FINAL DRAFT STATISTICAL ANALYSIS PLAN

Protocol
T817MAUS202 – Main Study

A Phase 2 multi-center, randomized, double blind, placebo-controlled, parallel group study to evaluate the efficacy and safety of T-817MA in patients with mild to moderate Alzheimer's Disease (US202)

Toyama Chemical Co., Ltd.
2-5, Nishishinjuku 3-chome,
Shinjuku-ku, Tokyo, 160-0023 Japan

Date/Version:09FEB2017 FINAL V1.0

A Phase 2 multi-center, randomized, double blind, placebo-controlled, parallel group study to evaluate the efficacy and safety of T-817MA in patients with mild to moderate Alzheimer's Disease (US202)

By signing below, all parties accept that the analysis methods and data presentations are acceptable and that this document is final.

Prepared By:

Noel Ellison, MS
Sr. Biostatistician, Consultant

Date

Reviewed By:

Suzanne Hendrix, PhD

Date

Reviewed By:

Ron Thomas, PhD

Date

Reviewed/Approved By:

Howard Feldman, MDCM, FRCP(C)

Date

Reviewed/Approved By:

Lon Schneider, MD

Date

Reviewed/Approved By:

Kenji Taguchi

Date

Table of Contents

List of Abbreviations

AChEI	Acetylcholinesterase inhibitor
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 Impressions of Change
AE	Adverse Event
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
ApoE4	Apolipoprotein E4 allele
ARH(1)	First-order heterogeneous autoregressive
AST	Aspartate aminotransferase
BUN	Blood Urea Nitrogen
CFB	Change from Baseline
CPK	Creatine Phosphokinase
CSF	Cerebro-spinal fluid
CT	Computed tomography
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders
eCRF	Electronic Case Report Form
ECG	Electrocardiogram
EDC	Electronic Data Capture system
FAQ	Functional Activities Questionnaire
γ-GTP	γ-Glutamyltransferase
HDL	High-density Lipoprotein
ITT	Intent-to-treat
LDL	Low-density Lipoprotein
LDH	Lactate dehydrogenase
LS	Least-squares
mITT	Modified Intent-to-treat
MMRM	Mixed model with repeated measures
MMSE	Mini-mental state examination
MAR	Missing at Random
MNAR	Missing Not at Random
MRI	Magnetic resonance imaging

N	Number of patients
NINCDS/ ADRDA	National Institute of Neurological and Communicative Diseases and Stroke/Alzheimer's Disease and Related Disorders Association
NPI	Neuropsychiatric Inventory
OC	Observed Cases
PMM	Pattern Mixture Model
PK	Pharmacokinetic
PP	Per Protocol
PPK	Population pharmacokinetics
RBC	Red blood cell
SAE	Serious Adverse Event
SD	Standard Deviation
TC	Total cholesterol
TG	Triglyceride
TSH	Thyroid stimulating hormone
T-817MA	Used interchangeably as Study drug and Active drug product; a maleate salt of T-817 [chemical name: 1-{3-[2-(1-benzothiophen-5-yl)ethoxy]propyl}azetidin-3-ol maleate]
T-817MAa	Notation of Study drug in previous trials
UN	Unstructured
vMRI	Volumetric magnetic resonance imaging
WHO	World Health Organization

DEFINITIONS

Safety Population	All subjects who receive at least one dose of study medication.
Modified Intent-to-Treat Population	All randomized subjects who receive any study treatment and have at least one post-baseline efficacy assessment.
Per Protocol Population	All subjects included in the mITT population who took the assigned medication for 52 weeks and did not have any major protocol violations.
Treatment	There are three treatment groups for this study: - 448 mg T-817MA - 224 mg T-817MA - Placebo

1 INTRODUCTION

The objective of the Statistical Analysis Plan is to ensure the credibility of all study findings by means of a predefined data analysis plan. This plan assumes familiarity with the study protocol and will provide further details of the summaries and analyses planned therein. This Statistical Analysis Plan was finalized via signatory prior to the treatment unblinding.

1.1 Alzheimer's Disease

Alzheimer disease's (AD) is a neurodegenerative disease characterized by progressive impairment in memory and cognitive function and is a major concern as a cause of dementia in our aging society.[1] More than 4 million people are afflicted with this disease in the United-states and in excess of 250,000 cases are diagnosed every year. Affecting twice as many females as males, AD is thought to be mainly due to genetic factors, however, several environmental factors are known to contribute to its evolution.[1]

Currently, two classes of drugs, acetylcholinesterase inhibitors (AChEI) [(donepezil (Aricept®), rivastigmine (Exelon®) and galantamine (Razadyne®)] and N-methyl-D- aspartate (NMDA) receptor antagonists [(memantine (Namenda™)], are available as symptomatic treatment for AD. Other drugs that target amyloid â-protein (Aâ) and neurofibrillary tangle, major hallmarks of AD, are actively being investigated. However, no drug has yet shown nor been approved for the delay and eventually prevention of AD progression.

Meanwhile, neurotrophic factors are known to be essential in maintenance and functional restoration in neural cells. T-817MA is a low molecular weight compound, which has neurotrophic effects to inhibit neural cell death and to promote neurite outgrowth, being developed by Toyama Chemical Co., Ltd. It could therefore be a valuable therapeutic option in AD and should be more fully developed to investigate its safety and efficacy.

1.2 T-817MA

Nonclinical pharmacology studies using in vitro models have shown that T-817MA has neuroprotective effects. T-817MA significantly inhibited the reduction in the number of neurons caused by Aß at doses of 0.1 and 1.0 µM, indicating a dose- dependent neuroprotective effect in an in vitro model for AD. Results in an in vitro model suggested that T-817MA has a beneficial effect of regeneration of neurons that have been damaged by AD.

Nonclinical data suggested that T-817MA not only improves symptoms but also prevents disease progression in AD. The pharmacological, toxicological and pharmacokinetic (PK) profiles of T-817MA, in the nonclinical studies, suggest that T- 817MA would be safe for oral administration in human.

To date, seven Phase 1 clinical trials and one Phase 2a clinical trial have been completed, and a total of 382 patients have received at least one dose of T-817MA. T-817MA has been safe and well tolerated at a single dose up to 896 mg, and at once daily doses of 224 mg up to six weeks, and 672 mg up to two weeks in healthy elderly subjects.

1.2.1 Study US201

Study US201 (Protocol AA437420) was a Phase 2a, multi-center, randomized, double-blind, placebo- controlled 52-week clinical trial in patients with mild to moderate Alzheimer's disease.

Patients were randomly assigned to receive either placebo or 224 mg T-817MA. Donepezil was maintained throughout the study period. A total of 373 patients were randomly assigned to study drug, 183 patients in the placebo group and 190 patients in the T-817MA group; of these, 249 patients (66.8%) completed the study, including 132 patients (72.1%) in the placebo group and 117 patients (61.6%) in the T-817MA group.

The primary objective was to evaluate the efficacy of T-817MA as measured by the Alzheimer's disease assessment scale-cognitive subscale (ADAS-cog) assessment. ADAS-cog score mean change from baseline values indicated a smaller degree of worsening in the T-817MA group than in the placebo group, and such an efficacy trend was consistent throughout the treatment period although the difference did not reach statistical significance. Secondary efficacy analyses for Alzheimer's Disease Cooperative Study- Clinical Global Impressions of Change (ADCS-CGIC) and Alzheimer's Disease Cooperative Study - Activities of Daily Living Inventory (ADCS-ADL) showed a similar pattern of efficacy.

A post-hoc analysis looking at patients entering the study with a baseline mini-mental state examination (MMSE) ≤ 20 was also performed. Results of the post-hoc analysis showed the ADAS-cog score mean change from Baseline to Week 52 for the per- protocol (PP) population, whose baseline MMSE was ≤ 20, was 5.46 (S.D. 8.98) for the placebo group and 2.03 (S.D. 7.36) for the T-817MA group, indicating that cognitive function showed better efficacy trend in the T-817MA group. Such a trend is consistent with the modified intent-to-treat (mITT) population at Week 52; with

ADAS-cog score mean change from Baseline 5.28 (S.D. 8.66) for the placebo group and 3.20 (S.D. 7.39) for the T-817MA group.

The most frequently occurring AEs in the T-817MA group (occurring in ≥ 5% of patients in that group) were diarrhea, urinary tract infection, nausea, dizziness, and headache. Of these AEs, those that occurred in a notably higher percentage of patients in the T-817MA group than in the placebo group included diarrhea (18.9% compared with 7.1%) and nausea (5.3% compared with 0.5%); those that occurred in only a slightly higher percentage of patients in the T-817MA group than in the placebo group included dizziness (5.3% compared with 3.3%) and headache (5.3% compared with 4.4%). Urinary tract infection occurred in a similar percentage of patients in each treatment group (10.0% compared with 10.4%).

Overall, T-817MA was well tolerated with diarrhea and nausea being the main AEs occurring frequently and more often in the T-817MA group than in the placebo group.

1.2.2 Study US108

Study US108 (Protocol AA99688) was a single and multiple 2-week dose study of 448, 672 and 896 mg in healthy elderly subjects. The study was designed to evaluate the safety, tolerability, and pharmacokinetics (PK) of T-817 following single and multiple 2-week oral administration of 3 dose levels (448, 672 and 896 mg) of T-817MA in healthy elderly subjects at 50 years of age or older. 24 subjects were enrolled and treated. There were no SAEs and no subject was discontinued due to an AE. No significant adverse events, such as vomiting or convulsions, were observed following multiple-doses of 448 and 672 mg, as well as single dose of 896 mg. Observed exposure by multiple 2- week doses of 672 mg was approximately a two-fold increase of observed exposure of 448 mg, which is planned to be used as the high dose level for this

Phase 2 study (US202). Plasma concentrations by multiple 2-week doses have been predicted to reach 1.3-fold of that achieved by single dosing.

The safety and tolerability of the intended doses (448 and 672 mg) have been demonstrated and therefore, there was no valid scientific reason to expose subjects to multiple-2-week doses of 896 mg, which could result in plasma concentrations close to those resulting in convulsions in monkeys. Toyama decided not to proceed to Cohort 4 (multiple 2-week doses of 896 mg) and terminated the US108 study.

T-817MA appeared to be well tolerated at 3 dose levels (multiple 2-week doses of 448 and 672 mg, in addition to single dose of 896 mg).

2 OBJECTIVES

2.1 Primary Objectives

The primary objective is to evaluate the efficacy of T-817MA as measured by ADAS-cog and ADCS-CGIC.

2.2 Secondary Objectives

The secondary objectives are:

- To evaluate the safety and tolerability of T-817MA as measured by clinical safety laboratories, physical examinations, ECGs and solicitation of adverse events;
- To evaluate the efficacy of T-817MA as measured by ADCS-ADL, FAQ, Neuropsychiatric Inventory (NPI) and Mini-mental State Examination (MMSE).

2.3 Additional Objectives

Population pharmacokinetics of T-817 will be analyzed (see PK SAP). Volumetric MRI (vMRI) scans will be evaluated. Biomarkers found in the cerebro-spinal fluid (CSF) will be evaluated in a subset of the study population.

3 STUDY DESIGN

3.1 Number of Subjects

In order to have 110 evaluable patients in each group, it is necessary to randomize at least 450 patients (at least 150 per group) to double-blind treatment, assuming a 26% dropout rate during the treatment period.

3.2 Sample Size Considerations

The change in the ADAS-cog score in the placebo group is expected to be 3.5 points in one year. Assuming that the difference in the ADAS-cog scale between each T-817MA treatment group and the placebo group is 2.5 points after one year, with a Standard Deviation of 6.5, the required number of cases to show a significant difference in the ADAS-cog is estimated at 110 per group. This calculation has 80% statistical power at the 5% significance level.

The power calculation of 80% as shown above corresponds to achieving significance on the ADAS-cog alone, which would be considered a successful outcome for a proof of concept study. Achieving significance on both co-primary endpoints is not fully powered, but would be considered a successful outcome for a pivotal study

3.3 Study Design

This is a Phase 2, multi-center, randomized, double-blind, placebo-controlled, parallel group study with a 52 week treatment period in patients with mild to moderate Alzheimer's disease who have been taking donepezil for at least 4 months and are on a stable dose for at least 3 months prior to Baseline.

Patients will be screened at the investigative site in order to determine eligibility for the study. Patients must be between the ages of 55 and 85 inclusive at screening to be eligible for enrollment. The patient must have a diagnosis of probable AD in accordance with the DSM-IV and NINCDS/ADRDA criteria. A brain CT or MRI consistent with the clinical diagnosis of probable AD is also required for eligibility in the study. A mini-mental state examination score of 12-22 inclusive at Screening, a modified Hachinski score of 4 or less at screening and the availability of a study partner are also among other inclusion criteria. Women must be post-menopausal or surgically sterile to be eligible for this study.

Donepezil or rivastigmine transdermal system, the concomitant treatment for Alzheimer's Disease, will be required to be taken as prescribed by the site principal investigator or other prescribing physician, and will be encouraged to be maintained throughout the study. Memantine will be allowed only when prescribed in combination with donepezil or rivastigmine transdermal system and patients taking it should be on it for at least 4 months and on a stable dose for at least 3 months prior to Baseline. Other medications for AD including galantamine, tacrine, huperizine A, oral rivastigmine, donepezil 23 mg tablets and donepezil 10 mg tablet BID are not permitted during the study.

A central review of screening results will then be conducted and, if appropriate, permission will be granted to randomize the patient.

At Baseline, patients who have been approved by the central review will be randomized to one of three treatment groups: placebo, 224 mg T-817MA or 448 mg T-817MA in a 1:1:1 allocation. A stratified permuted block randomization procedure will be used. The study medication will be taken once daily after breakfast. NOTE: On the first day of dosing (Day 1) the patient will receive their dose in the clinic after baseline cognitive testing (it is not necessary to give this first dose near a meal time). Beginning on Day 2, the patient must take the dose after breakfast.

Patients will be seen every four weeks for the first 12 weeks of the study, and then will be seen every six weeks thereafter until Week 36, then at Week 44 and Week 52. A final follow-up visit will be conducted at Week 56. Safety evaluations will be carried out at each visit. Efficacy will be assessed every 12 weeks beginning at Baseline through Week 36 then at Weeks 44 and 52 and finally during the follow up visit at Week 56. Volumetric MRI scans of the brain will be done at Screening and Week 52.

A subset of patients, who have consented to do so, will also have a lumbar puncture to obtain CSF at Weeks 0 and 52 to evaluate CSF biomarkers.

Blood will be taken from all the patients at Week 0, 12, 24, 36, 44 and 52 for population pharmacokinetic (PPK) assessments.

A final follow-up visit will be conducted at Week 56. Each patient may participate in the study for up to 62 weeks.

The table of the study visits and procedures to be performed at each visit can be found below in Table 1, the Clinical Trial Flow Chart.

Table 1: Clinical Trial Flow Chart

Visit Number	1	2	3	4	5	6	7	8	9	10	11 or Early Term	12
Stage of study	Screening (within 42 days prior to Randomization)	Baseline (Week 0)	Week 4 (±7 days)	Week 8 (±7 days)	Week 12 (±7 days)	Week 18 (±7 days)	Week 24 (±7 days)	Week 30 (±7 days)	Week 36 (±7 days)	Week 44 (±7 days)	Week 52 (±7 days)	Follow-Up Week 56 (±7 days)
Informed consent	X											
Demographics, medical history, education	X											
Inclusion / Exclusion criteria	X	X										
Height[1] & Weight	X	X			X		X		X		X	
Randomization		X										
Physical examination	X				X		X		X		X	X
Vital signs (Heart Rate and BP)	X	X	X	X	X	X	X	X	X	X	X	X
Concomitant medications	X	X	X	X	X	X	X	X	X	X	X	X
Adverse Events	X	X	X	X	X	X	X	X	X	X	X	X
12-lead ECG	X	X	X	X	X	X	X	X	X	X	X	X
Hematology, Serum chemistry[2]	X	X	X	X	X	X	X	X	X	X	X	X
Urinalysis	X	X	X	X	X	X	X	X	X	X	X	X
PPK blood draw[3]		X			X		X		X	X	X	
ApoE4 and CYP2D6		X										
Study drug dispense		X	X	X	X	X	X	X	X	X		
Study drug collection/compliance			X	X	X	X	X	X	X	X	X	
ADAS-cog	X	X			X		X		X	X	X	X
ADCS-CGIC		X			X		X		X	X	X	
ADCS-ADL		X			X		X		X		X	
MMSE	X	X			X		X		X	X	X	X
NPI		X			X		X		X		X	
FAQ		X			X		X		X		X	
C-SSRS Pediatric[4]		X			X		X		X		X	
vMRI or CT [5,6]	X										X	
CSF[7]		X									X	

[1] Height is done at baseline only
[2] B12 is done at Screening only
[3] PPK blood draws should be done before the first dose of study medication at Baseline, before dosing on the day of the Week 24 visit and 1-6 hours after patient's last dose of study medication at the other visits
[4] C-SSRS Ped Lifetime/Recent at Week 0 and C-SSRS Ped Since Last Visit at all others
5 CT at Screening if MRI is contraindicated and the patient has not had a brain CT within past year
6 Visit window for vMRI at Week 52 is 14 days before and up to14 days after visit. If patient is terminating early at 36 week or after, obtain vMRI. Neither vMRI nor CT is necessary for patients who have a contraindication to MRI.
7 Only for the CSF substudy. Visit windows for CSF are: up to 14 days prior to first dose of study medication and up to 14 days prior to Week 52 visit. If a patient in the CSF substudy is terminating early, they do not need to undergo lumbar puncture.
[8] Only applies to patients not going to the extension period.

4 RANDOMIZATION AND UNBLINDING PROCEDURES

4.1 Randomization

Each patient will be randomly allocated in a 1:1:1 ratio into one of the 3 groups: treatment with 448 mg or 224 mg of T-817 or placebo. A randomization schedule will be generated and incorporated into the Electronic Data Capture system (EDC) and the treatment group will be assigned as the site activates the patient. A centralized eligibility evaluation procedure will be applied for each patient. The Investigators will use the ADCS EDC system to enter screening information on each patient.

For those patients who qualified, the system will issue a medication kit number. Patients were randomized at baseline, after screening was completed and it was determined that the patient was eligible for the study. A stratified permuted block randomization procedure will be used. Site and baseline MMSE will be stratification factors.

4.2 Unblinding

Only in the case of an emergency, when knowledge of whether the patient has received the investigational product is essential for the clinical management or welfare of the patient, may the Investigator request to unblind a patient's treatment assignment. Unblinding at the study site for any other reason will be considered a protocol deviation. If the Investigator needs the blind to be unmasked for a patient for any reason, the Investigator must contact the ADCS Medical Monitor to obtain an approval. Breaking the blind must be reported, documenting the date, the site personnel exposed to the treatment assignment, and the reason the blind was broken.

Upon completion of the study and after the database is locked according to the Sponsor (or designee) operating procedures, the final SAP will be signed-off and the randomization codes provided to the study statistician at Pentara, Suzanne Hendrix, to unblind the study.

5 EFFICACY/SAFETY ASSESSMENTS

5.1 Primary Efficacy Endpoints

The co-primary efficacy endpoints are:

1. Change in the ADAS-cog from baseline to Week 52;
2. Change in ADCS-CGIC from baseline to Week 52.

These endpoints will be compared between each T-817MA treatment group (high dose or low dose) and the placebo group.

5.1.1 ADAS-cog

The ADAS-cog *(Rosen, Mohs, & Davis, 1984)* is a structured scale that evaluates memory (word recall, word recognition), reasoning (following commands), language (naming, comprehension), orientation, ideational praxis (placing letter in envelope) and constructional praxis (copying geometric designs). Ratings of spoken language, language comprehension, word finding difficulty, and ability to remember test instructions are also obtained. The test is scored in terms of errors, with higher scores reflecting poorer performance and greater impairment. Scores can range from 0 (best) to 70 (worse).

5.1.2 ADCS-CGIC

The ADCS-CGIC is a validated categorical measure of change in the patient's clinical condition between baseline and follow-up visits. It measures whether the effects of active treatment are substantial enough to be detected by a skilled and experienced clinician on the basis of a clinical interview and examination. It relies on both direct examination of the patient and an interview of the study partner. A skilled and experienced clinician who is blinded to treatment assignment rates the patient on a 7-point Likert scale, ranging from 1 (marked improvement) to 7 (marked worsening). It is suggested that the instrument has distinct clinical utility in assessing change in AD clinical trials *(Schneider et al. 1997)*.

5.2 Secondary Efficacy Endpoints

The secondary efficacy endpoints are:

1. The change in the ADAS-cog from Baseline to Weeks 12, 24, 36 and 44.
2. The change in the ADCS-CGIC from Baseline to Weeks 12, 24, 36 and 44.
3. The change in the MMSE from Baseline to Weeks 12, 24, 36, 44 and 52.
4. The change in the ADCS-ADL, FAQ and NPI from Baseline to Weeks 12, 24, 36 and 52.

5.2.1 ADCS-ADL

The ADCS-ADL scale is a validated tool for assessing instrumental and basic activities of daily living based on a 23-item structured interview of the study partner. The scale has a range of 0 to 78, with lower scores indicating greater impairment. (*Galasko et al. 1997)*.

5.2.2 MMSE

The MMSE (*Folstein, Folstein and McHugh, 1975)* is a validated, brief, frequently used screening instrument for Alzheimer's disease drug studies. The MMSE scale evaluates

orientation, memory, attention, concentration, naming, repetition, comprehension, and ability to create a sentence and to copy two overlapping pentagons. The MMSE is scored as the number of correctly completed items with a lower score indicative of poorer performance and greater cognitive impairment. The total score ranges from 0 (worse) to 30 (perfect performance).

5.2.3 FAQ

The FAQ, a 10-item questionnaire, is a validated interview which the study partner completes and which rates the study patient on their ability to carry out ten complex activities of daily living *(Pfeffer, Kurosaki, Harrah, Chance, & Filos, 1982)*: 1) manage finances, 2) complete forms, 3) shop, 4) perform games of skill or hobbies, 5) prepare hot beverages, 6) prepare a balanced meal, 7) follow current events, 8) attend to television programs, books or magazines, 9) remember appointments, and 10) travel out of the neighborhood. Each item is rated on a scale from 0 to 3. Scores are summed across items to provide a total disability score (higher scores = greater impairment; maximum score = 30).

5.2.4 NPI

The NPI is a validated, reliable, multi-item instrument to assess psychopathology in AD based on interview with the study partner *(Cummings 1997)*. The NPI evaluates both the frequency and severity of 12 neuropsychiatric features, including delusions, hallucinations, agitation/aggression, dysphoria, anxiety, euphoria, apathy, disinhibition, irritability/lability, and aberrant motor behavior, as well as evaluates sleep and appetite/eating disorders. Frequency assessments range from 1 (occasionally, less than once per week) to 4 (very frequently, once or more per day or continuously) as well as severity (1 = mild, 2 = moderate, 3 = severe). The overall score and the score for each subscale are the product of severity and frequency.

5.3 Additional Outcomes (Biomarkers)

Patients will undergo volumetric MRI (vMRI) scans of the brain at Screening and Week 52. And only a subset of the patients will undergo CSF collection to measure biomarkers at Baseline and Week 52.

5.3.1 Volumetric MRI

Whole brain volume, ventricular volume and hippocampal volume will be measured. MRI imaging of the brain will be performed in order to measure brain atrophy over time. Results from vMRI studies suggest that the patters of atrophy in AD can reliable be detected and tracked across time. Atrophy of the medial temporal lobe, including hippocampus and entorhinal cortex, has long been described in vMRI studies of AD. Hippocampal volume derived from MRI correlates with histological hippocampal volume and degree of neuronal loss and AD pathology, and entorhinal cortical thickness change appears to be an early and sensitive indicator of neurodegeneration associated with AD. Longitudinal MRI measures regional and whole brain volumetric change provide a valuable complement to cognitive measures in that they are not influenced by temporary symptomatic improvements, and they provide an early index of the study drug's ability to reach the target organ and have an effect on AD-related atrophy.

5.3.2 CSF Biomarkers

CSF will be taken at Baseline and Week 52 to measure CSF biomarkers such as Aβ 1-40. Aβ 1-42, tau and p-tau.

5.4 Safety Assessments

Safety assessments include the following:

- Adverse events (AEs);
- Vital signs;
- 12-lead ECG;
- Hematology, chemistry and urinalysis;
- Physical examinations
- Use of concomitant medications
- Pediatric Columbia-Suicide Severity Rating Scale (C-SSRS Ped)

5.4.1 Adverse Events

An adverse event (AE) or adverse experience is any untoward medical occurrence in a patient or clinical investigation patient who is administered a medicinal product that does not necessarily have a causal relationship with this treatment that occur after informed consent is signed and up to 28 days after the study drug has been discontinued. An AE can therefore be any unfavorable and unintended sign, symptom, or disease temporally associated with the use of a medicinal product, whether or not considered related to the medicinal product. Pre-existing conditions, which increase in frequency or severity or worsen in nature during or as a consequence of use of a drug in human clinical trials, will also be considered as adverse experiences. Adverse events may also include pre- or post-treatment complications that occur as a result of protocol-mandated procedures (e.g. invasive procedures such as biopsies).

All AEs must be promptly documented on the Adverse Event eCRF and assessed by the Investigator. Details of the event must include the dates of onset and resolution, severity, relationship to study drug, seriousness, and whether the event caused the patient to withdraw from the study, outcome and timing with regard to administration of the study drug. Severity will be recorded as mild, moderate or severe. Relationship to study drug will also be recorded in one of the following five categories: definite, probable, possible, unlikely and unrelated. The criteria for these classifications (found in the protocol), in addition to good clinical judgment, should be used as a guide for determining the causal assessment. If it is felt that the event is not related to study drug therapy, then an alternative explanation should be provided.

All events (serious and non-serious) that occur after informed consent is signed and up to 28 days after the study drug has been discontinued will be reported on the eCRF. All SAEs occurring more than 28 days after last study drug administration at Week 52 and considered at least possibly drug-related must also be reported.

An SAE is an AE from this study that results in any of the following outcomes:

- Death
- Life-threatening situation (patient is at immediate risk of death)
- Inpatient hospitalization or prolongation of existing hospitalization

- Persistent or significant disability/incapacity
- Congenital anomaly/birth defect in the offspring of a patient who received study drug
- Considered significant by the investigator for any other reason

Important medical events that may not result in death, be immediately life-threatening, or require hospitalization, may be considered an SAE when, based upon appropriate medical judgment, they may jeopardize the patient or may require medical or surgical intervention to prevent one of the outcomes listed in this definition.

All SAEs must be recorded on the AE eCRF as "serious" and submitted regularly during the course of the study.

A life-threatening AE is defined as any adverse experience that places the patient in the view of the Investigator, at immediate risk of death from the event as it occurred. This does not include an event that might have led to death, if it had occurred with greater severity.

Hospitalizations for study therapy, disease-related procedures, palliative or hospice care, elective procedures or placement of an indwelling catheter are not considered to be SAEs.

"Inpatient hospitalization" means the patient has been formally admitted to a hospital for medical reasons, for any length of time. Presentation and care within an emergency department does not necessarily constitute an SAE. Complications that occur during hospitalization are AEs. If a complication prolongs hospitalization, it is an SAE.

All deaths, regardless of cause, must be reported for patients on study (within 28 days of last study drug dose). The SAE term should reflect the event that leads to the death with "death" recorded as the outcome.

5.4.2 Vital Signs

The following vital signs will be measured at each visit: systolic blood pressure, diastolic blood pressure, and heart rate. Height will be measured at baseline only. Weight will be measured at screening, baseline, Week 12, Week 24, Week 36 and Week 52.

5.4.3 ECG

A resting 12-lead ECG will be conducted at each visit. If the inter-beat (RR) interval isn't in the EDC, it will be calculated in milliseconds using the following formula:

$$\frac{60{,}000}{Heart\ Rate\ in\ Beats\ Per\ Minute}.$$

5.4.4 Laboratory Assessment

The following safety laboratory parameters are to be measured in blood and urine samples:

1. Hematology: RBC, hemoglobin, hematocrit, total leukocyte count, differential leukocyte count (neutrophil, lymphocyte, monocyte, eosinophil, basophil), platelet count;
2. Chemistry: AST, ALT, γ-GTP, ALP, total bilirubin, LDH, BUN, creatinine, serum electrolytes (Na, K, Cl, Ca, P, CO_2), total protein, albumin, A/G ratio, total

cholesterol (TC), HDL cholesterol, LDL cholesterol, triglyceride (TG), creatine phosphokinase (CPK), uric acid, non-fasting glucose, TSH and vitamin B12 (Screening only);

3. Urinalysis: pH, specific gravity, protein, glucose, ketones, urobilinogen, bilirubin, blood.

An approximate blood volume of 9.5 ml at screening and 7 mL at all other visits will be collected for serum chemistry and hematology analyses. (If vitamin B12 is below normal and the Investigator thinks it necessary, an additional blood (5 ml) may be taken to measure serum methylmalonic acid and homocysteine levels during screening period.) Urine will be collected at each visit for urinalysis. The Central Laboratory will provide the investigational sites with all appropriate materials for specimen collection, sample processing, packaging and shipping. Full details of sampling (blood and urine), sample preparation and storage methods will be given in the laboratory manual.

5.4.5 Physical Examination

A complete physical examination is performed at Screening, Week 12, Week 24, Week 36, Week 52, and Follow-Up.

5.4.6 Concomitant Medications

Medications will be classified according to drug class and preferred term using the WHO dictionary. Medications with a start and end date occurring before the baseline visit will be identified as prior medications. Medications with an end date occurring on or after the baseline visit or that have unknown or on-going end dates will be identified as concomitant medications.

5.4.7 C-SSRS Pediatric

The C-SSRS is a scale that captures the occurrence, severity, and frequency of suicide-related thoughts and behaviors during the corresponding assessment period. The scale includes suggested questions to elicit the type of information needed to determine if a suicide-related thought or behavior occurred.

There is no version of the C-SSRS designed for use in a cognitively impaired population; therefore, the pediatric version will be used. Terms captured by the use of the C-SSRS pediatric version can be mapped to Columbia Classification Algorithm for Suicide Assessment (Posner et al. 2007).

The first time the scale is administered in this study, the C-SSRS pediatric Lifetime/Recent scale will be used, and the findings will constitute the baseline assessment. The C-SSRS Pediatric Since Last Visit scale will be used for all subsequent assessments.

The C-SSRS will be administered to the patient with the study partner present, after the cognitive and functional assessments. Responses from both the partner and patient will be considered when administering the scale. If a suicide-related thought or behavior is identified at any time during the study, a study physician will perform a thorough evaluation, and appropriate medical

care will be provided.

5.5 Other Evaluations

Additional evaluations include the following:

1. Demographics;
2. Baseline disease characteristics;
3. Medical history;
4. Prior medications.

6 ANALYSIS POPULATIONS AND GENERAL STATISTICAL PROCEDURES

6.1 Definition of Analysis Populations

Statistical analysis and data tabulation will be performed using the following subject populations unless specified otherwise:

1. Safety Population;
2. Modified Intent-to-Treat (mITT) Population;
3. Per Protocol (PP) Population.

The safety population will include all subjects who received at least one dose of study medication. Subjects in this population are analyzed based on the actual treatment they received. If a subject received different treatments throughout the course of the study, he or she will be analyzed based on the highest dose received.

The mITT population will include all randomized patients who took at least one dose of the study medication and who have at least one efficacy evaluation following baseline.

The PP population will include all mITT patients who took at least 70% of all doses of the assigned medication during the 52 weeks of treatment prior to the last co-primary efficacy assessment and did not have any significant deviations as determined separately by clinician review prior to unblinding the study.

Classification of deviations from the protocol as minor or major, and decisions regarding exclusion of patients and/or patient data from the statistical analyses, will be decided on a case-by-case basis without knowledge of the treatment assigned and before unblinding the study.

6.2 Application of Analysis Populations

The primary population for efficacy analysis is the mITT population. Analysis of primary and secondary efficacy endpoints will be performed in the mITT and PP populations. The safety population will be used for analyses of safety endpoints.

Subject enrollment, disposition, drug exposure, demographics and baseline disease characteristics will be shown for all populations.

Efficacy analyses will be performed on both the mITT and PP populations, unless otherwise stated.

All safety evaluations, medical history and medication use will be based on the safety population.

6.3 General Statistical Procedures

All analyses described in this plan are considered *a priori* analyses in that they have been defined prior to breaking the study blind. All other analyses, if any, defined subsequent to breaking the study blind will be considered *post hoc* analyses and will be applied using exploratory methodology. All *post hoc* analyses will be identified as such in the Clinical Study Report.

Descriptive statistics for continuous variables will include number of subjects (n), arithmetic mean, standard deviation (SD), median, minimum, maximum and first and third quartile limits unless otherwise noted. Frequency and percentage will be calculated for categorical variables. Unless stated otherwise, all summary tables will present descriptive statistics and/or frequencies either by treatment or overall, and all data listings will be sorted by subject number.

Unless otherwise specified, all significance testing will be 2-tailed using $\alpha = 0.05$. Tests will be declared statistically significant if the calculated p-value is ≤ 0.05. The primary study hypothesis will be deemed satisfied if the ADAS-cog scores show improvements in the 448 mg active treatment group relative to the placebo that are significant at the 0.05 level.

Change from baseline is calculated by subtracting the baseline score from the observed value at any subsequent visit. For safety summaries, the last pre-randomization measurement is defined as the baseline value. For efficacy measures baseline is defined as the last pre-randomization measurement. Early termination assessments and visits that occurred too early or too late will be included with the closest scheduled post-baseline visit that includes the efficacy assessment, based on number of days since Day 1. The end of the visit window will not be used for the week 52 visit. This convention results in sequential visit windows so that no data is excluded from analysis. If an early termination visit and a regular visit (other than baseline) both fall within the same visit window, any non-missing efficacy assessments will be averaged and a worst-case approach will be used for safety data.

Percentages are based on the number of subjects in each treatment group and overall in the given population for medical history, prior and concomitant medications and AE summary tables. For all other tables, percentages are based on the number of subjects with non-missing data in each treatment group and overall for the given population.

All analyses will be conducted with R v3.3.1 or SAS® v9.4 or later using procedures appropriate for the particular analysis. All data collected during the study will be analyzed and reported unless stated otherwise.

6.4 Procedures for Handling Missing Data

Patients who drop out will have all available post-baseline data included in the analysis. The primary analysis mixed model is based on an assumption of Missing at Random (MAR) and is designed to handle right censored data for subjects who drop out of the study.

Two additional analyses will be performed as sensitivity analyses in this study. They are both based on a z-score based pattern mixture model (PMM) approach. The first PMM will use an MAR assumption and the second will use an MNAR assumption. The MAR sensitivity analysis will use a subject's last observed value and the z-score of that observation as a carried forward value, assuming a pattern of progression similar to subjects within the same treatment group who completed each visit. At each subsequent visit, a value will be imputed such that it has the same z-score relative to that subject's treatment group mean and standard deviation for completers at that visit. The MNAR sensitivity analysis will use a similar approach except that for subjects in all treatment groups, the imputed z-score will be assigned relative to the placebo group mean and standard deviation of completers at each visit. The first analysis is intended to estimate the treatment effect expected if all subjects

continued on treatment and the second analysis is intended to estimate the treatment effect expected if active subjects continued to the end of the study with no treatment (or a placebo treatment). After imputation for both sensitivity analyses, the estimated change from baseline between active and placebo group will be assessed by fitting an analysis of covariance (ANCOVA) model including covariates and random effect described in section 8.1 (except for time effect) at each separate visit.

6.5 Interim Analysis

No interim analysis of the data was planned or performed.

7 SUBJECT DISPOSITION, DEMOGRAPHICS AND BASELINE CHARACTERISTICS EVALUATIONS

7.1 Subject Enrollment

Subject enrollment will be summarized by treatment and center for all populations. The number of subjects overall and at each center for each analysis population will be presented.

Study timelines will also be summarized by treatment and overall for all randomized patients. This summary will include the earliest and latest screening and dosing dates among subjects within each treatment group and overall. It will also list the last patient and the date of his/her last visit within each treatment group and overall. Study duration will be presented in weeks and will be calculated using the following formula:

$$\frac{(Latest\ Week\ 52\ Visit\ Date - Earliest\ Screening\ Date + 1)}{7}.$$

Enrollment information will be provided in a data listing by subject.

7.2 Subject Disposition

Subject disposition will be summarized overall and by treatment group for all populations. The number and percentage of subjects completing the study and discontinuing from the study will be presented by treatment and overall and by reason for termination.

Subject disposition will be provided in a data listing by subject.

7.3 Drug Exposure

Duration of exposure is defined as the total time a subject is exposed to any study drug. The duration of exposure in months will be calculated by dividing the total number of days from the first dose date (Day 1) to the last dose date by 365.25 days/year and multiplying by 12 months/year. If the last dose date is missing or a subject is lost to follow-up, but the study medication administration log confirms that the subject has taken study drug, the date of the last completed study medication administration will be used.

Extent of exposure to study drug will also be characterized by calculating the cumulative number of micrograms taken by subjects. The duration and extent of exposure to study drug will be summarized by treatment group for both the safety and mITT populations.

The duration of exposure will be summarized by exposure category in weeks (less than or equal to 4 weeks, 4-12 weeks, 12-24 weeks and 24-52 weeks). The extent of exposure will be summarized by number of 112 mg tablets administered with 4 or 5 categorical breakouts determined by distribution of the data. N and percentage of subjects in each population will be displayed. The duration and extent of exposure to study drug will be summarized using descriptive statistics.

7.4 Subject Demographic and Baseline Data

Subjects will be described using demographic information and baseline characteristics recorded during the screening phase.

Demographic information to be assessed is age, gender, racial group, smoking habits, level of education, height and weight. The following additional subject characteristics will be calculated: primary language, marital status, primary occupation, retirement status and type of residence, Subject demographics will be summarized by treatment for the safety, mITT and PP populations.

Racial group, gender, smoking habits, level of education and other categorical questions will be summarized using the number and percentage of subjects with a particular attribute. The denominators for calculating the percentages will be the number of subjects in each treatment for the safety, mITT, and PP populations. Age, weight, height and other numeric responses will be summarized using descriptive statistics.

Baseline disease characteristics will be provided in a separate summary table. Prior/current AD therapy, length of time on specified AD therapy, time since diagnosis and baseline efficacy variables will be summarized using descriptive statistics by treatment for the safety, mITT and PP populations.

Demographics and baseline disease characteristics will be provided in a data listing by subject.

7.5 Medical History

Medical history will be summarized by treatment for each System/Category for the safety population. The number and percentage of subjects with significant medical history will be presented for each system organ class and preferred term. The denominators for calculating the percentages will be based on the number of subjects in each treatment group in the mITT population.

Medical history will be provided in a data listing by subject.

7.6 Medications

Medication summaries will present the number and percentage of subjects taking medications for the safety population. Summaries will be presented for prior (prior to Day 1) medication use and concomitant (Day 1 or later) medication use, if applicable.

All summaries will present the number and percentage of subjects by treatment.

Prior, concomitant and AD medication will be provided in a data listing by subject.

7.7 Protocol Deviations

Major protocol deviations are defined to be those deviations that could potentially bias the conclusions of the study. Minor deviations are defined to be those deviations not deemed major.

Major protocol deviations may include but are not limited to:

- Subjects who entered the study even though they did not satisfy the entry criteria.

- An increase in drug dosage.

The protocol deviations summary will present the number and percentage of subjects with each deviation category and specific deviation term within each treatment group and overall. In this summary, the total number of protocol deviations and number of subjects with at least one protocol deviation will be tabulated by treatment group and overall.

Protocol deviations will be provided in a data listing by subject.

8 EFFICACY EVALUATIONS

8.1 Primary Efficacy Analyses

The primary efficacy analysis will be performed on the co-primary efficacy endpoints for the mITT population. The primary efficacy endpoints will be analyzed by comparing the change in efficacy outcome at endpoint between treatment groups using a mixed model with repeated measures (MMRM). The MMRM will compare the estimated change from baseline between active treatment and placebo in the primary efficacy endpoints. The model will use separate repeated measures longitudinal models for each efficacy endpoint. This analysis will assess whether or not there is a difference in estimated CFB values between treatment groups and placebo at 52 weeks using least squares means estimates from the MMRM model.

SAS® PROC MIXED will be used to fit an MMRM with CFB of the efficacy outcome (i.e. ADAS-cog and ADCS-CGIC) as the response variable and the following covariates and fixed effects:

- Age (covariate);
- Baseline Test Score of Efficacy Parameter (covariate);
- Baseline MMSE score (covariate);
- Center/ Site (random effect);
- Treatment (fixed effect);
- ApoE4 status (fixed effect, positive, negative or unknown);
- Time (fixed effect, time will be defined in terms of Visits);
- Time by Treatment Interaction (Time*Treatment);
- Baseline by Time Interaction (Baseline*Time).

The covariance structure for the repeated measures in this model will be unstructured (UN). If UN does not converge for the model, the MMRM model will be simplified to allow convergence as described in the following paragraph. Variance components will be used as the covariance structure for the random site effect in the model.

Any efficacy outcomes that do not converge using the specified primary model will be rerun using an first-order heterogeneous autoregressive (ARH[1]) covariance structure. If convergence is still not achieved then the model will be simplified to exclude the baseline by time interaction and the baseline by treatment interaction from the model, first with the UN and then with ARH[1] . If convergence is still not achieved, the following terms will be excluded from the model: ApoE4 status, Age, first with the UN and then with ARH[1]. No other modifications will be made to the primary models, however, the covariate models will follow the same procedure with the exception that the specified covariate for the model will not be removed. If convergence is still not achieved, then these models will also have the baseline MMSE score and the baseline test score of the efficacy parameter removed (as long as it is not the specified covariate) : leaving only center/site, treatment and time in the model in addition to the covariate specified for the analysis and the covariate by treatment interaction.

Least-squares means will be estimated at the endpoint for each primary outcome. The LS mean in at the endpoint is interpreted as the expected CFB in the efficacy outcome at Week 52 when the specified treatment is administered. Least squares means and standard errors will be estimated from the mixed model at Week 52 and will be shown for the active treatment groups and placebo. In addition, treatment differences, p-values, 95% confidence intervals for the difference, effect size compared to placebo, 95% confidence interval for the effect size, and an effect size based upon Cohen's D will be displayed for treatment comparisons. Effect size will be calculated by taking the difference of LSMEANS and dividing by the standard deviation (i.e. the standard error of the estimated difference multiplied by the squared degrees of freedom). The equations below show how effect size and Cohen's D effect size will be calculated. In the following formulas *t1* stands for treatment 1, *t2* for treatment 2, *SE* for standard error. In the formulas below, *t2* is the reference group (*t1-t2)*:

Cohen's d will be calculated using the following equation:

$$\text{Cohen's } D = \frac{M_{t1} - M_{t2}}{Pooled\ SD}.$$

where the pooled standard deviation (pooled SD) is the standard deviation for all three treatment groups. The three-group pooled SD will be the SD corresponding to the visit of the comparison, as obtained from the estimated covariance matrix of the mixed model.

The number of subjects with an observed efficacy outcome, mean, standard deviation, median, 25th percentile (Q1), 75th percentile (Q3), minimum and maximum will all be reported and accompany the estimates from the MMRM outlined in this section.

A gatekeeper strategy will be used to preserve the family-wise alpha error at the two-sided level of 0.05 between co-primary endpoints and the multiple doses. The order of the analysis is shown in the table below. The first stage of the analysis will use a MMRM to compare the high dose group to placebo group for ADAS-cog at an alpha level of 0.05. If the high dose for ADAS-cog is statistically significant, the high dose for CGIC will be formally compared to placebo group at an alpha level of 0.05. If both comparisons for the high dose are statistically significant, the study will be considered a pivotal study with co-primary endpoints, and then the low dose will be compared to placebo for co-primary endpoints using the same strategy as used for the high dose. In addition, an exploratory analysis comparing a pooled active treatment group to placebo will be performed.

Table: Priority order of analysis

Condition for performing test	Test
None	Gate 1 = Significance on ADAS-cog at 2-sided alpha=0.05 for high dose
Gate 1	Gate 2 = Significance on ADCS-CGIC at 2-sided alpha=0.05 for high dose

Gate 1 and Gate 2	Gate 3 = Significance on ADAS-cog at 2-sided alpha=0.05 for low dose
Gate 1, Gate 2 and Gate 3	Gate 4 = Significance on ADCS-CGIC at 2-sided alpha=0.05 for low dose

Center will be included in the model. Centers with fewer than 2 subjects in each treatment group will be pooled for analysis starting with the largest study centers and pooling until each pooled center meets the requirement for at least 2 subjects per treatment group. If the last pooled center does not have 2 subjects per treatment group, it will be combined with the next smallest pooled center.

LS mean line plots over time will be displayed for each efficacy outcome with separate lines for each treatment. Forest plots will also be provided in order to compare confidence intervals side-by-side between treatment groups for each efficacy outcome.

The analyses and figures in this section will be applied to the mITT and PP populations.

8.2 MMRM for Secondary Endpoints and vMRI

The model described in Section 8.1 will be run for the secondary endpoints and vMRI. For vMRI results, the model described in 8.1 will be run excluding the fixed effect "time" and all interactions with "time." Since brain volumes are only collected at Screening and Week 52 "time" cannot be included in the model. Therefore, the models for vMRI will not have repeated measures, estimates at any other time points than 52 weeks, or the need for a covariance structure. One additional analysis will be performed to present correlations for brain volumes and efficacy outcomes at 52 weeks. For this analysis Pearson correlations and related statistics will be reported.

The same summary results described in Section 8.1 will be provided for each of the secondary endpoints and vMRI. LSMEANs will be estimated for each of the outcomes at each of the specified endpoints (see Section 5.2). The interpretation of each LSMEAN estimate is the estimated CFB in the efficacy outcome, after adjusting for other factors, at the specified time point and specific treatment. LS mean line plots over time will be displayed for each efficacy outcome with separate lines for each treatment. These analyses and figures are to be performed on both the mITT and PP populations. These, as well as all the analyses that follow, are considered exploratory efficacy analyses.

8.3 MMRM Including Various Interaction Terms

The model from Section 8.1 will be run for the primary and secondary endpoints with the addition of a specified interaction term. The following quantitative and categorical items will be included, separately, in the model as a two-way interaction with treatment and a three-way interaction with time and treatment:

- Baseline Test Score of Efficacy Parameter (quantitative);
- Baseline MMSE Score (quantitative);

- Age (quantitative);
- Baseline Score Median Split (differs by outcome, better mild vs. worse mild, where the median is included in the "better mild" group);
- Baseline MMSE Group (>=20 vs. <20);
- Baseline Age Group (>=75 vs. <75);
- Median Age Group (>= median vs. < median);
- APOE4 Status (homozygotes vs. heterozygotes vs. non-carriers)
- APOE4 Homozygote Status (homozygotes vs. heterozygotes and non-carriers);
- APOE4 Carrier Status (carriers vs. non carriers);
- Gender (Male vs. Female);
- Memantine Use (memantine use vs. no memantine use);

For example, the model described in Section 8.1 will be run with the addition of Gender*Treatment and Gender*Time*Treatment interaction. This is an example of the interaction terms that will be included, but these two interaction terms will be added for each quantitative/categorical item listed above. Summaries will include the p-value for the two way and three-way interaction terms.

Collinearity is not an issue when terms are only be included as a means of correction in the model, i.e. the main MMRM (Section 8.1). However, when investigating the significance of covariate interaction terms it is important to address collinearity between effects in the model and the specified interaction term. This is because collinearity would account for some of the same variability as the specified interaction term and would therefore reduce the significance of this term. The following practice will be implemented to account for collinearity in the covariate interaction analyses: if the specified covariate is a different version of a main effect in the model, or highly correlated with the main effect, then the covariate will replace the main effect and any associated interaction terms.

For example, since baseline MMSE is highly correlated with baseline test score of the efficacy parameter, all main effects and interactions involving baseline test score of the efficacy outcome will be replaced with baseline MMSE score. Since baseline MMSE score is already in the model as a main effect, the main effect of baseline test score of the efficacy parameter will be removed from the model. However, the interaction between baseline test score and time will be replaced by the interaction between baseline MMSE score and time, since this interaction did not exist in the original model. When baseline test score of the efficacy parameter is the covariate of interest, the main effect of baseline MMSE will be removed from the model. When the covariate is categorical age, quantitative age in the MMRM will be replaced with categorical age. This same rule will be applied for all ApoE4 covariate interactions. Note that accounting for collinearity is not an issue for gender or memantine use since these terms are not included as effects in the main MMRM.

LS means and summary statistics will be provided for each level of the factor (e.g. Male, Female) at each visit. For example, separate summary statistics and LS means will be provided at each endpoint for males and females. The models assessing the interactions with covariates will display slopes for the covariate effect at each visit for each treatment group.

LS mean line plots over time will be displayed for each efficacy outcome with separate lines for each treatment.

ApoE4 status will be classified into 3 stratification groups: homozygote, heterozygote and non-carrier. Homozygotes are patients with 4,4 genotype. Heterozygote describes patients with (2,4) and (3,4) genotypes. Non-carriers encompass all remaining genotypes: (2,2); (2,3) and (3,3).

These analyses and figures are to be performed on both the mITT and PP populations.

8.4 Efficacy Analyses for CSF Biomarkers

Due to small sample sizes, the model to describe the effect of the treatment on CSF biomarkers has been simplified. An analysis of covariance will be performed with percent CFB of the CSF biomarker as the response and only two explanatory variables included in the model:

- Baseline value of CSF biomarker;
- Categorical variable to indicate active treatment or placebo.

Summary statistics will be provided for the estimated percent CFB in CSF biomarkers for those on active treatment vs. placebo. A p-value and 95% confidence interval for the estimated effect will be included in the summary. As with analyses described in Section 8.1, overall summary statistics and effect size calculations will be provided.

These analyses are to be performed on both the mITT and PP populations. This model will be applied to the following biomarker outcomes: Aβ 1-40, Aβ 1-42, tau and p-tau.

One additional analysis will be performed to present correlations for CSF biomarkers and efficacy outcomes at 52 weeks. For this analysis Pearson correlations and related statistics will be reported for the mITT and PP populations.

8.5 Relationship between Clinical Response and Treatment

A responder analysis will assess patients with a clinical outcome response on individual or multiple efficacy endpoints.

A clinical outcome responder analysis will be performed to illustrate the relationship between clinical outcomes and treatment for the mITT and PP populations. Improvement on an endpoint is defined as improvement on the outcome by the end of the study, when compared to the baseline evaluation (e.g. CFB>0 when a higher score indicates better performance). Stabilization on an endpoint is defined as an outcome that did not worsen or improve by the end of the study, when compared to the baseline evaluation (e.g. CFB=0).

For this analysis, the high and low doses of T-817MA will be grouped together and categorized as *active treatment*. The grouped active treatment will be compared to the placebo group in these analyses.

Active treatment and placebo will be compared for improvement/stabilization on each of the primary and secondary outcomes individually at study endpoint. There will be two variations of this analysis. The first variation will include completers and non-completers, however, non-completers will be counted as non-responders regardless of their last efficacy result(s). The second variation will only compare subjects with non-missing efficacy outcomes at 52 weeks.

In addition to analyzing the relationship between individual clinical responses and treatment multiple clinical responses will also be correlated with treatment using the two variations described above. Those who experienced the following combined clinical responses will be summarized in terms of counts and percent:

- Improvement or stabilization on none of the 6 primary/secondary endpoints;
- Improvement or stabilization on 1 of the 6 primary/secondary endpoints;
- Improvement or stabilization on at least 2 of the 6 primary/secondary endpoints;
- Improvement or stabilization on at least 3 of the 6 primary/secondary endpoints;
- Improvement or stabilization on at least 4 of the 6 primary/secondary endpoints;
- Improvement or stabilization on at least 5 of the 6 primary/secondary endpoints;
- Improvement or stabilization on all primary/secondary endpoints.

Treatment groups will be compared in terms of percent of patients who experienced the combined clinical response or response on the individual primary and secondary efficacy outcomes. Fisher's Exact tests will be performed to test if there is a significant difference in improving on individual and multiple clinical outcomes based on treatment group.

8.6 Subject Discontinuation Rate

Counts of subjects who discontinue from the study early will be compared between treatment groups for the safety, mITT and PP populations using Fisher's Exact tests. In addition, time to discontinuation will be displayed. These same analyses will be repeated for any of the following discontinuation reasons with sufficient numbers of patients:

- Adverse Event;
- Death;
- Lost to Follow-up;
- Screen Failure;
- Protocol Violation;
- Withdrawal by Subject;
- Other.

Time to discontinuation overall and by reason will be analyzed with a Gehan-Wilcoxon test and the corresponding Kaplan-Meier Plots will be displayed. Patients discontinuing for one of the other reasons will be censored and “time to event” will be used.

9 SAFETY EVALUATIONS

9.1 Adverse Events

AEs reported on CRFs will be coded into system organ classes and preferred terms using the Medical Dictionary for Regulatory Activities (MedDRA v16.1). A treatment-emergent adverse event (TEAE) is defined as an AE with an onset date on or after the start of dosing. The adverse event summary will include only TEAEs. Any AEs that are not considered treatment-emergent will be provided in data listings only.

The incidence of AEs will be summarized for the safety population. Although a preferred term or system organ class may be reported more than once for a subject, each subject will only be counted once in the incidence count for each category. If a subject has the same AE on multiple occasions, the highest severity (severe > moderate > mild) or drug relationship (definite > probable > possible > unlikely > unrelated) recorded for the event will be presented.

Severity levels include: mild, moderate and severe. Relationships will be grouped into two categories for analysis: related and unrelated. Unrelated and unlikely will be categorized as "unrelated." Possible, probable and definite will be categorized as "related." If severity or drug relationship is missing no data imputation will be performed and no category of missing will be presented.

Summary tables showing the number of patients and percent within each category will be generated for each of the following types of adverse events:

- All AEs;
- Fatal Adverse Events;
- AEs for Subjects who Died.

These summaries will present the number and percentage of subjects reporting an adverse event for each classification level. The denominators for calculating the percentages overall will be based on the number of subjects in the safety population. The denominators for calculating the percentages by treatment will be based on the number of subjects exposed to each treatment in the safety population. In addition to these summaries, all AEs will be summarized by action taken, seriousness, severity, and relationship to study drug.

All AEs that occurred in 5% or more of the combined T-817MA treatment groups will be tabulated by treatment group for the safety population. These results will be analyzed descriptively and their incidence rate and two-sided 95% confidence intervals will be summarized. In addition, the risk ratio and its 95% confidence intervals between active treatment groups and placebo will be calculated in order to estimate the occurrence of side effects and adverse events. AE occurrence rate will also be summarized by periods corresponding to different levels of exposure to study drug in order to identify any patterns in the timing of events. The level of exposure will be broken-up into four groups based on when the AE occurred: Week 0-4, Week 4-12, Week 12-24 and Week 24-52.

All SAEs, AEs leading to premature discontinuation from the study, AEs with fatal outcome, and AEs for subjects who died will also be provided in data listings by subject and preferred term.

9.2 Vital Signs

Each vital sign will be summarized for each treatment and overall by visit, using descriptive statistics (mean, median, SD, minimum, maximum, and number of subjects) for the safety population. Additionally, descriptive summaries will be provided for CFB values for each treatment by visit for vital sign measurements collected during the study.

The latest non-missing vital sign value collected prior to dosing will be used as the baseline values. The baseline values will usually be the vital signs recorded at the baseline visit. In the case of repeated vital signs, the last collected values within that visit will be used for the summary tables.

Vital signs will be provided in a data listing by subject, visit, and parameter.

9.3 Clinical Laboratory Evaluations

Continuous clinical laboratory analytes absolute values and change from baseline values will be summarized by analyte and visit using descriptive statistics (mean, median, SD, minimum, maximum, and number of subjects). Mean line plots over time will be displayed for each analyte with separate lines for each treatment. Categorical laboratory analytes, classified as normal or abnormal, will be summarized by analyte and visit using the number and percentage of subjects in each category. The denominators for calculating the percentages will be based on the number of subjects with non-missing assessments at a particular visit for the safety population. The latest non-missing clinical laboratory tests collected prior to dosing will be used as the baseline values.

Shifts to values outside of the normal range will be presented by analyte and will be summarized by the number and percentage of subjects with shifts. Shifts will be determined for analytes in which both the baseline value and the termination value are recorded. The denominators for calculating the percentages will be based on the number of subjects with non-missing assessments for a particular analyte.

If upon review of the safety analyses of unblinded data, a safety trend is noted or implied, we will define parameters for lab values that would, when applied to the data, yield a dataset of results that could potentially be considered as clinically significant. We will then run the safety analyses on this dataset to see if the safety trend is confirmed or proved insignificant.

Clinical laboratory results will be provided in data listings by subject, visit and analyte. Abnormal lab results will be provided in a separate listing by subject, center, analyte and visit.

9.4 Electrocardiogram

ECG values and change from baseline values will be summarized by visit using descriptive statistics. ECG abnormalities will be summarized as the count and percentage of patients in each treatment group. CFB will be summarized in a shift table crossing baseline and each visit result. The denominators for calculating the percentages will be the number of subjects in each

treatment group who have an evaluation for both the screening and each visit in the safety population. These results will be analyzed descriptively and their incidence rate and two-sided 95% confidence intervals will be summarized

9.5 Physical Exam

Physical examination findings will be summarized as the count and percentage of patients in each treatment group. CFB will be summarized in a shift table crossing baseline and each visit (Week 12, 24, 36 and 52) results. The denominators for calculating the percentages will be the number of subjects in each treatment group who have an evaluation for both the screening and each visit (Week 12, 24, 36 and 52) in the safety population. These results will be analyzed descriptively and their incidence will be summarized

9.6 C-SSRS Ped

The absolute value and change in the C-SSRS Ped score from Baseline to Weeks 12, 24, 36 and 52 will be summarized between each T-817MA treatment group (high dose and low dose) and the placebo treatment group.

10 OTHER LISTINGS

The following additional listings will be provided:

- Subjects excluded from the safety, mITT, and PP populations;
- Clinical laboratory results for hematology, blood chemistry and urinalysis;
- Abnormal laboratory results;
- Physical examination assessments;
- Dose administration dates and times.

11 ANALYSIS FOR EXTENSION PERIOD

Statistical Analysis Plan for the extension period of this study will be made separately.

12 REFERENCES

1. Dipiro J.T., Talbert R.L, Yee G.C., Matzke G.R., Wells B.G., Posey L.M. (1996) Pharmacotherapy: A Pathophysiologic Approach. Appleton and Lange Editions, third edition.
2. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 1984; 34:939-944.
3. WHO (2003) World Health Report2003-Shaping the future. Geneva: WHO. Available at http://www.who.int/whr/2003/en/index.html
4. Langra KM, Chernew ME, Kabeto MU, Herzog AR, Ofstedal MB, et al. (2001) National estimates of the quantity and cost of informal caregiving for the elderly with dementia. J Gen Intern Med 16:770-778