BIOSTATISTICS
RESEARCH CENTER

ALZHEIMER'S DISEASE COOPERATIVE STUDY (ADCS)

Rasagiline Rescue in Alzheimer's Disease (RAS) Clinical Trial

Statistical Analysis Plan (SAP) – DRAFT

CONFIDENTIAL DOCUMENT

September 9, 2014

Investigators: Kate Zhong, MD and Jeffrey Cummings, MD
Written by: ADCS RAS Biostatistics Team

Contents

1 Introduction

This document outlines the Statistical Analysis Plan (SAP) for the *Rasagiline Rescue in Alzheimer's Disease Clinical Trial.*

2 Study Design

This is a Phase II, randomized, double blind, placebo controlled, parallel group, proof of concept three-site study, to evaluate the effect of Rasagiline in the regional brain metabolism on FDG PET. The study consists of two phases: a 24 week double blind placebo controlled treatment period and a 4 week follow up period. Patients will be randomized in a 1:1 ratio at baseline to receive either Rasagiline or matching placebo.
The study drug will be given as 0.5 mg dose once daily for the 4 weeks, then increases to 1 mg daily for the next 20 weeks. A total of 50 subjects will be enrolled: 25 will receive Rasagiline and 25 will receive matching placebo for the 24-week treatment period.

3 Study Objectives and Hypotheses

3.1 Primary Objectives

To determine if exposure to 1 mg of Rasagiline daily is associated with improved regional brain metabolism in the treatment group compared to the placebo group in Alzheimer's Disease patients

3.2 Secondary Objectives

- To evaluate the efficacy of Rasagiline 1 mg once daily compared to placebo after 24 week treatment on cognition (ADAS-cog), activities of daily living (ADCS-ADL), global function (CIBIC+), and neuropsychiatric symptoms (NPI)
- To evaluate the efficacy of Rasagiline 1 mg once daily compared to placebo after 24 week treatment on measures of executive function (mazes and cancellation of the ADAS-Cog13, Digit Span test, and Controlled Oral Word Association Test [COWAT] for verbal fluency)
- To evaluate the safety and tolerability of Rasagiline as measured by incidence of adverse events/serious adverse events (AEs/SAEs) clinical lab test data, vital signs, 12 lead EKG data, brain MRI finding

4 Populations of Interest

- Intent-to-Treat (ITT) Population: All consented randomized subjects grouped according to the treatment assigned at randomization, regardless of any protocol violations or any crossovers.
- Per-Protocol (PP) Populations: All Intent-to-Treat subjects who meet the inclusion criteria and exclusion criteria specified for enrollment in the protocol, and complete the study (week 24), have ingested between 80% and 120% of the protocol prescribed study medication as measured by pill count, and with no protocol violations that affect the analysis.

5 Analyses Populations: Definitions

- FDG Analyses: The primary FDG analysis will include all patients who received both a screening and end of study scan of acceptable quality.
- Clinical Outcome Analyses: The primary clinical analyses will include the ITT population.
- Safety Analyses: The primary clinical analyses will include the ITT population.
- Responder analysis: Responders will be defined as those that have a significant improvement on brain metabolism on FDG for the whole brain or for any of the pre-specified regions.

6 Enrollment and Participant Flow

6.1 Accrual of the study

Tables will summarize accrual by study site, and figures will summarize the overall rate of accrual over calendar time The observed rate and projected rate of accrual will be displayed in a graph (the projected rate assumes uniform accrual over time).

6.2 Study Flow CONSORT Diagram

A description of participant flow per the CONSORT guidelines (Begg, 1996) will be performed. The diagram will describe study status from screening to the end of the study. At each stage, reasons for persons not moving forward will be summarized by frequency and category for reason. The diagram will include the following information:

- Number of subjects screened
- Number and reason for those who screen failed
- Number of subjects randomized
- Number of subjects who completed the week 24 assessments
- Number and reason for subjects who discontinued study before week 24
- Number of subjects who completed week 28 assessments
- Number and reason for subjects who discontinued study between week 24 and week 28

The CONSORT diagram is separated by study arm after the randomization step.

6.3 Premature Discontinuation of Study Evaluation

Premature discontinuation of the study will be defined as anyone leaving the study and/or discontinuing treatment prior to having a week 24 assessment. Proportions of subjects who prematurely discontinue study treatment within 24 weeks will be compared between the study arms using a Fisher's exact test. In addition, participants who discontinue between 24 and 28 weeks will be reported by study arm and overall. Similar tables and comparisons will be included for participants who prematurely discontinue from the study within 24 weeks. A table will summarize reasons for prematurely discontinuing study between week 24 and week 28 by study arm and overall.

7 Evaluation of Demographics and Baseline Characteristics

Tables will summarize the study population at baseline, overall and by study arm. Descriptive statistics will be presented as N, mean, standard deviation, minimum, 25th Q, median 75th Q and maximum for continuous variables and frequency tables (row, column percentages) for categorical variables. Statistical comparisons will be performed between randomized arms using Wilcoxon Rank Sum Test (for continuous variables) or Fisher's exact test (for categorical variables).

- Baseline Demographics: All variables collected in the Participant Demographics form
- Baseline Medical History: All variables collected in the Medical History Form
- Baseline Vital Signs: All variables collected in the Orthostatic Vital Signs Form
- Baseline Clinical Measures:
  - ADAS-Cog11
  - ADCS-ADL
  - NPI (individual item and total)
  - Digit Span
  - COWAT
  - QoL-AD (Study Partner)

8 Imaging Analysis (Extracted from the protocol)

8.1 Screening Classification

The Screening FDG PET scan of each subject will be evaluated using ADMdx's dementia classifier, to assess whether the subject expresses a pattern consistent with AD or alternatively, that of other dementias. The classifier has developed machine learning methods and a comprehensive set of reference data from ADNI and other sources. The FDG PET scan of a subject will be independently compared to a set of canonical variant patterns that in combination characterize the patterns of relative hypometabolism caused by different types of dementia. A probability will be assigned to determine whether the subject is AD-like or better characterized as a different dementia.
In addition, hypometabolism will be assessed in frontal cortex and occipital cortex, and parietal asymmetry checked, to identify possible atypical or comorbid presentations that may impact clinical attributes. The AD-like or non-AD-like status of each subject will be reported back within 7 days of image receipt to allow enrollment decisions.

8.2 Screening Characterization

The Screening FDG PET scan of each subject will be evaluated using ADMdx's AD Progression classifier, to assess the subject's disease-related hypometabolism status relative to reference subjects and other study subjects. The FDG PET scan of a subject will be independently compared to a set of canonical variant patterns that in combination characterize the patterns of relative hypometabolism caused by different stages of progression toward AD dementia. This numeric score will be used to project likely cognitive trajectory for comparison to actual results. It can also be used to stratify groups for sub-analyses, creating more homogeneous analysis groups at baseline.

8.3 Longitudinal Voxel-Based Evaluation

The spatially normalized longitudinal FDG PET scans of each subject will be analyzed using ADMdx's NPAIRS multivariate software. Classes will be defined according to treatment or placebo condition, and visit. Example analyses are shown in Table 1 below. The output of this evaluation will be a series of patterns showing relative increases and decreases in cerebral glucose metabolism, quantification of the placement of each subject at each time point relative to these patterns of effect, and quantification of the contribution of each pattern to the overall effect. Metrics of reproducibility and predictive power are also provided. At preference of Sponsor, group assignments may be provided in a blinded manner - that is, Group A and Group B, without designation as treatment or placebo. In the table below, to better illustrate the comparisons, they are referred to as Placebo and Treatment, but these may instead be "Group A" and "Group B".
SPM-t contrasts will also be performed of selected groups, at thresholds of $p¡0.005$, and a cluster extent threshold of 25 voxels. These are limited to paired contrasts (e.g. baseline vs. 6 months for treatment group), or to contrasts of the difference images between baseline and 6 months, treatment vs. placebo groups. Information is more limited with regard to individual subject distribution other than at specific voxel locations.

8.4 Longitudinal Region of Interest based Evaluation

The SUVrs calculated for each subject will be compared using t-tests, within groups and across groups, as shown in Table 4 below. While the pre-identified reference regions for testing will be whole brain, cerebellum, and pons, an alternate reference region may be identified through the use of NPAIRS multivariate analysis and applied.

9 Analysis of Clinical Measures

In general, analyses will incorporate the Intent-to-Treat principle, namely, all randomized participants will be included in the analysis. Analyses for all efficacy outcomes will be guided by exploratory analyses. All results will be reported as point estimates (odds ratios or mean differences across groups, as appropriate) and interval estimates (95% confidence intervals) with two-sided p-values denoting statistical significance. Since this is an early phase safety and preliminary efficacy study, no adjustments for multiple comparisons will be made and a p-value of 0.05 will be considered statistically significant.

All clinical measures (MMSE, ADAS-Cog11, ADCS-ADL, NPI, Digit-Span, COWAT and QoL) over 24 weeks will be analyzed using the linear mixed-effects regression model to assess the difference in the rate of change between the two treatment arms. The model will include scores as the dependent variable and time, treatment assignment, time-by-treatment interaction term and any variables determined to be confounders as independent variables. Time will be treated as a continuous variable. An unstructured variance-covariance structure will be used. A significant treatment effect will be concluded if the p-value for the time-by-treatment interaction term in the model is <=0.05. Sensitivity analyses will be conducted by analyzing the data using response profiles. Here, time will be treated as a categorical variable. The analysis of response profiles will allow us to characterize the patterns of change in the mean response over time in the treatment groups.
To assess the rate of change in the ADCS-CGIC over 24 weeks, a generalized estimating equations (GEE) model for binary data will be used. Because very few change scores were at the extreme ratings of marked worsening, moderate worsening, marked improvement, or moderate improvement are expected in this study, the ADCS-CGIC will be treated as a dichotomous dependent variable (no change/improved and worse).

The analyses will be repeated in the PP population.

10 Correlation between Imaging and Clinical Change

The relationship between FDG-PET measures and change in cognitive endpoints over 24 weeks will be evaluated using a multivariable regression model.

11 Evaluation of Safety Measures

- Event and Participant count of the following will be summarized overall and by treatment group
  - AE: Overall and by MedDRA System Organ Class
  - AE: MedDRA Preferred Term
  - SAE: Overall and by MedDRA System Organ Class
  - SAE: MedDRA Preferred Term
  - SAE Definitely Related to Study Drug: Overall and by System Organ Class
  - Hospitalization
  - Deaths
- Comparisons of the number of participants with at least one AE, SAE, SAE definitely related to study drug and Death will be done between treatment groups using the Fisher's Exact test.
- Change in vital signs (blood pressure, weight, pulse, temperature) will be compared between treatment group using the Wilcoxon Rank-Sum test

12 Software

Statistical software R (version 3.1.1) will be used *http://www.r-project.org*.

13 References

Begg, et al. (1996). Improving the quality of reporting of randomized controlled trials: The CONSORT statement, JAMA, 276(8), 637-639