# Statistical Analysis Plan (SAP)

Protocol Number: CTP2S1502HT6

Protocol Title: A Phase 2A Multicenter, Randomized, Double-Blind, Parallel Group, 26-Week, Placebo-Controlled Study of 50 mg and 100 mg of SUVN-502 in Subjects with Moderate Alzheimer's Disease Currently Treated with Donepezil Hydrochloride and Memantine Hydrochloride

NCT02580305

Date: 31 October 2019

[redacted]

Suven Life Science Limited

CTP2S1502HT6

# A PHASE 2A MULTICENTER, RANDOMIZED, DOUBLE-BLIND, PARALLEL GROUP, 26-WEEK, PLACEBO-CONTROLLED STUDY OF 50 MG AND 100 MG OF SUVN-502 IN SUBJECTS WITH MODERATE ALZHEIMER'S DISEASE CURRENTLY TREATED WITH DONEPEZIL HYDROCHLORIDE AND MEMANTINE HYDROCHLORIDE

Statistical Analysis Plan

[redacted]

[redacted]

[redacted]

| | [redacted] |
|---|---|
| [redacted] | [redacted] |

## TABLE OF CONTENTS

## LIST OF ABBREVIATIONS

| | |
|---|---|
| 5-HT6 | Serotonin Receptor Subtype 6 |
| ADAScog-11 | 11-item Alzheimer's Disease Assessment Scale for Cognition |
| ADCS-ADL | Alzheimer's Disease Cooperative Study Activity of Daily Living |
| AE | Adverse Event |
| ALT | Alanine Aminotransferase |
| APO-E | Apolipoprotein E |
| AST | Aspartate Aminotransferase |
| BDRM | Blinded Data Review Meeting |
| BID | Twice Daily |
| BMI | Body Mass Index |
| CDR | Washington University Clinical Dementia Rating Scale |
| CDR-SB | Clinical Dementia Rating- Sum of Boxes |
| CI | Confidence Interval |
| CM | Concomitant Medications |
| C-SDD | Cornell Scale for Depression and Dementia |
| CSR | Clinical Study Report |
| C-SSRS | Columbia Suicide Severity Rating Scale |
| CV | Coefficient of Variation |
| ECG | Electrocardiogram |
| eCRF | Electronic Case Report Form |
| EMA | European Medicines Agency |
| EP | Evaluable Population |
| FDA | Food and Drug Administration |
| GGT | Gamma Glutamyl Transpeptidase |
| HbA1c | Hemoglobin A1c or glycated hemoglobin |
| HCl | Hydrochloride |
| ICH | International Conference on Harmonization |
| ITT | Intent-to-Treat |
| IWRS | Interactive Web Response System |
| LOCF | Last Observation Carried Forward |
| LS | Least Squares |
| MCH | Mean Corpuscular Hemoglobin |
| MCV | Mean Corpuscular Volume |

| | |
|---|---|
| MedDRA | Medical Dictionary for Regulatory Activities |
| MHIS | Modified Hachinski Ischemic Score |
| mITT | Modified Intent-to-Treat |
| MMSE | Mini-Mental State Examination |
| MMRM | Mixed Model Repeated Measures |
| NINCDS-ADRDA | National Institute of Neurological and Communicative Diseases and Stroke/Alzheimer's Disease and Related Disorders Association |
| NMDA | N-methyl-D-aspartic acid |
| NPI | Neuropsychiatry Inventory |
| PDS | Protocol Deviation Specification |
| PK | Pharmacokinetics |
| PT | Preferred Term |
| QD | Once Daily |
| SAE | Serious Adverse Event |
| SAP | Statistical Analysis Plan |
| SD | Standard Deviation |
| SI | International System of Units |
| SOC | System Organ Class |
| SP | Safety Population |
| T4 | Thyroxine |
| TEAE | Treatment Emergent Adverse Event |
| ULN | Upper Limit of Normal |
| WHO | World Health Organization |

# 1 INTRODUCTION

Alzheimer's disease is a progressive and fatal neurodegenerative disorder manifested by cognitive and memory deterioration in addition to progressive impairment of activities of daily living. Currently available treatments, including acetylcholinesterase inhibitors (donepezil, rivastigmine, and galantamine) and the non-competitive *N*-methyl-D-aspartic acid (NMDA) receptor blocker (memantine) are seen as minimally effective, with only minor symptomatic improvements for a limited duration, and they do not slow the progression of the disease.

SUVN-502 is proposed as a novel, highly selective and orally active antagonist at a non-peripheral central nervous system serotonin receptor, the serotonin receptor subtype 6 (5-$HT_6$), intended for the treatment of cognitive disorders associated with Alzheimer's disease.

The 5-$HT_6$ receptor may be a viable target for pharmacologic intervention to improve the cognitive function of subjects with Alzheimer's disease. Antagonism of this receptor has been shown to improve learning and memory in animal models. SUVN-502 has been shown to improve learning and memory when tested in several models including the water maze test and novel object recognition task.

In animal testing, sub-therapeutic doses of SUVN-502 improved learning and memory in animals treated with sub-therapeutic dose of donepezil. SUVN-502 potentiated the increase in acetylcholine produced by the combination of donepezil and memantine in rat hippocampus as well as the procognitive effects of combined memantine and donepezil treatment in animal models. These data support the hypothesis that use of SUVN-502 in combination with donepezil and memantine might enhance the cognitive function of patients with Alzheimer's disease.

This study is designed to test the hypothesis that 50 mg and 100 mg of SUVN-502 improves the cognitive status of Alzheimer's disease subjects who are currently receiving donepezil hydrochloride (HCl) and memantine HCl, compared to placebo.

Animal toxicology data support the safety of the two dosage levels of SUVN-502, 50 mg and 100 mg, to be used in this study and data from a phase I single and multiple dose studies suggest SUVN-502 would be relatively safe and well tolerated in subjects with moderate Alzheimer's disease currently receiving donepezil HCl and memantine HCl.

The design of this study follows the recommendation of the Food and Drug Administration (FDA) and the European Medicine Agency (EMA) for studies of small molecule symptomatic drugs in Alzheimer's disease. The study is double-blind, placebo-controlled and follows an add-on design to the standard of care. Efficacy assessments include cognitive, functional, and global evaluations using well-recognized scales.

This Statistical Analysis Plan (SAP) is written in accordance with principles described in International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH): Guidance on Statistical Principles in Clinical Trials and is designed to guide the statistical analysis of Study CTP2S1502HT6, as a basis for evidence of the efficacy and safety of SUVN-502. It is based on version 2.0 of the protocol dated 02 March 2016. In the event of future amendments to the protocol, this SAP will be modified as necessary to account for the changes relevant to the statistical analysis.

## 2 STUDY OBJECTIVES

The primary objective of the study is to evaluate the efficacy of SUVN-502 at daily doses of 50 mg or 100 mg compared to placebo, as adjunct treatment in subjects with moderate Alzheimer's disease (Mini-Mental State Examination [MMSE] score of 12 to 20) currently treated with the acetylcholinesterase inhibitor, donepezil HCl, and the NMDA antagonist, memantine HCl. Efficacy will be assessed by the 11-item Alzheimer's Disease Assessment Scale for Cognitive Behavior (ADAScog-11) after 26 weeks of treatment.

The secondary objectives of the study are:

- To further evaluate the efficacy of these treatments using the following scales:
    - Clinical Dementia Rating Scale – Sum of Boxes (CDR-SB)
    - MMSE
    - Alzheimer's Disease Co-operative Study Activity of Daily Living (ADCS-ADL)
    - Neuropsychiatric Inventory (NPI) 12 item
    - Cornell Scale for Depression and Dementia (C-SDD)
- To evaluate the safety and therapeutic tolerability of these treatments using adverse events (AEs), laboratory evaluations, blood pressure, electrocardiograms

(ECGs), physical and neurological examination, and the Columbia Suicide Severity Rating Scale (C-SSRS)

- To evaluate the pharmacokinetics (PK) of SUVN-502 administered in combination with donepezil HCl and memantine HCl

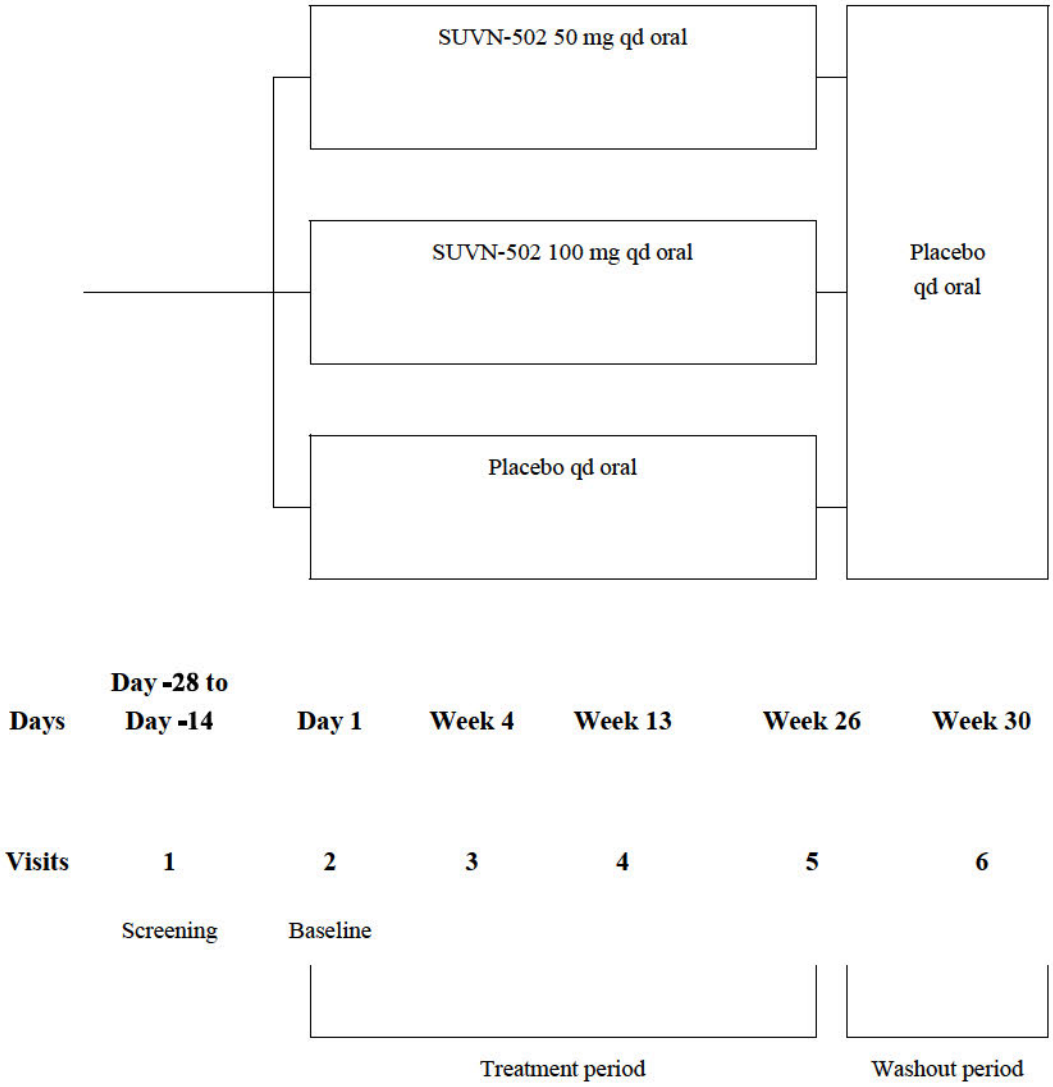
To explore the relationship between the efficacy of SUVN-502 and apolipoprotein E (APO-E) genotype, as well as analyze subjects more likely to have Alzheimer's disease.

# 3 INVESTIGATIONAL PLAN

## 3.1 Overall Study Design and Plan

This is a phase 2A, proof-of-concept, 26-week, double-blind, multicenter, randomized, parallel group, placebo-controlled study to compare the efficacy and safety of treatment with SUVN-502 (50 mg or 100 mg once daily [qd]) to placebo treatment in subjects with moderate Alzheimer's disease receiving donepezil HCl (10 mg qd) and either memantine HCl (10 mg twice daily [bid]), Namenda XR® (28 mg qd), or the combination therapy Namzaric™ containing 10 mg donepezil HCl and 28 mg Namenda XR®. Efficacy will be assessed by the ADAScog-11 after 26 weeks of treatment.

The study design is shown in Figure 1. The study consists of a two- to four-week screening period, followed by a 26-week double-blind treatment period, and a four-week single-blind placebo washout period.




*qd: once daily*

**Figure 1: Study Design**

A schedule of study procedures and evaluations is provided in Section 7.1 of the protocol.

Randomization of subject to treatment will occur at baseline (Visit 2) after all screening procedures have been performed and eligibility has been confirmed. Each randomized subject will receive a unique randomization number. Randomized subjects who terminate their study participation for any reason, regardless of whether study drug was taken or not, will retain their study number. For the randomization of subjects, the Investigator will use an IWRS. Randomization will be balanced by site with block size of 6.

The study population will include male or female subjects, 50 to 85 years of age, with moderate dementia due to probable Alzheimer's disease based on the National Institute of

Neurological and Communicative Diseases and Stroke/Alzheimer's Disease and Related and Related Disorders Association (NINCDS-ADRDA) criteria, diagnosed at least 1 year prior to the study and receiving stable doses for at least 3 months of donepezil HCl (10 mg qd) and either memantine HCl (10 mg bid), or Namenda XR® (28 mg memantine HCl extended-release qd) or the combination therapy, Namzaric™ (28 mg memantine HCl extended-release / 10 mg donepezil qd).

A total of 537 eligible subjects will be randomized in a 1:1:1 ratio to one of the following three study treatments groups (179 subjects per group):

- SUVN-502 (50 mg)
- SUVN-502 (100 mg)
- Placebo

During the course of the study, an Independent Data and Safety Monitoring Board (DSMB) will be assembled to periodically review and evaluate accumulated study data for subject safety, study conduct, and progress. The DSMB will have the responsibility to make recommendations concerning continuation, modification, or termination of the study.

There are no interim analyses planned for this study.

Details of the study design are given in the study protocol.

### 3.2 Efficacy and Safety Variables

The primary efficacy variable is the ADAScog-11 score at Week 26.

The secondary efficacy variables include:

- ADAScog-11 at Weeks 4 and 13
- CDR-SB at Weeks 4, 13, and 26
- MMSE at Weeks 4, 13, and 26
- ADCS-ADL at Weeks 4, 13, and 26
- NPI at Weeks 4, 13, and 26
- C-SDD at Weeks 4, 13, and 26

The safety variables include AEs reported throughout the study and blood pressure measurement, ECGs, physical and neurological examination, and assessment of suicidality using the C-SSRS at Weeks 4, 13, 26, and 30. Blood and urine samples for

laboratory evaluations will be collected at screening, baseline, Weeks 4 (Liver Function Test only), 13, 26 and 30.

# 4 STATISTICAL METHODS

## 4.1 Data Quality Assurance

All tables, figures and data listings to be included in the report will be independently checked for consistency, integrity, and in accordance with standard [redacted] procedures.

## 4.2 General Presentation Considerations

This section describes the analysis and data conventions that will be utilized for this study. All statistical analyses and summary information are to be generated according to this SAP. Any deviations from this SAP will be documented in the clinical study report (CSR).

### 4.2.1 General Definitions

Unless otherwise stated, baseline measurement is defined as the latest, valid pre-dose assessment available. 'End of Study' is defined as the last available post-treatment assessment. A subject who completes all study visits up to Week 30, will be considered to have completed the study. For evaluation of efficacy, a completed subject is defined as any subject who completes all visits up to Week 26.

### 4.2.2 Visit Windows

Assessments will be classified based on the visit information reflected on the eCRF. Assessments taken outside of protocol allowable windows will be assigned based on the closest scheduled visit. If there are assessments recorded on a day that is equidistant between two visits, the assessments will be attributed to the later visit. The only exception to this rule is for assessments recorded between Baseline and Week 4. Any assessment recorded after the Baseline visit will be mapped to Week 4 even if the day is closer to the Baseline visit. The timetable of visits is given in Section 7.1 in the protocol. Nominal study visits (Weeks relative to Baseline) will be used to summarize data presented by visit.

If subjects have more than one observation for a given time point, the average of the measurements will be used. This rule applies to efficacy data. For safety data, unless otherwise specified, if there is a repeated evaluation, the worst value will be used. If values are equal, the latest repeated (latest) value will be used.

If subjects withdraw early from the study, all study assessments at the early withdrawal visit will be assigned based on the rules described above. That is, they will be assigned to the closest scheduled visit. Data from subjects who discontinue but return after 4 weeks for the follow-up visit will be summarized with the Week 30 visits.

### 4.2.3 Handling of Outliers

Any outliers that are detected statistically during the blind review of the data will be investigated. If necessary, queries will be issued to the Investigator, to either correct or confirm the reason for the outlier. If any outlier is confirmed, the actual value will be included in the analysis.

### 4.2.4 Site Pooling Strategy

Site will be included in the model. Sites with fewer than 2 subjects in each treatment group will be pooled for analysis starting with the largest study site and pooling with the site that best balances that site's randomization until each pooled site meets the requirement for at least 2 subjects per treatment group. If the last pooled site does not have 2 subjects per treatment group, it will be combined with the next smallest pooled site.

### 4.2.5 Presentation of Descriptive and Inferential Statistics

Continuous data will be summarized in terms of the mean, standard deviation (SD), lower quartile, median, upper quartile, minimum, maximum and number of non-missing observations, unless otherwise stated. The minimum and maximum will be reported to the same number of decimal places as the raw data recorded in the database. The mean, median and quartiles will be reported to one more decimal place than the raw data recorded in the database. The SD (and standard error if applicable) will be reported to two more decimal places than the raw data recorded in the database. In general, the maximum number of decimal places reported shall be four for any summary statistic.

Categorical data will be summarized in terms of the number of subjects providing data at the relevant time point (n), frequency counts, and percentages. Any planned collapsing of categories will be detailed in the SAP text and the data displays.

Percentages will be presented to one decimal place. If the percentage is 100%, decimal will not be presented. Percentages will not be presented for zero counts. Percentages will be calculated using n from each treatment group as the denominator, unless otherwise specified in the data displays.

Changes from baseline in categorical data will be summarized using shift tables where appropriate.

Each SUVN-502 dose group comparison with the placebo group will be made at a 2-sided significance level of $\alpha$=0.05. For the secondary endpoints, each SUVN-502 dose group comparison with the placebo group will also be made at a 2-sided significance level of $\alpha$=0.05. All null hypotheses of treatment comparisons will be set as no treatment difference. Confidence intervals (CIs) produced will be 2-sided with 95% confidence level unless specifically stated otherwise.

In general, p-values greater than or equal to 0.001 will be presented to 3 decimal places. P-values less than 0.001 will be presented as "<0.001", and p-values greater than 0.999 will be presented as ">0.999". A check will be performed for p-values greater than 0.999.

Confidence intervals will be presented to 1 more decimal place than the mean.

Listings will be sorted by subject within each treatment group.

All analyses will be conducted using SAS® version 9.3 or a later version in a secure and validated environment.

## 4.3 Study Subjects

### 4.3.1 Disposition of Subjects

Information on the disposition of all subjects who enter the study will be provided, from screening to study completion (Week 30).

The following subject data will be presented by treatment group, including overall categories:

- The number of subjects screened (overall only)
- The number of subjects randomized

- The number and percent of subjects randomized who were intended to be treated with study medication
- The number and percent of subjects treated
- The number and percent who completed up to Week 26 and Week 30 follow-up
- The number and percent who completed up to Week 26 but not Week 30 follow-up
- The number and percent of subjects who withdrew from study prematurely but took placebo and completed Week 30 follow-up
- The number and percent of subjects who withdrew from study prematurely but completed Week 30 follow-up (retrieved dropouts)
- The number and percent of subject who withdrew from study prematurely and did not complete follow-up
- Time to discontinuation

Percentages of subjects will be based on the number of subjects randomized in each treatment group as 100%.

The number and percentage of subjects who prematurely withdrew during the study will be presented by reason for discontinuation. Time to study discontinuation will be presented by treatment group as the cumulative proportion of dropouts up to each visit.

In addition, subject listings will be provided for subjects who discontinued the study early (post-randomization) with reason for discontinuation.

### 4.3.2 Protocol Deviations

Protocol deviations will be identified on an ongoing basis by the clinical study team based on the protocol deviation specification document (PDS) and assessed as 'major' and 'minor' in consultation with the Sponsor. This assessment will be done during blinded data review meetings (BDRMs), which will take place when pre-determined accrual targets are met and before disclosure of randomization codes. Major deviations from the protocol that may lead to confounding of efficacy results will result in the exclusion of a subject from the Evaluable Population (EP). The final determination of major protocol deviations and the exclusion of subjects from each of the analysis populations will be made prior to database lock and unblinding.

Unless decided otherwise during the BDRMs, the following will be major protocol deviations:

- Informed consent violation
- Did not have any assessment of the primary endpoint at any of the visits at which the scales were scheduled to be assessed
- Protocol violations of certain inclusion/exclusion criteria
- Had unqualified raters or raters with substantial scoring errors for the primary measures
- Was not compliant with regard to study drug or concomitant Alzheimer's disease medications
- Use of prohibited medication during the double-blind period that may influence the outcome of the study for the subject
- Subject met withdrawal criteria but not withdrawn
- Subject failed to report SAE or pregnancy

Protocol deviations will be provided in a listing. Major protocol deviations will be also be summarized by treatment group for all randomized subjects.

### 4.4 Analysis Populations

Three populations are defined for the analyses:

- Modified intent to treat (mITT) population: All randomized subjects who receive at least 1 dose of study medication and have 1 post-baseline evaluation of the primary efficacy variable. Subjects will be assigned to the group they were randomized.
- Evaluable Population (EP): [redacted]
- Safety Population (SP): All randomized subjects who receive at least 1 dose of study medication. Subjects will be assigned to the treatment group based on the actual treatment received.

All efficacy analyses will be performed on the mITT population and the EP. The primary efficacy analysis of the primary endpoint will be based on the mITT population and a secondary analysis of the primary endpoint will also be performed based upon the EP. All secondary endpoints will be analyzed using both the mITT population and the EP.

Safety analyses will be performed on the SP. If a subject takes more than 1 study drug for any period of time, that subject will be included in the analyses with the highest dose group taken.

Demographic and baseline characteristics will be evaluated for the mITT population. Additional summaries will be presented using the Safety Population.

The number of subjects in the different analysis populations will be summarized by treatment group, and overall.

The subjects excluded from each analysis set will be listed with the reason for exclusion.

## 4.5 Demographic and Other Baseline Characteristics and Medical History

### 4.5.1 Demographic, Baseline, and Other Characteristics

Demographic and baseline characteristics will be summarized for the mITT population by treatment group and overall. Summaries will include descriptive statistics for continuous and categorical measures. Subject characteristics to be presented include:

- Age
- Age group: <65 years, ≥65 years and <75 years, ≥75 years
- Gender
- Height (m)
- Weight (kg)
- Body Mass Index (BMI) (weight (kg)/[height (m)]$^2$)
- Race
- APO-E4 carrier status (carrier one allele, carrier two alleles, non-carrier)
- Modified Hachinski Ischemic Score (MHIS) score
- Duration of AD (years)

  Duration of AD will be calculated as: (date of Screening – date of AD diagnosis + 1)/365.25

- Duration of memantine (years)
- Duration of donepezil (years)

  Duration of donepezil and memantine will be calculated as: (date of Screening – date of first dose + 1)/365.25
- Baseline severity of impairment as measured by ADAScog-11, CDR-SB, MMSE, ADCS-ADL, NPI, and C-SDD
- Baseline ADAScog-11 score (≤median, >median)
- Baseline MMSE score (12-15, 16-20)
- Concomitant Therapy: donepezil HCl and memantine HCl, donepezil HCl and Namenda XR®, Namzaric™

Age will be calculated as the number of complete years between a subject's birth date and the date of informed consent.

Subject listings of demographic and baseline characteristics will be provided.

### 4.5.2 Medical History

General medical history will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) version 18.1. A summary of medical history by system organ class (SOC) and preferred term (PT) will be presented as frequencies and percentages for each treatment group and overall.

Medical history will also be listed. The listing will be sorted by treatment group, subject identification number, SOC, PT, and reported term.

### 4.5.3 Concomitant Medications

Study medications will be coded using the World Health Organization (WHO) Drug dictionary version Sep2015 (WHODDE+HERBAL). Medications and all information collected will be reported as collected via the eCRF.

Prior medications are medications defined as those that stop before the first dose.

Concomitant medications (CMs) are medications taken during the study, including any medication that was ongoing at the time of first dose, medication taken on or after the first dose of study drug, or any medication started prior to the last treatment dose.

Partial date rules for flagging concomitant medications will follow the rules detailed in the AEs. In case of missing dates, a medication will be considered as concomitant.

A summary of CMs will be presented as frequencies and percentages for each treatment group and overall.

Prior and CMs will also be listed.

## 4.6 Treatment Compliance

### 4.6.1 SUVN-502 Compliance

Study medication compliance will be assessed during the Treatment Period.

Treatment Compliance is calculated as:

$$\frac{\text{Number of compliant days}}{\text{Expected number of compliant days}} \times 100\%$$

Number of compliant days is the number of days on which double-blind treatment was taken. Number of compliant days is calculated by the number of tablets dispensed less number of tablets returned at the next visit.

Expected number of compliant days is the number of days double-blind treatment is scheduled based on the subject's duration of participation in the study, excluding the follow-up phase.

In cases where the bottle is not returned at Week 4/Week 13/Week 30 and non-compliance was not reported by the subject or caregiver, the number of compliant days will be calculated based on the number of pills dispensed and capped at the number of pills they should have taken during the specific time period. That is, if there were 35 tablets dispensed at Baseline, and the subject does not return the bottle at Week 4, the number of compliant days will be 28 (if Week 4 occurs on Day 28). If non-compliance is reported when there are bottles not returned, the compliance for that subject will not be included in the summaries.

[redacted]

[redacted]

- [redacted]
- [redacted]
- [redacted]

The number of subjects considered compliant will be summarized for each timepoint by treatment group.

A listing of treatment compliance will be provided.

### 4.6.2 Concomitant Medication Compliance

Compliance will be summarized for each concomitant medication separately- donepezil HCl, memantine HCl, Namenda XR®, and Namzaric™.

Compliance for donepezil HCl and memantine HCl will be calculated in the same manner as the SUVN-502 study drug. Namenda XR® and Namzaric™ are provided by the subject on those medications, so the number of capsules taken will be recorded in the eCRF and used in the calculation.

Treatment compliance will be summarized for each concomitant medication by treatment groups between each visit using descriptive statistics, where the expected number of compliant days is the number of days between visits.

[redacted]

- [redacted]
- [redacted]
- [redacted]

The number of subjects considered compliant will be summarized for each timepoint by treatment group.

A listing of treatment compliance for each treatment will be provided.

[redacted]

- [redacted]
- [redacted]
- [redacted]

The compliance calculations for M1 of SUVN-502 subjects in the placebo group will not be calculated.

Week 26 plasma PK concentration data will also be used to determine inclusion into the EP in cases where accountability cannot be determined.

## 4.7 Efficacy Evaluation

### 4.7.1 Hypotheses

The primary and secondary analyses will compare each dose of SUVN-502 with placebo. The null hypothesis for each dose of SUVN-502 will be that there is no difference between SUVN-502 and placebo. The alternative hypothesis will be that there is a difference. Symbolically, this is expressed as follows:

$$Ho: \hat{\mu}_{SUVN-502\ 50\ mg} = \hat{\mu}_{Placebo}$$

$$Ha: \hat{\mu}_{SUVN-502\ 50\ mg} \neq \hat{\mu}_{Placebo}$$

and

$$Ho: \hat{\mu}_{SUVN-502\ 100\ mg} = \hat{\mu}_{Placebo}$$

$$Ha: \hat{\mu}_{SUVN-502\ 100\ mg} \neq \hat{\mu}_{Placebo}$$

where $\hat{\mu}$ is the mean change from baseline.

#### 4.7.1.1 Statistical Methods

Efficacy data will be analyzed using a mixed effects model for repeated measures (MMRM). The model will include fixed categorical factors for treatment, week, treatment-by-week interaction, and APO-E status (carrier- one allele, carrier- two alleles, and non-carrier), as well as a continuous covariate of baseline score, the baseline score-by-week interaction, age, and Baseline MMSE score. Pooled site, as defined in Section 4.2.4, will be included as a random effect. An unstructured covariance matrix will initially be used to model the variation within subjects for the repeated measures. The Least Square (LS) means of each treatment group, and the differences between each SUVN-502 dose group (50 mg and 100 mg) and placebo will be reported for Weeks 4, 13, and 26, along with the corresponding 95% confidence intervals and p-values. See Section 4.7.2 for further details.

#### 4.7.1.2 Handling of Dropouts or Missing Data

For the primary efficacy analyses missing total scores will be maintained as missing. Under the MMRM model, missing data is assumed to be missing at random. However, sensitivity analyses to assess the impact of missing data will also be conducted. These analyses will utilize the same method as described for the primary and secondary endpoints, however, missing total scores will be imputed using the z-score Last Observation Carried Forward (zLOCF) approach prior to analysis. That is, the missing value will be imputed as the score that is the same number of standard deviations from the treatment group mean at that time point as at the subject's last observed value. The missing values will be imputed using this method. Then the imputed values will be re-calculated using the newly imputed data in the calculation of the treatment group means. The value that is the same number of standard deviations from that treatment group mean will be used in the analysis. It is noted that the zLOCF approach is unbiased when dropouts are missing at random within the treatment groups or missing at random after taking into account the severity at the time of dropout.

If the proportion and/or pattern of missing data is found to be different to that expected, additional sensitivity analyses may be required. These will be documented in the CSR.

Handling of Missing Items in Calculating Totals

All total and subscale scores for efficacy will be derived from individual items. If any of the individual items are missing or unknown, every effort will be made to obtain the score for the missing item or items. If individual items that derive the total score are missing and fit the below criteria, then the remaining non-missing items will be used to impute the total score. This imputation is independent of the sensitivity analysis described at the beginning of this section and will be used for the primary and secondary analyses.

For the ADAScog-11, if less than 30% (fewer than 21 points) of the total score is missing, the total score (maximum 70) will be imputed using the available non-missing data. If the missing item is available at previous visits, then the missing data will first be imputed using straight line imputation. That is, a simple regression line will be fitted to interpolate or extrapolate the missing item. If the extrapolated value is outside of the range of the item, then the value will be cut off at the maximum/minimum value and will be used in the calculation of the total score. If there are items that have no non-missing values for that subject, then the total remaining items will be multiplied by a factor that includes the maximum score for the missing items. For example, if the first item, "Word Recall Task," which ranges from a score of 0 to 10 (maximum=10) is missing, and the second item "Commands," which ranges from a score of 0 to 5 (maximum=5) is missing, then the multiplication factor=70/(70-[10+5])=70/55=1.27. Thus, the total score for this example is the sum of the remaining 9 items multiplied by 1.27. The imputed number will be rounded up to the nearest integer. Specifically, if the sum of the items not including the missing "Word Recall Task" and "Commands," is 50, then the imputed score is 50*1.27=63.5. So, the imputed total score for this subject, would be 64. If more than 30% of the total score is missing, the total score for the ADAScog-11 at that visit will be considered missing.

The same approach will be used for missing items in the ADCS-ADL and CDR-SB. The only exception is that the imputed total score for CDR-SB will not be rounded up to the nearest integer.

For all other scales, if any item is missing, any total or sum involving that item will be considered missing. This includes MMSE, NPI, and C-SDD.

#### 4.7.1.3 Multiple Comparisons/Multiplicity

No adjustment for multiple comparisons is planned for the efficacy endpoints being monitored.

Safety results, including the summaries of AEs and changes in laboratory and vital signs data, will be interpreted on clinical grounds. No formal statistical hypothesis testing will be performed.

#### 4.7.1.4 Interim Analyses

No interim analysis is planned for this study.

#### 4.7.1.5 Examination of Subgroups

To assess the effects of various demographic and baseline characteristics on treatment outcome, subgroup analyses for the primary endpoint, ADAScog-11, will be performed based on the following variables:

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

[redacted]

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

[redacted]

- [redacted]
- [redacted]

[redacted]

- [redacted]
- [redacted]

[redacted]

- [redacted]
- [redacted]

[redacted]

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

[redacted]

[redacted]

#### 4.7.1.6 Exploratory Analysis of the Primary Endpoint

To test the overall efficacy of SUVN-502 over placebo regardless of dose, an exploratory analysis on the change from Baseline in ADAScog-11 score will be performed comparing placebo to the combined SUVN-502 treatment groups. This analysis will use the same methods as the primary efficacy analysis but the 50 mg and 100 mg SUVN-502 dose levels will be combined into one treatment group.

### 4.7.2 Primary Efficacy Endpoint- – ADAScog-11

The primary endpoint for the assessment of efficacy is the change from Baseline to Week-26 in the ADAScog-11 score.

The ADAScog-11 is a brief measure of cognition, including word recall, commands, constructional praxis, naming objects and fingers, ideational praxis, orientation, word recognition, remembering test instructions, spoken language ability, word-finding difficulty, and comprehension of spoken language.

*Word Recall*

The score for this item is the mean number of words not recalled on the three trials, for a maximum score of 10.

*Commands*

The score for this item is based on the total number of commands performed incorrectly, where the maximum score is 5. Scoring is summarized in the chart below:

| | |
|---|---|
| 0 | = all commands correct |
| 1 | = 1 command incorrect |
| 2 | = 2 commands incorrect |
| 3 | = 3 commands incorrect |
| 4 | = 4 commands incorrect |
| 5 | = 5 commands incorrect |

*Constructional Praxis*

The score for this item is based on the total number of forms drawn incorrectly, where the maximum score is 5. Scoring is summarized in the chart below:

| | |
|---|---|
| 0 | = all 4 drawings correct |
| 1 | = 1 form drawn incorrectly |
| 2 | = 2 forms drawn incorrectly |
| 3 | = 3 forms drawn incorrectly |
| 4 | = 4 forms drawn incorrectly |
| 5 | = no figures drawn |

*Naming Objects and Fingers*

The score for this item is based on the number of different objects and fingers that the subject named incorrectly. The maximum score for this item is 5 and is summarized in the chart below:

| | |
|---|---|
| 0 | = 0-2 items (objects and fingers) named incorrectly |
| 1 | = 3-5 items (objects and fingers) named incorrectly |

| | |
|---|---|
| 2 | = 6-8 items (objects and fingers) named incorrectly |
| 3 | = 9-11 items (objects and fingers) named incorrectly |
| 4 | = 12-14 items (objects and fingers) named incorrectly |
| 5 | = 15-17 items (objects and fingers) named incorrectly |

*Ideational Praxis*

The score for this item is based on the total number of components the subjects are not able to perform. The maximum score is 5. Scoring is summarized in the chart below:

| | |
|---|---|
| 0 | = all components performed correctly |
| 1 | = failure to perform 1 component |
| 2 | = failure to perform 2 components |
| 3 | = failure to perform 3 components |
| 4 | = failure to perform 4 components |
| 5 | = failure to perform 5 components |

*Orientation*

The score for this item is based on the total number of responses incorrect. The maximum score is eight. Scoring is summarized in the chart below:

| | |
|---|---|
| 0 | = all responses correct |
| 1 | =1 response incorrect |
| 2 | =2 responses incorrect |
| 3 | =3 responses incorrect |
| 4 | =4 responses incorrect |

| | |
|---|---|
| 5 | =5 responses incorrect |
| 6 | =6 responses incorrect |
| 7 | =7 responses incorrect |
| 8 | =8 responses incorrect |

*Word Recognition*

The score for this item is the total number of words not recalled. The maximum score for this item is 12.

*Remembering the Instructions*

The score for this item is based solely on the subject's performance on the Word Recognition Task. The score corresponds to the number of reminders given after the first two rounds only reminders after the second word are included in the scoring range of 0-22. Scoring for this item is summarized below, with a maximum score of 5.

| | | |
|---|---|---|
| 0 | None | = subject never needs extra reminders of instructions |
| 1 | Very mild | = forgets once |
| 2 | Mild | = must be reminded 2 times |
| 3 | Moderate | = must be reminded 3 or 4 times |
| 4 | Moderately severe | = must be reminded 5 or 6 times |
| 5 | Severe | = must be reminded 7 or more times |

*Comprehension*

The score for this item is based on the evaluation of the subject's ability to understand speech (what is being said to them). Scoring is summarized in the chart below.

| | | |
|---|---|---|
| 0 | None | = subject understands |

| | | |
|---|---|---|
| 1 | Very mild | = one or two instances of misunderstanding |
| 2 | Mild | = 3-5 instances of misunderstanding |
| 3 | Moderate | = requiring several repetitions and rephrasing |
| 4 | Moderately severe | = subject only occasionally responds correctly; i.e. yes/no questions |
| 5 | Severe | = subject rarely responds to questions appropriately, not due to poverty of speech |

*Word Finding Difficulty*

The score for this item is based on rating impairment in expressive language (difficulty in finding the desired word in spontaneous speech). The maximum score is 5. Scoring is summarized in the chart below:

| | | |
|---|---|---|
| 0 | None | = subject speaks clearly and/or is understandable |
| 1 | Very mild | = 1 instance of lack of understandability |
| 2 | Mild | = subject has difficulty less than 25% of the time |
| 3 | Moderate | = subject has difficulty 25-50% of the time |
| 4 | Moderately severe | = subject has difficulty more than 50% of the time |
| 5 | Severe | = one or two word utterance; fluent, but empty speech; mute |

*Spoken Language Ability*

The score for this item is a global rating of the quality of the subject's verbal communication. The score range is 0-5, as summarized in the chart below.

| | | |
|---|---|---|
| 0 | None | = subject speaks clearly and/or is understandable |
| 1 | Very mild | = 1 instance of lack of understandability |

| | | |
|---|---|---|
| 2 | Mild | = subject has difficulty less than 25% of the time |
| 3 | Moderate | = subject has difficulty 25-50% of the time |
| 4 | Moderately severe | = subject has difficulty more than 50% of the time |
| 5 | Severe | = one or two word utterance; fluent, but empty speech; mute |

The total score for the ADAScog-11 is the sum of the above 11 items. The maximum score is 70 points. A lower score indicates a better performance.

If any item is not administered because the subject is too cognitively impaired to complete, then the subjects receives the maximum amount of points for that item. Any other reason for not administering the item on the eCRF will be left as missing. Handling of missing items in the ADAScog-11 is summarized in Section 4.7.1.2.

**Summary**

ADAScog-11 scores will be summarized by treatment group and visit in terms of absolute values and changes from baseline, both for the observed case and LOCF analyses.

A listing of observed and change from Baseline value will be provided.

**Analysis**

A MMRM will be used to test the hypotheses stated in Section 4.7.1 that there is no difference between SUVN-502 and placebo in the mean change from baseline at Week 26 for the ADAScog-11 score. The LS means of each treatment group, and the differences between each SUVN-502 dose group and placebo will be reported for Weeks 4, 13, and 26, along with the corresponding 95% confidence intervals and p-values.

The change from baseline score at each scheduled post-baseline visit (according to the Study Schedule) during the treatment period will be the dependent variable. The model will include the fixed effects of treatment, week, treatment-by-week interaction, and APO-E status (carrier- one allele, carrier- two alleles, and non-carrier), as well as a continuous covariates of baseline score, the baseline score-by-week interaction, age, and Baseline MMSE score. Pooled site, as defined in Section 4.2.4, will be included as a random effect. Week will be considered a categorical variable with values equal to the visit numbers at which the scales were assessed. An unstructured covariance matrix will

initially be used to model the within-subject variance-covariance errors. The following SAS code will be used to perform the analyses:

```
PROC MIXED;
CLASS treat week subjid apoe site;
MODEL outcome = treat week treat*week base base*week apoe mmse_base age /
        DDFM=kr;
RANDOM site;
REPEATED week / SUBJECT=subjid TYPE=UN;
LSMEANS treat/ cl diff;

RUN;
```

If the unstructured covariance structure matrix results in a lack of convergence, the following covariance structures will be used in sequence: heterogeneous compound symmetry covariance structure and compound symmetry covariance structure, until the model converges. The Kenward-Roger approximation will be used to estimate the denominator degrees of freedom.

A plot showing the LS mean change from baseline in ADAScog-11 over time within each treatment group will be provided.

Sensitivity analyses to investigate the impact of the choice of method for handling missing data and the choice of analysis population will be performed as described in Section 4.7.1.2.

The homogeneity of the treatment effect for a number of important subgroups will be investigated as described in Section 4.7.1.5.

The combined treatment effect of SUVN-502 regardless of dose will be investigated as described in Section 4.7.1.6.

Additional ADAScog-11 analyses will be performed on subsets of the mITT Population and EP:

- [redacted]
- [redacted]
- [redacted]

### 4.7.3 Secondary Efficacy Endpoints

Secondary efficacy endpoints will be analyzed using the same MMRM model as for the primary efficacy endpoint (ADAScog-11). Changes from Baseline to Weeks 4, 13, and 26 in CDR-SB score, MMSE score, ADCS-ADL score, total NPI score, NPI distress score and C-SDD score will be assessed using the output from the MMRM analyses. Each secondary endpoint will be tested separately at $\alpha$=0.05(2-sided).

*Clinical Dementia Rating Scale – Sum of Boxes*

The CDR is a global assessment of cognition and function rated in 6 domains: memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care, which is used to assess the severity of dementia.

Each domain is rated on a 5-point scale of functioning as follows:

- 0: no impairment
- 0.5: questionable impairment
- 1: mild impairment
- 2: moderate impairment
- 3: severe impairment

Personal care is scored on a 4-point scale without a 0.5 rating available.

The CDR-SB score will be obtained by summing each of the domain box scores, with scores ranging from 0 to 18. Details of handling missing domains in the CDR-SB are summarized in Section 4.7.1.2.

**Summary**

CDR-SB total scores will be summarized by treatment group and visit in terms of absolute values and changes from baseline, both for the observed case and LOCF analyses.

A plot showing the adjusted mean change from baseline in CDR-SB over time within each treatment group will be provided.

The homogeneity of the treatment effect for a number of important subgroups will be investigated as described in Section 4.7.1.5.

A listing of subjects' observed and change from Baseline value will be provided.

*Mini-Mental State Examination*

The MMSE is a brief cognitive measure and assesses orientation to time and place, immediate and delayed recall of words, attention and calculation, language (naming, comprehension, and repetition), and spatial ability (copying a figure).

Subjects receive 1 point for each correct response to the questions. The total score is sum of the number of points awarded for each item, with a max score of 30.

**Summary**

MMSE scores will be summarized by treatment group and visit in terms of absolute values and changes from baseline, both for the observed case and LOCF analyses.

A plot showing the adjusted mean change from baseline in MMSE over time within each treatment group will be provided.

The homogeneity of the treatment effect for a number of important subgroups will be investigated as described in Section 4.7.1.5.

A listing of subjects' observed and change from Baseline value will be provided.

*Alzheimer's Disease Co-operative Study Activity of Daily Living*

The ADCS-ADL is a validated scale for measuring functional abilities in activities of daily living in subjects with Alzheimer's disease. It measures 6 basic and 17 instrumental activities of daily living and was specifically developed as a sensitive tool to track changes in functional performance in Alzheimer's disease over time. The basic activities include self-care tasks such as eating, walking, toileting, bathing, grooming, and dressing. The instrumental activities are more complex skills that are required to successfully live independently and include shopping, keeping appointments, traveling outside of home, making a meal or snack, reading, and writing.

The ADCS-ADL has 23 questions and the total score is calculating by summing the scores of individual questions in the scale. Scores range from 0 to 78, with lower scores indicating greater impairment. Handling of missing items in the ADCS-ADL is summarized in Section 4.7.1.2.

If a subject has 4 or more "don't know" responses at a particular visit, then the total score for that visit will be set to missing for the primary analysis. Exploratory analysis may be conducted that includes all scores regardless of the number of "don't know" responses.

**Summary**

ADCS-ADL scores will be summarized by treatment group and visit in terms of absolute values and changes from baseline, both for the observed case and LOCF analyses.

A plot showing the adjusted mean change from baseline in ADCS-ADL over time within each treatment group will be provided.

The homogeneity of the treatment effect for a number of important subgroups will be investigated as described in Section 4.7.1.5.

A listing of subjects' observed and change from Baseline value will be provided.

*Neuropsychiatric Inventory*

The NPI is an interview-based tool for assessing behavioral domains common in dementia (delusions, hallucinations, agitation/aggression, depression/dysphoria, anxiety, elation/euphoria, apathy/indifference, disinhibition, irritability, aberrant motor behavior, sleep, appetite and eating disorders) in terms of frequency, severity, and caregiver distress.

Frequency is rated as:

1. Occasionally- less than once per week
2. Often- about once per week
3. Frequently- several times per week but less than every day
4. Very frequently- once or more per day

Severity is rated as:

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

The score for each domain is: domain score = frequency x severity.

The Total NPI score is the sum of each domain. There are 12 items in the NPI, so the maximum score is 144 (12 for each domain).

NPI subtotals will be based on the following:

- NPI Domains: Delusions + Hallucinations will sum the delusions and hallucinations domains. The maximum score is 24.

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

Additionally, the total distress score for the NPI is the sum of the scores of the individual caregiver distress questions.

Caregiver distress is scored as:

0. Not at all
1. Minimally
2. Mildly
3. Moderately
4. Severely
5. Very severely or extremely

The maximum score is 60 (5 for each domain) for the total caregiver distress score.

**Summary**

Total NPI scores and NPI distress scores will be summarized by treatment group and visit in terms of absolute values and changes from baseline, both for the observed case and LOCF analyses.

A plot showing the adjusted mean change from baseline in the total NPI score and NPI distress score over time within each treatment group will be provided.

NPI domain scores will also be summarized individually using the same methods as the NPI total score.

The homogeneity of the treatment effect for a number of important subgroups will be investigated as described in Section 4.7.1.5.

Additional NPI domain and subtotal analyses will be performed on subsets of the mITT Population and EP:

- [redacted]
- [redacted]

A listing of subjects' observed and change from Baseline value will be provided.

*Cornell Scale for Depression and Dementia*

The C-SDD was specifically developed to assess signs and symptoms of major depression in demented subjects.

The scale contains nineteen questions, divided into 5 sub-scales, with each question given a score ranging from 0 (absent) to 2 (severe):

- Mood-related signs: anxiety, sadness, a minimal reaction to pleasant events and irritability
- Behavioral disturbance: agitation, psychomotor retardation, multiple physical complaints and loss of interest
- Physical signs: loss of appetite, weight loss and lack of energy
- Cyclic functions: diurnal variation in mood, difficulty falling asleep, multiple awakening during the night and early morning awakening
- Ideational disturbance: suicidal tendencies, low self-esteem, pessimism and mood-congruent delusions

The total score for the C-SDD is the sum of each item in each section labeled "rater's opinion." The maximum score for this scale is 38. If the question is marked 'unable to evaluate' on the eCRF, the response will be replaced with a score of 0. However, the

score will only be calculated if there are at most 20% of the C-SDD items (3 out of 19) with the rating "unable to evaluate." Otherwise, the total score will be missing.

**Summary**

C-SDD scores will be summarized by treatment group and visit in terms of absolute values and changes from baseline, both for the observed case and LOCF analyses. If the range of total scores is less than ten, an additional decimal place will be added to the continuous summaries. Otherwise, 1 decimal place will be used.

A plot showing the adjusted mean change from baseline in C-SDD over time within each treatment group will be provided.

A listing of subjects' observed and change from Baseline value will be provided.

## 4.8 Safety Evaluation

All safety summaries and analyses will be based upon the SP as defined in Section 4.4.

### 4.8.1 Extent of Exposure

Exposure as measured by duration of treatment will be summarized by treatment group. Extent of exposure will be calculated for the Treatment Period (up to Week 26).

Days of exposure will be calculated for each subject: (date of last dose of treatment drug – date of first dose + 1).

The duration of study participation including the follow-up period will also be summarized by treatment group.

The duration of study participation will be calculated as (the latest of (last visit/follow-up date or study termination date) – randomization date + 1. For subjects who die during the study, time of study will be calculated as (date of death – randomization date) + 1.

### 4.8.2 Adverse Events

AEs will be coded using MedDRA (Version 18.1) with reference to SOC and PT. All of the displays will be summarized by SOC and PT unless otherwise specified.

The following AE summaries will be presented by treatment group:

- Number and percentage of TEAE
- Number and percentage of treatment related TEAE

- Number and percentage of TEAE by SOC, PT, and severity
- Number and percentage of treatment related TEAE by SOC, PT, and severity
- Number and percentage of TEAE by SOC, PT, and relatedness
- Number and percentage of Serious TEAE
- Number and percentage of Serious treatment related TEAE
- Number and percentage of TEAE leading to discontinuation
- Number and percentage of TEAE with fatal outcome

All AE summaries will provide the number of subjects reporting at least 1 adverse event and the total number of events reported.

General Rules for Adverse Events

1. An AE will be considered a TEAE if it begins on or after the first study drug dosing or that worsens in severity after at least 1 dose of study drug has been administered. In case of insufficient information to determine if the event occurred before, during or after study drug dosing, the AE will be considered treatment-emergent.

   The imputation method for the handling of missing or partial dates will be as follows:

   Start dates

   - If the day is missing, but month and year are present and the same as the month and year of the first dose of study drug, then day will be set to the same day as the start of study drug. Otherwise this will be imputed as the first day of the month
   - If the month is missing, but year is present and the same as the year of the first dose of study drug, then month will be set to the same month as the start of study drug. Otherwise this will be imputed as January
   - If the year is missing, the event will be considered as treatment emergent and no imputation of the date will occur

Stop dates

- If the day is missing, this will be imputed as the last day of the month
- If the month is missing, this will be imputed as December
- If the year is missing, the adverse event will be considered as ongoing and no imputation of the date will occur

2. Subjects will be classified as having withdrawn from the study due to an AE if the subject had a study drug action taken recorded as "drug permanently withdrawn" on the Adverse Events page of the eCRF or subject had the Subject Status item recorded as 'Dropout' with 'AE' as the primary reason for premature study termination on the Final Subject Disposition page of the eCRF.
3. Events occurring during the pre-treatment phase will not be reported/summarized but will be listed.
4. If a subject experiences the same AE (i.e. same PT) more than once, all occurrences will be counted only once under the count for PT.
5. If a subject experiences more than one AE in a particular SOC, they will only be included once in the count for the SOC, but will appear in the count for each appropriate PT within the SOC (unless it is the same PT).
6. AEs related to study drug tables will include only those AEs with a relationship to study drug of 'probably related', 'possibly related', 'not assessable', or if there is a missing relationship on the AE page of the eCRF.
7. For AE severity, when there is more than one AE of the same PT, the worst severity will be considered in the summary tables by severity

The AEs will be ordered by decreasing frequency, then alphabetically, for total subjects for each SOC and PT within an SOC.

An overall summary of AEs will be provided by treatment group and overall. The summary will include incidences for the following:

- Any TEAE
- Any Treatment Related TEAE
- Any Serious TEAE
- Any Treatment Related Serious TEAE
- Any TEAE leading to Study Discontinuation

- Any TEAE with Fatal Outcome

A listing of all TEAEs will be provided. Non-treatment-emergent AEs will be shown in a separate listing. These listings will be presented by treatment group and will include: site, subject identifier, adverse event (SOC, PT, and verbatim term), date of onset, date of resolution, duration, severity, seriousness, action taken, outcome and causality.

### 4.8.3 Deaths, Serious Adverse Events, and Other Significant Adverse Events

An overview of AEs, including the number and percentage of subjects who died, reported serious adverse events (SAEs), and discontinued due to AEs will be provided. The incidence of treatment emergent SAEs will be presented by SOC, PT, and treatment group.

Listings of all SAEs and deaths as well as AEs leading to discontinuation will be provided.

### 4.8.4 Clinical Laboratory Evaluation

All laboratory values will be reported in International System of Units (SI). Summary displays for laboratory safety data (continuous and categorical parameters) will be produced by visit and treatment group. Summary statistics will be produced for both observed and change from baseline values of the different laboratory variables.

If there are multiple records of laboratory measurements at any post-baseline visit, the worst valid record will be used. It is noted that invalid laboratory data may not be used (from hemolyzed samples, mishandled samples, quantity not sufficient, or other conditions that would render values invalid). Missing data will be maintained as missing for all summaries.

The following laboratory variables will be determined at screening, baseline, Weeks 4 (liver function tests only), 13, 26, and 30 as indicated in the protocol and Table 2 below. It is noted that reducing substance was included in the protocol as part of the urinalysis assessments but was not performed and is not included in the summaries.

**Table 2: Laboratory Assessments**

| | | | |
|---|---|---|---|
| Hematology: | erythrocytes<br>hemoglobin<br>hematocrit<br>MCV<br>MCH<br>neutrophils<br>eosinophils<br>basophils<br>lymphocytes<br>monocytes<br>platelets<br>leukocytes | Endocrinology: | free T4<br>pregnancy test[a] |

| | | | |
|---|---|---|---|
| Clinical chemistry: | sodium<br>potassium<br>chloride<br>calcium<br>phosphorus<br>glucose<br>creatinine<br>urea<br>uric acid<br>total bilirubin [b]<br>direct bilirubin [b]<br>indirect bilirubin [b]<br>alkaline phosphatase [b]<br>AST [b]<br>ALT [b]<br>GGT [b]<br>albumin [b]<br>total protein [b]<br>triglycerides<br>cholesterol<br>calculated creatinine clearance<br>vitamin B12<br>HbA1c | Urinalysis (macroscopy): | appearance<br>bilirubin<br>color<br>glucose<br>ketone<br>leukocytes<br>nitrite<br>occult blood<br>protein<br>reaction pH<br>specific gravity<br><br>Microscopy will be done if macroscopy is abnormal |

*ALT=alanine aminotransferase; AST=aspartate aminotransferase; GGT=gamma-glutamyl transpeptidase; HbA1c=glycated hemoglobin; MCV=mean corpuscular volume; MCH=mean corpuscular hemoglobin; T4=thyroxine.*

*a. At screening and Week 26 for women of childbearing potential, a urine pregnancy test will be performed. Positive result will be confirmed by a serum pregnancy test*

*b. At Week 4, only liver function tests will be performed.*

Laboratory values will be classified as normal, low, or high based on normal ranges supplied by the central laboratory. A summary of shift from baseline classifications will be provided by visit and treatment group for the hematology and clinical chemistry parameters.

The number and percentage of subjects with elevated alanine aminotransferase (ALT), aspartate aminotransferase (AST), or total bilirubin, meeting any 1 of the following criteria will be summarized by treatment group for each visit and overall:

- ALT or AST ≥ 8x upper limit of normal (ULN)
- ALT or AST ≥5x ULN for more than 2 weeks
- ALT or AST ≥ 3x ULN and total bilirubin ≥ 2x ULN at the same visit
- ALT or AST ≥ 3x ULN with the appearance of symptoms indicating hepatitis (e.g., worsening fatigue, nausea, vomiting, right upper quadrant pain or tenderness, fever, rash, or eosinophilia)

Due to an equipment malfunction at the analytic laboratory, serum creatinine results for some subjects may have increased by approximately 0.3-0.4 mg/dL between October 1, 2016 and January 13, 2017. This affected both the serum creatinine results and the calculated creatinine clearance, since calculated creatinine clearance is derived from creatinine. These results will be excluded from the primary lab summaries but will be included in the listing. A second summary for serum creatinine will be presented where the values of all the affected samples are decreased by 0.35 mg/dL. No correction will be performed for calculated creatinine clearance.

### 4.8.5 Vital Signs, Physical Findings and Other Observations Related to Safety

#### 4.8.5.1 Vital Signs

Summaries of blood pressure will be presented by visit and treatment group. Summary statistics will be produced for both observed and change values from baseline for each systolic and diastolic blood pressure.

If there are multiple records of vital sign measurements, the worst record will be used. Any repeated vital signs measurements will be excluded from summaries but included in listings. Missing data will be maintained as missing.

#### 4.8.5.2 Electrocardiogram

Summary displays for ECG parameters (heart rate, PR mean, RR interval, QRS duration, duration, QT interval, QTcB interval, and QTcF interval) will be produced. ECG parameters should be performed 3 hours +/- 1 hour after the drug administration per protocol. Two separate summaries will be produced: one excluding results for ECGs performed outside that window and one including all results regardless of whether the

results were obtained within that window. Summary statistics of observed and change from baseline values will be tabulated by visit and treatment group.

If there are multiple records of ECG measurements at any post-baseline visit, the worst result will be used. Missing data will be maintained as missing.

#### 4.8.5.3 Physical and Neurological Examinations

Results of physical examination at each visit will be summarized by body system, treatment group, and visit. Physical examination results will also be listed.

Treatment-emergent abnormal findings are the values which are normal at baseline and fall into the abnormal categories post-baseline. This will also include values which are abnormal post-baseline, but have no baseline value. The number and percentage of subjects experiencing a treatment-emergent abnormal physical examination finding, by body system and treatment group will be summarized.

Results of a general neurological exam will also be summarized for each parameter in each module by treatment group and visit. The general neurological exam can be found in Appendix 7 of the protocol. Neurological examination results will also be listed.

#### 4.8.5.4 Columbia Suicide Severity Rating Scale

The results of the C-SSRS will be summarized using number of subjects and percentages with (i) events in suicidal behavior, (ii) suicidal ideation, and (iii) suicidal behavior and ideation.

Suicidal ideation is defined as an event in any of the following 4 categories:

- Wish to be dead (Q1)
- Non-specific active suicidal thoughts (Q2)
- Active suicidal ideation with any methods (not plan) without intent to act (Q3)
- Active suicidal ideation with some intent to act, without specific plan (Q4)
- Active suicidal ideation with specific plan and intent (Q5)

Suicidal behavior is defined as an event in any of the following categories:

- Actual attempt
- Interrupted attempt
- Aborted attempt
- Preparatory acts or behavior

- Suicidal Behavior
- Complete Suicide

Complete suicide is not supplied for the lifetime questionnaire but is supplied only for the previous 6 months.

The incidence of subjects with suicidal behavior and self-injurious behavior will be summarized by treatment group and visit. A by-subject listing of the C-SSRS questionnaire data will be provided.

### 4.8.6 Safety Monitoring (Data and Safety Monitoring Board [DSMB])

A DSMB will be assembled to periodically review and evaluate accumulated study data for subject safety, study conduct and progress. The DSMB will have the responsibility to make recommendations concerning the continuation, modification, or termination of the study based on their independent evaluation of all study related safety data. The DSMB will consist of a chairperson and at least 2 additional members, including 1 statistician. The approved DSMB charter enumerates the roles of the DSMB members, the frequency with which it meets, and the structure of their meetings. A separate SAP for analyses associated with the DSMB lists out the specific analyses that the DSMB will review.

## 4.9 Other Analyses

### 4.9.1 Pharmacokinetics Analyses

Concentration data (trough levels) will be summarized for SUVN-502 and M1 of SUVN-502 for each visit by treatment group.

Concentration data for donepezil and memantine (trough levels) will be summarized for each visit by treatment group.

Summaries will include the geometric mean, arithmetic mean, minimum, maximum, SD, and % coefficient of variation (CV). The summaries will be based on the SP.

Concentration data for SUVN-502 and M1 of SUVN-502 as well as donepezil and memantine will be excluded from the analyses if the sample of blood was drawn after the subject received the daily dose of the particular medication.

## 4.10 Determination of Sample Size

A total of 537 subjects will be randomized into one of three treatment groups, SUVN-502 (50 mg), SUVN-502 (100 mg) or placebo (179 subjects per group). With a sample size

of 537, there is at least 80% power to detect a 2-point drug-placebo difference on the ADAScog-11 with a standard deviation of 6, assuming a 2-sided 5% significance level and a drop-out rate of 20% or less.

### 4.11 Changes in the Conduct of the Study or Planned Analysis

Changes in the conduct of the study may be instituted through a protocol amendment. Planned analyses will be revised as appropriate. Similarly, planned analyses may be changed as a result of planned blinded data reviews. Changes will be finalized prior to database lock.

## 5 REFERENCES

Borza, T, Engedal, K, Bergh, S, Barca, M, Benth, J, Selbaek, G. The course of depressive symptoms as measured by the Cornell scale for depression in dementia over 74 months in 1158 nursing home residents. Journal of Affective Disorders 175 (2015) 209-216.

Guidelines for Industry: Statistical Principles for Clinical Trials (E9), International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, February 1998.

Leontjevas,R.,Gerritsen,D.L.,Vernooij-Dassen,M.J.,Smalbrugge,M.,Koopmans,R.T., 2012.Comparativevalidationofproxy-basedMontgomery-Asbergdepression ratingscaleandcornellscalefordepressionindementiainnursinghome residentswithdementia.Am.J.Geriatr.Psychiatry20,985–993.

[redacted]