Statistical Analysis Plan v2 I7X-MC-LLCF

Effect of LY3202626 on Alzheimer's Disease Progression as Measured by Cerebral 18F-AV-1451 Tau-PET in Mild Alzheimer's Disease Dementia

NCT02791191

Approval Date: 16-Aug-2018

1. Statistical Analysis Plan for LLCF: (I7X-MC-LLCF): Effect of LY3202626 on Alzheimer's Disease progression as Measured by Cerebral 18F-AV-1451 Tau PET in the Mild Alzheimer's Disease Dementia

LY3202626 Alzheimer's Disease

Multicenter, randomized, double-blind, placebo-controlled, Phase 2 study comparing 3 mg and 12 mg of LY3202626 with placebo over 52 weeks in approximately 500 patients with mild Alzheimer's disease dementia.

Eli Lilly and Company
Indianapolis, Indiana USA 46285

[Protocol I7X-MC-LLCF]
[Phase 2]

Statistical Analysis Plan electronically signed and approved by Lilly on date provided below.

Approval Date: 16-Aug-2018 GMT

2. Table of Contents

Section **Page**

Table of Contents

Table **Page**

Table of Contents

Appendix **Page**

3. Revision History

Statistical Analysis Plan (SAP) Version 1 was approved prior to unblinding.

Statistical Analysis Plan (SAP) Version 2 was approved prior to final database lock. Due to early termination of the study, certain analyses were removed or simplified, while others were clarified. The changes in SAP version 2 are summarized below:

- Clarified that change from baseline to last observation is to be used in ANCOVA analysis if Week 52 observation is not available (Section 5.1)
- Removed the exploratory analysis on subjects with study partner change during the study (previously Section 5.11 in Version 1)
- Clarified the analysis for changes from baseline in NPI total score and BASQID total score will be ANCOVA instead of MMRM because they were collected only once during post-baseline visits (Section 5.11.2)
- Specified the cut-off for defining Tau subgroups and removed the subgroup analysis regarding race and country (Section 5.11.3)
- Removed sensitivity analysis on Tau PET using percentage changes from baseline (previously Section 5.12.4 in Version 1)
- Simplified analysis of CSF biomarkers to listing of values by visit by treatment due to very limited number of subjects participated in the CSF addendum (Section 5.12.2)
- Added analysis on annualized change in imaging biomarkers (vMRI, AV-45 PET, AV-1451 PET) because majority of the subjects would not have Week 52 observation due to early termination of the study (Section 5.12.5)
- Added an incidence summary for treatment-emergent CSSRS findings (Section 5.13.8.1)
- Simplified and clarified the analysis for categorical eye exam and eye imaging data (Section 5.13.8.3)

4. Study Objectives

4.1. Primary Objective

To assess the change from baseline in 18F-AV-1451 positron emission tomography (PET) after treatment with LY3202626 3 mg or 12 mg per day compared with placebo for 52 weeks among patients with mild Alzheimer's disease (AD) dementia and evidence of brain amyloid.

The endpoint used to assess this will be the change in standardized uptake value ratio (SUVr) of ^{18}F-AV-1451 PET from baseline to 52 weeks. The change will be compared between the LY3202626 doses (3mg and 12mg) and placebo.

4.2. Secondary Objectives

Objectives	Endpoints

Secondary
• To evaluate the safety and tolerability of LY3202626 (3 mg and 12 mg per day) compared with placebo.
	• Standard safety assessments:
	  o spontaneously reported adverse events (AEs)
	  o clinical laboratory tests
	  o vital sign and body weight measurements
	  o 12-lead electrocardiograms (ECGs)
	  o physical and neurological examinations
	• magnetic resonance imaging (MRI; amyloid-related imaging abnormality [ARIA] and emergent radiological findings)
	• Skin examination
	• Eye examination
	• Columbia-Suicide Severity Rating Scale (C-SSRS)

• To assess peripheral pharmacokinetic (PK) and pharmacodynamic (PD) of LY3202626 (3 mg and 12 mg per day) over 52 weeks.	• Plasma PK of LY3202626 • Plasma PD $A\beta_{1-40}$, $A\beta_{1-42}$, $A\beta_{1-x}$
• To assess change from baseline in cognition after 52 weeks of treatment among patients treated with LY3202626 (3 mg and 12 mg per day) compared with placebo.	• 13-item Alzheimer's Disease Assessment Scale – Cognitive subscale (ADAS-Cog13)
• To assess change from baseline in function after 52 weeks among patients treated with LY3202626 (3 mg and 12 mg per day) compared with placebo.	• Alzheimer's Disease Cooperative Study-Activities of Daily Living inventory (ADCS-iADL)
• To assess change from baseline in a composite measure of cognition and function after 52 weeks among patients treated with LY3202626 (3 mg and 12 mg per day) compared with placebo.	• integrated Alzheimer's Disease Rating Scale (iADRS)

4.3. Exploratory Objectives

Objectives	Endpoints
Exploratory	
• To assess change in cognition after 52 weeks among patients treated with LY3202626 (3 mg and 12 mg per day) compared with placebo	• Montreal Cognitive Assessment (MoCA)
• To assess change in function after 52 weeks among patients treated with LY3202626 (3 mg and 12 mg per day) compared with placebo	• Functional Activities Questionnaire (FAQ)
• To assess change in cognition after 52 weeks among patients treated with LY3202626 (3 mg and 12 mg per day) compared with placebo	• Mini-Mental State Examination (MMSE)
• To assess change in everyday functioning after 52 weeks among patients treated with LY3202626 (3 mg and 12 mg per day) compared with placebo	• Everyday Cognition (ECog)
• To assess change in neuropsychiatric symptoms after 52 weeks among patients treated with LY3202626 (3 mg and 12 mg per day) compared with placebo	• Neuropsychiatric Inventory (NPI)
• To assess change in subjective quality of life after 52 weeks among patients treated with LY3202626 (3 mg and 12 mg per day) compared with placebo	• Bath Assessment of Subjective Quality of Life in Dementia (BASQID)
• To assess change from baseline in rCBF after 52 weeks among patients treated with LY3202626 (3 mg and 12 mg per day) compared with placebo.	• SUVr cerebral perfusion from early post-injection florbetapir PET scan
• To assess change from baseline in brain amyloid after 52 weeks among patients treated with LY3202626 (3 mg and 12 mg per day) compared with placebo	• Florbetapir PET SUVr
• To assess change from baseline in brain and brain regional volumes after 52 weeks among patients treated with LY3202626 (3 mg and 12 mg per day) compared with placebo.	• Volumetric MRI

Abbreviations: MRI = magnetic resonance imaging; PET = positron emission tomography; rCBF = regional cerebral blood flow; SUVr= standardized uptake value ratio.

5. A Priori Statistical Methods

5.1. General Considerations

All tests of treatment effects of biological efficacy or clinical efficacy will be conducted at a 1-sided α=.10, unless otherwise stated. Safety assessments will be conducted at a 2-sided α=.05.

A repeated measures analysis refers to a restricted maximum likelihood (REML)-based, mixed effects repeated measures analysis using all the longitudinal observations at each post-baseline visit.

All analyses will follow the intent-to-treat (ITT) principle unless otherwise specified. An ITT analysis is an analysis of data by the groups to which subjects are assigned by random allocation, even if the subject does not take the assigned treatment, does not receive the correct treatment, or otherwise does not follow the protocol.

When change from baseline is assessed, subjects will be included in the analysis only if both a baseline and a postbaseline measure are available. Unless otherwise defined, Visit 3 is the baseline visit, and baseline value is defined as last value available before or at Visit 3.

All total and subscale scores will be derived from individual items. If any individual item is missing, the corresponding total and subscale scores will be considered missing, except for the ADAS-Cog13 and ADCS-iADL, which will be imputed.

A repeated measures analysis model will include the fixed, categorical effects of treatment, visit, and treatment-by-visit interaction, as well as the continuous, fixed covariates of baseline, baseline-by-visit interaction, and age at baseline.

When an analysis of covariance (ANCOVA) model is used to analyze a continuous efficacy variable, the model will contain the main effects of treatment with baseline value and age at baseline added as covariates. For ANCOVA analysis on change from baseline to Week 52 endpoint, change from baseline to last observation during treatment period will be used if patients do not have Week 52 observation available.

Categorical comparisons between treatment groups will be performed using Fisher's exact tests or logistic regression where appropriate.

Any change to the data analysis methods described in the protocol will require an amendment to the protocol only if it changes a principal feature of the protocol. Any other change to the data analysis methods described in the protocol, and the justification for making the change, will be described in the clinical study report (CSR). Additional exploratory analyses of the data will be conducted as deemed appropriate.

5.2. Adjustments for Covariates

The repeated measures models will include the fixed, categorical effects of treatment, visit (categorical covariate), treatment-by-visit interaction, as well as the continuous, fixed covariate of baseline, baseline-by-visit interaction, and age at baseline.

When an ANCOVA model is used to analyze a continuous efficacy or safety variable, the model will contain the main effects of treatment and appropriate baseline value and age at baseline included as covariates.

5.3. Handling of Dropouts or Missing Data

5.3.1. Handling Missing Data from Subject Dropouts

A likelihood-based mixed effects model for repeated measures will be used to handle missing data. The model parameters are simultaneously estimated using restricted likelihood estimation incorporating all of the observed data. Estimates have been shown to be unbiased when the missing data are missing at random and when there is ignorable non-random missing data.

5.3.2. Handling Missing Items in Calculating Totals

All total and subscale scores for safety, efficacy, and health outcomes measures will be derived from individual items. If any of the individual items are missing or unknown, every effort will be made to obtain the score for the missing item or items.

For ADAS-Cog_{13} and ADCS-iADL, if >30% of the items are missing, the total score at that visit will be considered missing. If <30% of the items are missing, the total score will be imputed by prorating the nonmissing items. The imputed number will be rounded up to the nearest integer. For example, if the first item of ADAS-Cog_{13}, "Word-Recall Task," which ranges from a score of 0 through 10 (maximum = 10) is missing, and the second item "Commands," which ranges from a score of 0-5 (maximum = 5) is missing, then the multiplication factor = 85/(85 - [10 + 5]) = 85/70 = 1.21 (85 is the total score for the ADAS-Cog13). Thus, the total score for this example will be the sum of the remaining 11 items multiplied by 1.21.

For all other scales, if any item is missing, any total or sum involving that item will be considered missing.

5.4. Multicenter Studies

This study will be conducted by multiple investigators at multiple sites internationally. A listing including country, investigator site with address, number of subjects enrolled (randomized) by each site and unique subject IDs will be presented.

5.5. Multiple Comparisons/Multiplicity

No adjustments will be made for multiple comparisons.

5.6. Population for Analysis

The Efficacy population is defined as all randomized participants who take at least 1 dose of double-blind study treatment and have at least 1 postdose efficacy measurement, which is used to assess the effectiveness of the study drug regimen; all observed data will be included (including data gathered after a subject discontinues study drug).

In order to assess the study drug treatment effect, data gathered after a subject discontinues study drug will not be included in the analysis of per protocol population.

Additionally, when change from baseline is assessed, subjects will be included in the analysis only if both a baseline and at least 1 valid post-baseline measure are available.

All subjects who received at least 1 dose of randomized study treatment (LY3202626 or placebo) will be included in the safety analysis set. If an LY3202626 dose group is dropped following an interim analysis, subjects switched to the remaining LY3202626 dose and remaining on treatment will be accounted for as a separate group.

For the Lumbar Puncture addendum, the analysis sets will include all those subjects whose eligibility was determined by the respective biomarker, who signed consent for addendum, and who have a baseline and at least 1 valid post-baseline cerebrospinal fluid (CSF) assessment.

The PK analysis set is defined as all subjects in the safety population who have at least 1 post-dose PK assessment. The population PK analyses will be performed using this analysis set.

The primary and secondary efficacy measures will be analyzed using the efficacy population unless otherwise specified. In addition, the primary measure will be analyzed using per-protocol populations to verify the robustness of the results. Summaries and analyses for safety measures will be based on the safety population.

Table LLCF.SAP.1 below defines each of the analysis populations used in this study. Table LLCF.SAP.2 lists the study measures that will be summarized and/or analyzed for each population.

Table LLCF.SAP.1. Analysis Populations for Study I7X-MC-LLCF

Population Name	Description of Population
Enrolled	All participants who have been assigned to treatment (randomized).
Efficacy	All randomized participants who take at least 1 dose of double-blind study treatment and have at least 1 postdose efficacy measurement.
Safety	All randomized participants who take at least 1 dose of double-blind study treatment.
Per-Protocol	All subjects in the Full Efficacy population who also: • had an assessment of the primary endpoint at each scheduled visit completed • had no violations of inclusion/exclusion criteria • had no study dosing algorithm violation (such as if subjects randomized to treatment A were given treatment B or subjects randomized to treatment A never received the assigned study drug) • had no unqualified raters and no raters with substantial scoring errors for the primary measure • was not considered non-compliant with regard to study drug

Table LLCF.SAP.2. Efficacy and Safety Measures Summarized and/or Analyzed for Each Analysis Population

Population Name	Variables Assessed
Enrolled	Listings
Efficacy	Tables, Listings and/or Figures of the following: subject disposition, subject characteristics, pre-existing conditions, significant historical diagnoses ADAS-Cog_{13}, ADCS-iADL (and factor analyses), NPI, FAQ, ECog, BASQID, MoCA, MMSE, plasma Aβ parameters, vMRI parameters, CSF parameters, AV-45 and Tau PET parameters and concomitant medications.
Safety Population	Tables, Listings and/or Figures of the following: Compliance, adverse events, laboratory results, vital signs, weight, ECG, MRI, skin examination, eye examination, and C-SSRS.
Per-Protocol Population	Tables, Listings and/or Figures of 18F-AV-1451 PET SUVr.

Abbreviations: ADAS-Cog_{13} = 13-item Alzheimer's Disease Assessment Scale – Cognitive subscale; ADCS-iADL = Alzheimer's Disease Cooperative Study-Activities of Daily Living inventory; BASQID = Bath Assessment of Subjective Quality of Life in Dementia; CSF = cerebrospinal fluid, C-SSRS = Columbia-Suicide Severity Rating Scale; ECG = electrocardiogram; ECog = Everyday Cognition; FAQ = Functional Activities Questionnaire; MMSE = Mini-Mental State Examination; MoCA = Montreal Cognitive Assessment; MRI = magnetic resonance imaging; NPI = Neuropsychiatric Inventory; PET = positron emission tomography; SUVr = standardized uptake value ratio.

5.7. Subject Disposition

The protocol for this study makes the distinction between subject withdrawal and subject discontinuation of study treatment. Subject withdrawal occurs when a subject will no longer participate in the study (no longer taking study treatment and no longer being assessed). Subject discontinuation of study treatment occurs when a subject agrees to continue being assessed but will no longer take study treatment.

Reasons for discontinuation for all patients will be tabulated for treatment groups, and comparisons between treatment groups will be assessed by Fisher's exact test. Additionally, the median time to all-cause discontinuation will be compared between the treatment groups. The null hypothesis is that the median time to discontinuation estimated using the Kaplan-Meier product limit method will be identical for the LY3202626 dose groups and placebo. The log rank test will be used to test the null hypothesis against the alternate hypothesis that the median time to discontinuation is not the same between the groups.

5.8. Subject Characteristics

Standard baseline characteristics of gender, age, and race will be summarized for all patients. Treatment group comparisons will be made using Fisher's exact test for categorical data and an ANOVA with an independent factor for treatment for continuous data. Diagnosis, illness characteristics, and baseline efficacy and safety measures will be analyzed in a similar manner.

Baseline characteristics will be summarized for the efficacy and per-protocol populations by treatment group and overall. Summaries will include descriptive statistics for continuous and categorical measures. Fisher's exact test or Pearson's chi-square test will be used for treatment-group comparisons of categorical data. For continuous data, analysis of variance (ANOVA),

with independent factors for treatment, will be used. Subject characteristics to be presented include:

- Baseline AV-45 SUVr
- Baseline Tau PET SUVr
- age
- age group: 55-64, 65-74, ≥75
- gender
- race
- ethnicity
- height
- body weight
- body mass index (weight (kg) / [height (m)]^2)
- region
- tobacco use
- alcohol use
- years of education
- work status
- ApoE4 carrier status (carrier [ε2/ε4, ε3/ε4, ε4/ε4], noncarrier [ε3/ε3, ε2/ε2, ε3/ε2])
- ApoE4 genotype (ε2/ε4, ε3/ε4, ε4/ε4, no ε4)
- having 1 or more first degree relatives with AD
- AChEI and/or memantine use at baseline
- Baseline severity of impairment as measured by ADAS-Cog$_{13}$, ADCS-iADL instrumental (ADCS-iADL) and , MMSE, MoCA, NPI, ECog, BASQID , and FAQ.

Baseline characteristics will also be listed.

5.9. Treatment Compliance

Only subjects who consume at least 80% of the prescribed daily dose during this study will be considered compliant. The percentage of compliant subjects will be compared across treatment groups using Fisher's exact test.

5.10. Concomitant Therapy

Prior medications are defined as those that stop before randomization (Visit 3). Concomitant medications are defined as those being taken on or after randomization (Visit 3). A summary of

concomitant medications will be presented as frequencies and percentages for each treatment group. Fisher's exact test will be used to test for treatment differences between groups.

Medications will be coded using the World Health Organization (WHO) drug dictionary.

5.11. Efficacy Analyses

5.11.1. Primary Outcome and Methodology

The primary objective of this study is to assess the PD effect of 2 different LY3202626 doses (3 mg and 12 mg) in patients with mild AD dementia compared with placebo, measured by change from baseline in 18F-AV-1451 PET SUVrs at Week 52.

The primary statistical analysis is the ANCOVA, which will be used to statistically evaluate change in composite SUVr from baseline at 52 weeks post-dose. The ANCOVA will include the fixed, categorical effects of treatment dose, and the continuous, fixed covariate of baseline Tau PET SUVr and age at baseline.

Change from baseline analyses will be conducted on SUVrs computed from the following 3 composite regions with the bimodal white matter serving as the reference region: MUBADA (primary), Delta_Z, and the frontal lobe region defined as a weighted average (weights being the number of voxels in the region) of SUV counts, averaged bilaterally, in the following Region of Interests (ROIs):

- cingulum anterior
- Inferior frontal gyrus, opercular part
- Inferior frontal gyrus, orbital part
- Inferior frontal gyrus, triangular part
- Middle frontal gyrus
- Middle frontal gyrus, orbital part
- Superior frontal gyrus
- Superior frontal gyrus, medial
- Superior frontal gyrus, orbital part

5.11.2. Key Secondary and Exploratory Outcomes

The additional clinical and outcome measurements listed below will be analyzed separately using a mixed-effect model for repeated measures (MMRM) analysis.

The change from baseline score at each scheduled post-baseline visit (according to the Study Schedule - data collected at unscheduled visits, including early discontinuation visits if not matching the study schedule, will not be included in this analysis) during the treatment period will be the dependent variable. The model will include the fixed effects of treatment, visit

(categorical covariate), treatment-by-visit interaction, baseline age, baseline score (continuous covariate) and baseline-by-visit interaction. Visit will be considered a categorical variable with values equal to the visit numbers at which the scales were assessed. An unstructured covariance matrix will be used to model the within-subject variance-covariance errors. If the unstructured covariance structure matrix results in a lack of convergence, the following tests will be used in sequence: heterogeneous Toeplitz covariance structure, heterogeneous autoregressive covariance structure, heterogeneous compound symmetry covariance structure, and compound symmetry covariance structure. The Kenward-Roger approximation will be used to estimate the denominator degrees of freedom. Significance tests will be based on LSMean using a 1-sided Alpha of 0.10. Analyses will be implemented using SAS PROC MIXED. Longitudinal graphs of the LSMean estimates at each visit from the model will be plotted and the LSMean and 95% confidence intervals (CIs) results will be summarized in a summary table.

The secondary outcomes that will be analyzed are the following:

- Change from baseline in ADAS-Cog_{13}.
- Change from baseline in ADCS-iADL.
- Change from baseline in iADRS.

The exploratory outcomes that will be analyzed are the following:

- Change from baseline in MoCA total scores.
- Change from baseline in FAQ.
- Change from baseline in MMSE total scores.
- Change from baseline in ECog.

The following exploratory outcomes will be analyzed using the ANCOVA model described in Section 5.1:

- Change from baseline in NPI total scores.
- Change from baseline in BASQID total scores.

5.11.3. Subgroup Analyses

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

- Tau PET subgroup (MUBADA SUVr at baseline <1.1, 1.1-1.46, and >1.46)
- Age: 55-64, 65-74, ≥75
- ApoE4 carrier status (carrier [ε2/ε4, ε3/ε4, ε4/ε4], noncarrier [ε3/ε3, ε2/ε2, ε3/ε2])

The primary outcome measure will be modeled using an ANCOVA approach. This general model will include terms for baseline, treatment, subgroup, subgroup-by-treatment interaction.

5.12. Bioanalytical and Pharmacokinetic/Pharmacodynamic Methods

5.12.1. Analysis of Plasma Aβ

To evaluate the change in plasma Aβ analytes (including assayed plasma $A\beta_{1-40}$, $A\beta_{1-42}$, $A\beta_{1-x}$ after treatment, an MMRM will be used to compare percent change from baseline to week 52. Log transformation will be applied to data prior to analysis and result will be back-transformed to get percent change from baseline.

The model for the fixed effects will include terms for the following independent effects: log-transformed baseline plasma Aβ, treatment, visit, treatment-by-visit interaction. The dependent variable is the change from baseline in log transformed plasma Aβ analytes. Visit will be considered a categorical variable with values equal to the visit numbers at which plasma Aβ is assessed. A similar MMRM analysis will be performed on untransformed percent change from baseline in plasma Aβ analytes if the residual analysis suggests the log transformed analysis is inappropriate.

5.12.2. Analysis of CSF Data

Due to very limited number patients participated in the addendum for CSF sample collection, CSF biomarkers (including CSF $A\beta_{1-40}$, CSF $A\beta_{1-42}$, CSF $A\beta_{1-x}$, CSF sAPP alpha, CSF sAPP beta, CSF total tau, and CSF p-tau from lumbar puncture) will be listed by treatment by visit.

5.12.3. Analyses of AV-45 PET Data

The analyses of change from baseline to the post-baseline visit of the composite summary SUVr of AV-45 (amyloid imaging) normalized will be conducted using an ANCOVA model with baseline AV-45 result and baseline age as covariates and treatment as categorical effect.

Change from baseline analysis will also be conducted for SUVr of the following regions (average of right and left) normalized (based on AVID guidelines): cingulum anterior, cingulum posterior, frontal cortex, mesial temporal cortex, occipital cortex, orbitofrontal cortex, parietal cortex, and temporal cortex.

The complete listing of AV-45 parameters obtained is provided in Appendix 4.

5.12.4. Analyses of perfusion PET Data

An additional perfusion-weighted image will be acquired at each florbetapir PET scanning visit, with no additional tracer injection, by utilizing the initial wash-in of florbetapir to the brain. These images will provide a perfusion (or blood flow) map of the brain at each time point. In AD, cerebral perfusion is reduced, especially in temporal and parietal areas, and this pattern of hypoperfusion closely mirrors the hypometabolism pattern observed using 18F-fluorodeoxyglucose (FDG)-PET. As such, the perfusion PET images provide a biomarker of resting brain function. Changes in florbetapir perfusion PET between the baseline and follow-up scans will be compared across treatment groups and to total exposure to LY3202626.

The analysis of change from baseline to the post-baseline visit of the composite summary SUVr of perfusion scan will be performed using an ANCOVA model with baseline biomarker result and baseline age as covariates and treatment as categorical effect.

The analyses of change from baseline will also be conducted for SUVr of each individual Region of Interest (ROI).

5.12.5. Analyses of Volumetric MRI Data

Brain atrophy is one of the signature biological changes associated with AD, with shrinkage of the medial temporal lobe (including the hippocampus) and ventricular enlargement particularly prominent. The volume of different brain structures and measures of atrophy can be accurately quantified from vMRI scans. Magnetic resonance imaging changes in brain volume from baseline to after 52 weeks of treatment (or early discontinuation) will be quantified. Measurements of brain structural changes, including, but not limited to, whole brain, ventricles and hippocampus will be evaluated. Brain structural changes will be compared across treatment groups and to total exposure to LY3202626.

Analyses of the following vMRI parameters will be conducted:

- Right hippocampal volume (mm^3)
- Left hippocampal volume (mm^3)
- Right hippocampal volume (mm^3) + Left hippocampal volume (mm^3)
- Right entorhinal cortex (mm^3)
- Left entorhinal cortex (mm^3)
- Atrophy of Total whole brain volume (cm^3)
- Enlargement of Ventricular volume (cm^3)

To evaluate the changes in vMRI data after treatment, an ANCOVA model will be used to compare change from baseline to 52 weeks. The change from baseline to the endpoint visit will be the dependent variable. The model for the fixed effects will include terms for the following independent effects: baseline vMRI value, age at baseline, and treatment.

In addition, annualized change in imaging biomarkers (vMRI, AV-45 and AV-1451) for each patient will be calculated using the change at the last post-baseline visit. The annualized change will be compared among the treatment groups with the same ANCOVA model described above.

5.13. Safety Analyses

5.13.1. Extent of Exposure

Summary statistics will be provided for the total number of subjects who were compliant by treatment group during the whole study period. The proportion of subjects who were compliant will be compared using Fisher's exact test. Only subjects who consume at least 80% of the prescribed daily dose during this study will be considered compliant.

Summary statistics will be provided for the total number of days of exposure by treatment.

5.13.2. Adverse Events

Treatment-emergent adverse events (TEAEs) will be defined as events that first occurred or worsened after the randomization date (visit 3 date). Should there be insufficient data for AE start date, stop date, and time to make this comparison, the AE will be considered treatment-emergent.

Treatment-emergent adverse events will be coded according to established Medical Dictionary for Regulatory Activities (MedDRA) terms and summarized by MedDRA System Organ Class and Preferred Term.

Summaries of AEs by decreasing frequency of preferred term within system organ class will be provided for the following:

- Preexisting conditions
- TEAEs
- TEAEs by maximum severity
- TEAEs occurring in greater than or equal to 5% of subjects by preferred term
- Serious adverse events (SAEs)
- AEs reported as reason for study treatment discontinuation

These summaries will include number and percentages of subjects with TEAEs. Treatment comparisons will be carried out using Fisher's exact Test. A summary of TEAEs by visit will also be provided.

Preexisting conditions, TEAEs, SAEs, and discontinuations due to AEs will be listed.

5.13.3. Deaths, Other Serious Adverse Events, and Other Notable Adverse Events

An overview of AEs, including the number and percentage of subjects who died, suffered SAEs, discontinued due to AEs and who suffered TEAEs, will be provided. Comparison between treatments will be performed using Fisher's exact Test.

In addition, the proportion of subjects within specific clusters of TEAEs will be summarized and treatment comparisons will be conducted using Fisher's exact Test. Clusters will be created from MEDRA's High Level Group Terms (HGLT) (Table LLCF.SAP.3). Each of the 3 groups with multiple HLGT's will be combined to form one cluster. Additionally, the 2 HLGT's in the skin disorder group will be analyzed separately.

Table LLCF.SAP.3. Adverse Events of Special Interest

AE Groups of Interest	MedDRA Search Strategy
Nervous System Disorders	Neuromuscular Disorders HLGT
Eye Disorders	Retinal disorders SMQ
Skin Disorders	Hypopigmentation disorders HLT
	Pigmentation changes, NEC HLT
	Epidermal and Dermal Conditions HLGT
Liver Disorders	Drug related hepatic disorders – comprehensive search SMQ
Cardiovascular-type events - Arrhythmic	Arrhythmia related investigations, signs and symptoms SMQ
	Cardiac arrhythmia terms (incl. bradyarrhythmias and tachyarrhythmias) SMQ
	TdP/QT prolongation SMQ
Cardiovascular-type events - Ischemic	Ischaemic heart disease SMQ
Cardiovascular-type events – Stroke	Central nervous system vascular disorders SMQ
Vascular Disorders	Decreased and Nonspecific Blood Pressure Disorders and Shock HLGT

Abbreviations: AE = adverse event; HLGT = high-level group term; HLT = high-level term; MedDRA = Medical Dictionary for Regulatory Activities; NEC = not elsewhere classified; SMQ = Standardized MedDRA Query' TdP = Torsades de Pointes.

5.13.4. Clinical Laboratory Evaluation

Laboratory measurements will be analyzed using continuous data (change from baseline) and categorical or ordinal data (proportion of treatment-emergent abnormalities).

If there are multiple records of laboratory measurements at baseline or postbaseline visit, the last record will be used.

Summaries and analyses of continuous data (change from baseline) will be performed using both conventional and International System of Units (SI units).

Change from baseline to each post-baseline visit at which laboratory measurements are taken and change from baseline to the last observation will be summarized and compared between treatment groups using an ANCOVA model. For each lab analyte, the rank-transformation will be applied to the change from baseline for all subjects and all visits prior to analysis. Similarly, an independent rank-transformation will be applied to the baseline values prior to analysis. The model for the fixed effects will include terms for baseline value, treatment and age. This analysis will be done separately for each laboratory analyte.

Treatment differences in the proportion of subjects with treatment-emergent high or treatment-emergent low or treatment-emergent abnormal laboratory values at (1) anytime and (2) each post-baseline visit will be assessed using Fisher's exact test. Treatment-emergent high or low laboratory abnormality will be based on SI unit. For each laboratory analyte, only subjects who were low or normal at baseline and have at least one post-baseline measurement will be included in the denominator when computing the proportion of subject with treatment-emergent high values. Similarly, only subjects who were high or normal at baseline and have at least one post-

baseline measurement will be included in the denominator when computing the proportion of subject with treatment-emergent low values. In addition, treatment differences in the proportion of subjects who were normal at baseline and change to abnormal high or abnormal low values at any post-baseline visit will be summarized.

For urinalysis parameters, baseline to post-baseline shifts will be summarized at each visit. Likelihood ratio chi-square tests will be used to compare increase, no change and decrease shifts in urinalysis parameters between treatment groups at each visit.

For all laboratory analytes, frequencies of subjects with notable changes (that is, increases or decreases of a prespecified amount unique to each analyte) from baseline to each postbaseline visit were also summarized for all subjects and stratified by low, normal, or high at baseline.

The proportion of subjects with treatment-emergent clinically significant changes from a low value or normal value at all baseline at any time in alanine aminotransferase (ALT), alkaline phosphatase (ALP), and total bilirubin will be summarized by treatment group. Abnormal criteria for treatment-emergent clinically significant changes are presented in Appendix 2. Significant changes of interest for hepatic laboratory values are: ALT ≥3 x upper limit of normal (ULN), total bilirubin ≥2 x ULN, alkaline phosphatase (ALP) >2X ULN, ALT ≥5 x ULN, ALT ≥10 x ULN, and total bilirubin ≥2 x ULN. Additionally, Hy's Law analysis will be conducted by comparing treatment groups with regard to the proportion of patients with (ALT ≥3 x ULN OR aspartate aminotransferase (AST) ≥ 3 x ULN) AND total bilirubin ≥ 2 x ULN. Comparisons between treatment groups will be carried out using Fisher's exact test.

5.13.5. Vital Signs and Other Physical Findings

Vital sign measurements and weight will be analyzed using continuous data (change from baseline) and categorical data (proportion of potentially clinically significant changes).

If there are multiple records of vital sign or weight measurements at baseline or postbaseline visit, the last record will be used. Summary statistics will be presented for observed values at baseline and for change from baseline results at each scheduled postbaseline visit. Systolic and diastolic blood pressure and pulse (collected in sitting position), orthostatic diastolic and orthostatic systolic blood pressures and orthostatic pulse (measurement after 5 minutes in the supine position minus that after 3 minutes in the standing position), temperature, and weight by treatment group for all subjects in the safety population will be summarized.

With the large number of the scheduled vital sign visits, the MMRM model is not suitable for the change from baseline comparison of treatments due to computational challenge. Change from baseline to each post-baseline visit at which vital signs are taken will be assessed using an ANCOVA model with treatment and as independent factors and baseline value and age as covariates in the model. This analysis will be done separately for each vital sign parameter and weight.

The incidence of treatment-emergent abnormal high or low vital signs and weight will be presented by treatment group and visit. Treatment-emergent vital sign evaluations are defined for evaluations collected after the initiation of study medication. Abnormal criteria for

postbaseline vital signs and weight are presented in Appendix 3. Any vital sign or weight meeting the criteria will be considered abnormal. Treatment differences in the proportion of subjects with treatment-emergent abnormal high or low vital signs and weight will be assessed between treatment groups using Fisher's exact test at (1) any time (2) post-baseline visit.

For each vital sign at each post-baseline visit, only subjects who had a baseline result and had a nonmissing result at that post-baseline visit will be included in the denominator when computing the proportion of subjects with treatment-emergent high, low, or abnormal values.

Summary and analyses of change from baseline in weight will be provided. The proportion of subjects with a weight gain or loss of greater than or equal to 7 percent of baseline body weight will be compared between treatment groups using Fisher's exact test at each visit and at any time. For those subjects exhibiting a weight loss of greater than or equal to 7 percent of baseline body weight, body mass index (BMI) will be tabulated by treatment and compared using an ANOVA model with factors for treatment. In addition, categories of BMI (underweight = <18.5, normal = ≥18.5 - <25, overweight = ≥25 to <30, obese = ≥30) will be compared across treatment groups using Pearson's chi-square test at each visit and at any time.

A listing of treatment-emergent abnormal vital signs and weight will also be presented by subject and visit.

5.13.6. Electrocardiograms

Electrocardiogram measurements will be analyzed using continuous data (change from baseline) and categorical data (proportion of treatment-emergent abnormalities).

Since ECG is measured in triplicates, the average of triplicates will be used at baseline and each visit. If there are multiple records after averaging ECG triplicates within a visit, the last record of averages will be used.

The analysis will be done for the following ECG measurements: heart rate, PR, QT, QTc, and RR intervals and QRS duration. All analyses of QTc will be carried out using the Fridericia correction (QTcF) method. These summaries will include data from each visit ECG measures are performed. Change from baseline to each post-baseline visit at which ECG measurements are taken and change from baseline to last observation will be assessed using an ANCOVA model. The model for the fixed effects will include terms for the following independent effects: baseline age, baseline ECG score, and treatment. This analysis will be done separately for each ECG parameter.

Incidence of treatment-emergent abnormal ECGs will be assessed by comparisons at (1) anytime and (2) each post-baseline visit between treatment groups with Fisher's exact test. For analyses of treatment-emergent abnormal ECGs, baseline will be considered as all visits before the initiation of drug dose.

Abnormal ECG criteria and criteria for abnormal QTcF prolongation are presented in Appendix 1.

Treatment-emergent high ECG parameters (heart rate, PR interval, QRS duration, QT and QTcF intervals) are the values which are low or normal at all baseline visits and fall into the high

abnormal categories post-baseline. Similarly, treatment-emergent low ECG parameters (Heart Rate, PR interval, QRS duration) are the values which are high or normal at all baseline visits and fall into the low abnormal categories above.

In additional, treatment differences in the proportion of subjects who have normal baselines with a change to abnormal high or abnormal low values at any post-baseline visits will be summarized.

5.13.7. Analyses of MRI Data

Magnetic resonance imaging findings believed to have potential clinical significance will be reported to the investigator and sponsor. Any new or aggravated clinically relevant abnormal MRI finding as compared with the baseline MRI should be reported to Lilly or its designee as an AE. If a clinically meaningful change in MRI is noted during the study, an additional, full neurological exam will be performed as soon as possible, along with any other medical follow-up deemed necessary by the investigator.

To evaluate any changes in MRI data following treatment, Pearson's chi-square tests will be used to compare frequencies of responses in the MRI parameters.

Frequencies and percentages of the following amyloid-related imaging abnormality – edema (ARIA-E, also known as vasogenic edema) and ARIA – hemorrhage (ARIA-H, also known as microhemorrhage) parameters will be summarized:

- ARIA-E:
    - Severity (mild, moderate, severe, or no presence)
    - Status compared to the previous MRI(s) (increased, unchanged, partial resolution, or complete resolution)
- ARIA-H:
    - Number of ARIA-H (1, 2 to 5, 6 to 10, >10, or no presence),
    - Baseline to endpoint changes (increase in size of pre-existing ARIA-H, increase in number of ARIA-H, no change, partial resolution, or complete resolution)

A listing of abnormal MRI finding will also be presented by subject and visit.

5.13.8. Additional Safety Concerns

5.13.8.1. Columbia Suicide Severity Rating Scale

Suicidal ideation, suicidal behavior, and self-injurious behavior without suicidal intent, based on the C-SSRS, will be listed by subject and visit. Only subjects that show suicidal ideation/behavior or self-injurious behavior without suicidal intent will be displayed (i.e., if a subject answers are all 'no' for the C-SSRS, then that subject will not be displayed). However, if a subject reported any suicidal ideation/ behavior or self-injurious behavior without suicidal intent at any time point then all their ideation and behavior will be displayed, even if not positive.

In addition, the number and percent of patients who experienced at least one of various comparative measures during treatment will be presented and compared. These include treatment-emergent suicidal ideation compared to recent history, treatment-emergent serious suicidal ideation compared to recent history, emergence of serious suicidal ideation compared to recent history, improvement in suicidal ideation at endpoint compared to baseline, and emergence of suicidal behavior compared to all prior history.

5.13.8.2. Skin Examination

Skin color will be reported at baseline by using Fitzpatrick Scale Rating. The frequencies of the outcomes (I, II, III, IV, V, and VI) will be displayed by treatment group.

Any hypopigmentation will be assessed by location, percentage of body surface area involvement (BSA), degree (partial/decreased pigmentation to complete depigmentation), and other findings in or around the hypopigmentation area (e.g., redness or induration). A static physician's global assessment (sPGA) will be used to determine the subject's overall hypopigmentation severity at a given timepoint using a visual analog scale (VAS) ranging from 0 to 100. In addition, subjects noted to have evidence of hypopigmentation will be asked to record how bothersome they find the hypopigmentation on a VAS ranging from 0 to 100.

Frequency tables and summary statistics for the continuous parameters will be given by treatment group.

In order to display any changes/deteriorations during treatment, the following will be reported: number of subjects for whom no hypopigmentation was observed at baseline, but for whom at least once after randomization hypopigmentation was observed during treatment period; summary statistics for the difference of maximum value during treatment minus baseline value for percentage BSA of hypopigmentation, overall severity (sPGA) and "how bothered is the subject"; shift table of baseline vs. maximum value during treatment for degree of overall lesion severity.

5.13.8.3. Eye Examination

Summary statistics will be produced for the following continuous parameters: thickness and volume of eye retinal thickness grid for each left and right eye: left eye total visual acuity score, right eye total visual acuity score (scores expressed as logMAR calculated as the negative log (base 10) of the decimal scores), as well as left eye intraocular pressure and right eye intraocular pressure (both in mmHg). In addition, summary statistics for the difference of maximum value during treatment minus baseline value for left eye total visual acuity score, right eye total visual acuity score, left eye intraocular pressure (mmHg), and right eye intraocular pressure (mmHg) will also be produced.

In order to assess any changes/deteriorations during treatment, number and proportion of subjects with any treatment-emergent abnormal finding (in slit lamp examination, dilated fundus examination, color vision examination, fundus photography and ocular coherence tomograph) will be summarized by visit and at any time during treatment period, and compared among treatment groups using Fisher exact test. Treatment-emergent abnormal findings include any

"new" or "worsened" findings in slit lamp, dilated fundus, and color vision exams, and any severity category increase in fundus photography and ocular coherence tomograph parameters at any post-baseline visits. Potentially clinically significant findings in slit lamp, dilated fundus, and color vision exams will also be summarized separately.

5.14. Safety Follow-up Visit

Follow-up visit (Week 56) is not required for patients who discontinue study treatment early and who continue in the study unless study drug was discontinued within the last 4 weeks of the study.

Adverse events, concomitant medications, vital signs, and ECGs will be collected at these visits. These data will be summarized by treatment received during placebo-controlled treatment period.

5.15. Protocol Violations

Listings of subjects with significant protocol violations will be provided for the Full Efficacy population. The following list of significant protocol violations will be determined from the clinical database and from the clinical/medical group:

- Did not have an assessment of either the primary endpoint at any of the visits at which the scales were scheduled to be assessed.
- Protocol violations of inclusion/exclusion criteria.
- Had a study dosing algorithm violation (such as if subjects randomized to treatment A were given treatment B or subjects randomized to treatment A never received the assigned study drug.)
- Had unqualified raters or raters with substantial scoring errors for the co-primary measures.
- Was not compliant with regard to study drug.

The analyses for the primary objective will be repeated for the per protocol subject population.

5.16. Interim Analyses and Data Monitoring

An assessment committee (AC) will have the responsibility to review accumulating study data at a pre-specified interim analysis and make recommendations to protect the safety of subjects. The approved AC charter enumerates the roles of the AC members and the structure of their meetings. A statistical analysis plan (SAP) for analyses associated with the AC list out the specific analyses that the AC will review.

One scheduled interim analyses will be conducted during this study. The AC charter will specify the final details of the interim analyses. Only the AC is authorized to evaluate unblinded interim safety analyses. Study sites will receive information about interim results only if they need to know for the safety of their subjects.

5.16.1. Interim Analysis

The objective of the interim analysis (IA) is to conduct a safety analysis. One safety interim analysis conducted by the AC is planned when data from approximately 105 qualified patients at 12 weeks and 75 patients through 24 weeks have been obtained. The AC may be asked to conduct additional safety interim analyses if safety concerns are observed during blinded safety monitoring. In the event that blinded safety monitoring uncovers an issue that needs to be addressed by unblinding at the group level, only members of the AC can conduct additional analyses of the safety data.

A limited number of pre-identified individuals may gain access to the unblinded PK/PD data, as specified in the unblinding plan, prior to the interim or final database lock. Information that may unblind the study during the analyses will not be reported to study sites or blinded study team until the study has been unblinded.

5.17. Clinical Trial Registry Analyses

Analyses provided for the Clinical Trial Registry requirements will include the following:

Summary of AEs, provided as a dataset which will be converted to an XML file. Both SAEs and 'Other' AEs are summarized: by treatment group, by MedDRA preferred term.

- An AE is considered 'Serious' whether or not it is a TEAE.
- An AE is considered in the 'Other' category if it is both a TEAE and is not serious. For each SAE and 'Other' AE, for each term and treatment group, the following are provided:
    - the number of participants at risk of an event
    - the number of participants who experienced each event term
    - the number of events experienced.
- Consistent with www.ClinicalTrials.gov requirements, 'Other' AEs that occur in fewer than 5% of subjects in every treatment group may not be included if a 5% threshold is chosen (5% is the minimum threshold).

6. Appendices

Appendix 1. Potentially Clinically Significant ECGs

Parameter	Unit	Low PCS Criteria	High PCS Criteria
QRS Interval	msec	NA	≥ 120
PR Interval	msec	NA	≥ 220
QTcF	msec	NA	> 500
QTc interval: change from baseline	> 60 msec at any time after randomization		

Abbreviations: NA = not applicable; PCS = potentially clinically significant; QTc = corrected QT interval; QTcF = Fridericia's corrected QT interval.

Appendix 2. Potentially Clinically Significant Laboratory Values

Parameter	Conventional Unit	Low PCS Criteria	High PCS Criteria	SI Unit	Low PCS Criteria	High PCS Criteria
			Hematology (whole blood)			
Hemoglobin (male)	g/dL	<11	>18	gL	110	180
Hemoglobin (female)	g/dL	<10	>17	gL	100	170
Hematocrit	%	<30	>50 (F) >55 (M)	L/L	<0.3	>50 (F) >55 (M)
Leukocyte (WBC Count)	X 10^3/uL	≤2.8	≥15	X 10^9/L	≤2.8	≥15
Neutrophils	Cells/mm^3	1500	NA	Proportion of 1.0	≤0.15	NA
Platelet Count	X 10^3/uL	≤75	≥700	10^9/L	≤75	≥700
			Chemistry (serum or plasma)			
ALT (SGPT)	U/L	NA	≥3 X ULN	U/L	NA	≥3 X ULN
AST (SGOT)	U/L	NA	≥3 X ULN	U/L	NA	≥3 X ULN
Total Bilirubin	mg/dL	NA	≥1.5 X ULN	umol/L	NA	≥1.5 X ULN
BUN	mg/dL	NA	≥30	mmol/L	NA	≥10.71
Creatinine Kinase (CK)	U/L	NA	≥3 X ULN	U/L	NA	≥3 X ULN
Sodium	mEq/L	≤125	≥155	mmol/L	≤125	≥155
Potassium	mEq/L	≤3.0	≥5.5	mmol/L	≤3.0	≥5.5
Calcium	mg/dL	Corrected Value ≤7.5	Corrected value ≥12.0	mmol/L	≤1.875	≥3
Alkaline Phosphatase	U/L	NA	≥3 X ULN	U/L	NA	≥3 X ULN
Albumin	g/dL	≤2.6	≥6.0	g/L	≤26	≥60

Parameter	Conventional Unit	Low PCS Criteria	High PCS Criteria	SI Unit	Low PCS Criteria	High PCS Criteria
Chloride	mEq/L	≤85	≥120	mmol/L	≤85	≥120
Glucose (random)	mg/dL	≤45.1	≥200.0	mmol/L	≤2.48	≥11
Serum Creatinine	mg/dL	NA	>1.5 X ULN	umol/L	NA	>1.5 X ULN
Gamma-Glutamyl Transferase (GGT)	U/L	NA	≥3 X ULN	U/L	NA	≥3 X ULN
TSH	uU/L	below normal range	above normal range	mIU/L	below normal range	above normal range
			Urinalysis			
Hb/RBCs/Blood					NA	≥ + 2
Protein/Albumin					NA	≥ + 2
Glucose					NA	≥ + 2

Abbreviations: ALT = alanine aminotransferase; AST = aspartate aminotransferase; F = female; Hb = hemoglobin; M = male; NA = not applicable; PCS = potentially clinically significant; RBCs = red blood cells; SGOT = serum glutamic oxaloacetic transaminase; SGPT = serum glutamic pyruvic transaminase; SI = Système International; TSH = thyroid-stimulating hormone; ULN = upper limit of normal; WBC = white blood cell.

Appendix 3. Criteria of Potential Clinical Significant Vital Signs and Weight

Parameter	Unit	Observed Value	And/Or	Change Relative to Baseline	Change from Supine to Standing
Systolic Blood Pressure (supine or sitting)	mmHg	≥180	And	Increase of ≥20	-
		≤90	And	Decrease of ≥20	-
Diastolic Blood Pressure (supine or sitting)	mmHg	≥105	And	Increase of ≥15	-
		≤50	And	Decrease of ≥15	-
Systolic Blood Pressure (supine to standing)	mmHg	≤80	And	-	Decrease of ≥20
Pulse (supine or sitting)	bpm	≥120	And	Increase of ≥15	-
		≤50	And	Decrease of ≥15	-
Orthostatic hypotension (Systolic)	Decrease in systolic blood pressure when going from 5 minutes supine to 3 minutes standing of ≥20 mm Hg or the inability to stand quickly for the measurements due to symptoms of orthostasis				
Orthostatic hypotension (Diastolic)	Decrease in diastolic blood pressure when going from 5 minutes supine to 3 minutes standing of ≥10 mm Hg or the inability to stand quickly for the measurements due to symptoms of orthostasis				
Orthostatic pulse (Tachycardia)	Increase in heart rate when going from 5 minutes supine to 3 minutes standing of ≥30 bpm				
Weight		Not Applicable		Increase of ≥7% Decrease of ≥7%	

Appendix 4. Complete List of AV-45 (Florbetapir) Parameters

SUVr will be obtained for the regions listed below normalized (based on AVID guidelines) and to cerebellar gray matter:

composite summary	lateral temporal cortex left	pons
caudate left	lateral temporal cortex right	putamen right
caudate right	mean cerebellum gray matter	putamen left
cerebellar cortex left	mean whole cerebellum	rectus left
cerebellar cortex right	mesial temporal cortex left	rectus right
cerebellar white matter	mesial temporal cortex right	subcortical white matter
cingulum anterior left	occipital cortex left	temporal cortex left
cingulum anterior right	occipital cortex right	temporal cortex right
cingulum posterior left	orbitofrontal cortex left	thalamus left
cingulum posterior right	orbitofrontal cortex right	thalamus right
frontal cortex left	parietal cortex left	
frontal cortex right	parietal cortex right	