STATISTICAL ANALYSIS PLAN

(SAP for Part 1)

Title: Randomized, double-blind, placebo controlled, multicenter studies to evaluate the safety and efficacy of azeliragon as a treatment for subjects with mild Alzheimer's disease and impaired glucose tolerance

Protocol: TTP488-305, 29 April 2019

Study Drug: TTP488 (azeliragon)

Sponsor: vTv Therapeutics LLC

Date: 08 December 2020

Randomized, double-blind, placebo controlled, multicenter studies to evaluate the safety and efficacy of azeliragon as a treatment for subjects with mild Alzheimer's disease and impaired glucose tolerance

Prepared by:

Jessie Nicodemus Johnson 12/09/2020

Jessie Nicodemus Johnson, Ph.D. Date
Senior Statistician, Consultant
Pentara Corporation, Inc

Approved By:

Suzanne Hendrix 12/09/2020

Suzanne Hendrix, PhD Date
CEO, Consultant Statistician TTP488-305
Pentara Corporation, Inc

12/09/2020

Aaron Burstein, PharmD Date
SVP Clinical Development
vTv Therapeutics LLC

Table of Contents

1 Introduction

Alzheimer's disease (AD) is a neurodegenerative disease and the leading cause of dementia in the aging population. Neuropathological changes in AD consist of the formation and deposition of amyloid plaques and neurofibrillary tangles. The principal components of amyloid plaques are aggregated and insoluble forms of amyloid beta (Aβ) peptides ending predominantly at amino acid residues 40 and 42. Furthermore, surrounding the amyloid plaque are astrocytes expressing a calgranulin protein, S100b, which is a cytokine associated with chemoattractant activity for monocytes and activation of inflammatory cells of the myeloid lineage. The reactive microglia that surrounds plaques increases the expression of pro-inflammatory cytokines and complement receptors.

TTP Translational Technology®, TransTech's (now vTv Therapeutics) proprietary drug discovery engine, was employed to develop TTP488, which is an orally bioavailable antagonist of the Receptor for Advanced Glycation Endproducts (RAGE). This product is being developed as a potential treatment for AD.

These studies are designed to evaluate the safety and efficacy of azeliragon as a treatment for subjects with mild Alzheimer's disease and impaired glucose tolerance.

This study is planned with an enrichment strategy, following the draft regulatory guidance depicted in the FDA guidance *Enrichment Strategies for Clinical Trials to Support Determination of Effectiveness of Human Drugs and Biological Products* (March 2019). The statistical analysis is planned to conform to FDA guidance provided in Draft Guidance Early *Alzheimer's Disease: Developing Drugs for Treatment* (February 2018), review templates provided by the FDA, and *Guidance for Industry: Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products* (CDER&CBER, 1998).

The goals of this statistical analysis plan (SAP) are the following:

- To describe the approach of statistical analysis in order to support scientifically sound study conclusions for Part 1 of TTP488-305.
- To conform to regulatory guidance to facilitate the use of the statistical evaluation for support in the regulatory review of study results for purposes of product registration.
- To guide the analysis and reporting of study results to enable the production of valid and accurate depictions of study data.

The approach will conform to FDA guidance as described in Guidance for Industry Statistical Principles for Clinical Trials (1998).

1.1 Study Objectives

The primary objective of this study is:

- **To evaluate the impact of 6 months of treatment with oral azeliragon on cognitive performance**

The secondary objectives of this study are:

- **To evaluate the efficacy of azeliragon treatment on measures of function and activities of daily living**
- **To evaluate the efficacy of azeliragon treatment on complications of diabetes**
- **To evaluate the safety and tolerability of 6 months of azeliragon treatment**
- **To evaluate the effect of azeliragon treatment on biomarkers and markers of inflammation**

1.2 Study Rationale

Analysis of study TTP488-301 demonstrated potential beneficial effects of azeliragon that increased with higher baseline HbA1c value. This finding is consistent with the hypothesis that the TTP488-301 RAGE antagonist may have beneficial effects in patients who have overexpression of RAGE. HbA1c is being used as a marker for increased RAGE expression because research has shown a correlation of increased HbA1c with increased RAGE in a group of diabetic elderly patients with MCI (as well as a correlation of increased RAGE with increased AGEs). Further, patients with diabetes and HbA1c have increased AGE accumulation where AGEs bind to RAGE and promote increased expression of RAGE. (Gorska-Ciebiada 2015)

This study is an adaptive learn-and-confirm study to confirm TTP488-301 diabetes subgroup findings at 6 months (in Part 1) as well as use the information in Part 1 to refine the protocol for Part 2.

1.3 Subject Population

The study population includes subjects with a clinical diagnosis of probable mild AD (screening MMSE between 21 and 26, inclusive; baseline MMSE between 19 and 27, inclusive; CDR-global of 0.5 or 1 at screening and baseline; ADAS-cog14 score of at least 10 at screening and baseline) and screening HbA1c of 6.5% - 9.5%. The study population includes subjects who have a reliable caregiver with regular contact (i.e., 10 hours a week as combination of face-to-face visits and telephone contact acceptable) who will facilitate the subject's full participation in the study. Caregivers must have sufficient subject interaction to be able to provide meaningful input into the rating scales administered in this study where caregiver input is required, in particular the CDR, A-IADL, and FAQ and evidence of this should be documented in source

documentation. Participants who reside in assisted living facilities are permitted provided that they meet caregiver criteria. Subjects are between 50 and 85 years of age, inclusive.

1.4 Randomization

A total of 100 subjects were originally planned to be randomized into Part 1 of this study, but the total enrolled was 43 due to a slow enrollment and sponsor decision to stop enrollment due to the additive impact of the COVID pandemic on enrollment rates . Subjects were enrolled and randomized according to a fixed randomization scheme blocked by study investigative site.

Randomization is intended to have balanced allocation (1:1) between active and placebo. Dropouts will not be replaced.

1.5 Study Design

TTP488-305 is a protocol that consists of two sequential, multi-center, randomized, double-blind, placebo-controlled, parallel group studies operationally conducted under a single protocol. The purpose of the study is to evaluate the safety and efficacy of 5 mg/day oral azeliragon relative to placebo in subjects with mild Alzheimer's disease (screening MMSE 21-26, baseline MMSE 19-27 and ADAS-cog14 score of ≥ 10) and impaired glucose tolerance (Screening HbA1c 6.5% - 9.5%, inclusive). Eligible participants will be randomly assigned to azeliragon or placebo in a 1:1 randomization for a double-blind dosing period of 6 months (Part 1) or 18 months (Part 2). The Part 1 and Part 2 studies will have independently randomized, unique study populations.

The Part 1, 6-month, double-blind treatment portion of the study will include 5 clinic visits. The primary objective of the Part 1 portion is to evaluate the efficacy of 6 months of oral azeliragon on a measure of cognition (ADAS-cog14). Secondary objectives of Part 1 will be to evaluate the efficacy of 6 months of oral azeliragon on multiple measures of function (CDR, Amsterdam-IADL, FAQ) to inform selection of a functional endpoint to serve as co-primary in Part 2. Additionally, Part 1 will be used to inform the sample size required to adequately power Part 2 of this study. The Schedule of Activities for Part 1 is described in Section 14 of this SAP.

Part 1 is conducted at multiple sites in the U.S and Canada.

The overall structure of the study includes the following periods:

Screening period: Screening procedures occur within 60 days of start of study drug administration, and they include obtaining informed consent and evaluations to determine eligibility for study participation.

Baseline Study Period: (Day -7 — Day 1, pre-dose): Subjects will have baseline assessments for efficacy and safety measures, including imaging and PK assessments.

Treatment Period: (Day 1 — Month 6): Subjects are treated with double-blind study medication. Efficacy and safety measures are taken. Blood samples are taken and subjects have assessments for safety and PK.

Follow-up period: (Month 9): Subjects return to the site for final safety assessments including vital signs, adverse events, electrocardiography and concomitant medications are taken. Blood draws are taken for safety evaluations and for pharmacokinetics evaluations.

1.6 Sample Size and Power Considerations

Original Sample size considerations for Part 1 (from protocol):

Assuming a standard deviation of the change from Baseline to Month 6 in ADAS-cog14 of 4, using alpha = 0.05, a total sample size of approximately 82 subjects in balanced allocation (41 subjects per group) provides at least 90% power to detect a difference between treatment groups of 2.9 using a 2-sided, 2-sample t-test. Assuming a dropout rate of 18% or less, randomization of 100 subjects provides adequate statistical power for Part 1 to demonstrate superiority of azeliragon over placebo based on the ADAS-cog14. (Section 9.2).

Recalculated power considerations for Part 1:

ADAS-Cog: Estimates and standard deviations were drawn from a model that fit change in ADAS-Cog11 over time using the data from the STEADFAST A-study and B-study combined (N=880). The model utilized all individuals in the study. Estimates were drawn at an HbA1C of 7.2 (average from present TTP488-305 study; Part 1). The estimates of standard deviation and effect size for the change from baseline in ADAS-Cog was 3.4 and -1.54 respectively. Given a sample size of 43, a two-sided alpha of 0.05 would result in 30% power to detect a significant effect. If we observe the estimated effect size and standard deviation at the given sample size in Part 1 of the current study (TTP-305) we would expect to observe a p-value of 0.144 which could be considered evidence of a trend.

Previous studies (Siemers et al 2016) suggest that mean to standard deviation ratio between ADAS-Cog 11 and 14 are similar, therefore it is expected that the estimates drawn using ADAS-Cog11 will be applicable to ADAS-Cog14 data.

Cognitive Composite: The association of treatment was also tested using a composite that was derived from STEADFAST A and B studies. Composites allow the global assessment of multiple aspects of disease as well as exclusion of items that do not contribute to progression and thus add noise to the data. The composite is composed of the ADAS-Cog comprehension, word finding difficulty, and remembering test instructions variables. The composite will be calculated by summing the within visit z-scores for each item across individuals. Z-scores are calculated by subtracting the individual item score at visit x from the baseline mean item score and dividing this value by the item standard deviation at baseline.

The ADAS-Cog variables included in the model were chosen using PLS regression where treatment status at 6 months was the outcome variable and ADAS-Cog 11 and CDR-SB individual items were predictor variables. The PLS was modeled using

STEADFAST study B as this dataset showed a smaller ADAS-Cog effect with treatment.

PLS estimates and standard deviations were drawn from the STEADFAST data (A-study and B-study combined) as above for ADAS-cog. The model fit change in PLS over time with estimates drawn at an HbA1C of 7.2. The estimates of standard deviation and effect size for the change from baseline were -0.92 and 1.5, respectively. Given a sample size of 43, an alpha of 0.05 would result in 49% power to detect a significant effect. If we observe the estimated effect size and sd at the given sample size in the present study (Part 1) we would expect to observe a p-value of 0.052 indicating that the composite is somewhat more sensitive to treatment effects than the ADAS-cog 11.

Modified Cognitive Composite: The STEADFAST analysis indicates that a portion of the observed change may be due to age-related cognitive decline. To assess this, a modified composite that incorporated aspects associated with cognitive decline due to aging, specifically trail making test (TMT A/B) and Controlled Oral Word Association Test (COWAT) was generated. As these variables are not available in ELEVAGE, historic data (ADNI) was used to identify the best surrogate for each. All surrogates were found to be within ADAS-Cog individual items. Word finding difficulty was found to be a suitable surrogate for TMTA, while delayed word recall and ideational praxis were surrogates for TMTB (r=0.5, 0.6, and 0.6 respectively). The COWAT surrogates were orientation and word recognition (r=0.5 at 12 months). (delayed recall, ideational praxis, orientation, and word recognition).

The model fit change in PLS over time with estimates drawn at an HbA1C of 7.2. The estimates of standard deviation and effect size for the change from baseline were -1.73 and 2.9, respectively. Given a sample size of 43, a two-sided alpha of 0.05 would result in 47% power to detect a significant effect. If we observe the estimated effect size and sd at the given sample size in the present study (Part 1) we would expect to observe a p-value of 0.059 indicating that the composite is somewhat more sensitive to treatment effects than the ADAS-cog 11.

1.7 Early Stopping, Data Monitoring, and Interim Analysis

There will be no interim analyses.

An adjustment for alpha is included in the original protocol, as a conservative measure to accommodate unblinded evaluations by the independent data monitoring committee (IDMC). The alpha adjustment is applied to accommodate up to 10 analyses by the IDMC because the safety evaluation will include review of ADAS-Cog data from the safety perspective. Alpha = 0.0001 is apportioned to each of no more than 10 analyses by the IDMC that includes review of ADAS-Cog data.

There will be two database locks: one for the double-blind treatment period, and one for the follow-up period. Final analysis will be performed on the interim database after all subjects complete the treatment period (Month 6 or Early Termination) and all Treatment Period data is locked. The final database lock will occur after all Follow-up Visits are complete.

1.8 Conformance to Regulatory Standards

An objective of this SAP is to comply with regulatory standards. The following guidance documents were specifically consulted in the preparation of this SAP:

- ICH E9: Statistical principles for clinical trials (September 1998); ICH and FDA
- ICH E9(R1): Estimands and sensitivity analysis in clinical trials (draft; October 2017); ICH and FDA.
- ICH E14: Clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs (October 2012); ICH and FDA
- E14 Q&A (R3) ICH guideline E14: the clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs (R3)—questions and answers (26 January 2016); EMA
- Guideline on adjustment for baseline covariates in clinical trials (26 February 2015); EMA.
- Guideline on the clinical investigation of medicines for the treatment of Alzheimer's disease (Feb 2018); EMA.
- Good Review Practice: Statistical review template (30 July 2012). Office of Biostatistics; CDER; MAPP 6010.4
- Good Review Practice: Clinical review template (10 December 2010). Office of New Drugs; CDER; MAPP 6010.3
- Guidance: Providing clinical evidence of effectiveness for human drug and biological products (May 1998); CDER; CBER.
- Draft Guidance: Demonstrating Substantial Evidence of Effectiveness for Human Drug and Biological Products (Draft, December 2019); CDER, CBER.
- Good Review Practice: Reviewer guidance: Conducting a clinical safety review of a new product application and preparing a report on the review (February 2005); CDER.
- Draft guidance: Early Alzheimer's Disease: Developing drugs for treatment (February 2018); CDER, CBER.
- Draft guidance: Adaptive designs for clinical trials of drugs and biologics (November 2019); CDER, CBER.
- Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design (18 October 2007); EMA
- Draft guidance: Enrichment strategies for clinical trials to support Determination of effectiveness of human drugs and biological products (March 2019); CDER, CBER.
- Guidance: Drug-induced liver injury: premarketing clinical evaluation (July 2009); CDER, CBER.
- Guidance: Patient-reported outcome measures: Use in medical product development to support labeling claims (December 2009); CDER, CBER, CDRH.
- ICH and MSSO. (March 2019). MedDRA Introductory Guide Version 22.0.
- FDA Guidance on Conduct of Clinical Trials of Medical Products during COVID-19 Public Health Emergency: Guidance for Industry, Investigators, and Institutional Review Boards. March 2020 . Updated on September 2020.

- FDA Guidance: Statistical Considerations for Clinical Trials During the COVID-19 Public Health Emergency Guidance for Industry. June 2020; CDER, CBER, CDRH, CVM.

The planned statistical analysis of this study is intended to comply with regulatory expectations and standards as depicted in the above-listed guidance documents.

1.9 Modifications from the Statistical Section of the Protocol

The power calculation was updated from the protocol and the composite score analysis was added. Secondary variables (MMSE and biomarkers) and exploratory assessment of ADAScog11 progression were added. In addition, the statistical analysis plan adds more detail and analyses and supersedes the statistical section of the protocol.

1.10 Adaptive Study

This study is an adaptive trial, which includes 2 studies under the same protocol. This SAP is focused on Part 1 of this adaptive study.

This SAP conforms to standards set forth in *Adaptive Designs for Clinical Trials of Drugs and Biologics Guidance for Industry* (CDER, CBER, November 2019). The intention of the adaptive nature of this protocol is to leverage potential advantages related to statistical analysis as depicted in the FDA final guidance.

2 Statistical Hypotheses

Part 1 has a single primary endpoint (ADAS-Cog14).

The testing sequence will be as follows for strict significance, however proof of concept will be attained by significance in any of the outcomes below (Figure 1):

The first set of hypotheses to be tested is as follows:

H_{01}: The mean change in ADAS-cog14 in the active group is equal to that of the placebo group*.

H_{11}: The mean change in ADAS-cog14 in the active group is not equal (is superior) to that of the placebo group*.

Conditional on statistical significance, testing will proceed hierarchically:

H_{02}: The mean change in A-IADL in the active group is equal to that of the placebo group.

H_{12}: The mean change in A-IADL in the active group is not equal (is superior) to that of the placebo group[§].

Conditional on statistical significance, testing will proceed:

H_{03}: The mean change in CDR-sb in the active group is equal to that of the placebo group.

H_{13}: The mean change in CDR-sb in the active group is not equal (is superior) to that of the placebo group[§].

Conditional on statistical significance, testing will proceed:

H_{04}: The mean change in FAQ in the active group is equal to that of the placebo group.

H_{14}: The mean change in FAQ in the active group is not equal (is superior) to that of the placebo group[§].

H_{05}: The mean change in the cognitive composites or GST in the active group is equal to that of the placebo group.

H_{15}: The mean change in the cognitive composites or GST in the active group is not equal (is superior) to that of the placebo group.

* ADAS-Cog will also be run using LZCF. If neither of these are significant, the cognitive composites and global statistical test (GST) will be assessed for significance to demonstrate proof of concept. It is acknowledged that the present study is underpowered to replicate the clinically meaningful effect size observed in the STEADFAST combined A- and B-study diabetes subgroup. Therefore, if the observed effect size (treatment difference) or percent slowing (treatment difference divided by the placebo change) is equal to or greater than the corresponding values from the STEADFAST pooled analysis (-0.92; -61%), this will be considered proof of concept. Lack of strict statistical significance is expected with this range of effect size due to the small n, but this result, in the context of the consistency of the effect seen in STEADFAST Part A with the STEADFAST Part B data would provide proof of concept for this compound.

§ All secondary variables will also be run using LZCF.

3 Estimand Specification

This SAP is constructed to conform to regulatory guidance. In particular, this section is included as recommended by ICH E9 (R1) Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials (30 August 2017).

In accordance with ICH E9 Addendum, the specification of the estimand is included in this SAP with the 4 attributes detailed in ICH E9 Addendum, Section A.3.1:

A. The Population, that is, the subjects targeted by the scientific question:
   For this study, the population consists of subjects with AD and HbA1c of 6.5% - 9.5% who meet eligibility requirements for the study, receive double-blind medication, and have at least one valid post-baseline measure of ADAS-cog(14).

B. The variable (or endpoint) to be obtained for each patient, that is required to address the scientific question:
   For this study, the primary endpoint is Change in ADAS-cog(14).

C. The specification of how to account for intercurrent events to reflect the scientific question of interest:
   Due to the small sample size intercurrent events will not be accounted for in this analysis.

D. The population-level summary for the variable that provides, as required, a basis for a comparison between treatment conditions:
   For this study, the population-level summary for the variable is the least-squares mean (LSM) of the treatment difference in mean change from baseline in ADAS-cog(14).

4 Populations of Analysis

In accordance with guidance from the International Conference of Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) guideline ICH E9 Statistical Principles for Clinical Trials (1998), the following population will be used for all statistical analysis:

- The full analysis set (FAS) includes all randomized subjects, whether or not they receive any study medication, and who have at least one post-Baseline assessment.
- The per protocol set (PPS) includes all subjects in the FAS, except for those who are excluded because of major protocol violations, where a major protocol violation is one that may affect the interpretation of study results (e.g., taking less than 50% of prescribed study medication during participation).

  Final determinations of the PPS will be made at the masked data review meeting held in accordance with ICH E9 prior to the lock data. The complete specification (made blinded to treatment arm) will be documented in the minutes to the blind data review meeting held prior to lock the data.

- The safety set (SAF) includes all subjects who receive any study medication.

The FAS will be used for all hypothesis tests of efficacy. Analysis of superiority using the PPS will also be done for supportive analyses. If the PPS does not differ from the FAS by at least 15%, the PPS is considered optional.

The SAF will be used for safety analyses.

5 Variables of Analysis

5.1 Efficacy Variables

Efficacy evaluation will include the primary endpoint (ADAS-cog14 for Part 1), secondary measures, and other efficacy markers.

Statistical analysis will be done on each primary and secondary variable independently.

5.1.1 Primary Efficacy Variable

The primary analysis will include assessment of the following variables.

- Mean change from Baseline to Month 6 in ADAS-cog14*

5.1.2 Secondary Variables of Analysis

- Mean change from baseline in the Amsterdam-IADL at Month 6.
- Mean change from baseline in CDR-sb at Month 6.
- Mean change from baseline in the FAQ at Month 6.
- Mean change from baseline in the MMSE at Month 6.
- Mean change of treatment on biomarkers and markers of inflammation

5.1.3 Exploratory Variable

- Mean change from baseline in ADAS-cog11 at Month 6.

5.1.4 Proof of Concept

- Mean change from baseline in Cognitive Composite at Month 6
- Mean change from baseline in Modified Cognitive Composite at Month 6
- Mean change from baseline in GST at Month 6

*If the ADAS-Cog is not significant ($P > 0.05$) the cognitive composites and global statistical test (GST) will be assessed for significance to demonstrate proof of concept. No correction will be performed for these comparisons since they are all testing the overall effect of treatment on progression of disease and are expected to be highly correlated.

5.1.5 Efficacy Scales

Generation of instrument total scores will be calculated from individual items. The following instruments that are used in this study are subject to statistical analysis:

5.1.5.1 Alzheimer's Disease Assessment Scale - Cognitive Subscale 90 point (ADAS-cog):

- Range: The scale range is 0 to 90 with higher scores indicating greater cognitive impairment.

- Brief description: The 14-Item Alzheimer's Disease Assessment Scale - Cognitive Subscale (ADAS-cog14) is a structured scale (approximately 45 minutes to complete) that evaluates memory, orientation, attention, reasoning, language, constructional praxis, delayed recall, digit cancellation, and maze-completion. The ADAS-cog14 scoring range is from 0 to 90, with higher scores indicating greater cognitive impairment (Mohs et al. 1997). The ADAS-Cog should always be administered prior to other cognitive measures. The ADAS-Cog will be administered at Screening, Baseline, Months 3 and 6 or in the event of early termination. Participants must have scores of ≥10 at Screening and Baseline to be eligible for participation in the study.

5.1.5.2 Mini Mental State Examination (MMSE):

- Range: The scale range is 0 to 30 with lower scores indicating greater cognitive impairment.

- Brief description: The MMSE is a brief 30-point test that is used to assess cognition (Folstein, 1975). It is commonly used to screen for dementia. In the time span of about 10 minutes, it samples various functions, including arithmetic, memory and orientation. The MMSE will be administered at Screening, Baseline, Months 3 and 6 or in the event of early termination. Participants must have scores of 21-26 at Screening and 19-27 at Baseline to be eligible for participation in the study.

5.1.5.3 Functional Activities Questionnaire (FAQ):

- Range: The score ranges from 0 to 30. Higher scores are associated with greater impairment.

- Brief description: The FAQ is an inventory of instrumental activities of daily living developed by (Pfeffer 1982). The FAQ is administered by asking the informant to rate patient's ability (0-3) on 10 different activities of daily living. The FAQ will be administered at Screening, Baseline, Months 3 and 6 or in the event of early termination.

5.1.5.4 Amsterdam-Instrumental Activities of Daily Living Questionnaire (A-IADL):

- Range: scoring range from 20 to 80, with lower scores indicating greater impairment.

- Brief description: The A-IADL is an adaptive questionnaire designed to assess impairments in instrumental activities of daily living in early dementia.(Sikkes 2013) The questionnaire is completed by a caregiver / informant. The original questionnaire consists of 70 items in seven categories and the administration time is approximately 20-25

minutes. The A-IADL total score is calculated using an item response theory method of scoring (Embretson and Reise 2000, Reise and Waller). Higher scores indicate greater functional impairment. The A-IADL will be administered at Screening, Baseline, Months 3 and 6 or in the event of early termination.

5.1.5.5 Clinical Dementia Rating Scale (CDR):

- Global CDR Range:
    - 0 = normal; healthy individuals
    - 0.5 = questionable dementia
    - 1 = mild dementia
    - 2 = moderate dementia
    - 3 = severe dementia.

- CDR sum-of-boxes (CDR-sb) range:

  CDR-sb scores range from 0 to 18 with higher scores indicating greater cognitive impairment.

- Brief description: The CDR scale is used as a global measure of dementia and is completed by a clinician in the setting of detailed knowledge of the individual patient collected from interviews with the patient and caregiver (Berg, 1988). The CDR describes 5 degrees of impairment in performance on each of 6 categories including memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care. Much of the information will therefore already have been gathered, either as part of normal clinical practice or as part of a research study. The interview takes approximately 40 minutes to administer.

  The scores for each category can also be summed and this is known as the sum of box score (CSR-SB). The CDR will be conducted at Baseline and at Months 3 and 6 or in the event of early termination. To avoid patient fatigue, the CDR may be performed on a separate, but proximate, visit day than the other cognitive tests.

5.1.5.6 Global Statistical Test (GST):

A GST allows assessment of a global change in disease status/trajectory by standardizing and then combining measures. The GST will be a combination of 3 z-scores (ADAS-COG, CDR-SB, and A-IADL) with standardization calculated using the baseline mean and standard deviation for each score. The GST will be calculated for each subject at each timepoint by averaging the three z-scores and this GST will then be analyzed as a new outcome by calculating change from baseline to each visit.

5.1.5.7 Cognitive Composite:

Composites allow the global assessment of multiple aspects of disease as well as exclusion of items that do not contribute to progression and thus add noise to the data. The cognitive composite is composed of the ADAS-Cog comprehension, word finding

difficulty, and remembering test instructions variables. The composite will be calculated at each timepoint by summing the within visit z-scores for each item across individuals. Z-scores are calculated by subtracting the individual item score at visit x from the mean item score at visit x and dividing this value by the item standard deviation at visit x.

5.1.5.8 Modified Cognitive Composite:

The modified composite incorporates the cognitive composite variables as well as aspects associated with cognitive decline due to aging, specifically trail making test (TMT A/B) and Controlled Oral Word Association Test (COWAT). As these variables are not available in ELEVAGE; historic data (ADNI) was used to identify the best surrogate for each. All surrogates were found to be within ADAS-Cog individual items. Word finding difficulty was found to be a suitable surrogate for TMTA, while delayed word recall and ideational praxis were surrogates for TMTB (r=0.5, 0.6, and 0.6 respectively). The COWAT surrogates were orientation and word recognition (r=0.5 at 12 months).

5.1.5.9 Biomarker:

Blood for plasma biomarkers will be collected at baseline as well as months 3, 6, and follow-up.

5.2 Safety Variables

Safety is monitored in this study by collection of adverse events, vital signs, electrocardiography, C-SSRS, and clinical laboratory measures. It is noted that all untoward events or experiences are reported as adverse events regardless of whether they are identified by clinical observation, subject reporting, physical examination, clinical laboratory test result, electrocardiography, or any other examination or test. It is noted that listings are reviewed by medically qualified individuals including vital signs, laboratory data, electrocardiograms (ECG), and all safety data from all other sources to ensure that safety signals are identified and reported in the analysis of the safety data from this study.

5.2.1 Treatment-emergent Adverse Events

Adverse events will be coded using Medical Dictionary for Regulatory Activities (MedDRA) Version 16.0 or above. Adverse event coding will be done to the lowest level term (LLT). Adverse events will be summarized by System Organ Class (SOC) and preferred terms (PT).

Definitions:

- A treatment-emergent adverse event (TEAE) is an event that is observed or reported after administration of study medication that was not present prior to study medication administration or an event that represents the exacerbation of a pre-existing event.
- An adverse withdrawal is a subject who withdrew from the study due to an adverse event.
- A serious adverse event (SAE) is an AE that is classified as serious according to the criteria specified in the study protocol.

Safety and tolerability variables based on adverse events include the following:

- Proportions of subjects with TEAEs by Preferred Term and decreasing frequency of AEs

- Proportions of subjects with TEAEs by System Organ Class and Preferred Term
- Proportions of subjects with related TEAEs
- Proportions of subjects with severe TEAEs
- Proportions of subjects with treatment-emergent SAEs.
- Subjects with TEAEs that result in study termination

5.2.2 The Columbia Suicide Severity Rating Scale (C-SSRS)

- Range: The following outcomes are C-SSRS categories and have binary responses (yes/no). The categories have been re-ordered from the actual scale to facilitate the definitions of the composite and comparative endpoints, and to enable clarity in the presentation of the results.
    - Category 1 – Wish to be Dead
    - Category 2 – Non-specific Active Suicidal Thoughts
    - Category 3 – Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act
    - Category 4 – Active Suicidal Ideation with Some Intent to Act, without Specific Plan
    - Category 5 – Active Suicidal Ideation with Specific Plan and Intent
    - Category 6 – Preparatory Acts or Behavior
    - Category 7 – Aborted Attempt
    - Category 8 – Interrupted Attempt
    - Category 9 – Actual Attempt (non-fatal)
    - Category 10 – Completed Suicide

Self-injurious behavior without suicidal intent is also a C-SSRS outcome (although not suicide-related) and has a binary response (yes/no).

The following outcome is a numerical score derived from the C-SSRS categories. The score is created at each assessment for each patient and is used for determining treatment emergence.

Suicidal **Ideation** Score: The maximum suicidal ideation category (1-5 on the C-SSRS) present at the assessment. Assign a score of 0 if no ideation is present.

Endpoints:

Composite endpoints based on the above categories are defined below.

- Suicidal **ideation**: A “yes” answer at any time during treatment to any one of the five suicidal ideation questions (Categories 1-5) on the C-SSRS.

- Suicidal **behavior**: A “yes” answer at any time during treatment to any one of the five suicidal behavior questions (Categories 6-10) on the C-SSRS.

- Suicidal **ideation or behavior**: A “yes” answer at any time during treatment to any one of the ten suicidal ideation and behavior questions (Categories 1-10) on the C-SSRS.

Comparative endpoints of interest are defined below. “Treatment emergence” is used for outcomes that include events that first emerge or worsen. “Emergence” is used for outcomes that include events that first emerge.

- Brief description: The Columbia Suicide Severity Rating Scale (C-SSRS) is a joint interview with the caregiver and patient that systemically assesses suicidal ideation and suicidal behavior (Posner et al, 2007). This scale will be administered at Screening Visit to evaluate lifetime suicide attempt, suicide behaviors, and other non-suicidal self-injuries. Positive responses on the C-SSRS will be mapped to the Columbia Classification Algorithm of Suicide Assessment (C-CASA) for classification and reporting using a standard algorithm.)

5.2.3 Vital Signs

Vital signs measurements consist of blood pressure and pulse rate. Variables of analysis will include means, mean changes over time, and proportions of subjects meeting criteria for potential clinical concern in vital signs:

- Mean and mean changes in systolic and diastolic blood pressures
- Mean and mean changes in pulse
- Proportions of subjects with treatment-emergent values or changes of potential clinical concern.

5.2.4 ECG

12-lead ECGs are obtained at the Screening, Baseline, Month 3, Month 6/ET, and Follow-up visits. Triplicate measurements are obtained at Baseline; all other study visits collect single measurements.

ECG results will be obtained from the Central Reader. Interpretation of results as normal, abnormal not clinically significant or abnormal clinically significant will also be collected from the investigator or designee.

5.2.5 Clinical laboratory assessments

Routine clinical laboratory data, including eGFR, will be summarized with descriptive statistics on assessment values and change from Baseline. Routine clinical laboratory data will also be evaluated on the basis of laboratory-defined reference ranges and clinically important values or changes, as defined in this document (Section 10). Routinely collected laboratory safety data that are not required by the protocol will be considered source data, and will not be captured for inclusion into the study database.

Definition:

- A treatment-emergent abnormal value (TEAV) is a laboratory value that is abnormal after administration of study medication that was normal prior to study medication administration.

- A TEAV of potential clinical concern is a value or change that meets criteria specified in this SAP in Section 10.

Laboratory analytes in this study include:

- Hematology
  - Hemoglobin
  - Hematocrit
  - Erythrocyte (RBC) count
  - Platelet count
  - Total leukocyte (WBC) count
  - MCV
  - Leukocyte differential (percent and total)
    - Total neutrophils
    - Eosinophils
    - Monocytes
    - Basophils
    - Lymphocytes
- Clinical chemistry
  - BUN
  - Creatinine
  - Glucose
  - HbA1c
  - Calcium
  - Sodium
  - Potassium
  - Chloride
  - Bicarbonate
  - AST
  - ALT
  - GGT
  - LDH
  - Total bilirubin
  - Alkaline phosphatase
  - Uric acid
  - Albumin
  - Total protein
  - Phosphorus

Safety variables of analysis include:

- Subjects with TEAVs
- Subjects with TEAVs of potential clinical concern (based on criteria in this SAP)
- Subjects with a TEAV in a liver function test (LFT)

5.3 Pharmacokinetics and Pharmacodynamics Variables

Blood samples for pharmacokinetic (PK) analysis will be collected prior to dosing at Baseline, Months 3, 6, and at Early Termination.

All participants with at least one dose of study medication will be included in the pharmacokinetic and pharmacokinetic/pharmacodynamic analyses, as appropriate.

5.4 Adverse Events of Special Interest

Adverse events related to signs and symptoms of AD potentially pertain to efficacy as well as safety of the compound. Descriptive analysis will include TEAEs classified by MedDRA into the SOC for psychiatric disorders (at the SOC level within the hierarchy of MedDRA) and high level group terms (HLGTs) of anxiety disorders and symptoms (at the HLGT level within the hierarchy of MedDRA).

Standardized MedDRA queries (SMQs) will be used for dynamic analysis of any TEAE that emerges of potential interest.

Conditions associated with substance abuse are included in the SOC Psychiatric Disorders in high level term (HLT) "Substance-related disorders," which will be used for identification of terms pertaining to substance abuse. Variables of interest will include incidence estimates for each treatment group for the HLT and PTs within the HLT. It is noted that terms that refer to a person as a "drug abuser" would be captured in the SOC Social Circumstances, which is not intended for contributing to variables of analysis.

6 Statistical Methodology

General statistical methods will be applied as appropriate for out-patient clinical studies. Safety evaluations will rely on descriptive statistics. Randomization groups will be examined for homogeneity between treatment groups.

Continuous variables will be summarized using mean, median, standard deviation, minimum, maximum, and number of subjects available for analysis. Categorical variables will be summarized using frequency, proportion, and number of subjects available for analysis.

SAS version 9.3 or later will be used.

6.1 Pooling of Investigator Sites

Pooled sites will not be included in the model due to small N overall.

6.2 Statistical Methodology for Efficacy Analysis

The primary analysis will use the ITT methodology and a mixed-model measures (MMRM) methodology with a random slope and intercept for each individual. The MMRM will utilize an

unstructured covariance with only one parameter for the correlation between the slope and intercept. Time (ADY) is considered as a continuous variable. Analysis will include baseline risk score, treatment, time, as well as HbA1C by treatment, treatment-by-time, baseline by treatment, and baseline by time interactions as fixed effects. Subject will be included as a random effect. Baseline visit will be included in the model. Baseline risk score is derived from the historic STEADFAST data and calculated as a linear calculation of:

age-0.03582 + hba1c*-0.2335 + mmsebl*-0.2683 + base*-0.09342*

First the baseline risk score will be subtracted from the change score for each individual. Then the mean baseline value of the risk adjusted change score will be calculated separately for each treatment group. Finally, the risk adjusted change scores for each individual will have their group mean added back in to ensure an average score of 0 at baseline for each group. This adjusted change score will be used in the model in place of change from baseline. Model assumptions will be evaluated. If the parametric assumptions are evaluated as inappropriate or violated, rank analogues will be advanced.

6.3 Per Protocol Set

If the PPS differs from the FAS by more than 15%, the analyses will be replicated on the PPS. If the PPS and the FAS do not differ by more than 15%, analysis may not be done on the PPS. Final judgments will be made at the blind data review meeting in accordance with ICH E9.

6.4 Adjustments for multiplicity

Multiplicity of the primary efficacy analyses is controlled by using a conditional sequence of hypotheses. The study will be considered to demonstrate statistical significance if the primary analysis has a resulting p-value less than 0.049 for both primary endpoints for both A-study and B-study. No other adjustment for multiplicity is required.

6.5 Subgroup Analyses

6.5.1 Investigator Sites

Blind data review (recommended by ICH E9) may reveal many sites with few subjects randomized. Understanding the influence of site effects on analysis conclusions is essential and analyses to explore this are recommended in ICH E9 for key variables of analysis. A forest plot will be generated excluding each site one at a time to assess the impact of site for all efficacy measures.

6.5.2 Covariate and Categorical Interaction Analyses

The following variables will be assessed for interactions with the primary efficacy variable and time using the MMRM as described above:

- baseline score (for outcome being analyzed)
- baseline MMSE
- baseline HbA1C (continuous)
- ApoE4(carrier/non-carrier; discrete)
- ApoE4(number of E4 alleles; discrete)

6.6 Statistical Methodology for Primary Efficacy Analysis

We will use alpha = 0.0499, which includes an adjustment of 0.0001 for interim analyses performed by the Independent Data Monitoring Committee (IDMC).

In addition to the primary MMRM model, a last z-score carried forward (LZCF), will be run for the primary variable. Z-scores will be calculated for each timepoint. For individuals with missing data the last z-score will be carried forward. Let Z be the imputed z-score, while x is the last observation at timepoint t_1 and μ and σ are the mean and standard deviation of the next timepoint (t_2) for which the data is missing.

$$Z_{t2}=(x_{t1} - \mu_{t2})/\sigma_{t2}$$

Z-scores imputed relative to the group mean and standard deviation at each timepoint, will better preserve the slope of each arm and thus is more robust to differences in dropout rate across treatment. The model using LZCF will be the primary analysis for the GST and the composite scores..

Descriptive summaries will be produced of the observed values and change from Baseline in ADAS-cog14 by treatment group at each individual time point and at endpoint (final on-treatment assessment for each subject).

For statistical analyses, 95% confidence intervals will be produced for the least-squares means (LSM) in each treatment group, as well as the LSM differences as compared to placebo. For MMRM and ANCOVA, two-sided p-values will be displayed for the comparison against placebo.

The primary analysis controls alpha through the conditional sequence of hypothesis testing. No further adjustment for multiplicity is necessary.

If the PPS differs from the FAS by more than 15%, the analyses will be replicated on the PPS. If the PPS and the FAS do not differ by more than 15%, analysis may not be done on the PPS. Final judgments will be made at the blind data review meeting in accordance with ICH E9.

6.6.1 PK

The trough PK concentration data collected prior to each dose will be used in a concentration response relationship analysis to assess correlation with each outcome which will mirror the primary analysis but replace the treatment variable with PK concentration.

6.6.2 Missing Data

An essential component of the thorough analysis includes an assessment of the impact of missing data on study conclusions. As a part of assessing this impact, supportive analysis will include:

- observed cases by assessment time.

To ensure robustness of analysis conclusions against missingness, multiple imputation methods will also be done to cope with missing data as a key supportive analysis using the final on-treatment assessment of treatment failures as non-missing data at endpoint. Multiple imputations (MI) will be used as a supportive analysis with 100 invocations (acknowledging that more invocations are needed with more missing data). Monte Carlo methods are planned.

In accordance with the recommendations of the report from the National Academy of Science (NAS) panel "*The Prevention and Treatment of Missing Data in Clinical Trials*," (National Research Council, 2010), missingness is classified as "missing at random" (MAR) or "missing not at random" (MNAR). When subjects are withdrawn due to treatment failure, this event is considered in this sensitivity analysis to be MNAR, and the endpoint value is taken as the hard endpoint (imputed for progression over time) in analysis. Data that are missing for other reasons are considered to be MAR, and standard multiple imputation methods are planned.

6.7 Statistical Methodology for Secondary Analysis

Secondary and exploratory endpoints that are measurement variables will use similar statistical methodology to the methodology used for primary analysis and will also use similar figures as those described in the associated TOC. Efficacy variables that are proportions will be analyzed using Mantel-Haenszel test. For the analysis of the categories for responders, a Cochran-Mantel-Haenszel test will be used.

Secondary endpoints will be performed using LZCF as described above. Composite and GST scores will only be run as LZCF.

Methodologies for variables that are categorical and summarized with proportions will include construction of confidence intervals for each treatment group and for the difference between groups. For analysis, Fisher's exact test will be used for single-population analysis, and a Mantel-Haenszel test will be used for analyses combining data across strata.

Methodologies for time-to-event variables will include construction of 95% confidence intervals for each group separately and for the difference between groups. For analysis, a Wilcoxon test will be used for single-population analysis.

A graphical representation will be defined where the y-axis will show the cumulative percentage of subjects who achieved the specific measure of improvement or worsening in the ADAS-cog14 total score shown on the x-axis with a separate cumulative percentage curve for each treatment group.

6.8 Disposition, Demography, and Baseline Characteristics

A tabulation of subject disposition will be presented, including the number screened, the number randomized in each population group, the number dosed in each population group, the number who withdrew prior to completing the study, and reasons for withdrawal.

Demographic and baseline characteristics (disease history, medical history, and prior treatments for AD) will be summarized for all randomized subjects and for the FAS. No formal statistical comparisons will be performed. Summaries of continuous variables will include number of subjects, mean, median, minimum, maximum, and standard deviation. Summaries of categorical variables will include numbers of subjects in each category.

The variables to be summarized will include:

- Age, gender, race, and ethnicity
- Weight, height, BMI
- Years since diagnosis of AD
- Education level
- Previous medication for AD (AChEI, Memantine)
- Hypertension Y/N
- Diabetic medication type (ATC4 classification)
- Acetylcholinesterase inhibitor usage (Y/N)
- Memantine usage (Y/N)
- Acetylcholinesterase inhibitor usage (Y/N) and Memantine usage (Y/N)

Summaries will be provided overall and stratified by treatment group.

Additionally, baseline characteristics will be compared for efficacy measures of MMSE, ADAS-Cog14, CDR-sb, FAQ, Amsterdam IADL, HbA1c. No formal statistical comparisons will be performed.

6.9 Analysis of Safety Data

6.9.1 Adverse Events

Adverse events reported in this study will be coded using MedDRA®, Version 16.0, or later. Coding will be to the lowest level terms (LLT). The verbatim text, the preferred term, and the primary SOC will be listed in subject listings. Summaries that include frequencies and proportions of subjects reporting AEs will include the preferred terms and the SOCs. Summaries will include TEAEs by severity and by relationship to study medication.

Adverse event summaries will be constructed displaying AEs in decreasing order of total frequency according to the numbers of subjects reporting the AE (not the number of reports).

- Number (percent) of subjects reporting TEAE by treatment and overall in accordance with variables listed in Section 4.1.1.

Supportive listings will be constructed that includes the subject identification, the treatment group, TEAEs, MedDRA terms, seriousness, severity, causality, elapsed time to onset, duration, and outcome.

Methodology for AESIs will follow the same methodology for TEAEs.

6.9.2 Vital Signs

Subjects with vital signs meeting the criteria in Section 10 of this SAP for values of potential clinical concern will be identified and summarized. Proportions of subjects with vital signs of potential clinical concern will be examined for the treatment groups.

No formal inferential statistics will be applied to the vital signs data.

For each post-Baseline assessment, descriptive statistics are provided for the assessment value and the change from Baseline to the assessment with mean, median, standard deviation, minimum, and maximum at each assessment time.

6.9.3 APOE genotyping

APoE4 promotes the accumulation of β-amyloid proteins that are the cause of the characteristic plaques seen in the brains of AD subjects. The ApoE genes have consistently been associated with risk of AD dementia and will be
assessed as indicated in Table 1. The genotypes will be designated as 2/2, 2/3, 3/3, 3/4, or 4/4. Data will be summarized by treatment group.

6.9.4 Routine Clinical Laboratory Measurements

Because “exacerbations” of pre-existing abnormalities in laboratory analytes are examined using clinical judgment and are reported as TEAEs, additional analysis on TEAVs is limited to subjects with values that are normal prior to dosing and abnormal after dosing. Potentially

clinically significant abnormalities in laboratory analytes are to be reported as TEAEs and summarized as clinical TEAEs.

Subjects with clinical laboratory data meeting the criteria in Section 10 of this SAP for values of potential clinical concern will be identified and summarized. Proportions of subjects with clinical laboratory values of potential clinical concern will be examined for the treatment groups.

For each post-Baseline assessment, descriptive statistics are provided for the assessment value and the change from Baseline to the assessment with mean, median, standard deviation, minimum, and maximum at each assessment time. Liver Function Tests

Liver function tests have additional analysis views for this study in accordance with current regulatory guidance. To explore the potential for drug-induced liver injury consistent with *Guidance for Industry "Drug-induced liver injury: premarketing clinical evaluation"* (CDER, CBER, July 2009), subjects will be summarized and listed who meet the following criteria:

(1) Elevations in either AST or ALT of at least 3-times the upper limit of normal, and
(2) An accompanying abnormal bilirubin.

6.9.5 Electrocardiography

The heart rate, QT, PR, and QRS intervals will be recorded at each assessment time. No formal inferential statistics will be applied to the ECG data.

For each post-Baseline assessment, descriptive statistics are provided for the assessment value and the change from Baseline to the assessment with mean, median, standard deviation, minimum, and maximum at each assessment time.

In addition, the number of subjects with corrected and uncorrected QT values >500 msec will be summarized. Values from individual tracings within the triplicate measurement on Day 1, time 0 hour that are >500 msec will not be included in the categorical analysis unless the average from that triplicate measurement is also >500 msec.

Electrocardiography data will be summarized by Baseline and Change from Baseline to each scheduled assessment time with descriptive statistics.

Corrections to QT intervals will be made by Fridercia's method (QTcF). Categorical analysis will be done consistent with ICH E14, "Clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs" (October 2005).

In accordance with ICH E14, subjects will be categorized and summarized as described above according to:

- Absolute QTc interval prolongation:
  - QTc interval > 450
  - QTc interval > 480
  - QTc interval > 500
- Change from Baseline in QTc interval:
  - QTc interval increases from Baseline > 30
  - QTc interval increases from Baseline > 60.

6.10 Pharmacokinetics Analysis

Data resulting from blood sampling for trough concentrations of TTP488 and metabolites will be analyzed descriptively.

No formal inferential statistics will be applied to the pharmacokinetic data.

6.11 Pharmacodynamic Analysis

Plasma inflammatory biomarker change from baseline to 3 and 6 months will be evaluated using an MMRM containing terms for baseline value, treatment, time, and treatment-by-time interaction. Analysis will be done separately for each biomarker.

6.12 Concomitant Medications

Concomitant medications will be summarized by drug category and listed. Summaries will also be prepared for individuals taking medications for the treatment of hypertension and diabetes. Hypertension summaries will be stratified by Y/N while diabetes summaries will be stratified according to type of medication classified under the concomitant medications ATC4 variable. Use of these medications in addition to, Acetylcholinesterase and Memantine usage at baseline will be included in baseline tables. Summaries will be provided overall and stratified by treatment group.

7 COVID-19 Data and Analysis Conventions

Regulatory guidance was issued by the FDA and the EMA to acknowledge the potential impact of the COVID-19 pandemic on clinical trials, including missed visits or evaluations done in a manner different from the original intention of the protocol. This SAP includes recommended plans for coping with statistics and data issues related to COVID-19.

From the FDA guidance, the following excerpt applies to this SAP:

> Sponsors should describe in appropriate sections of the clinical study report (or in a separate study-specific document):
>
> 1. Contingency measures implemented to manage study conduct during disruption of the study as a result of COVID-19 control measures.
> 2. A listing of all participants affected by the COVID-19 related study disruption by unique subject number identifier and by investigational site, and a description of how the individual's participation was altered.
>
> 3. Analyses and corresponding discussions that address the impact of implemented contingency measures (e.g., trial participant discontinuation from investigational product and/or study, alternative procedures used to collect critical safety and/or efficacy data) on the safety and efficacy results reported for the study.
>
> Robust efforts by sponsors, investigators, and IRBs/IECs to maintain the safety of trial participants and study data integrity are expected, and such efforts should be documented. As stated above, FDA recognizes that protocol modifications may be required, including unavoidable protocol deviations due to COVID-19 illness and/or COVID-19 control measures. Efforts to minimize impacts on trial integrity, and to document the reasons for protocol deviations, will be important.

The following modifications to the primary analysis are intended to reflect the original analysis plan accommodating emergency modifications to visit schedules:

- Visits were intended to result in the primary endpoint evaluation at Month 6. Due to COVID-19, the endpoint visit may be later than Month 6. Three approaches will be used:
  - In the primary analysis, the Month 6 value can be predicted from earlier or later time points. Multiple imputation methods will be employed to predict all missing Month 6 visits, based on available time points for each patient.
  - Supportive analysis will be done using MMRM, which copes with extra data or mis-timed data.
  - Supportive analysis will also be done using the latest on-treatment value available, regardless of whether it is earlier or later than Month 6.
- Homogeneity analysis will be done comparing results prior to COVID-19 with those after impact of COVID-19 started. COVID-19 impact start date will be 13 March 2020.

8 Data Conventions

The following analysis conventions will be used in the statistical analysis.

8.1 Definition of Baseline

For safety evaluations, the Baseline assessment for all measurements will be the latest available, valid measurement taken prior to the initiation study medication, except for electrocardiography (ECG) values: The Baseline assessment for ECG include a triplicate assessment with the baseline value intended to be the arithmetic mean of the values taken at the Baseline assessment that are taken in triplicate. Baseline values for efficacy outcomes will be the average of screening and baseline visits where available. For the GST, the z-scores at each visit will be calculated using the mean over all patients of the average of screening and baseline, and using the standard deviation pooled across screening and baseline.

8.2 Missing Data

In general, missing clinical data will not be imputed, except where explicitly defined. In the event of study withdrawal, a subject's final assessment will be used as the endpoint value, if applicable.

Data are considered to be "on-treatment" if the assessment or collection follows the first administration of study medication and if the assessment occurs within 45 days following the final administration of study medication, justified based on the long half-life of this drug.

Dates with missing fields will not have days imputed.

In accordance with the recommendations of the report from the National Academy of Science (NAS) panel "*The Prevention and Treatment of Missing Data in Clinical Trials*," (National Research Council, 2010), sensitivity analyses will be done to ensure that study conclusions are robust against missing data. Last observation carried forward (LOCF) and baseline observation carried forward (BOCF) methods will be used as appropriate in sensitivity analysis. Missing data types will be examined, and statistical methodologies for appropriate sensitivity analyses will be finalized during blind data review as patterns emerge as the study progresses. Primary methodologies following the ITT principle are not subject to change.

The three types of missing data (Rubin, 1976) are considered to examine robustness of analysis conclusions against missingness and type of missingness:

- MCAR (missing completely at random) assumes the event of missingness is independent both of observable values and of unobservable values, and the missingness is entirely at random.
- MAR (missing at random) assumes that the missingness can be fully accommodated by accounting for variables where data are available.
- MNAR (missing not at random) the missingness may be related to the reason the data are missing.

The robustness of analysis conclusions against missingness is managed through supportive and sensitivity analyses, employing techniques MAR MI for placebo and MAR MI with delta adjustment (jump-to-placebo) for the group randomized to TTP488 (O'Kelly and Ratitch, *Clinical Trials with Missing Data: A Guide for Practitioners*, 2014).

8.3 Dropouts

If a subject withdraws from the study, if the date of an adverse event is not available and a determination of whether or not the event is treatment emergent cannot be made, by convention the event will be considered to be treatment emergent.

8.4 Visit Windows

Assessments taken outside of protocol allowable windows will be displayed according to the case report form (CRF) assessment recorded by the investigator.

8.5 Early Termination Assessments and Follow-up Assessments

When a patient withdraws from the study prematurely for any reason, an early termination (ET) visit is to be performed, except in the case when a patient withdraws consent (and the protocol specifically states that no further data would be collected). An ET visit should be "on treatment," and efficacy data collected at the ET visit are suitable for use in statistical analysis. Any data collected 45 or more days after last dose of investigational product (IP)_will be considered "off treatment".

8.6 Unscheduled Assessments

Extra assessments (laboratory data or vital signs associated with non-protocol clinical visits or obtained in the course of investigating or managing adverse events) will be included in listings, but not summaries. If more than one laboratory value is available for a given visit, the observation identified by the investigator to be the scheduled assessment will be used in summaries; all observations will be presented in listings.

Data from unscheduled assessments will be included in longitudinal data summaries (e.g., minimum, maximum over time). Data planned at scheduled times for which unscheduled assessments are taken will not be reflected in by-visit summaries.

8.7 Values Below LLQ

Concentrations for TTP488 and metabolites below the lower limit of quantification (LLQ) for the assay will be set to zero for the purpose of analysis.

8.8 Analysis Deviations

All statistical analyses and summary information will be generated according to this analysis plan. Any deviation from this plan will be documented in the clinical study report. Exploratory analyses are permitted per protocol; those analyses will not be considered to be deviations.

9 Software

All analyses will be done using SAS Version 9.3 or later.

10 Data Displays

The plans for displaying data will conform to this guideline and also to ICH E3 (*Structure and Content of clinical Study Reports)*, and to ICH E9 (*Statistical Principles for Clinical Trials*).

The following excerpt from ICH E3 (1996) applies to this SAP:

> "The guidance provided below is detailed and is intended to notify the applicant of virtually all of the information that should routinely be provided so that post-submission requests for further data clarification and analyses can be reduced as much as possible. Nonetheless, specific requirements for data presentation and/or analysis may depend on specific situations, may evolve over time, may vary from drug class to drug class, may differ among regions and cannot be described in general terms; it is therefore important to refer to specific clinical guidelines and to discuss data presentation and analyses with the reviewing authority, whenever possible. Detailed written guidance on statistical approaches is available from some authorities.
>
> "Each report should consider all of the topics described (unless clearly not relevant) although the specific sequence and grouping of topics may be changed if alternatives are more logical for a particular study" (p. 2).

In accordance with FDA and ICH guidance, tables, figures, and key data listings are planned. Exploratory and data-driven data displays may also be generated.

In addition to the data summaries and listings, the set of displays will include clinical laboratory normal ranges (with units) and also a display of the MedDRA mapping glossary, which will display the verbatim text as reported by the investigator, the LLT, PT, and SOC.

11 Values of Potential Clinical Concern

Criteria for Safety Values of Potential Clinical Concern

Hematology

Assay	Lower limit of the normal reference range (LLN)	Upper limit of the normal reference range (ULN)
Hemoglobin	<0.8 times	>1.2 times
Hematocrit	<0.8 times	>1.2 times
RBC	<0.8 times	
Platelets	<0.5 times	>1.75 times
WBC	<0.6 times	>1.5 times
Total Neutrophils (abs)	<0.8 times	>1.2 times
Eosinophils (abs)		>1.2 times
Monocytes (abs)		>1.2 times
Basophils (abs)		>1.2 times
Lymphocytes (abs)	<0.8 times	>1.2 times

Chemistry

Assay	Lower limit of the normal reference range (LLN)	Upper limit of the normal reference range (ULN)
Total bilirubin		>1.5 times
AST		>3 times
ALT		>3 times
GGT		>3.0 times
Alkaline Phosphatase		>3.0 times
Creatinine		>1.5 times
BUN		>1.3 times
Glucose	<0.6 times	>1.5 times
Uric acid		>1.5 times
Sodium	<0.95 times	>1.05 times
Potassium	<0.9 times	>1.1 times
Calcium	<0.9 times	>1.1 times
Albumin	<0.8 times	>1.2 times
Total protein	<0.8 times	>1.2 times
Bicarbonate	<0.9 times	>1.1 times
Chloride	<0.9 times	>1.1 times

12 References

- Berg, L. Clinical Dementia rating (CDR). Psychopharmacol Bull. 1988;24:637-9.
- Folstein, MF, Folstein, SE; McHugh, PR. Mini-mental State. A practical method for grading the cognitive state of participants for the clinician. Journal of Psychiatric Research 1975;12:189-98.
- Little, RJA, and Rubin, DB. (1987). Statistical analysis with missing data. New York: John Wiley and Sons.
- O'Kelly, M. Ratitch, B., 2014. *Clinical Trials with Missing Data: A Guide for Practitioners.* https://books.google.com/books?isbn=1118762533
- Rubin,l D.B. (1976, December). Inference and Missing data. Biometrikia, Volume 63, Issue 3 (Dec., 1976), 581-592.
- ICH E9: Statistical principles for clinical trials (September 1998); ICH and FDA
- ICH E9(R1): Estimands and sensitivity analysis in clinical trials (draft; October 2017); ICH and FDA.
- ICH E14: Clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs (October 2012); ICH and FDA
- E14 Q&A (R3) ICH guideline E14: the clinical evaluation of QT/QTc interval prolongation and proarrhythmic potential for non-antiarrhythmic drugs (R3)—questions and answers (26 January 2016); EMA
- Guideline on adjustment for baseline covariates in clinical trials (26 February 2015); EMA.
- Guideline on the clinical investigation of medicines for the treatment of Alzheimer's disease (Feb 2018); EMA.
- Good Review Practice: Statistical review template (30 July 2012). Office of Biostatistics; CDER; MAPP 6010.4
- Good Review Practice: Clinical review template (10 December 2010). Office of New Drugs; CDER; MAPP 6010.3
- Guidance: Providing clinical evidence of effectiveness for human drug and biological products (May 1998); CDER; CBER.
- Draft Guidance: Demonstrating Substantial Evidence of Effectiveness for Human Drug and Biological Products (Draft, December 2019); CDER, CBER.
- Good Review Practice: Reviewer guidance: Conducting a clinical safety review of a new product application and preparing a report on the review (February 2005); CDER.
- Draft guidance: Early Alzheimer's Disease: Developing drugs for treatment (February 2018); CDER, CBER.
- Draft guidance: Adaptive designs for clinical trials of drugs and biologics (November 2019); CDER, CBER.
- Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design (18 October 2007); EMA
- Draft guidance: Enrichment strategies for clinical trials to support Determination of effectiveness of human drugs and biological products (March 2019); CDER, CBER.
- Guidance: Drug-induced liver injury: premarketing clinical evaluation (July 2009); CDER, CBER.
- Guidance: Patient-reported outcome measures: Use in medical product development to support labeling claims (December 2009); CDER, CBER, CDRH.

- ICH and MSSO. (March 2019). MedDRA Introductory Guide Version 22.0.
- FDA Guidance on Conduct of Clinical Trials of Medical Products during COVID-19 Public Health Emergency: Guidance for Industry, Investigators, and Institutional Review Boards. March 2020 . Updated on September 2020.
- FDA Guidance: Statistical Considerations for Clinical Trials During the COVID-19 Public Health Emergency Guidance for Industry. June 2020; CDER, CBER, CDRH, CVM.
- Morris, J.C., Heyman, A., Mohs, R.C., Hughes, J.P., van Belle, G., Fillenbaum, G., Mellitis, E.D., Clark, C. "The Consortium to Establish a Registry For Alzheimer's Disease (CERAD). Part 1. Clinical and neuropsychological assessment of Alzheimer's disease." Neurology 1989; Sep; 39 (9): 1159-65.
- Reitan R. Validity of the trail making test as an indicator of organic brain disease. Perceptual and Psychomotor Skills 1958;8:271 6.
- Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. American Journal of Psychiatry 1984;141:1356-64.
- Mohs RC, Knopman D, Petersen RC, Ferris SH, Ernesto C, Grundman M, Sano M, Bieliauskas L, Geldmacher D, Clark C, Thal LJ. Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the Alzheimer's Disease Assessment Scale that broaden its scope. The Alzheimer's Disease Cooperative Study. Alzheimer Dis Assoc Disord. 1997;11 Suppl 2:S13-21.
- Pfeffer RI, Kurosaki TT, Harrah CH Jr, Chance JM, Filos S. Measurement of functional activities in older adults in the community. J Gerontol 1982; 37: 323-9.
- Sikkes SA, Knol DL, Pijnenburg YA, de Lange-de Klerk ES, Uitdehaag BM, Scheltens P. Validation of the Amsterdam IADL Questionnaire- _, a new tool to measure instrumental activities of daily living in dementia. Neuroepidemiology 2013;41:35–41.
- FDA Guidance on Conduct of Clinical Trials of Medical Products during COVID-19 Public Health Emergency: Guidance for Industry, Investigators, and Institutional Review Boards. March 2020 . Updated on June 3, 2020.
- Gorska-Ciebiada M, Saryusz-Wolska M, Borkowska A, Ciebiada M and Loba J (2015) C-reactive protein, advanced glycation end products, and their receptor in type 2 diabetic, elderly patients with mild cognitive impairment. Front. Aging Neurosci. 7:209. doi: 10.3389/fnagi.2015.00209
- Siemers ER, Sundell KL, Carlson C, Case M, Sethuraman G, Liu-Seifert H, Dowsett SA, Pontecorvo MJ, Dean RA, Demattos R. Phase 3 solanezumab trials: Secondary outcomes in mild Alzheimer's disease patients. Alzheimers and Dementia 2016; 12:110-120.
- Reise SP and Waller NG. Item response theory and clinical measurement. Annual Review of Clinical Psychologists. 2009:27-48.
- Embretson SE and Reise SP. Item response theory for psychologists. Ed. Erlbaum, Mahwah NJ (2000).

13 Schedule of Time and Events

SCHEDULE OF ACTIVITIES

Part 1 Schedule of Activities

	Screening	Baseline	Treatment Period[a]							Follow-up[b]
Protocol Activity	**Day -60 to Day -1**	**Day -7 to Day 1**	**Day 1[c]**	**Month 1**	**Month 2**	**Month 3**	**Month 4**	**Month 5**	**Month 6 / ET**	**3 months after last dose**
Study Days	-60 to -1	-7 to 1	1	30	60	90	120	150	180 or ET	270
Window				±7d	±7d	±7d	±7d	±7d	±7d	±7d
Sign informed consent / provide assent[d]	X									
Assessment of continued consent/assent		X				X			X	X
Assign Screening ID	X									
Randomization			X							
Demographic information	X									
Review Inclusion/Exclusion Criteria	X	X	X							
Medical History/Surgical History / Drug, Alcohol, Tobacco Use History	X									
Complete Neuro Exam & Physical Exam	X									
Height	X									
ApoE genotyping		X								
Telephone Contact				X	X		X	X		
Body weight	X	X				X			X	X
BMI calculation	X									
Brief Neuro & Physical Exams		X				X			X	X
Review Concomitant Medications	X	X		X	X	X	X	X	X	X
Blood Pressure and Pulse Rate (supine) [e]	X	X[e]				X			X	X
12 Lead ECG[e]	X[f]	X[e]				X			X	X

	Screening	Baseline	Treatment Period[a]							Follow-up[b]
Protocol Activity	**Day -60 to Day -1**	**Day -7 to Day 1**	**Day 1[c]**	**Month 1**	**Month 2**	**Month 3**	**Month 4**	**Month 5**	**Month 6 / ET**	**3 months after last dose**
Study Days	-60 to -1	-7 to 1	1	30	60	90	120	150	180 or ET	270
Window				±7d	±7d	±7d	±7d	±7d	±7d	±7d
Adverse Events Assessment		X		X	X	X	X	X	X	X
Dispense / Return Study Drug / Drug Accountability			X			X			X	
Study Drug Dosing[a,c]			X	→	→	→	→	→		
MMSE[g]	X	X				X			X[h]	X
ADAS-cog14[g]	X	X				X			X[h]	X
CDR[g]	X	X				X			X[h]	X
FAQ		X				X			X[h]	X
Amsterdam-IADL		X				X			X[h]	X
Columbia Suicide Severity Scale (C-SSRS)[i]	X	X				X			X	X
Fasting Central Lab Blood and Urine Collection[j]	X	X				X			X	X
Urine Drug Screen	X									
Provide meal/snack[k]	X	X				X			X	X
Blood glucose measurement[l]	X	X				X			X	X
Brain MRI[m]	X									
Blood Sample for PK, PD and Plasma Retention and Storage		X				X			X	X

a Participants will receive azeliragon 5 mg once daily, or matching placebo during the treatment period.

b Follow-up visit required for all participants who dosed with study medication. A Follow-up Visit should be completed 3 months after a subject completes Month 6 for those participants who complete the study; or at least 45 days after the last dose of study drug if the subject discontinues the study prior to Month 6.

c Subject begins taking study medication in the clinic as soon as baseline procedures are completed, and eligibility confirmed. Date of first dose is noted as Day 1.

d Where assent is in accordance with local laws, regulations and ethics committee policy.

e Vitals and ECG will be measured in triplicate at Baseline and as single values at all other times. An average of the three measurements for ECG and blood pressure will be used for eligibility determination.

f ECG at Screening visit repeated if QTc > 470 msec.

g Neuropsychological assessments are not to be performed with fasted participant; ensure participant has been fed and meets blood glucose measurement requirements prior to any neuropsychological assessments.

h Neuropsychological assessments are performed at the Early Termination Visit only when the visit is within 30 days of the last dose of study drug.

i C-SSRS is administered to subject jointly with caregiver.
j Blood Chemistry sample is collected in a fasted state (8-hour fast).
k Provide a low carbohydrate/high protein meal or snack *after* fasting blood samples are drawn and *before* any neuropsychological assessments are performed.
l Blood glucose measurement is an in-office point of care measurement after an appropriate period following the meal/snack, and prior to neuropsychological assessments are performed.
m CT scan may be utilized for subjects with contraindication to MRI. MRI or CT should be done at least 14 days prior to the baseline visit to allow time for MRI report to support eligibility decision.

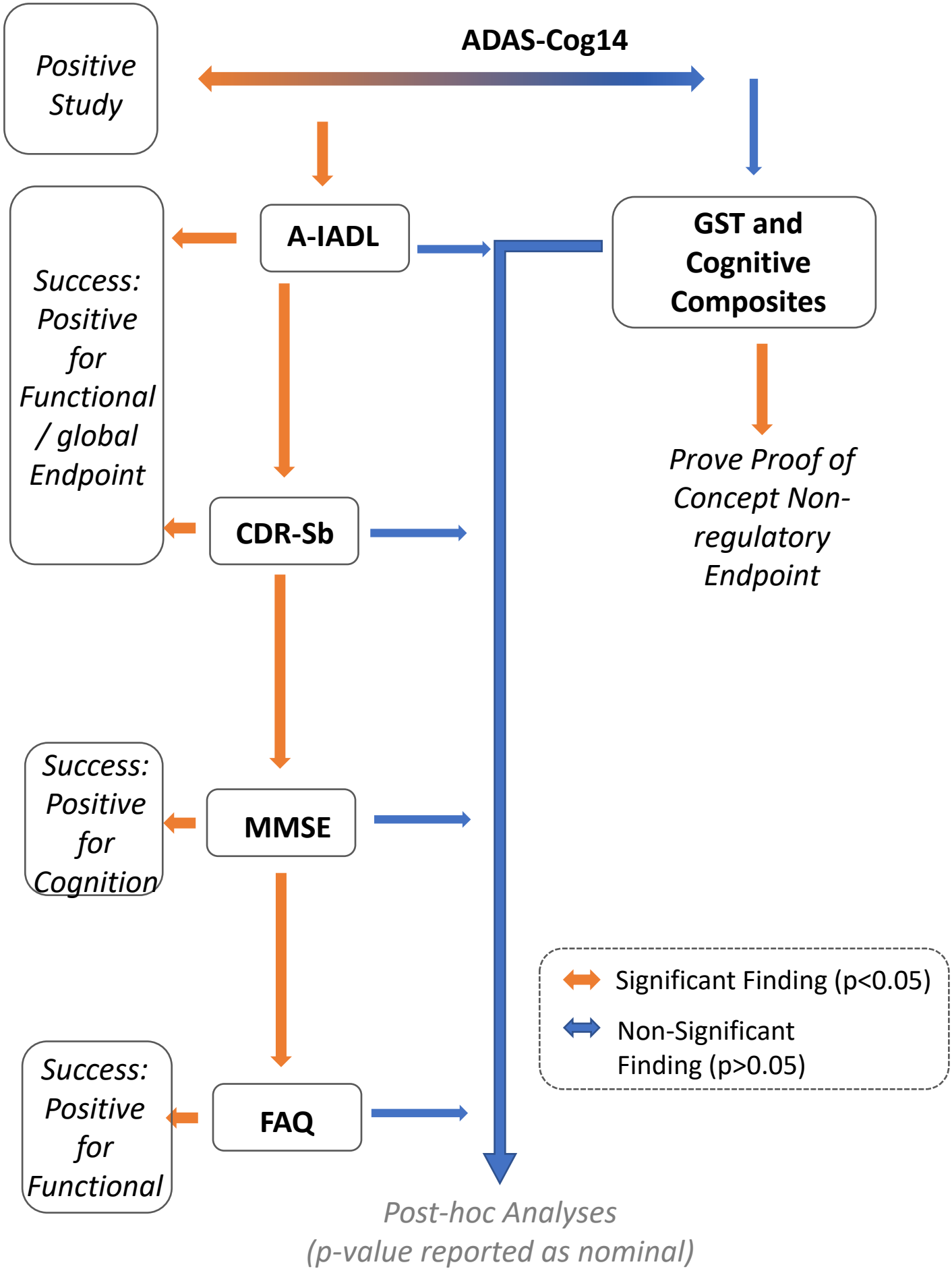

Figure 1. Statistical Hypothesis Schematic