN	mg/dL	No lower limit	≥ 30.0
Creatinine Kinase (CK)	U/L	No lower limit	≥ 3 ULN
Sodium	mEq/L	≤ 125	≥ 155
Potassium	mEq/L	≤ 3.0	≥ 5.5
Calcium	mg/dL	Corrected value ≤ 7.0	Corrected value ≥ 12.0
Lactate Dehydrogenase (LDH)	U/L	No lower limit	≥ 3X ULN
Alkaline Phosphatase	U/L	No lower limit	≥ 3X ULN
Total Cholesterol	mg/dL	No lower limit	≥ 301
Uric Acid (Male)	mg/dL	No lower limit	≥ 10.5
Uric Acid (Female)	mg/dL	No lower limit	≥ 8.5
Albumin	g/dL	≤ 2.6	≥ 6.0
Total Protein	g/dL	≤ 5.0	≥ 10.0
Chloride	mEq/L	≤ 85	≥ 120
Glucose	mg/dL	≤ 45.1	≥ 200.0
Carbon Dioxide	mmol/L	≤ 9	> 40
Serum Creatinine	mg/dl	NA	> 1.5 ULN
Phosphorus	mg/dl	≤ 1.4	> 12
Magnesium	mg/dl	< 0.9	> 3
Triglycerides	mg/dl	NA	> 300
Gamma-Glutamyl Transferase [GGT]	U/L	NA	≥ 3 ULN

### 5.7.6 ECG

ECG parameters, general findings, heart rate, QRS complex, PR Interval, QTc interval (including corrections due to Bazett and Fridericia) are collected at Screening, Baseline, and Visits 2 through 7. Data will be summarized by treatment group at each time and changes from Baseline calculated and displayed where appropriate. The number and percentage of patients meeting the criteria for PR Interval and QTcF as defined in Table 7 below will be tabulated. These summaries will be done on both the machine-read/cardiologist readings and the semi-automated readings. For the machine-read/cardiologist data, the cardiologist reading will override the machine-read reading when they are different.

The machine read/cardiologist values are based on an automated algorithm to calculate the value and may be inaccurate due to poor quality of the ECG or abnormal waveform variations. The machine-read/cardiologist values were used for the immediate safety status of the subject.

In the semi-automated process, an iCardiac ECG analyst reviews each ECG for a more precise measurement. The iCardiac software (COMPAS) is used to determine the ECG interval measurements (QT, RR, PR, QRS). The ECG Analyst chooses 3 consecutive beats and verifies that the measurements are correct for each ECG interval or changes the measurement as appropriate. If the analyst cannot find 3 consecutive beats of good quality, a review is then done of all beats in the ECG. If there are at least 3 beats of good quality the ECG is reported. If the analyst cannot find 3 beats of good quality, then no measurement is reported for that ECG.

**Table 7: ECG Criteria**

ECG Parameter	Data Used in Analysis	Criteria
PR Interval (msec)	Observed <sup>a</sup>	> 200 msec
		> 220 msec
QTcF Interval (msec)	Change from Baseline	≥ 30 msec to < 60 msec increases
		≥ 60 msec increases
	Observed <sup>a</sup>	Total (≥ 30 msec) increases
		> 450 msec
		> 480 msec
		> 500 msec

<sup>a</sup> A patient will be included in each criterion he or she satisfies, eg, a patient with QTc = 490 msec will be counted in the > 450 msec and > 480 msec categories.

### 5.7.7 Study Discontinuation due to AE/Lack of Efficacy

The number and percent of patients who discontinue from the study due to adverse event or lack of efficacy will be summarized by treatment group.



### 5.7.8 Extent of Exposure

Exposure to treatment will be calculated in terms of duration of treatment. Duration of treatment will be calculated as the number of days the patient has taken randomized study drug. Data will be summarized using n, mean, std, minimum, median, and maximum as described in [Appendix 3 Treatment Group Rationale](#) by Table Type.

### 5.7.9 Prior and Concomitant Medications

Prior medications are those taken by patients before entry into this study, and not continued while on study medication. Concomitant medications are defined as non-study medications with a start date on or before the final study visit, and that are either ongoing at the end of the study or have a stop date on or after the date of first dose of study drug. Section 5.5 of the study protocol identifies medications taken prior to study entry and the conditions under which they may be continued on a concomitant basis, as well as medications which are to be discontinued before study entry, including timing of discontinuations, and those medications prohibited to be taken during the study.

All non-study medications, including rescue medications, will be coded (WHO-Drug June 2012 Version) and tabulated by study treatment using counts and percentages.

Refer to [Appendix 6](#) for details of imputing partial start dates of prior/concomitant medications.

## 5.8 Pharmacokinetic/Pharmacodynamic Analysis

Celerion, Inc. will perform pharmacokinetic parameter ( $C_{max}$  and  $AUC_{0-12}$ ) estimation using plasma concentrations of DM, Dextromethorpan (DX), and Q obtained from blood samples collected on Days 36 (Visit 4) and 70 (Visit 7) of this study.

All subjects enrolled in the study will undergo a blood draw at Day 36 (Visit 4) and at Day 70 (Visit 7) for concentration analysis. Plasma concentrations of DM, DX, and quinidine (Q) will be summarized descriptively at each visit and overall and by metabolizer subgroup.

Predicted PK parameter estimations for  $C_{max}$  and  $AUC_{0-12}$  for DM, DX, and Q received from Celerion, Inc. will also be summarized descriptively, by metabolizer status. Metabolizer status will be categorized as shown in Section 5.2, Table 4 (“Normal” and “Normal or Intermediate” metabolizers will be considered “Extensive” metabolizers).

In addition, analyses exploring the association between selected PK parameters ( $C_{max}$  and  $AUC_{0-12}$ ) and selected clinical outcomes (NPI Agitation/Aggression domain score, NPI4D, and NPI4A) will be performed.

## 6.0 GENERAL CONSIDERATIONS

### 6.1 Interim Analysis

There will be no interim analysis for efficacy.

Interim safety analyses will be conducted as per the DSMB charter. All necessary safeguards as described in the charter will be put into place to avoid unblinding of study staff or introduction of bias.

The DSMB tables consisted of demographics, Adverse Events, Serious Adverse Events, Physical and Neurological Examinations, Laboratory summaries by visit, Potentially Clinically Significant Laboratory Abnormalities, Treatment Compliance, Scored Assessment Inclusion Criteria, Stage 1 Randomization and NPI, ADAS-cog and MMSE by visit.

## **6.2 Changes from Protocol**

The protocol version dated 28-Jan-2014 states in Section 8.3.4: “Additional efficacy analyses will be carried out using the ITT population.” In order to enable meaningful comparisons between the results of specific exploratory analyses and the primary analysis, several exploratory analyses will be performed on the MITT population.

## **6.3 Pooling of Centers**

There will be no consideration of study center in the formal analysis models.

## **6.4 Multiple Comparisons/Multiplicity**

Because there is a single, pre-specified primary efficacy analysis, no corrections will be made to the Type I error.

## **6.5 Statistical Software Used for Analysis**

All statistical analyses will be performed using SAS<sup>®</sup> version 9.1 or higher.

## **6.6 Treatment Group Labels**

Treatment group labels used for summarization are Only Placebo, Only AVP-923, Placebo/AVP, All Placebo, and All AVP-923. See [Appendix 3](#) for further details and rationale.

## **7.0 CHANGES FROM THE PREVIOUS VERSION**

The following changes were made to the previous version of the Statistical Analysis Plan (August 7, 2014):

1. List of Abbreviations: Added International System of Units (SI units)
2. Section 4.2: Reference to Caregiver Distress Score was clarified. Rationale: To provide clarification around the efficacy outcome score.
3. Section 5.5.1: Addition of “Agitation” to ADCS-CGIC parameter; addition of ADCS-CGIC Overall Clinical Status parameters. Rationale: To clarify the ADCS-CGIC outcome measures and add additional analyses parameters.
4. Section 5.5.1: Addition of NPI Caregiver Distress Score outcomes to the list of secondary efficacy endpoints. Rationale: to evaluate NPI Caregiver Distress as a secondary outcome.
5. Section 5.6.2: Addition of sensitivity analysis. Rationale: To see whether removing patients who completed the NPI-NH version (rather than the standard NPI version) at all visits or at some visits has an effect on the primary results.

6. Section 5.6.3: Addition of “Agitation” to ADCS-CGIC parameter. Rationale: To clarify the outcome measure. Addition of analysis of NPI Caregiver Distress Score. Rationale: to add additional analysis parameter.
7. Section 5.6.4: Addition of two descriptive analyses. Rationale: to examine the level of “responders” in each treatment group for two different patient groupings.
8. Section 5.7.5: Added that SI units will be used for laboratory summaries. Rationale: to clarify laboratory data analysis
9. Section 5.8: Added description of metabolizer subgroups and re-mapping of reported metabolizer status. Rationale: to clarify summaries of PK data by metabolizer status.
10. Section 5.7.9: WHO-Drug version added.
11. Appendix 4: Changed “NPI Total Score” to Total NPI Score” for consistency; new analyses indicated above were added to this table for completeness.

## 8.0 REFERENCE(S)

- 1 Yeh-Fong Chen, Yang Y., James Hung HM, Wang S-J. Evaluation of performance of some enrichment designs dealing with high placebo response in psychiatric clinical trials. *Contemporary Clinical Trials* 2011;32:592-604.
- 2 Roy N Tamura, Huang X. An examination of the efficacy of the sequential parallel design in psychiatric clinical trials. *Clinical Trials* 2007;4:309-17.
- 3 Guidance for Industry, E14: Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs. US Department of Health and Human Services, Food and Drug Administration, October 2005.
- 4 Statistical Review and Evaluation of Memantine, 12/19/2002: on-line at [fda.gov/ohrms/dockets/ac/03/briefing/3979B1\\_05-FDA](http://fda.gov/ohrms/dockets/ac/03/briefing/3979B1_05-FDA)
- 5 CDISC SDTM ADAS-Cog Questionnaire Supplement (Version 1.0 ): on-line at [cdisc.org/stuff/contentmgr/files/0/f42fd061dc3854fba6c982920dc603/misc/sdtm\\_qs\\_\\_adas\\_cog\\_v1\\_2011\\_12\\_27\\_final.pdf](http://cdisc.org/stuff/contentmgr/files/0/f42fd061dc3854fba6c982920dc603/misc/sdtm_qs__adas_cog_v1_2011_12_27_final.pdf)
- 6 Gheorghe Doros, Pencina M, Rybin D, Meisner A, Fava M. A repeated measures model for analysis of continuous outcomes in sequential parallel comparison design studies. *Statistics in Medicine* 2013;32:2767-2789.
- 7 Sheung-Tak Cheung, Kwok T, Lam, LCW. Neuropsychiatric symptom clusters of Alzheimer's disease in Hong Kong Chinese: prevalence and confirmatory factor analysis of the Neuropsychiatric Inventory. *International Psychogeriatrics* 2012;24:1465-1473.
- 8 Marco Canevelli, Adali N, Voisin T, Soto ME, Bruno G, Ccesari M, Vellas B. Behavioral and psychological subsyndromes in Alzheimer's disease using the Neuropsychiatric Inventory. *International Journal of Geriatric Psychiatry* 2013;28:795-803.
- 9 Lea T Drye, Ismail Z, Porsteinsson AP, Rosenberg PB, Weintraub D, Marano C, Pelton G, Frangakis C, Rabins PV, Munro CA, Meinert CL, Devanand DP, Yesavage J, Mintzer JE, Schneider LS, Pollock BG, Lyketsos CG. Citalopram for agitation in Alzheimer's disease: Design and methods. *Alzheimer's & Dementia* 2012;8:121-130.

## 9.0 APPENDICES

### Appendix 1: Study Schedule

Procedure	Visit:	Stage 1					Stage 2			
		Study Day:	Screening	Baseline	Visit 2 <sup>1</sup>	Visit 3 <sup>2</sup>	Visit 4 <sup>2</sup>	Visit 5 <sup>1</sup>	Visit 6 <sup>2</sup>	Visit 7 <sup>2</sup> /ET <sup>3</sup>
			Day -28 to -1	Day 1	Day 8	Day 22	Day 36	Day 43	Day 57	Day 70
			Week -4 to -1	Week 1	Week 2	Week 4	Week 6	Week 7	Week 9	Week 10
Informed consent forms signed		X								
Medical history		X								
Review of inclusion and exclusion criteria		X	X							
Randomization (Stage 1)			X							
Re-Randomization (Stage 2)						X				
Physical and neurological examination		X				X			X	
Record vital signs		X	X	X	X	X	X	X	X	
Clinical Global Impression of Severity of Illness (CGI-S)		X	X			X	X		X	
Resting 12-lead ECG		X	X	X	X	X	X	X	X	
Review of adverse events			X	X	X	X	X	X	X	
Review previous and concomitant medication		X	X	X	X	X	X	X	X	
Mini-Mental State Examination (MMSE)		X				X			X	
Neuropsychiatric Inventory (NPI)			X	X	X	X	X	X	X	
Alzheimer's Disease Assessment Scale-cognitive subscale (ADAS-cog)			X			X			X	
The Cornell Scale for Depression in Dementia (CSDD)		X				X			X	
Activities of Daily Living Inventory (ADCS-ADL)			X			X			X	
Caregiver Strain Index (CSI)			X	X	X	X	X	X	X	
Patient Global Impression of Change (PGI-C) rated by the caregiver						X			X	
Baseline ADCS-CGIC Evaluation			X							
Clinical Global Impression of Change (ADCS-CGIC)						X			X	
Quality of Life-AD (QoL-AD)			X			X			X	
Administer first dose of study medication in clinic			X			X				
Administer last dose of study medication in clinic									X	
Chemistry, hematology, and urinalysis		X				X			X	
Urine dipstick for females of childbearing potential only			X			X			X	
Blood sample for PK assay						X			X	
Blood sample for CYP2D6 genotyping			X							
Dispense study medication			X	X	X	X	X	X		
Review and return unused study medication and diary				X	X	X	X	X	X	

<sup>1</sup> Visit 2 and Visit 5 have a + 3 days window

<sup>2</sup> All study visits have a ± 3 days window (except Visit 2 and Visit 5)

<sup>3</sup> Early Termination visit for patients who withdraw prior to study completion

## Appendix 2: Algorithms for Study Instruments

Following is a table summarizing different study instruments. Exceptions to the stated rules will be noted and their impact on outcome detailed.

Individual item values and overall scores are based on the Case Report Form entries.

Instrument Name	Computational Algorithm(s), if any	Exception/Note(s)
Clinical Global Impression of Severity (CGIS)	Single number with value 1-7	A score of 0 means the instrument was not completed. Do NOT use values of 0 in any calculation.
PGIC (Patient Global Impression of Change)	Single number with value 1-7	A score of 0 means the instrument was not completed. Do NOT use values of 0 in any calculation.
Clinical Global Impression of Change (CGIC)	Single number with value 1-7	Selections are phrases which correspond exactly with numbered outcomes in PGIC.
Caregiver Strain Index (CSI)	Sum of "YES" responses ("YES" has value = 1). Range of outcomes is 0-13	If any response is missing, then total is sum of remaining responses.
Quality of Life AD (QoL-AD)	Two parts, one for Patient, the other for Caregiver. In each part the final score is the sum of individual item scores. Each item can range in value from 1 to 4 yielding maximum total scores of 52.	If "Patient not done" is checked, there should be no Patient QoL-AD score. Similarly for Caregiver if "Caregiver not done" is checked. It is possible that an individual item is not scored. The sum will then be of non-missing items only.
Mini-Mental State Examination (MMSE)	Sum of scores of 11 items, with individual items scored 0, 1, 2, 3, or 5. Maximum possible sum is 30, and minimum sum is 0 if the patient cannot successfully respond to any item.	The sum of scores can be missing if no items scored. Otherwise the sum must be $\geq 0$ .
Alzheimer's Disease Cooperative Study - Activities of Daily Living Inventory (ADCS-ADL)	Sum of scores of 19 items. Item Scores can be 0, 1, 2, 4, or 5. Range of sum is 0 to 54.	A score of 0 can have different interpretations. For items 7 through 15, there are two possible meanings: "No" or "Don't Know." A score of 0 is valid to use in calculations. Missing individual items will be assigned a value of 0. If more than 4 of the 19 items have missing outcomes, then the sum of scores is set to missing and not used.

Instrument Name	Computational Algorithm(s), if any	Exception/Note(s)
Cornell Scale for Depression in Dementia	Sum of scores from 19 items, each item can have a value of 0, 1, or 2. Range of possible sum scores is 0-38.	An item scored as “a” or “A” is not to be included in the sum. “a” or “A” means “unable to evaluate” and is to be interpreted as a missing value. The sum is computed using non-missing scores.
Alzheimer’s Disease Assessment Scale - Cognitive Subscale (ADAS-cog)	Sum of scores from 10 items. Items have different individual scoring strategies. The first item, Word Recall Task, is scored as the mean number of incorrect responses from 3 trials of 10 words each. The mean <b>MUST</b> appear to 2 decimal places, ie, 7.17 and <b>not</b> 7.2. Potential total scores can range from 0 to 70.	Missing items are replaced by the "worst" score for that item. For item 1 the "worst" score is 10; items 2-5 and 8-10 the "Worst" score is 5; item 6 then "Worst" score is 8; item 7 the "Worst" score is 12.
Neuropsychiatric Inventory (NPI)	Twelve domains, either from NPI or NPI-NH (Nursing Home Version). Domain scores are the product of frequency (scored as 1 to 4) and severity (scored as 1, 2, or 3). The domain scores are summed for the total score. In this study the total score will include all 12 domains.	Each domain permits the option of checking “N/A” for Not Applicable; this is to be regarded as a missing value. There is also an option for a score of 0, meaning “Absent.” If either NA or 0 appears for a domain, there can be no non-missing frequency and severity scores, their product, or “Occupational Distress/Disruption.” A domain checked as 0 is a valid domain score and is used in calculations. The first 10 domains (A-J) are required for all patients and total scores come from them. Domains K and L are optional, and their scores are kept separate from required domains.

### Appendix 3: Treatment Group Rationale by Table Type

The following table provides a general summary of table and analysis contents by broad data groupings, including appropriate stage and randomized treatments at that stage with reasoning for that display. Previous sections of this document provide greater detail on populations eligible for analysis and formal statistical methodologies to be employed.

Table Type	Rx Groups Displayed and Definition of Rx Groups		Rationale
Demographics and Baseline Characteristics (and other data assessed at Screening or Baseline)	Stage 1 Placebo	Patients randomized to Placebo at Stage 1	These summaries display data collected once, at the start of the study. Therefore the groups displayed will be the groups that patients were originally randomized to in Stage 1.
	Stage 1 AVP-923	Patients randomized to AVP-923 at Stage 1	
Demographics and Baseline Characteristics (and other data assessed at Screening or Baseline)	Stage 2 Placebo	Patients re-randomized to placebo at Stage 2	These summaries display data collected once, at the start of the study. The groups displayed will be the groups that patients were re-randomized to in Stage 2.
	Stage 2 AVP-923	Patients re-randomized to AVP-923 in Stage 2	
NPI change from Baseline (and all other efficacy summaries)	Placebo	Patients randomized to Placebo at Stage 1 who were responders at Stage 1 and Placebo non-responders from Stage 1 who were re-randomized to Placebo at Stage 2	Per the SPCD design
	AVP-923	Patients randomized to AVP-923 at Stage 1 and Placebo non-responders from Stage 1 who were re-randomized to AVP-923 at Stage 2	



Table Type	Rx Groups Displayed and Definition of Rx Groups		Rationale
AE and PCS Lab summaries	Only Placebo	Patients who only received Placebo	Summarizes all AEs that occurred in the study under the drug taken at the time. Note that Patients in the Placebo/AVP group will be counted in both the All Placebo and All AVP-923 columns.  Similar rationale as the AE summaries, the PCS summaries includes all incidences of PCS criteria that occurred in the study under the drug taken at the time. Note that Patients in the Placebo/AVP group will be counted in both the All Placebo and All AVP-923 columns.
	Only AVP-923	Patients who only received AVP-923	
	Placebo/AVP-923: Placebo	For Patients in the Placebo/AVP-923 group, AEs that occurred while on Placebo	
	Placebo/AVP-923: AVP-923	For Patients in the Placebo/AVP group, AEs that occurred while on AVP-923	
	All Placebo	Patients who took Placebo at either stage – All AEs that occurred on Placebo	
	All AVP-923	Patients who took AVP-923 at either stage – All AEs that occurred while on AVP-923	
Patient Disposition	Only Placebo	Patients who only received Placebo	Similar to AE groupings. However, while AEs can occur in both stages of the study, a patient can only discontinue from the study once. Therefore all withdrawals or completions in the Placebo/AVP-923 group will occur while on AVP-923. An All Placebo column is not needed since it would be the same as the 'only' Placebo column. The first 3 columns will add up to the Overall column.
	Only AVP-923	Patients who only received AVP-923	
	Placebo/AVP-923	Patients re-randomized to AVP-923 in Stage 2	
	All AVP-923	All patients who received AVP-923 during the study. This will be a sum of the AVP-923 and Placebo/AVP columns.	
	Overall	All Patients	
Patient Status	Stage 1 AVP-923	Randomized group in Stage 1	Provides a summary of patient flow in the study, similar to a cohort diagram. May not need the Stage 1 and Stage 2 identifiers because apparent from the table.
	Stage 1 Placebo	Randomized group in Stage 1	
	Stage 2 Placebo (subgroup under Placebo column)	Randomized group in Stage 2	
	Stage 2 AVP-923 (subgroup under Placebo column)	Randomized group in Stage 2	
Change from Baseline, Physical and Neurological Examination Summaries	Only Placebo	Patients who only received Placebo	Stage 1 study medication will represent Baseline for both the Only Placebo and Only AVP-923 treatment groups. As these patients remained on the same medication
	Only AVP-923	Patients who only received AVP-923	
	Placebo/AVP-923	Patients re-randomized to AVP-923 in Stage 2	

Table Type	Rx Groups Displayed and Definition of Rx Groups		Rationale
Lab Shift Tables			<p>throughout the study, the original Stage 1 Baseline assessment is used to compare all post-Baseline assessments.</p> <p>For the Placebo/AVP-923 treatment group, Baseline is the last assessment prior to receiving study medication in a given stage. Since patients receive both Placebo and AVP-923, Baseline changes will depend on last assessment prior to stage randomization.</p>
	Only Placebo AVP-923	Patients who only received Placebo Patients randomized to either AVP-923 at either stage.	Baseline will be defined as the last assessment prior to receiving study medication. For Patients receiving Placebo during Stage 1 and re-randomized to AVP-923 at Stage 2, Baseline will be defined as last assessment prior to re-randomization and summarized as AVP-923.

#### Appendix 4: Efficacy Analyses

Efficacy Endpoint	Comparison	Data Grouping (Figure 1)	Population	Treatment Groups Displayed on Table	Statistical Method	Interpretation	SAP Section
NPI Agitation/Aggression Domain	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Primary analysis	5.4.1 – 5.4.2
NPI Agitation/Aggression Domain	Change from Stage 2 Baseline to Visits 5 and 6	C x D	MITT	Placebo, AVP-923	ANCOVA	Secondary	5.5.1
NPI Agitation/Aggression Domain	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	Week 4 Evaluable	Placebo, AVP-923	SPCD	Exploratory	5.6.3
NPI Agitation/Aggression Domain	Change from Stage 1 Baseline to Visits 2 and 3, and change from Stage 2 Baseline to Visits 5 and 6	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Secondary/evaluation of 20/10 effect	5.5.1
NPI Agitation/Aggression Domain	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Exploratory/alternative definition of Responder	5.6.3
NPI Agitation/Aggression Domain	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and (C + E x D + F)	MITT	Placebo, AVP-923	SPCD	Exploratory/including Placebo Responders	5.6.4
NPI Agitation/Aggression Domain	Change from Stage 2 Baseline to Visits 5 and 6	C x D	MITT	Placebo, AVP-923	ANCOVA	Exploratory/alternative definition of Responder	5.6.3
NPI Agitation/Aggression Domain	Change from Stage 2 Baseline to Visits 5 and 6	(C + E) x (D + F)	MITT	Placebo, AVP-923	ANCOVA	Exploratory/including Placebo responders	5.6.4
NPI Agitation/Aggression Domain	Change from Stage 1 Baseline to Visits 2 and 3, and Stage 2 Baseline to Visits 5 and 6	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Exploratory/alternative definition of Responder	5.6.3
NPI Agitation/Aggression Domain	Change from Stage 1 Baseline to Visits 2 and 3, and Stage 2 Baseline to Visits 5 and 6	A x B and (C + E x D + F)	MITT	Placebo, AVP-923	SPCD	Exploratory/including Placebo responders	5.6.4
NPI Agitation/Aggression Domain	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	ITT/CitAD patient subset	Placebo, AVP-923	SPCD	Exploratory	5.6.5

CONFIDENTIAL

Page 43 of 56

Efficacy Endpoint	Comparison	Data Grouping (Figure 1)	Population	Treatment Groups Displayed on Table	Statistical Method	Interpretation	SAP Section
NPI Agitation/Aggression Domain	Change from Stage 2 Baseline to Visits 5 and 6	C x D	ITT/CitAD patient subset	Placebo, AVP-923	ANCOVA	Exploratory	5.6.5
NPI Agitation/Aggression Domain	Change from Stage 1 Baseline to Visits 2 and 3, and change from Stage 2 Baseline to Visits 5 and 6	A x B and C x D	ITT/CitAD patient subset	Placebo, AVP-923	SPCD	Exploratory	5.6.5
NPI Agitation/Aggression Domain	Change from Stage 1 Baseline to each visit	A x B and C x D	MITT	Placebo, AVP-923	Repeated measures mixed model	Exploratory Sensitivity	5.6.2
NPI Agitation/Aggression Domain	Change from Stage 1 Baseline to each visit	(A + C + E) x (B + D + F + G)	ITT	Placebo, AVP-923	Repeated measures mixed model	Exploratory Sensitivity	5.6.2
NPI Agitation/Aggression Domain	Actual scores at each visit	(A + C + E) x (B + G)	ITT	Only Placebo, Only AVP-923	Longitudinal analysis/linear mixed effects model	Exploratory	5.6.4
NPI Agitation/Aggression	Change from Baseline to Visit 7/LOCF	(A + C + E) x (B + G)	ITT	Only Placebo, Only AVP-923	ANCOVA	Exploratory	5.6.4
NPI Agitation/Aggression	Change from Stage 1 Baseline to Visit 4	A x B	MITT	Placebo, AVP-923	ANCOVA	Exploratory	5.6.4
NPI Agitation/Aggression	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT, CGI-S =4 vs. >4 at Stage 1 Baseline	Placebo, AVP-923	Descriptive	Exploratory	5.6.5
NPI Agitation/Aggression	Percent Decrease at Visit 4 from Stage 1 Baseline >=25%, >=50%, >=75%	(A + C + E) x (B + G)	MITT	Placebo, AVP-923	Descriptive	Exploratory	5.6.4
NPI Agitation/Aggression	Percent Decrease at Visit 7 from Stage 2 Baseline >=25%, >=50%, >=75%	(C + E) x (D + F)	MITT	Placebo, AVP-923	Descriptive	Exploratory	5.6.4
NPI Agitation/Aggression Domain	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT, excluding patients using different versions of NPI	Placebo, AVP-923	SPCD	Exploratory Sensitivity analysis	5.6.2

Efficacy Endpoint	Comparison	Data Grouping (Figure 1)	Population	Treatment Groups Displayed on Table	Statistical Method	Interpretation	SAP Section
NPI Agitation/Aggression Domain	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT, excluding patients using NPI-NH throughout	Placebo, AVP-923	SPCD	Exploratory Sensitivity analysis	5.6.2
NPI4D	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Secondary	5.5.1
NPI4D	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Exploratory/alternative definition of Responder	5.6.3
NPI4D	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and (C + E x D + F)	MITT	Placebo, AVP-923	SPCD	Exploratory/including Placebo responders	5.6.4
NPI4D	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	ITT/CitAD patient subset	Placebo, AVP-923	SPCD	Exploratory	5.6.5
NPI4D	Actual scores at each visit	(A + C + E) x (B + G)	ITT	Only Placebo, Only AVP-923	Longitudinal analysis/linear mixed effects model	Exploratory	5.6.4
NPI4D	Change from Baseline to Visit 7/LOCF	(A + C + E) x (B + G)	ITT	Only Placebo, Only AVP-923	ANCOVA	Exploratory	5.6.4
NPI 4D	Change from Stage 1 Baseline to Visit 4	A x B	MITT	Placebo, AVP-923	ANCOVA	Exploratory	5.6.4
NPI4D	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT, CGI-S =4 vs. >4 at Stage 1 Baseline	Placebo, AVP-923	Descriptive	Exploratory	5.6.5
NPI4D	Percent Decrease at Visit 4 from Stage 1 Baseline >=25%, >=50%, >=75%	(A + C + E) x (B + G)	MITT	Placebo, AVP-923	Descriptive	Exploratory	5.6.4
NPI4D	Percent Decrease at Visit 7 from Stage 2 Baseline >=25%, >=50%, >=75%	(C + E) x (D + F)	MITT	Placebo, AVP-923	Descriptive	Exploratory	5.6.4

Efficacy Endpoint	Comparison	Data Grouping (Figure 1)	Population	Treatment Groups Displayed on Table	Statistical Method	Interpretation	SAP Section
NPI4D	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT, excluding patients using different versions of NPI	Placebo, AVP-923	SPCD	Exploratory Sensitivity analysis	5.6.2
NPI4D	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT, excluding patients using NPI-NH throughout	Placebo, AVP-923	SPCD	Exploratory Sensitivity analysis	5.6.2
NPI4A	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Secondary	5.5.1
NPI4A	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Exploratory/alternative definition of Responder	5.6.3
NPI4A	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and (C + E x D + F)	MITT	Placebo, AVP-923	SPCD	Exploratory/including Placebo responders	5.6.4
NPI4A	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	ITT/CitAD patient subset	Placebo, AVP-923	SPCD	Exploratory	5.6.5
NPI4A	Actual scores at each visit	(A + C + E) x (B + G)	ITT	Only Placebo, Only AVP-923	Longitudinal Analysis/linear mixed effects model	Exploratory	5.6.4
NPI4A	Change from Baseline to Visit 7/LOCF	(A + C + E) x (B + G)	ITT	Only Placebo, Only AVP-923	ANCOVA	Exploratory	5.6.4
NPI4A	Change from Stage 1 Baseline to Visit 4	A x B	MITT	Placebo, AVP-923	ANCOVA	Exploratory	5.6.4
NPI4A	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT, CGI-S =4 vs. >4 at Stage 1 Baseline	Placebo, AVP-923	Descriptive	Exploratory	5.6.5
NPI4A	Percent Decrease at Visit 4 from Stage 1 Baseline >=25%, >=50%, >=75%	(A + C + E) x (B + G)	MITT	Placebo, AVP-923	Descriptive	Exploratory	5.6.4

CONFIDENTIAL

Page 46 of 56

Efficacy Endpoint	Comparison	Data Grouping (Figure 1)	Population	Treatment Groups Displayed on Table	Statistical Method	Interpretation	SAP Section
NPI4A	Percent Decrease at Visit 7 from Stage 2 Baseline $\geq 25\%$ , $\geq 50\%$ , $\geq 75\%$	(C + E) x (D + F)	MITT	Placebo, AVP-923	Descriptive	Exploratory	5.6.4
NPI4A	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT, excluding patients using different versions of NPI	Placebo, AVP-923	SPCD	Exploratory Sensitivity analysis	5.6.2
NPI4A	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT, excluding patients using NPI-NH throughout	Placebo, AVP-923	SPCD	Exploratory Sensitivity analysis	5.6.2
Total NPI Score	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Secondary	5.5.1
Total NPI Score	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Exploratory/alternative definition of Responder	5.6.3
Total NPI Score	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and (C + E x D + F)	MITT	Placebo, AVP-923	SPCD	Exploratory/including Placebo responders	5.6.4
Total NPI Score	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	ITT/CitAD patient subset	Placebo, AVP-923	SPCD	Exploratory	5.6.5
Total NPI Score	Actual scores at each visit	(A + C + E) x (B + G)	ITT	Only Placebo, Only AVP-923	Longitudinal Analysis/linear mixed effects model	Exploratory	5.6.4
Total NPI Score	Change from Baseline to Visit 7/LOCF	(A + C + E) x (B + G)	ITT	Only Placebo, Only AVP-923	ANCOVA	Exploratory	5.6.4
Total NPI Score	Percent Decrease at Visit 4 from Stage 1 Baseline $\geq 25\%$ , $\geq 50\%$ , $\geq 75\%$	(A + C + E) x (B + G)	MITT	Placebo, AVP-923	Descriptive	Exploratory	5.6.4
Total NPI Score	Percent Decrease at Visit 7 from Stage 2 Baseline $\geq 25\%$ , $\geq 50\%$ , $\geq 75\%$	(C + E) x (D + F)	MITT	Placebo, AVP-923	Descriptive	Exploratory	5.6.4

CONFIDENTIAL

Page 47 of 56

Efficacy Endpoint	Comparison	Data Grouping (Figure 1)	Population	Treatment Groups Displayed on Table	Statistical Method	Interpretation	SAP Section
Total NPI Score	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT, excluding patients using different versions of NPI	Placebo, AVP-923	SPCD	Sensitivity analysis	5.6.2
Total NPI Score	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT, excluding patients using NPI-NH throughout	Placebo, AVP-923	SPCD	Sensitivity analysis	5.6.2
All Other NPI Domains separately	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Secondary	5.5.1
All Other NPI Domains separately	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Exploratory/alternative definition of Responder	5.6.3
All Other NPI Domains separately	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and (C + E x D + F)	MITT	Placebo, AVP-923	SPCD	Exploratory/including Placebo responders	5.6.4
All Other NPI Domains separately	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	ITT/CitAD patient subset	Placebo, AVP-923	SPCD	Exploratory	5.6.5
All Other NPI Domains separately	Actual scores at each visit	(A + C + E) x (B + G)	ITT	Only Placebo, Only AVP-923	Longitudinal Analysis/linear mixed effects model	Exploratory	5.6.4
All Other NPI Domains separately	Change from Baseline to Visit 7/LOCF	(A + C + E) x (B + G)	ITT	Only Placebo, Only AVP-923	ANCOVA	Exploratory	5.6.4
NPI Caregiver Distress Score for NPI Agitation/Aggression Domain	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Secondary	5.6.3
NPI Caregiver Distress Score for NPI Agitation/Aggression Domain	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Exploratory/alternative definition of Responder	5.6.3
NPI Caregiver Distress Score for NPI Agitation/Aggression Domain	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and (C + E x D + F)	MITT	Placebo, AVP-923	SPCD	Exploratory/including Placebo responders	5.6.4

CONFIDENTIAL

Page 48 of 56



Efficacy Endpoint	Comparison	Data Grouping (Figure 1)	Population	Treatment Groups Displayed on Table	Statistical Method	Interpretation	SAP Section
NPI Caregiver Distress Score for NPI Agitation/Aggression Domain	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	ITT/CitAD patient subset	Placebo, AVP-923	SPCD	Exploratory	5.6.5
NPI Caregiver Distress Score for Total NPI	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Secondary	5.6.3
NPI Caregiver Distress Score for Total NPI	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Exploratory/alternative definition of Responder	5.6.3
NPI Caregiver Distress Score for Total NPI	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and (C + E x D + F)	MITT	Placebo, AVP-923	SPCD	Exploratory/including Placebo responders	5.6.4
NPI Caregiver Distress Score for Total NPI	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	ITT/CitAD patient subset	Placebo, AVP-923	SPCD	Exploratory	5.6.5
NPI Caregiver Distress Score for NPI4D	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Secondary	5.6.3
NPI Caregiver Distress Score for NPI4D	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Exploratory/alternative definition of Responder	5.6.3
NPI Caregiver Distress Score for NPI4D	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and (C + E x D + F)	MITT	Placebo, AVP-923	SPCD	Exploratory/including Placebo responders	5.6.4
NPI Caregiver Distress Score for NPI4D	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	ITT/CitAD patient subset	Placebo, AVP-923	SPCD	Exploratory	5.6.5
NPI Caregiver Distress Score for NPI4A	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Secondary	5.6.3
NPI Caregiver Distress Score for NPI4A	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Exploratory/alternative definition of Responder	5.6.3
NPI Caregiver Distress Score for NPI4A	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and (C + E x D + F)	MITT	Placebo, AVP-923	SPCD	Exploratory/including Placebo responders	5.6.4
NPI Caregiver Distress Score for NPI4A	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	ITT/CitAD patient subset	Placebo, AVP-923	SPCD	Exploratory	5.6.5

Efficacy Endpoint	Comparison	Data Grouping (Figure 1)	Population	Treatment Groups Displayed on Table	Statistical Method	Interpretation	SAP Section
CSI	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Secondary	5.5.1
CSI	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Exploratory/alternative definition of Responder	5.6.3
CSI	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and (C + E x D + F)	MITT	Placebo, AVP-923	SPCD	Exploratory/including Placebo responders	5.6.4
CSI	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	ITT/CitAD patient subset	Placebo, AVP-923	SPCD	Exploratory	5.6.5
CSI	Actual scores at each visit	(A + C + E) x (B + G)	ITT	Only Placebo, Only AVP-923	Longitudinal Analysis/linear mixed effects model	Exploratory	5.6.4
CSI	Change from Baseline to Visit 7/LOCF	(A + C + E) x (B + G)	ITT	Only Placebo, Only AVP-923	ANCOVA	Exploratory	5.6.4
ADCS-ADL	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Secondary	5.5.1
ADCS-ADL	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Exploratory/alternative definition of Responder	5.6.3
ADCS-ADL	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and (C + E x D + F)	MITT	Placebo, AVP-923	SPCD	Exploratory/including Placebo responders	5.6.4
ADCS-ADL	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	ITT/CitAD patient subset	Placebo, AVP-923	SPCD	Exploratory	5.6.5
ADCS-ADL	Change from Baseline to Visit 7/LOCF	(A + C + E) x (B + G)	ITT	Only Placebo, Only AVP-923	ANCOVA	Exploratory	5.6.4
QoL-AD	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Secondary	5.5.1
QoL-AD	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Exploratory/alternative definition of Responder	5.6.3

Efficacy Endpoint	Comparison	Data Grouping (Figure 1)	Population	Treatment Groups Displayed on Table	Statistical Method	Interpretation	SAP Section
QoL-AD	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and (C + E x D + F)	MITT	Placebo, AVP-923	SPCD	Exploratory/including Placebo responders	5.6.4
QoL-AD	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	ITT/CitAD patient subset	Placebo, AVP-923	SPCD	Exploratory	5.6.5
QoL-AD	Change from Baseline to Visit 7/LOCF	(A + C + E) x (B + G)	ITT	Only Placebo, Only AVP-923	ANCOVA	Exploratory	5.6.4
CSDD	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Secondary	5.5.1
CSDD	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Exploratory/alternative definition of Responder	5.6.3
CSDD	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and (C + E x D + F)	MITT	Placebo, AVP-923	SPCD	Exploratory/including Placebo responders	5.6.4
CSDD	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	ITT/CitAD patient subset	Placebo, AVP-923	SPCD	Exploratory	5.6.5
CSDD	Change from Baseline to Visit 7/LOCF	(A + C + E) x (B + G)	ITT	Only Placebo, Only AVP-923	ANCOVA	Exploratory	5.6.4
MMSE	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Secondary	5.5.1
MMSE	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Exploratory/alternative definition of Responder	5.6.3
MMSE	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and (C + E x D + F)	MITT	Placebo, AVP-923	SPCD	Exploratory/including Placebo responders	5.6.4
MMSE	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	ITT/CitAD patient subset	Placebo, AVP-923	SPCD	Exploratory	5.6.5
MMSE	Change from Baseline to Visit 7/LOCF	(A + C + E) x (B + G)	ITT	Only Placebo, Only AVP-923	ANCOVA	Exploratory	5.6.4
ADCS-CGIC Agitation	Summary of Scores at Visit 4 and Visit 7 (Relative to Stage 1 Baseline)	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Secondary	5.5.1

Efficacy Endpoint	Comparison	Data Grouping (Figure 1)	Population	Treatment Groups Displayed on Table	Statistical Method	Interpretation	SAP Section
ADCS-CGIC Agitation	Summary of Scores at Visit 4 and Visit 7(Relative to Stage 1 Baseline)	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Exploratory/alternative definition of Responder	5.6.3
ADCS-CGIC Agitation	Summary of Scores at Visit 4 and Visit 7(Relative to Stage 1 Baseline)	A x B and (C + E x D + F)	MITT	Placebo, AVP-923	SPCD	Exploratory/including Placebo responders	5.6.4
ADCS-CGIC Agitation	Summary of Scores at Visit 4 and Visit 7(Relative to Stage 1 Baseline)	A x B and C x D	ITT/CitAD patient subset	Placebo, AVP-923	SPCD	Exploratory	5.6.5
ADCS-CGIC Agitation	Summary of Scores at Visit 4 and Visit 7(Relative to Stage 1 Baseline)	A x B and C x D	Week 4 Evaluable	Placebo, AVP-923	SPCD	Exploratory	5.6.3
ADCS-CGIC Agitation	Summary of Scores at Visit 4 and Visit 7(Relative to Visit 4)	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Secondary	5.5.1
ADCS-CGIC Agitation	Summary of Scores at Visit 4 and Visit 7(Relative to Visit 4)	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Exploratory/alternative definition of Responder	5.6.3
ADCS-CGIC Agitation	Summary of Scores at Visit 4 and Visit 7(Relative to Visit 4)	A x B and (C + E x D + F)	MITT	Placebo, AVP-923	SPCD	Exploratory/including Placebo responders	5.6.4
ADCS-CGIC Agitation	Summary of Scores at Visit 4 and Visit 7(Relative to Visit 4)	A x B and C x D	ITT/CitAD patient subset	Placebo, AVP-923	SPCD	Exploratory	5.6.5
ADCS-CGIC Agitation	Summary of Scores at Visit 4 and Visit 7(Relative to Visit 4)	A x B and C x D	Week 4 Evaluable	Placebo, AVP-923	SPCD	Exploratory	5.6.3
ADCS-CGIC Agitation	Summary of Scores at Visit 4 and Visit 7	A x B and C x D	MITT	Placebo, AVP-923	Proportional Odds Regression	Exploratory	5.6.3
PGI-C	Summary of Scores at Visit 4 and Visit 7 (Relative to Baseline)	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Secondary	5.5.1
ADCS-CGIC Overall Clinical Status	Summary of Scores at Visit 4 and Visit 7 (Relative to Stage 1 Baseline)	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Secondary	5.5.1
ADCS-CGIC Overall Clinical Status	Summary of Scores at Visit 4 and Visit 7(Relative to Visit 4)	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Secondary	5.5.1

Efficacy Endpoint	Comparison	Data Grouping (Figure 1)	Population	Treatment Groups Displayed on Table	Statistical Method	Interpretation	SAP Section
PGI-C	Summary of Scores at Visit 4 and Visit 7(Relative to Baseline)	A x B and C x D	MITT	Placebo, AVP-923	SPCD	Exploratory/alternative definition of Responder	5.6.3
PGI-C	Summary of Scores at Visit 4 and Visit 7(Relative to Baseline)	A x B and (C + E x D + F)	MITT	Placebo, AVP-923	SPCD	Exploratory/including Placebo responders	5.6.4
PGI-C	Summary of Scores at Visit 4 and Visit 7(Relative to Baseline)	A x B and C x D	ITT/CitAD patient subset	Placebo, AVP-923	SPCD	Exploratory	5.6.5
PGI-C	Summary of Scores at Visit 4 and Visit 7	A x B and C x D	MITT	Placebo, AVP-923	Proportional Odds Regression	Exploratory	5.6.3
ADAS-Cog	Change from Stage 1 Baseline to Visit 4 and Stage 2 Baseline to Visit 7	A x B and C x D	ITT	Placebo, AVP-923	SPCD	Exploratory	5.6.1
Changes in Allowed Psychotropic Medications		(A + C + E), (B + G), (A + D + F), (B + D + F + G)	MITT	Only Placebo, Only AVP-923, Placebo/AVP-923, All AVP-923, Overall	Descriptive	Secondary	5.5.2
Rescue Medication (Oral Lorazepam) Use		(A + C + E) x (B + G)	MITT	Only Placebo, Only AVP-923, Placebo/AVP-923, All AVP-923, Overall	Descriptive	Secondary	5.5.2

## Appendix 5: Adverse Event Start/Stop Date Imputation

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

Parameter	Missing	Additional Conditions	Imputation
Start date for AEs	D	M and Y same as M and Y of first dose of study drug	Date of first dose of study drug
		M and/or Y not same as date of first dose of study drug	First day of month
	D and M	Y same as Y of first dose of study drug	Date of first dose of study drug
		Y prior to Y of first dose of study drug but same as Y of screening date	Date of screening date
	D, M, Y	None - date completely missing	Date of first dose of study drug
Stop date for AEs	D	M and Y same as M and Y of last dose of study drug	Date of last dose of study drug
		M and/or Y not same as date of last dose of study drug	Use last day of month
	D and M	Y same as Y of last dose of study drug	Date of last dose of study drug
		Y not same as Y of last dose of study drug	Use Dec 31
	D, M, Y	None - date completely missing	No imputation, but assume ongoing

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

Similarly, if the estimated stop date is before a complete or imputed start date, use the last day of the start day month.

## Appendix 6: Prior and Concomitant Medication Start Date Imputation

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

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

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

Similarly, if the estimated stop date is before a complete or imputed start date, use the last day of the start day month.

## Appendix 7: Medical History Start Date Imputation

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

Parameter	Missing	Additional Conditions	Imputation
Start date for medical history	D only	M and Y prior to screening date	Last day of month
		M and Y same as screening date	Day prior to screening date
	M and D	Y prior to year of screening date	Jan 1 of Y
		Y is same as screening year	Day prior to screening date
M, D, and Y	None - date completely missing	Keep as missing	