I5T_MC_AACN Statistical Analysis Plan Version 1

A Phase 3, Open-Label, Parallel-Group, 2-Arm Study to Investigate Amyloid Plaque Clearance with Donanemab Compared with Aducanumab-avwa in Participants with Early Symptomatic Alzheimer's Disease

NCT05108922

Approval Date: 24-Sep-2021

Statistical Analysis Plan:

Protocol Title: A Phase 3, Open-Label, Parallel-Group, 2-Arm Study to Investigate Amyloid Plaque Clearance with Donanemab Compared with Aducanumab-avwa in Participants with Early Symptomatic Alzheimer's Disease

Protocol Number: I5T-MC-AACN

Compound Number: LY3002813

Short Title: A Study to Investigate Amyloid Plaque Clearance with Donanemab Compared with Aducanumab-avwa in Participants with Early Symptomatic Alzheimer's Disease

Acronym: TRAILBLAZER-ALZ-4

Sponsor Name: Eli Lilly and Company

Legal Registered Address: Indianapolis, Indiana, USA 46285

Regulatory Agency Identifier Number(s): IND 109157

Registry	ID
ClinicalTrials.gov	TBD

Approval Date: 24-Sep-2021 GMT

Table of Contents

Version history

This Statistical Analysis Plan (SAP) for study AACN is based on the protocol dated 17AUG2021.

Table 1 SAP Version History Summary

SAP Version	Approval Date	Change	Rationale
1		Not Applicable	Original version

1. Introduction

Study I5T-MC-AACN (AACN) is a Phase 3, randomized, open-label, head-to-head study to evaluate the superiority of donanemab to aducanumab in reaching amyloid plaque clearance in participants with early symptomatic AD.

The duration of the treatment period of the study is 76 weeks and includes up to 72 weeks of treatment. Florbetapir F18 PET scans are collected at baseline and throughout the treatment period (Weeks 24, 52, and 76) to assess the effect of donanemab and aducanumab on amyloid plaque removal, which is a known hallmark pathology of AD, and hypothesized to contribute to the cognitive and functional decline in people with AD.

1.1. Objectives, Endpoints, and Estimands

Objectives	Endpoints
Co-Primary	
• To assess the superiority of donanemab versus aducanumab on complete brain amyloid plaque clearance in participants with early symptomatic AD.	• Superiority of percentage of participants who reach complete amyloid plaque clearance on florbetapir F18 PET scan on donanemab versus aducanumab at 6 months.
• To assess the superiority of donanemab versus aducanumab on complete brain amyloid plaque clearance in the intermediate tau subpopulation of participants with early symptomatic AD.	• Superiority of percentage of participants who reach complete amyloid plaque clearance on florbetapir F18 PET scan in the intermediate tau subpopulation on donanemab versus aducanumab at 6 months.
Key Secondary	
• To assess the superiority of donanemab versus aducanumab on degree of brain amyloid plaque reduction.	Superiority of mean absolute change from baseline in brain amyloid plaque on florbetapir F18 PET scan for donanemab versus aducanumab at: • 6 months • 12 months, and • 18 months.

• To assess the superiority of donanemab versus aducanumab in time to reach complete amyloid plaque clearance.	• Time to reach complete amyloid plaque clearance on donanemab versus aducanumab.
Other Secondary	
• To assess the superiority of donanemab versus aducanumab on degree of brain amyloid plaque reduction.	Superiority of mean percent change from baseline in brain amyloid plaque on florbetapir F18 PET scan for donanemab versus aducanumab at: • 6 months • 12 months, and • 18 months.
• To assess the non-inferiority of donanemab versus aducanumab on degree of brain amyloid plaque reduction.	Non-inferiority of mean absolute change from baseline in brain amyloid plaque on florbetapir F18 PET scan at: • donanemab 6 months vs aducanumab 12 months, and • donanemab 6 months vs aducanumab 18 months.
• To assess the superiority of donanemab versus aducanumab on complete brain amyloid plaque clearance.	Superiority of percentage of participants who reach complete amyloid plaque clearance on florbetapir F18 PET scan on donanemab versus aducanumab at: • 12 months, and • 18 months.
• To assess the superiority of donanemab versus aducanumab on complete brain	Superiority of percentage of participants who reach complete amyloid plaque clearance on

amyloid plaque clearance in the intermediate tau subpopulation.	florbetapir F18 PET scan in the intermediate tau subpopulation on donanemab versus aducanumab at: • 12 months, and • 18 months.
• To assess the superiority of donanemab versus aducanumab on degree of brain amyloid plaque reduction in the intermediate tau subpopulation.	Superiority of mean absolute change from baseline in brain amyloid plaque on florbetapir F18 PET scan in the intermediate tau subpopulation for donanemab vs aducanumab at: • 6 months • 12 months, and • 18 months.
• To assess the superiority of donanemab versus aducanumab on degree of brain amyloid plaque reduction.	• Superiority of change from baseline in brain amyloid plaque reduction on florbetapir F18 PET scan for donanemab 6 months vs aducanumab 12 months.
• To evaluate the safety and tolerability of donanemab vs aducanumab.	Characterization of standard safety assessments: • Spontaneously reported AEs • Clinical laboratory tests • MRI (ARIA and emergent radiological findings) • Serious hypersensitivity reactions • C-SSRS

Coprimary estimands

The primary clinical question of interest is: What is the difference in brain amyloid plaque clearance after 6 months of intervention in participants with early symptomatic Alzheimer's Disease regardless of discontinuation for any reason and regardless of initiation of rescue intervention or change in background intervention in the entire study population regardless of tau level. Additionally, the same clinical question will be addressed in participants with early symptomatic Alzheimer's Disease and intermediate levels of tau proteins in the brain as measured by flortaucipir F18 PET scans.

The estimand is described by the following attributes:

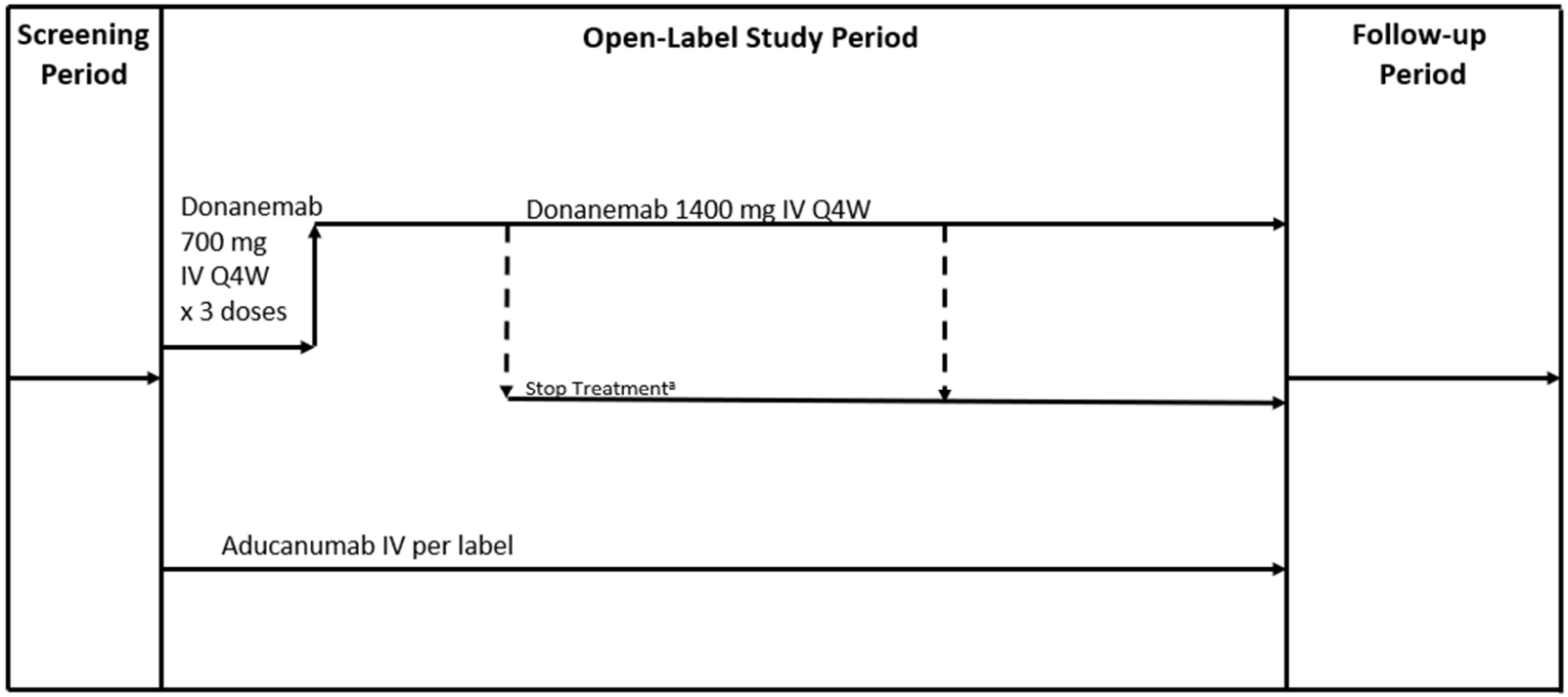
Population: participants with early symptomatic Alzheimer's Disease.

Endpoint: Proportion of participants with complete brain amyloid plaque clearance as measured by florbetapir F18 PET scans

Treatment condition: the randomized treatment with or without change in background medication (treatment policy strategy).

Population-level summary: estimated odds of complete amyloid plaque clearance for donanemab versus aducanumab (odds ratio).

1.2. Study Design



Abbreviations: IV = intravenous; mg = milligrams; Q4W = every 4 weeks.

Note: V601 is optional. For participants who do not complete V601, procedures will be included in V1. Randomization occurs at V2.

[a] Participants whose amyloid plaque reduction as measured by florbetapir F18 PET scans at Visit 8 (Week 24) or Visit 15 (Week 52) meets criteria will stop donanemab infusions but continue all other assessments for the remaining duration of the open-label period and short-term follow-up period. These dose reduction rules are defined by the sponsor

2. Statistical Hypotheses

The primary objective is to demonstrate that donanemab is superior to aducanumab in achieving complete brain amyloid clearance at 6 months. Thus, the null hypothesis to be tested in relation to the primary estimand is as follows:

- Null hypothesis: Donanemab is not different from aducanumab with respect to the percentage of participants achieving complete amyloid plaque clearance at 6 months.

The same null hypothesis will be applied to the entire study population and to the subpopulation with intermediate tau levels (co-primary objectives). Refer to Section 3 for definitions of the analysis sets.

2.1. Multiplicity Adjustment

The statistical comparisons for the co-primary efficacy endpoints and the key secondary endpoints will be carried out such that strong control of Type I error will be achieved at a 2-sided level of 5%. Since both co-primary endpoints need to be met for the study to be successful, each of the co-primary endpoints will be tested at the full 2-sided significance level of 0.05. If both these co-primary endpoints are statistically significant at the 5% level, the key secondary endpoints will be tested in a gated manner in the following order at the 5% level:

i) the mean change from baseline in amyloid plaque levels at 6 months
ii) the mean change from baseline in amyloid plaque levels at 12 months
iii) time to amyloid plaque clearance tested at the end of the study
iv) the mean change from baseline in amyloid plaque levels at 18 months

When data for 6 months PET scans are available, the co-primary objectives will be assessed. If both these null hypotheses for the co-primary objectives are rejected, the mean change from baseline in amyloid plaque levels at 6 months will be compared between donanemab and aducanumab. If the null hypothesis is rejected for the 6-month comparison, the mean change comparison between donanemab and aducanumab will be assessed when 12-month PET scans are available. The next two key secondary endpoints (in order) are time to amyloid plaque clearance, and mean change from baseline in amyloid plaque levels at 18 months, which will be assessed at the end of the study when data for all time points are available.

The testing scheme will follow this gating pattern until all four key secondary hypotheses are tested or until the null hypothesis is not rejected for an endpoint, at which point, any further testing would stop for the key secondary objectives.

If an interim analysis is performed, the alpha spending approach specified in Section 4.9 will be followed to control the overall Type I error rate for the co-primaries under a two-sided level of 5%. The testing approach for the key secondary endpoints should an interim be performed will be similar as described in this section. However, the alpha available for testing the key secondary endpoints in the event an interim analysis is performed is described in Section 4.9.

3. Analysis Sets

For the purposes of analysis, the following analysis sets are defined:

Participant Analysis Set	Description
Full analysis set / ITT analysis set	• All randomized participants with baseline and post-baseline florbetapir F18 PET scan results
Intermediate tau analysis set	• All randomized participants with baseline and post-baseline florbetapir F18 PET scan results and a baseline flortaucipir F18 PET scan meeting the intermediate tau criteria
Safety analysis set	• All participants who are exposed to study intervention. Participants will be included in the analyses according to the intervention they actually received.

The full analysis set and the intermediate tau analysis set are used to analyze endpoints related to the biomarker objectives, and the safety analysis set is used to analyze the endpoints and assessments related to safety.

4. Statistical Analyses

4.1. General Considerations

Statistical analysis of this study will be the responsibility of the Sponsor or its designee.

Unless otherwise noted, tests of treatment effects will be conducted at a 2-sided alpha level of 0.05 (or equivalently, a 1-sided 0.025 alpha level); 2-sided CIs will be displayed with a 95% confidence level. If relevant, tests of interactions between treatment and other factors will be conducted at an alpha level of 0.05.

Participant assignment to treatment will be stratified by ApoE ε4 genotype, and baseline amyloid level. These factors will be included in analysis models unless otherwise stated.

All amyloid plaque analyses will follow the ITT principle unless otherwise specified. An ITT analysis is an analysis of data by the groups to which participants are assigned by random allocation, even if the participant 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, participants will be included in the analysis only if both a baseline and a postbaseline measure are available. Unless otherwise defined, a baseline measure is the last non-missing observation collected prior to the first administration of study medications. Endpoint is the last non-missing post-baseline measurement.

To allow for sequential testing, database locks and the testing of corresponding primary and key secondary objectives are expected to occur after all randomized participants have had a chance to complete the scheduled 24-week PET scan, the scheduled 52-week PET scan, and the scheduled 76-week PET scan. If an interim analysis occurs, a database lock will additionally occur after a sufficient number of participants (for example, approximately 75-125) have had a chance to complete the scheduled 24-week PET scan. Data collected during the immunogenicity and safety follow-up period will be summarized and analyzed separately.

Any change to the data analysis methods described in the protocol will require a protocol amendment 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 SAP and the clinical study report. Additional exploratory analyses of the data will be conducted as deemed appropriate.

4.2. Participant Dispositions

Reason(s) for study discontinuation will be summarized by treatment group for all randomized participants. The percentage of participants discontinuing from each treatment group will be compared between groups using Fisher's exact test. The comparisons will be done for the overall percentage of participants who discontinue and for select specific reasons for discontinuation. The reason(s) will be collected and will be summarized by treatment group for all randomized participants. The percentage of participants discontinuing from each treatment group will be compared between groups using Fisher's exact test. The comparisons will be done for the overall percentage of participants who discontinue and for select specific reasons for discontinuation.

4.3. Patient Characteristics

Baseline characteristics will be summarized for the randomized population by treatment group and overall. Summaries will include descriptive statistics for continuous and categorical measures. Patient characteristics to be presented include:

- age
- gender
- race
- ethnicity
- height
- body weight
- body mass index (weight (kg) / [height (m)]2)
- tau PET burden (various measures)
- 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)
- AChEI and/or memantine use at baseline
- Baseline CDR
- Baseline MMSE
- Baseline Amyloid PET centiloids

4.4. Co-Primary Endpoints Analysis

The primary objectives of this study are to test the hypothesis that participants treated with donanemab will have a significantly greater proportion of participants who reach amyloid plaque clearance compared to participants treated with aducanumab at the 6-month timepoint in both the ITT population and the intermediate tau subpopulation.

As the 6-month timepoint is the first scheduled opportunity to collect a post-baseline florbetapir F18 PET scan, the comparison of donanemab to aducanumab in complete brain amyloid clearance at 6 months will be performed using logistic regression. The logistic regression model will include fixed effect for treatment, ApoE ε4 genotype, baseline amyloid level, and baseline age. The test for this endpoint will be based on the contrast for the main effect of treatment from this logistic regression model. This analysis will be conducted on the full ITT population and the intermediate tau subpopulation. In case of quasi-complete separation of data points which may be possible due to small sample sizes, a Firth's penalized logistic regression model will be used to make inference on treatment efficacy.

Treatment effects from these logistic regression models will be summarized using estimated probabilities within each treatment group, along with odds ratios and a 100*(1-α)% confidence intervals, where α is the two-sided significance level for the primary endpoints.

4.4.1. Definition of endpoint(s)

Complete brain amyloid plaque clearance is a binary outcome and is defined as a centiloid value <24.1 from the florbetapir F18 PET scan.

4.4.2. Main analytical approach

As described above, the co-primary endpoints will be analyzed using logistic regression with the specified models for the full analysis set and for the intermediate tau analysis set. A missing 6-month PET scan will result in the participant being excluded from the analyses.

4.4.3. Sensitivity Analyses

A last-observation-carried-forward (LOCF) analysis for complete amyloid plaque clearance will be conducted. For participants without PET scan results at either 12 or 18 months, the centiloid measurement from their last available post-baseline PET scan will be used to determine complete amyloid plaque clearance for the missing assessments.

4.5. Key Secondary Endpoints Analysis

The mean change from baseline in amyloid plaque levels at months 6 and 12, the time to complete amyloid plaque clearance, and mean change from baseline in amyloid plaque levels at month 18 are the four key secondary endpoints. It is expected that these endpoints will be analyzed sequentially as data for them are available. That is, for any efficacy analyses for 6-month/12-month/18-months endpoints, data will be available at baseline and for all timepoints up to and including the time point in question.

4.5.1. Key Secondary Endpoints

4.5.1.1. Definition of endpoint(s)

For the 6-, 12-, and 18-month mean change analyses, change from baseline will be defined as the specific time point PET centiloid value minus the baseline PET centiloid value. For the time to complete amyloid plaque clearance analysis, participants will be categorized as either meeting criterion for complete amyloid plaque clearance based on florbetapir F18 PET centiloid values or not meeting this criterion. Time to event will be defined as the difference between the date of the PET scan which shows complete amyloid plaque clearance and the date of the first dose of study drug. Participants who do not meet this criterion will be censored at the time of their final PET scan.

4.5.1.2. Main analytical approach

As there is only one post-baseline PET assessment within the first 6 months of the trial, mean change from baseline in amyloid plaque levels at Month 6 will be analyzed using an analysis of covariance model with the treatment group as fixed effect variable of interest, and APOE e4 genotype, baseline amyloid levels, and baseline age as the covariates. For the analysis of mean change from baseline at 12 and 18 months, a separate mixed model repeated measures (MMRM) analysis with data up to and including that specific time point will be used. This model will include the fixed effects of treatment, time, and treatment by time interaction, and APOE e4 genotype, and baseline amyloid level as covariates. Correlation among repeated measures on the same subject will be accounted for by enforcing an unstructured variance covariance matrix among the residuals. If the unstructured covariance structure matrix results in a lack of convergence, the following structures will be used in sequence until convergence is reached: 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.

Time to amyloid plaque clearance between the two treatment groups will be analyzed and compared using a log-rank test for survival data. Kaplan-Meier estimates of percent of participants reaching amyloid plaque clearance and the p-value from the log-rank test will be used to determine the superiority of donanemab vs aducanumab in lowering time to plaque clearance.

4.5.1.3. Sensitivity Analyses

Observed values for mean changes in centiloid values will be reported for 6-, 12-, and 18- month PET scans. Additionally, a last-observation-carried-forward (LOCF) mean change analysis in centiloid values will be reported.

Time to plaque clearance will also be analyzed using a Cox proportional-hazards model with the treatment group as fixed effect variable of interest, and APOE e4 genotype, baseline amyloid levels, and baseline age as covariates.

4.5.1.4. Other secondary analyses

Mean percent change from baseline in brain amyloid plaque is defined as the specific time point PET centiloid value minus the baseline PET centiloid value, divided by the baseline PET centiloid value. These values will be modelled with the same approach as outlined for the mean change analyses.

Non-inferiority analyses will compare the changes in brain amyloid plaque levels for donanemab at 6 months to the changes in brain amyloid plaque levels for aducanumab at 12 months and at 18 months. The non-inferiority margin of interest is 5 centiloids. The MMRM models specified for the key secondary comparisons at 12 and 18 months will be used for the non-inferiority test, with appropriate contrasts and confidence intervals for the treatment-by-time interaction effects. For these non-inferiority analyses, if the upper bound of the 95% confidence interval for the donanemab vs. aducanumab contrasts is <5 centiloids, donanemab at 6 months will be considered to be non-inferior to aducanumab at that particular time point. An additional secondary objective to test the superiority of donanemab in amyloid plaque reduction at 6 months versus aducanumab at 12 months will also use this MMRM model and contrast.

For both the full study population and the intermediate tau subpopulation, the 12-month and 18-month comparison of proportion of participants with amyloid plaque clearance for donanemab versus aducanumab will include repeated post-baseline amyloid plaque clearance status (Yes/No) measurements and will be analyzed using separate generalized linear mixed model (PROC GLIMMIX) as pseudo-likelihood-based mixed effects repeated measures analysis assuming binomial distribution with a logit link. These models will be analyzed using data for outcomes up to and including the time point in question. The GLIMMIX model will include fixed effects for treatment, month, treatment-by-month interaction, APOE e4 genotype, baseline amyloid level, and baseline age as covariates. An unstructured variance-covariance matrix will be used to model the correlation among repeated measures on the same individual. If model convergence is not reached lesser restrictive variance-covariance structures will be employed until model convergence is attained.

For the mean change from baseline in amyloid plaque levels for the intermediate tau subpopulation, the statistical models will be identical to that of the full population, with analysis of covariance applied to the 6-month data and an MMRM analysis applied to the 12- and 18-month data.

4.6. Exploratory Endpoints Analysis and PK/PD Analyses

Exploratory biomarker samples are being collected throughout the study. These analyses will be performed in the overall study population and in the intermediate tau population. In addition to the analyses specified below, additional analyses of these data will be conducted as appropriate.

4.6.1. Analysis of Neurofilament Light Chain (NfL)

To evaluate the change from baseline in Neurofilament Light chain (NfL), an MMRM analysis will be used to compare change from baseline to scheduled collection times in the full analysis set. The model will include the fixed, categorical effects of treatment, visit, and treatment-by-visit interaction, as well as continuous effects of baseline NfL and age at baseline. Visit will be considered a categorical variable with values equal to the visit numbers at NfL is assessed. The null hypothesis is that the difference in means between donanemab and aducanumab equals zero. The values for NfL may be log transformed to fit the normality assumption of the model.

4.6.2. Analysis of Plasma Tau

To evaluate the change in plasma tau analytes (including assayed plasma total tau and p-tau) after treatment, an MMRM will be used to compare change from baseline to scheduled collection times in the full analysis set. The model will include the fixed, categorical effects of treatment, visit, and treatment-by-visit interaction, as well as the continuous effects of baseline plasma tau and baseline age. Visit will be considered a categorical variable with values equal to the visit numbers at which plasma tau is assessed. The null hypothesis is that the difference in means between donanemab and aducanumab equals zero.

4.6.3. PK/PD Analyses

Compartmental modeling of donanemab PK data using nonlinear mixed effects modeling or other appropriate methods may be explored, and population estimates for clearance and central volume of distribution may be reported. Depending on the model selected, other PK parameters may also be reported. Exploratory graphical analyses of the effect of dose level or demographic factors on PK parameters may be conducted. If appropriate, data from other studies of donanemab may be used in this analysis.

The PK/PD relationships between plasma donanemab concentration and florbetapir F18 PET parameters or other markers of PD activity may be explored graphically. The relationship between the presence of antibodies to donanemab and PK, PD, safety, and/or efficacy may be assessed graphically. If warranted, additional analysis may be explored to evaluate potential interactions for ADA, PD, and other endpoints (PET scan, safety, etc.).

4.6.4. Clinical Trial Registry Analyses

Additional analyses will be performed for the purpose of fulfilling the Clinical Trial Registry (CTR) requirements.

Analyses provided for the CTR requirements include the following:

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

- An adverse event is considered 'Serious' whether or not it is a treatment emergent adverse event (TEAE).
- An adverse event is considered in the 'Other' category if it is both a TEAE and is not serious. For each Serious AE 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 participants/subjects in every treatment group may not be included if a 5% threshold is chosen (5% is the minimum threshold).
- AE reporting is consistent with other document disclosures for example, the CSR, manuscripts, and so forth.

4.7. Safety Analyses

4.7.1. Extent of Exposure

Days of exposure will be calculated for each participant (date of last dose – date of first dose + 28). Additionally, days in study will also be calculated for each participant (end date of study participation – date of first dose). Summary statistics will be provided for the total number of days and participant-years of exposure by treatment.

The number and percentages of participants that stopped dosing due to meeting centiloid stopping criteria will be reported for donanemab participants at the time points associated with the post-baseline florbetapir F18 PET scans.

4.7.2. Adverse Events

Treatment-emergent adverse events (TEAEs) will be defined as events that first occurred or worsened after the randomization date (Visit 2 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. The MedDRA lower-level term (LLT) will be used in the treatment-emergent computation. The maximum severity for each lower-level term (LLT) during the baseline period will be used as baseline.

An overview of AEs, including the number and percentage of participants who died, suffered serious adverse events (SAEs), discontinued due to AEs and who suffered TEAEs, will be provided. Comparison between treatments will be performed using Fisher's Exact Test.

Summaries of AEs by decreasing frequency of PT within SOC will be provided for the following:

- Preexisting conditions
- TEAEs
- TEAEs by maximum severity
- TEAEs occurring in greater than or equal to 2% of participants by PT
- Serious adverse events
- Adverse events reported as reason for study treatment discontinuation

These summaries will include number and percentages of participants with TEAEs. Treatment comparisons will be carried out using Fisher's Exact Test.

Hypersensitivity and Infusion-Related Reactions (IRR) will be summarized and compared between treatment groups using Fisher's Exact test. Hypersensitivity and IRR will be broken out between Potential Immediate (defined as TEAEs occurring on the date of infusions) and Potential Non-Immediate (defined as TEAEs not occurring on the date of infusions but prior to the administration of a subsequent infusion). The following will be used to identify such TEAEs:

- Anaphylactic reaction SMQ (20000021; narrow, algorithm per SMQ guide, and broad)
- Hypersensitivity SMQ (20000214; narrow and broad)
- Angioedema SMQ (20000024; narrow and broad)
- Event maps to Preferred Term (PT) of Infusion related reaction (10051792)

The number and percentage of participants who experienced a TEAE for the following will be analyzed:

- Any narrow or algorithmic term from any one of the 3 SMQs indicated above (that is, combined search across narrow and algorithmic portions of all 3 SMQs)
- Any narrow scope term within each SMQ, separately (that is, narrow SMQ search)
- Any term within each SMQ, separately (that is, broad SMQ search)

4.7.3. 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 post-baseline visit at which laboratory measurements are taken will be compared between treatment groups using an MMRM model on the safety analysis set. The model will include the fixed, categorical effects of treatment, visit, and treatment-by-visit interaction as well as the continuous effects of ranked baseline value and age at baseline. This analysis will be done separately for each laboratory analyte.

Treatment differences in the proportion of participants with treatment-emergent high or treatment-emergent low or treatment-emergent abnormal laboratory values at (1) anytime and (2) each scheduled 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 participants who were low or normal at baseline and have at least 1 post-baseline will be included in the denominator when computing the proportion of participants with treatment-emergent high. Similarly, only participants who were high or normal at baseline and have at least 1 post baseline will be included in the denominator when computing the proportion of participants with treatment-emergent low. In addition, treatment differences in the proportion of participants who have normal baselines with a change to abnormal high or abnormal low values at any post-baseline visits will be summarized.

The proportion of participants with treatment-emergent clinically significant changes from a low value or normal value at all baselines at any time in ALT and total bilirubin will be summarized by treatment group. Clinically significant changes of interest at any time are: ALT ≥3 x upper limit of normal (ULN) and total bilirubin ≥2 x ULN, AST ≥3 x 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 participants with (ALT ≥3 x ULN OR AST ≥3 x ULN) AND total bilirubin ≥2 x ULN at any time. Comparisons between treatment groups will be carried out using Fisher's Exact test. When criteria are met for hepatic evaluation and completion of the hepatic safety case report form (CRF), investigators are required to answer a list of questions pertaining to the participant's history, relevant pre-existing medical conditions, and other possible causes of liver injury. A listing of the information collected on the hepatic-safety CRF will be generated.

4.7.4. Vital Signs and Other Physical Findings

Although measured at the site, vital signs are not recorded in the CRF. No summaries of vital signs are planned.

4.7.5. Safety MRIs

The frequency and percentage of participants with significant treatment-emergent MRI findings, especially Amyloid Related Imaging Abnormalities (ARIA) events such as vasogenic edema (ARIA-E) or microhemorrhage (ARIA-H) will be reported. Fisher's exact test will be used for treatment comparisons.

To evaluate white matter changes over time, a shift table will be created from the following categories:

- 0 = No lesions
- 1 = Focal lesions
- 2 = Beginning confluence of lesions
- 3 = Diffuse involvement of entire region

A listing of MRI data will also be presented.

4.7.6. Immunogenicity

The frequency and percentage of participants with preexisting (baseline) ADA, ADA at any time after baseline, and TE-ADAs to donanemab will be summarized. If no ADAs are detected at baseline, TE-ADAs are defined as those with a titer 2-fold (1 dilution) greater than the MRD of the assay. For samples with ADA detected at baseline, TE-ADA are defined as those with a 4-fold (2 dilutions) increase in titer compared to baseline. For the TE-ADA subjects, the

distribution of maximum titers will be summarized. The frequency of subjects with neutralizing antibodies (subset of the TE-ADA participants) will also be summarized.

4.7.7. Columbia Suicide Severity Rating Scale

Suicidal ideation, suicidal behavior, and self-injurious behavior without suicidal intent occurring during treatment, based on the Columbia-Suicide Severity Rating Scale (C-SSRS), will be summarized by treatment. In particular, for each of the following events, the number and percent of participants with the event will be enumerated by treatment: completed suicide, nonfatal suicide attempt, interrupted attempt, aborted attempt, preparatory acts or behavior, active suicidal ideation with specific plan and intent, active suicidal ideation with some intent to act without specific plan, active suicidal ideation with any methods (no plan) without intent to act, nonspecific active suicidal thoughts, wish to be dead, and self-injurious behavior without suicidal intent. Although not suicide-related, the number and percent of participants with non-suicidal self-injurious behavior occurring during the treatment period will also be summarized by treatment.

In addition, the number and percent of participants who experienced at least one of various composite measures during treatment will be presented and compared. These include suicidal behavior (completed suicide, non-fatal suicidal attempts, interrupted attempts, aborted attempts, and preparatory acts or behavior), suicidal ideation [active suicidal ideation with specific plan and intent, active suicidal ideation with some intent to act without specific plan, active suicidal ideation with any methods (no plan) without intent to act, non-specific active suicidal thoughts, and wish to be dead], and suicidal ideation or behavior.

The number and percent of participants 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.

Specifically, 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).

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.

The following outcome is a numerical score derived from the C-SSRS categories. The score is created at each assessment for each participant 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.

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.

- Treatment-emergent suicidal ideation compared to recent history:

  An increase in the maximum suicidal ideation score during treatment (Visits Y1-Y2) from the maximum suicidal ideation category during the screening and lead-in periods (C-SSRS scales taken at Visits X1-X2). Recent history excludes "lifetime" scores from the Baseline C-SSRS scale or Baseline/Screening C-SSRS scale.

- Treatment-emergent serious suicidal ideation compared to recent history: An increase in the maximum suicidal ideation score to 4 or 5 on the C-SSRS during treatment (Visits Y1-Y2) from not having serious suicidal ideation (scores of 0-3) during the screening and lead-in periods (C-SSRS scales taken at Visits X1-X2). Recent history excludes "lifetime" scores from the Baseline C-SSRS scale or Baseline/Screening C-SSRS scale.
- Emergence of serious suicidal ideation compared to recent history:

  An increase in the maximum suicidal ideation score to 4 or 5 on the C-SSRS during treatment (Visits Y1-Y2) from no suicidal ideation (scores of 0) during the screening and lead-in periods (C-SSRS scales taken at Visits X1-X2). Recent history excludes "lifetime" scores from the Baseline C-SSRS scale or Baseline/Screening C-SSRS scale.

- Improvement in suicidal ideation at endpoint compared to baseline:
  A decrease in suicidal ideation score at endpoint (the last measurement during treatment; Visits Y1-Y2) from the baseline measurement (the measurement taken just prior to treatment; (Visit X2). This analysis should only be performed for a non-lifetime baseline measurement (i.e., having improvement from the worse event over a lifetime is not clinically meaningful). A specific point in time can be used instead of endpoint.
- Emergence of suicidal behavior compared to all prior history:

  The occurrence of suicidal behavior (Categories 6-10) during treatment (Visits Y1-Y2) from not having suicidal behavior (Categories 6-10) prior to treatment (Visits X1-X2). Prior to treatment includes "lifetime" and/or "screening" scores from the Baseline C-SSRS scale, Screening C-SSRS scale, or Baseline/Screening C-SSRS scale, and any "Since Last Visit" from the Since Last Visit C-SSRS scales taken prior to treatment.

Participants who discontinued from the study with no postbaseline C-SSRS value will be considered unevaluable for analyses of suicide-related events. Only evaluable participants will be considered in the analyses. Fisher's exact test will be used for treatment comparisons.

4.8. Other Analyses

4.8.1. Treatment Compliance

Because dosing occurs at study visits, participants who attend all visits and successfully receive donanemab or aducanumab infusions are automatically compliant with this treatment. Any infusion at which 75% (approximately 105 mL) or more of the infusion solution is given will be considered a complete infusion.

Summary statistics for treatment compliance will be provided for the total number of complete infusions received, duration of complete infusion, and volume of complete infusion by treatment group. Frequencies and percentages of reasons why infusion was stopped will also be presented.

4.8.2. Concomitant Therapy

Prior medications are defined as those that stop before randomization (the day prior to the first administration of study drug). Concomitant medications are defined as those being taken on or after randomization (the day prior to the first administration of study drug). 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. If the start or stop dates of therapies are missing or partial to the degree that determination cannot be made of whether the therapy is prior or concomitant, the therapy will be deemed concomitant. Medications will be coded using the World Health Organization (WHO) drug dictionary.

4.8.3. Subgroup analyses

To assess the effects of genetic factors and clincal severity of illness on treatment outcome, subgroup analyses for the primary endpoint, complete brain amyloid clearance, and mean changes in amyloid reduction will be conducted:

- Age - <65, ≥65.
- APOE4 Carrier Status – Carrier defined as E2/E4, E3/E4, or E4/E4 genotype; No-Carrier defined as all other genotypes
- Clinical staging at baseline – MCI (defined as 27 ≤ MMSE ≤ 30) or mild AD (defined as 20 ≤ MMSE ≤ 26)

The statistical models will be the same as those stated for the co-primary and key secondary objectives, with the addition of factors for the subgroup variable, the subgroup-by-treatment interaction, and the subgroup-by-visit interaction where appropriate.

4.9. Interim Analyses

A planned interim analysis may be conducted for Study AACN as deemed by Lilly or its designee to assess for early efficacy of the two co-primary endpoints. No changes to the conduct of the study will result from executing an interim analysis, nor will the results of the interim impact the conduct of the study. Should an interim analysis be deemed necessary, it will be

performed by the independent data monitoring committee (IDMC), consisting of a minimum of 3 external members, including a physician with expertise in AD and a statistician who are independent experts not otherwise involved in the study, when approximately at least a sufficient number of participants (75-125) overall have had a chance to receive their 6-month amyloid PET scan. The analysis method for the primary efficacy endpoint described in Section 4.4 Co-Primary Endpoint Analysis will be used for the interim analysis.

Alpha Spending Approach for the Potential Interim Analysis Based on the group sequential design with the Pocock alpha spending approach, a 2-sided alpha of 0.03100566 will be allocated to the interim analysis for evaluating the efficacy of each of the two co-primary endpoints. If either of the null hypothesis for the two co-primary endpoints are not rejected at the interim analysis, both co-coprimary endpoints will be evaluated at the originally scheduled time point when all randomized participants have had a chance to complete their 6-month amyloid PET scan. This analysis would be performed at a 2-sided alpha level of 0.02774 to maintain strong control of an overall 2-sided Type I error rate of 5%.

The 2-sided alpha levels specified above were calculated assuming an information fraction of 0.50 or when data is available on 50% of participants. Depending on the actual information fraction at the time of the interim analysis, the alpha levels for the interim and final time points will be adjusted accordingly, but the Pocock spending function will still be utilized. The information fraction will be defined as the number of patients who were randomized prior to a specified date (patients included in the interim) divided by the total number of patients randomized.

If the interim analysis is deemed not necessary to be conducted by the sponsor, the primary efficacy analysis will be conducted when all participants have had a chance to receive their 6-month amyloid PET scan. This will be conducted at the two-sided significance level of 5%.

If an interim analysis is performed and both co-primary endpoints/null hypotheses are rejected, the 2-sided alpha available to test the four key secondary endpoints will be at least as much as 0.03100566 yet will be lower than 0.05. The alpha level will be determined using the approach outlined in Hung, Wang, and O'Neill (2007). If an interim analysis is performed and we fail to reject both co-primary endpoints, the co-primary endpoints will be re-tested (when all patients have their 6 month PET scan) and key secondary endpoints will be tested in a sequential/gated manner at a 2-sided alpha level of 0.02774 as determined by the Pocock alpha spending approach. The order of testing and the gating strategy for the key-secondary endpoints will be the same as detailed in Section 2.1.

4.10. Changes to Protocol-Planned Analyses

There are no changes to the protocol planned analyses at this time.

5. Sample Size Determination

This sample size will provide at least 98% power to demonstrate that donanemab is superior to aducanumab in achieving complete amyloid brain plaque clearance individually at 6 months in the overall population and the intermediate tau subpopulation. Power estimates were obtained using specific contrasts for the 6-month time point of treatment by time interaction effect within the framework of generalized linear mixed effect model assuming a logit link.

The assumptions for this power calculation are that 35.7%, 50.1%, and 54.6% of donanemab treated participants reach complete brain amyloid clearance at 6 months, 12 months, and 18 months, respectively. The corresponding complete brain amyloid clearance percentages for aducanumab treated participants are assumed to be 3.0%, 22.8%, and 27.8%, respectively. The simulation for the power calculation and sample size determination was carried out in SAS Enterprise Guide Version 7.15 and all sample size estimates were obtained assuming a 2-sided significance level of 0.05.

6. Supporting Documentation

6.1. Appendix 1: Protocol Deviations

Listings of patients with significant protocol violations will be provided for randomized participants. The following list of significant protocol violations will be determined from the clinical database and from the clinical/medical group:

- Informed consent violation detected as a missing date of informed consent.
- Did not have a florabetapir F18 PET scan at any of the visits at which the procedure scheduled to be collected.
- Incomplete infusions (any infusion at which less than 75%, approximately 105 mL, of the infusion solution is given).

The following list of significant protocol violations will be determined by clinical/medical group:

- Protocol violations of inclusion/exclusion criteria.
- Had a study dosing algorithm violation.

Other protocol deviations reported through the monitoring process will be reviewed by the study team and if judged to be significant, will be added to the final reported listing.

Summaries of significant protocol deviations will also be displayed by treatment group.

7. References

H. M. James Hung, Sue-Jane Wang & Robert O'Neill (2007) Statistical Considerations for Testing Multiple Endpoints in Group Sequential or Adaptive Clinical Trials, Journal of Biopharmaceutical Statistics, 17:6, 1201-1210

Leo Document ID = a3097bab-164d-418d-b7ff-b3ea9e7a2cd3

Approver: PPD

Approval Date & Time: 24-Sep-2021 17:08:06 GMT

Signature meaning: Approved