# I5T-MC-AACI Statistical Analysis Plan

**Protocol Title:** Assessment of Safety, Tolerability, and Efficacy of Donanemab in Early Symptomatic Alzheimer's Disease

**Protocol Number:** I5T-MC-AACI

**Compound Number:** LY3002813

**Short Title:** Donanemab in Early Symptomatic Alzheimer's Disease

**Acronym:** TRAILBLAZER-ALZ 2

**Sponsor Name:** Eli Lilly and Company

**Legal Registered Address:** Indianapolis, Indiana, USA 46285

**Regulatory Agency Identifier Number(s)**

IND: 109157

EudraCT: 2020-000077-25



**Document ID:** VV-CLIN-075800

Template Version: 1.0



Approved on 20 Apr 2023 GMT

## Table of Contents

# Version history

**Table AACI.1.1. SAP Version History Summary**

| SAP Version | Approval Date | Change | Rationale |
|---|---|---|---|
| 1 | 01 Nov 2021 | Not Applicable | Original version |
| 2 | 29 Mar 2022 | Updated study to Phase 3 and added analyses planned for addendum 9 part | Updated by including the analysis planned for addendum 9 cohort before the first patient visit for addendum 9 |
| 3 | See Date on Page 1 | Primary analysis method changed to natural cubic spline with 2 degrees of freedom | Update based on protocol amendment e approved on 10 November 10 2022 |

# 1. Introduction

This version of statistical analysis plan (SAP) is drafted based on the AACI study protocol amendment e, approved 10 November 10 2022. Efficacy and safety analyses for placebo-controlled, double-blind phase of this study are described in this document. The analyses for pharmacokinetics (PK), immunogenicity, open label safety addendum cohort and long-term extension (LTE) are described in separate SAPs.

A set of secondary objective analyses to evaluate the disease progression status by treatment using time-PMRM (progression model with repeat measures) model are added to the SAP, which were not included in protocol at the time of amendment e. The details of these analyses are described in Section 4.4.2.

Table, figure, and listing (TFL) specifications are contained in a separate document.

## 1.1. Objectives, Endpoints, and Estimands

| Objectives | Endpoints |
|---|---|
| Primary | |
| To assess the effect of donanemab versus placebo on clinical progression in participants with early symptomatic AD | iADRS change from baseline through Week 76 in at least one of<br>• the low-medium (or intermediate) tau pathology population or<br>• the overall population |
| Secondary | |
| To assess the effect of donanemab versus placebo on clinical progression in participants with early symptomatic AD | Change from baseline through Week 76 in at least one of<br>• the low-medium tau pathology population or<br>• the overall population<br>as measured by:<br>• CDR-SB<br>• ADAS-Cog13 score<br>• ADCS-iADL score<br>• MMSE score |
| To assess the effect of donanemab versus placebo on brain amyloid deposition | Change in brain amyloid plaque deposition from baseline through Week 76 as measured by florbetapir F18 PET scan in at least one of |

| Objectives | Endpoints |
|---|---|
| | • the low-medium tau pathology population or<br>• the overall population |
| To assess the effect of donanemab versus placebo on brain tau deposition | Change in brain tau deposition from baseline through Week 76 as measured by flortaucipir F18 PET scan in at least one of<br>• the low-medium tau pathology population or<br>• the overall population |
| To assess the effect of donanemab versus placebo on brain region volumes | Change in volumetric MRI measures from baseline through Week 76 |
| To evaluate safety and tolerability of donanemab | Standard safety assessments:<br>• Spontaneously reported AEs<br>• Clinical laboratory tests<br>• Vital sign and body weight measurements<br>• 12-lead ECGs<br>• Physical and neurological examinations<br>• MRI (ARIA and emergent radiological findings)<br>• Infusion related reactions<br>• C-SSRS |
| To assess peripheral PK and presence of anti-donanemab antibodies | Plasma PK of donanemab<br>ADAs against donanemab including<br>• treatment-emergent ADAs<br>• • neutralizing antibodies |
| Tertiary/Exploratory | |
| To assess the effect of donanemab versus placebo on blood-based biomarkers | Plasma in at least one of<br>• the low-medium tau pathology population or<br>• the overall population<br>• NfL<br>• GFAP<br>• P-tau |

| Objectives | Endpoints |
|---|---|
| | • Ab levels |
| To assess the effect of donanemab versus placebo on cognition | Change in DSST - Medicines Version from baseline through Week 76 in at least one of<br>• the low-medium tau pathology population or<br>• the overall population |
| To assess the efficacy of donanemab to prolong time in the current disease state | CDR global score<br>CDR-SB in at least one of<br>• the low-medium tau pathology population or<br>• the overall population |
| To assess the effect of donanemab versus placebo on time progression of the disease in participants with early symptomatic AD | Slowing in time progression of the disease through week 76 in at least one of<br>• the low-medium tau pathology population or<br>• the overall population<br>as measured by<br>• iADRS<br>• CDR-SB |

Abbreviations: Aβ = amyloid beta; AD = Alzheimer's disease; ADA = anti-drug antibody; ADAS-$Cog_{13}$ = Alzheimer's Disease Assessment Scale – Cognitive subscale; ADCS-ADL = Alzheimer's Disease Cooperative Study – Activities of Daily Living; ADCS-iADL = Alzheimer's Disease Cooperative Study – instrumental Activities of Daily Living; AE = adverse event; ARIA = amyloid-related imaging abnormalities; CDR = Clinical Dementia Rating Scale; CDR-SB = Clinical Dementia Rating Scale – Sum of Boxes; C-SSRS = Columbia Suicide Severity Rating Scale; DSST = Digit Symbol Substitution Test; ECG = electrocardiogram; GFAP = glial fibrillary acidic protein; iADRS = integrated Alzheimer's Disease Rating Scale; MMSE = Mini Mental State Examination; MRI = magnetic resonance imaging; NfL = neurofilament light chain; PET = positron emission tomography; PK = pharmacokinetics; P-tau = phosphorylated tau; QOL-AD = Quality of Life in Alzheimer's Disease; RUD-Lite = Resource Utilization in Dementia – Lite Version.

**Primary estimand/coprimary estimand**

The primary clinical question of interest for study AACI is: What is the intervention difference in slowing of progression of AD relative to placebo across 76 weeks of intervention in participants with early symptomatic AD, regardless of initiation or change to standard of care medications and regardless of whether a participant stops taking study drug. Therefore, the estimand is described by the following attributes:

Population: Participants with early symptomatic AD either with intermediate tau level at baseline, or entire randomized participants including those with high tau value at baseline.

Endpoint**:** Integrated Alzheimer's Disease Rating Scale (iADRS) values at each visit through Week 76.

Treatment condition: The randomized treatment, donanemab or placebo, regardless of initiation or change to standard of care medications and regardless of whether a participant stops taking study intervention (treatment policy strategy).

Intercurrent events: The 2 intercurrent events 'initiation or change to standard of care medications' and 'discontinuation of donanemab' are both addressed by the intention-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. No other intercurrent events are considered.

Population-level summary: the difference of adjusted mean change from baseline (CFB) values at 76 weeks between donanemab arm and the placebo arm.

Rationale for estimand: This estimand is based on the intent to treat principle, and it aims at reflecting how patients with early symptomatic AD are treated in clinical practice. The primary analysis will use a natural cubic spline model with 2 degrees of freedom (NCS2) to compare the cognitive and functional decline as measured by iADRS between treatment groups at 76 weeks.

## 1.2. Study Design

Study AACI is a multicenter, randomized, double-blind placebo-controlled, Phase 3 study of donanemab in participants with early symptomatic AD. Participants who meet entry criteria will be randomized in a 1:1 ratio to one of the following treatment groups:

- Donanemab: 700 mg IV Q4W for first 3 doses and then 1400 mg IV Q4W
- Placebo

The randomization is stratified by intermediate or high tau level as decided by tau PET at screening, and the study sites. After 76 weeks, participants will enter long-term extension (LTE) part of the study and will be assigned to donanemab or placebo based on criteria described in Section 4.1.3 of protocol amendment e.

The maximum total duration of study participation for each participant, including screening and the post-treatment follow-up periods, is up to 205 weeks:

- Lead-In: any time prior to complete screening
- Complete Screening: up to 7 weeks
- Double-Blind: 76 weeks
- Extension: 78 weeks
- Follow-Up: up to 44 weeks

The maximum duration of treatment is 150 weeks.

Scheduled Reduction of Donanemab to Placebo

Participants whose amyloid plaque reduction as measured by florbetapir F18 PET scans at Visit 8 (Week 24), Visit 15 (Week 52), Visit 21 (Week 76), Visit 28 (Week 102), or Visit 35 (Week 130) meets dose reduction criteria will have a double-blind dose reduction of donanemab to IV placebo for the remaining duration of the study.

These dose reduction rules are defined by the sponsor, that is, amyloid plaque level was <11 centiloid at any single amyloid PET scan, or 11≤CL<25 from two consecutive amyloid PET scans.

This SAP covers the analyses of data collected through double-blind phase, that is, up to and including visit 21 (week 76). The analyses of LTE phase are described in a separate LTE SAP.

# 2. Statistical Hypotheses

The primary objective of this study is to test the hypothesis that IV infusion of donanemab will slow the cognitive and/or functional decline of AD as measured by iADRS score compared with placebo in the population of participants with intermediate tau pathology at baseline or the overall population. Thus, the null hypothesis to be tested in relation to the primary estimand is as follows:

H0: Least square (LS) mean change from baseline of iADRS score at 76 weeks from donanemab treated group is not different from the LS mean change from baseline of iADRS score at 76 weeks from placebo treated group, neither from participants with intermediate tau pathology at baseline, nor from overall population

The null hypotheses corresponding to the secondary objectives are as follows:LS mean change of CDR-SB score at 76 weeks from donanemab treated group is not different from the LS mean change of CDR-SB score at 76 weeks from placebo treated group, neither from participants with intermediate tau pathology at baseline, nor from overall population.

- LS mean change of ADAS-Cog13 score at 76 weeks from donanemab treated group is not different from the LS mean change of ADAS-Cog13 score at 76 weeks from placebo treated group, neither from participants with intermediate tau pathology at baseline, nor from overall population.
- LS mean change of iADL score at 76 weeks from donanemab treated group is not different from the LS mean change of iADL score at 76 weeks from placebo treated group, neither from participants with intermediate tau pathology at baseline, nor from overall population.
- LS mean change of MMSE score at 76 weeks from donanemab treated group is not different from the LS mean change of MMSE score at 76 weeks from placebo treated group, neither from participants with intermediate tau pathology at baseline, nor from overall population.

The null hypotheses for biomarker analyses are:

- LS mean change of amyloid burden as measured by amyloid PET centiloid values at 76 weeks from donanemab treated group is not different from that from placebo treated group
- LS mean change of brain tau deposition as measured by flortaucipir PET standard uptake value ratio (SUVR) values at 76 weeks from donanemab treated group is not different from that from placebo treated group
- LS mean change of brain regional volumes as measured by volumetric MRI at 76 weeks from donanemab treated group is not different from that from placebo treated group

The hypotheses for PK and anti-donanemab antibodies analyses will be described in a separate SAP.

## 2.1. Multiplicity Adjustment

The primary efficacy objective of Study AACI is to demonstrate donanemab slows clinical decline in AD as measured by iADRS comparing to placebo within 76 weeks in at least 1 of the following populations: the overall population or the participants with intermediate tau burden at baseline.

A prespecified hypothesis testing plan is developed that employs Bretz's graphical approach (Bretz et al. 2009, 2011) to provide a strong control of the study-wise Type I error rate for the primary and key secondary hypotheses at 2-sided level α=0.05. For the primary objective hypothesis testing, the initial 2-sided alpha level is set to 0.04 for baseline intermediate tau level population and 0.01 for overall population. The hypothesis testing scheme, alpha recycle and weight, are described in detail in Figure AACI.2.1.

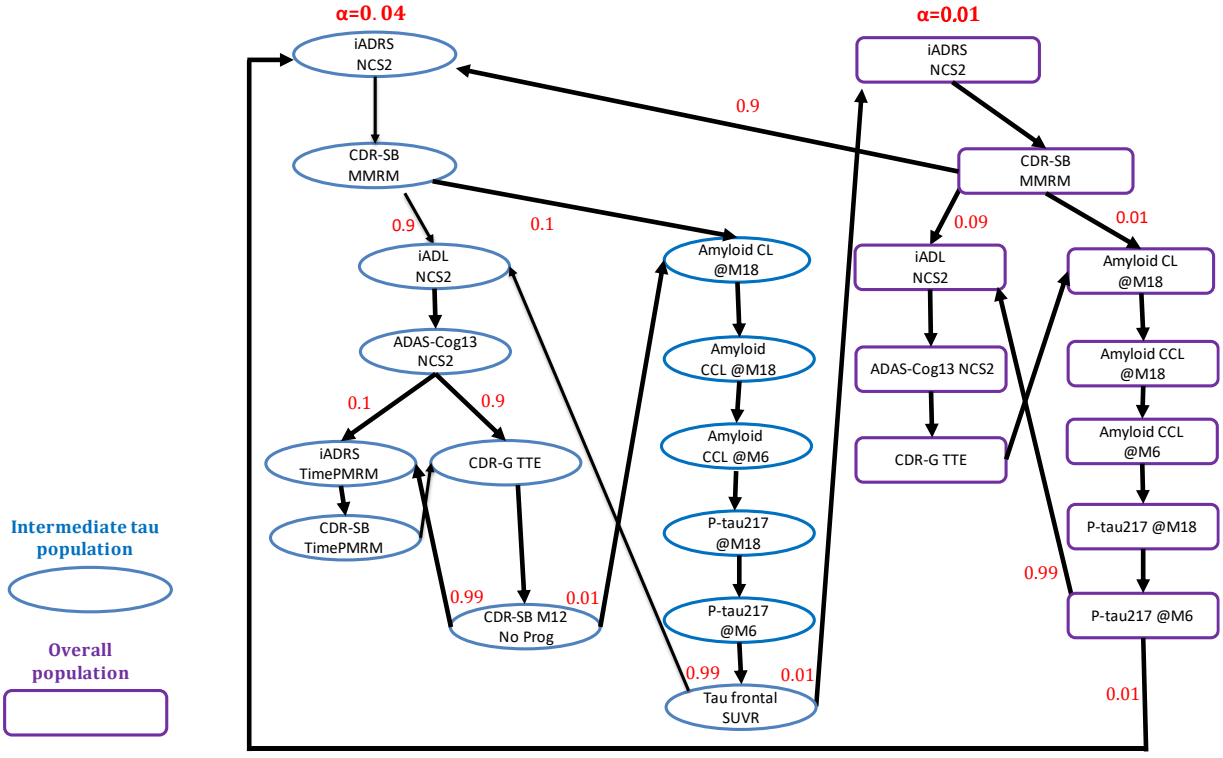


**Figure AACI.2.1. Hypothesis testing scheme for controlling study-wise type I error rate at 2-sided 5%.**

The hypothesis tested in Figure AACI.2.1 are detailed in Table AACI.2.1.

**Table AACI.2.1. Hypothesis Included in Graphical Testing Scheme**

| | Hypothesis to test |
|---|---|
| iADRS NCS2 | iADRS score change LS mean differences at Week 76, tested with NCS model with 2 degree-of-freedom |
| CDR-SB MMRM | CDR-SB score change LS mean differences at Week 76, tested with MMRM |
| iADL NCS2 | iADL score change LS mean differences at Week 76, tested with NCS model with 2 degree-of-freedom |
| ADAS-Cog13 NCS2 | ADAS-Cog13 score change LS mean differences at Week 76, tested with NCS model with 2 degree-of-freedom |
| iADRS time-PMRM | Disease progression time saved at Week 76 as measured by iADRS, tested with time-PMRM model |
| CDR-SB time-PMRM | Disease progression time saved at Week 76 as measured by CDR-SB, tested with time-PMRM model |
| CDR-G TTE | Difference in hazard of progressing to first meaningful clinical worsening event defined by CDR-global score, tested with Cox proportional hazard model |
| CDR-SB wk 52 No Prog | Difference in probability of "no progression" as defined by CDR-SB at Week 52. Tested with GLIMM model |
| Amyloid CL | Amyloid centiloid change LS mean difference at Week 76, tested with MMRM |
| Amyloid CCL @ Week 24 | Probability of amyloid complete removal (centiloid <24.1) among donanemab treated arm at Week 24, tested with binomial test |
| Amyloid CCL @ Week 76 | Probability of amyloid complete removal (centiloid <24.1) among donanemab treated arm at Week 76, tested with binomial test |
| P-tau217 @ Week 24 | P-tau217 change LS mean difference at Week 24, tested with MMRM |
| P-tau217 @ Week 76 | P-tau217 change LS mean difference at Week 76, tested with MMRM |
| Tau frontal SUVR | Tau PET frontal SUVR change LS mean difference at Week 76, tested with ANCOVA analysis |

Abbreviations: ADAS-Cog13 = Alzheimer's Disease Assessment Scale – Cognitive subscale; ANCOVA = analysis of covariance; CDR-G = Clinical Dementia Rating Scale -Global Score; CDR-SB = Clinical Dementia Rating Scale – Sum of Boxes; CL = centiloid; iADRS = integrated Alzheimer's Disease Rating Scale; LS = least squares; PMRM = progression model with repeat measures; NCS2 = natural cubic spline model with 2 degrees of freedom; MMRM = mixed-effect model for repeated measures; PET = positron emission tomography; SUVR = standard uptake value ratio; TTE = time-to-event.

# 3. Analysis Sets

Analysis sets are defined in Table AACI.3.1.

**Table AACI.3.1. Study AACI Analysis Sets**

| Participant Analysis Set | Description |
|---|---|
| Entered | All participants who sign informed consent |
| Randomized | All entered participants who are randomized to study treatment |
| Evaluable Efficacy (EES) | All randomized participants with a baseline and at least one post-baseline efficacy scale |
| Safety | All randomized participants who are exposed to study drug. Participants will be summarized according to the treatment group to which they were randomized |
| Per-Protocol | All subjects in the Evaluable Efficacy set who also:<br>signed the inform consent form<br>had an assessment of the primary endpoint at each scheduled visit completed<br>had no violations of inclusion/exclusion criteria<br>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)<br>were not considered non-compliant with regard to study drug |
| Completers | All randomized subjects who have completed the placebo controlled double blinded phase |

Efficacy and safety measures summarized and/or analyzed by these analysis sets are presented in Table AACI.3.2.

**Table AACI.3.2. Efficacy and Safety Measures by Analysis Set**

| Participant Analysis Set | Variables Assessed and Outputs |
|---|---|
| Entered | Listings |
| Randomized | Tables and listings for patient characteristics, baseline severity, and patient disposition |
| Evaluable Efficacy | Tables, listings, and figures of the following: iADRS, CDR-SB, ADAS-Cog13, ADCS-ADL (basic, instrumental, and total), MMSE, CDR-Global, Digit Symbol Substitution Test (medicines version), plasma GFAP, plasma p-tau, amyloid PET centiloid, flortaucipir SUVR values, volumetric MRI measurements, and concomitant medications |
| Safety | Tables, listings, and figures of the following: compliance, adverse events, laboratory results, vital signs, weight, ECG, safety MRIs, C-SSRS |
| Per-Protocol | Tables, listings, and figures of the following: iADRS, CDR-SB, ADAS-$Cog_{13}$, ADCS-ADL (basic, instrumental, and total), MMSE |
| Completers | Tables, listings, and figures of the following: iADRS, CDR-SB, ADAS-$Cog_{13}$, ADCS-ADL (basic, instrumental, and total), MMSE, DSST (medicines version), plasma total tau, plasma p-tau, florbetapir parameters, flortaucipir parameters, and volumetric MRI measurements |

# 4. Statistical Analyses

## 4.1. General Considerations

The protocol calls for a Data Monitoring Committee (DMC) charged with making decisions regarding patient safety and study futility. This analysis plan describes analyses planned for the double blinded phase clinical study report, interim analysis for safety and all interim analyses for the DMC. Analyses planned for AACI long term extension (LTE) phase or for open label safety addendum part are described in separate SAPs.

Unless otherwise noted, all pairwise tests of treatment effects will be conducted at a 2-sided alpha level of 0.05; 2-sided confidence intervals (CIs) will be displayed with a 95% confidence level. All tests of interactions between treatment and other factors will be conducted at an alpha level of 0.05.

Unless otherwise noted baseline is defined as the last measurement prior to dosing. When change from baseline is assessed, subjects will only contribute to the analysis if both a baseline and a post-baseline measurement are available. Endpoint is the last non-missing post-baseline measurement within the time period for the given analysis. For natural cubic spline (NCS), mixed-effect model for repeated measures (MMRM), and disease progression (DPM) models, observations collected at unscheduled visits will not be included in the analyses (Andersen and Millen 2013).

This study will be conducted by multiple investigators at multiple sites internationally. In the event that any investigator has an inadequate number of subjects (defined as 1 or 0 randomized subjects per treatment group) for the planned analyses, data from all such sites will be pooled. The pooling will be done first within a country. If the resulting pool within a country is still inadequate (1 or 0 randomized subjects to 1 or more treatment arms), no further pooling will be performed. In addition, a listing including country, investigator site with address, number of patients enrolled (randomized) by each site and unique subject IDs will be presented.

Any change to the data analysis methods described in the protocol will require an 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 within this SAP and/or clinical study report.

## 4.2. Participant Dispositions

Because this is a long-term study in a patient population that is elderly with multiple comorbidities, patient withdrawal is of particular concern. Additional efforts will be undertaken to reduce patient withdrawals and to obtain information on patients who are initially categorized as lost to follow-up.

From the randomized population, the percentage of patients withdrawing from each treatment group will be summarized. From the safety population, the percentage of patients withdrawing from each treatment group will be compared between groups using Fisher’s exact test.

Comparisons using Fisher's exact test will be done for the overall percentage of patients who withdraw and also for each specific reason for withdrawal.

The median time to discontinuation will also be compared between treatment groups using the Kaplan-Meier product limit estimator. For any-cause study discontinuation as well as any-cause treatment discontinuation, comparisons of time-to-discontinuation will be conducted using the Kaplan-Meier product limit estimator and the associated log-rank test.

## 4.3. Primary Endpoint Analysis

The primary objective of this study is to test the hypothesis that IV infusion of donanemab will slow the clinical decline of AD as measured by integrated Alzheimer's Disease Rating Scale (iADRS) score compared with placebo in at least one of the low-medium tau pathology population or the overall population.

### 4.3.1. Definition of Endpoint(s)

The iADRS assesses the impact of cognitive loss on the ability to conduct everyday activities and provides a measure of global AD severity as a single summary score. The iADRS comprises 2 underlying domains ("cognitive ability" and "functional ability"), with each representing related but separate concepts. The iADRS integrates the items that make up both domains into a single overall score that is conceptually distinct from either domain assessed individually. The combination score of the iADRS captures commonalities across its domains, minimizing noise that exists within each domain individually.

The ADAS-Cog13 and the ADCS-ADL will be the actual scales administered to participants. If any of the individual items for ADAS-Cog13 or ADCS-ADL are missing or unknown, every effort will be made to obtain the score for the missing item or items. For ADAS-Cog13, if 3 or fewer of a total of 13 items are missing, the total score (maximum = 85) will be imputed as follows: the total from remaining items will be multiplied by a factor that includes the maximum score for the missing items. For example, if the first item, "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 to 5 (maximum = 5), is missing, then the multiplication factor = 85/(85 - [10 + 5]) = 85/70 = 1.21. Thus, the total score for this example will be the sum of the remaining 11 items multiplied by 1.21. The imputed number will be rounded up to the nearest integer. If more than 3 items are missing, the total score for ADAS-Cog13 at that visit will be considered missing.

For the ADCS-iADL, if <30% of the items are missing, the total score will be imputed. For the 3 questions with sub-questions (that is; Q8, 18 and 19), each sub-question is considered a separate item. If the response to the parent question is "no" or "don't know," the sub-questions should not be considered missing. The sum of the non-missing items will be prorated to the sum of total items like described above. The imputed number will be rounded up to the nearest integer. If the nearest integer is greater than the maximum possible score, the imputed score will be equal to the maximum score. If >30% of the items are missing, the total score for ADCS-iADL at that visit will be considered missing. The same imputation technique will be applied to the ADCS-ADL total score. Note that, depending on the specific item responses that are missing,

it is possible to have an imputed total score for both the ADCS-iADL and the ADCS-ADL, an imputed total score for one but not the other, or both total scores missing.

The iADRS score is calculated as:

ADCS-iADL score-ADAS-Cog13 score + 85.

If either ADAS-Cog13 or ADCS-iADL is missing, iADRS score will be considered missing.

The same imputation technique will be applied to the CDR-SB. If only 1 box (of 6) of the CDR is missing, the sum of the boxes will be imputed by prorating the sum from the other 5 boxes. If the score from more than 1 box is not available, the CDR-SB at that visit will be considered missing.

### 4.3.2. Main Analytical Approach

A NCS analysis (Donahue et al. 2023) with 2 degrees of freedom (NCS2) will be used to assess the difference between treatment groups in iADRS score at Week 76. For this NCS2 model applied to primary analysis, 3 knots over the observation time will be placed: 2 at the boundaries (minimum and maximum observation time), and 1 internal knot at the median observation time. The baseline estimates are restricted to be the same for treatment and placebo groups. The model will be estimated using restricted maximum likelihood method.

The iADRS score at baseline and at each of the scheduled post-baseline visits (according to Schedule of Activities [SoA]) will be included in model as a dependent variable. Study visit will be treated as a continuous variable with values equal to weeks between baseline and post-baseline exam dates, and the NCS basis function will be derived using these visits in weeks. The model will include these fixed effects: NCS basis expansion terms (two terms), NCS basis expansion term-by-treatment interaction (two terms), baseline age, concomitant AchEI and/or memantine use at baseline (yes/no), and pooled investigator. Baseline tau category will also be included as a covariate to the model applied to overall population. An unstructured variance-covariance structure matrix will be used to within-subject variance-covariance errors. If the unstructured variance-covariance structure matrix results in a lack of convergence, the following structures will be used in sequence:

- Heterogeneous Toeplitz covariance structure
- Heterogeneous autoregressive order 1 covariance structure
- Heterogeneous compound symmetry covariance structure, and
- Compound symmetry covariance structure

Mean change from baseline values, and the comparisons between change from baseline values by treatment arms will be estimated through the proper contrast set up. The primary time point for treatment comparison will be at Week 76. The Kenward-Roger approximation will be used to estimate the denominator degrees of freedom.

Percent slowing comparing to placebo group will be calculated as the LS estimates of differences in change from baseline between treatment groups at Week 76, divided by the LS estimates of

mean change from baseline value from placebo group. A 95% confidence interval (CI) for this percent slowing is calculated based on a Delta method (Beyene et al. 2005).

### 4.3.3. Sensitivity Analyses

Numerous sensitivity analyses are planned as detailed below.

#### 4.3.3.1. Mixed Model with Repeated Measures (MMRM) Analysis

For MMRM analysis, the change from baseline score on the iADRS at each scheduled postbaseline visit (according to the SoA) during the treatment period will be included as the dependent variable. The model for the fixed effects will include the following terms: baseline iADRS score, baseline score-by-visit interaction, pooled investigator, treatment, visit, treatment-by-visit interaction, concomitant AChEI and/or memantine use at baseline (yes/no), and age at baseline. Baseline tau category will also be included as a fixed effect to the model applied to overall population. Visit will be considered a categorical variable. 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
- compound symmetry covariance structure

The Kenward-Roger approximation will be used to estimate the denominator degrees of freedom.

For MMRM, the primary time point for treatment comparison will be at Week 76. The treatment group contrast in least-squares mean progression and its associated p-value and 95% CI will be calculated for the treatment comparison of donanemab versus placebo using the MMRM model specified above.

#### 4.3.3.2. Disease Progression Model (DPM)

Bayesian Disease Progression Model (DPM) will be applied to evaluate possible slowing of disease progression with treatment of donanemab relative to placebo. The primary purpose of the DPM is to estimate a quantity known as the disease progression ratio (DPR), which measures the proportion of disease progression in donanemab-treated participants relative to placebo-treated participants. A DPR value less than 1 indicates the donanemab arm is slowing the disease progression relative to placebo, and a DPR value greater than 1 indicates the donanemab arm is worsening the disease progression relative to placebo.

The key assumption of the DPM model is that it assumes that the treatment effect of donanemab is proportional to placebo over the course of the study. The proportionality assumption is similar to what is made in proportional hazards modeling of time to event data. The model includes diffuse priors on all parameters; therefore, the prior distributions have very little impact on the posterior distributions.

The DPM model is as follows:

$$Y_{ij} = \gamma_i + e^{\theta T_i} \sum_{v=0}^{j} \alpha_v + \varepsilon_{ij}$$

where $Y_{ij}$ denotes the clinical outcome at visit *j* for participant *i*; the clinical outcome score for a participant at baseline (prior to treatment) is $Y_{i0}$. The value $\gamma_i$ (*i*=1, 2, …, *k*) represents a subject specific random effect. The parameter $T_i$ denotes the treatment arm for participant *i*, where $T_i$ has a value of 1 if a participant is randomized to donanemab, and a value of 0 if the participant is randomized to placebo. The parameter $\alpha_v$ is the change in mean clinical outcome score for placebo from visit *v*-1 to *v*, and $\varepsilon_{ij}$ is the error term. The DPR for donanemab relative to placebo is provided by the parameter $e^{\theta}$. Covariates of the model include concomitant AChEI and/or memantine use at baseline (yes/no), age at baseline, and pooled investigator. Baseline tau category will also be included as a fixed effect to the model applied to overall population.

The DPM will be fit using prior distributions based on the assumption of no information or knowledge of the effect of donanemab from previous studies. The Bayesian posterior probability of the donanemab treatment arm being superior to placebo will be calculated by a margin of interest with 15%, 25% or 30% slowing of placebo progression.

In addition to the Bayesian DPM, a frequentist version of the model will be fit using the same model structure as the Bayesian DPM described above, including the same modeling terms. The model will be fit using an unstructured covariance matrix.

#### 4.3.3.3. NCS with 3 Degree of Freedom

NCS with 3 degrees of freedom model (NCS3) will also be applied. This model assumes two internal knots which were placed at the equidistant percentiles of the scheduled study visit time. The model has same set of covariates as described for NCS2 model, with the exception that three basis functions are included in the model as opposed to two.

#### 4.3.3.4. Censoring Post Amyloid Related Imaging Abnormality – Edema and Infusion Related Reaction Events

The occurrence of Amyloid Related Imaging Abnormality – Edema (ARIA-E) and infusion related reaction (IRR) potentially may lead to functional unblinding of the study treatment. To evaluate the impact from these events, a sensitivity analysis is arranged with iADRS measurements censored post the first occurrence of ARIA-E (by MRI findings and TEAE cluster as defined in Section 4.6.3) and/or IRR (based on CRF reports). The NCS2 model will be applied to this censored dataset, with the same modeling details as described in Section 4.3.2.

#### 4.3.3.5. Analysis Evaluating the Impact from Death

Another sensitivity analysis will be imputing the worst possible iADRS score 0 as measurements post death for the death cases, until Week 76. The NCS2 model will be applied to this censored dataset, with the same modeling details as described in Section 4.3.2.

#### 4.3.4. Supplementary Analyses

The following analyses are planned as supplementary analyses.

##### 4.3.4.1. Completer Analysis

The primary efficacy outcome, iADRS, from the dataset of those patients who remained in the study and on treatment through Week 76 ("completers" for placebo-controlled double blinded phase) will be analyzed using NCS2 analysis. The model setup and included covariates will be the same as those described for NCS2 in Section 4.3.2. Baseline tau category will also be included as a fixed effect to the model applied to overall population.

##### 4.3.4.2. Per Protocol Analysis

The primary efficacy outcome, iADRS, from the per-protocol dataset will be analyzed using the NCS2 analysis. The model setup and included covariates will be the same as those described for NCS2 in Section 4.3.2 .

##### 4.3.4.3. Amyloid Related Imaging Abnormality – Edema Adjusted Analysis

ARIA-E events potentially may lead to functional unblinding. To assess the impact of ARIA-E on treatment effect evaluation, the donanemab treated subjects will be divided into two groups: with or without ARIA-E. The primary efficacy outcome, iADRS will be analyzed using the NCS2 analysis by this ARIA-E adjusted 3-level treatment group: donanemab treated with ARIA-E, donanemab treated without ARIA-E, and placebo treated. The model setup and included covariates will be the same as those described for NCS2 in Section 4.3.2.

### 4.4. Secondary Endpoints Analysis

#### 4.4.1. Confirmatory Secondary Endpoints

Additional clinical and outcome measurements listed below will be analyzed separately using NCS2 or MMRM analysis on both the overall population and the intermediate baseline tau subpopulation. Family wise type I error will be controlled for the analyses included in the graphical testing scheme, as described in Section 2.1.

##### 4.4.1.1. Definition of Endpoint(s)

The clinical and outcome endpoints measurements included in the confirmatory secondary analyses are listed below. The details of these endpoint measurements are described in AACI protocol amendment e Section 8.1.2.1 – 8.1.2.4.

- CDR-SB
- ADAS-Cog13 total score
- ADCS-iADL score
- MMSE

##### 4.4.1.2. Main Analytical Approach

MMRM analysis will be applied as the main analytical approach for CDR-SB, with similar model details as described in Section 4.3.3.1. Other than CDR-SB, NCS2 analysis will be applied to the rest of endpoint measurements as the main analytical approach on both the overall

population and the intermediate baseline tau subpopulation separately. The models setup and adjusting covariates included to models will be identical to what described in Section 4.3.2. In addition, CDR-SB will also be tested using NCS2.

### 4.4.2. Slowing in Time of Disease Progression

Time progression models for the repeated measures (Time-PMRM) (Raket 2022) will be used to estimate the slowing of the time progression of the disease due to donanemab treatment, as compared to the time progression in the placebo group. The model will be parametrized by a single parameter describing the proportional time slowing of time progression of the disease in donanemab treated patients. The null hypothesis is that there is no slowing of the time progression of the disease in donanemab treated patients as compared to the patients in the placebo arm. For this analysis, baseline and post-baseline endpoint measurements at the scheduled visits will be used as dependent variables, and the model will include the baseline age, concomitant AChEI or memantine use at baseline (yes/no), and pooled investigator as covariates. Baseline tau category will also be included as a covariate to the model applied to overall population. Planned visit in weeks from randomization will be included as a continuous variable. The intercepts are constrained to be the same between treatment arms considering of the adequate randomization. A natural cubic spline model with internal knots at each planned visit will be used to interpolate the disease progression between the planned visits for the placebo arm and the donanemab treatment trajectory will be estimated assuming the mean disease progression of the treatment group at a given visit can be estimated by the mean disease progression of the placebo group at another time point. Model parameters will be estimated using maximum likelihood estimation, and significance testing will be done using likelihood ratio tests. The assumption of proportional time slowing will be tested and if the assumption is not met, a model similar to the above, but without proportionality assumption, instead having individual time slowing parameters estimated separately at each post-baseline visit will be fitted. This model will be applied to iADRS, CDR-SB, ADAS-Cog13, iADL and MMSE in both the intermediate and overall populations.

### 4.4.3. Biomarker Secondary Endpoints

All the analyses described in this section will be performed on both the overall population and the intermediate tau subpopulation.

#### 4.4.3.1. Analysis of Amyloid PET Scan

Participants' brain amyloid deposition will be measured by amyloid PET imaging, either florbetapir F18, or florbetaben F18 at visits of screening, 24, 52 and 76 weeks. Both scan measurements will be standardized to amyloid centiloid following the specific formula for each tracer below, with details described in the Independent Review Charter (IRC) from PET imaging vendor.

$^{FBP}$ CL = 183.07 * $^{FBP}$ SUVr -177.26

$^{FBB}$ CL = 156.06 * $^{FBB}$ SUVr – 148.13,



Approved on 20 Apr 2023 GMT

Where $^{FBP}$ CL = florbetapir centiloid, $^{FBB}$ CL = florbetaben centiloid, $^{FBP}$ SUVr = florbetapir SUVr, and $^{FBB}$ SUVr = florbetaben SUVr.

The change from baseline to the post-baseline visit of the amyloid imaging centiloid will be evaluated using a MMRM model which includes the fixed, categorical effects of treatment, visit, and treatment-by-visit interaction, as well as continuous effects of baseline centiloid, baseline centiloid-by-visit interaction and age at baseline. Baseline tau category will also be included as a fixed effect to the model applied to overall population. Visit will be considered a categorical variable with values equal to the visit numbers at which amyloid imaging is assessed.

To assess the relationship of biomarker with cognition and function with treatment, Spearman's rank correlation coefficient will be obtained on change from baseline at each follow up visit between centiloid change and change from baseline to Week 76 for iADRS, CDR-SB, ADAS-Cog13, ADCS-iADL, and MMSE. Correlation analyses will be conducted by including patients from both treatment groups, as well as by treatment groups.

#### 4.4.3.2. Analysis of Tau PET Scan

Participant's brain tau deposition will be measured using flortaucipir F18 PET scans. Global tau will be measured as MUBADA (Muti-block Bayrecentric Discriminant Analysis) SUVr, an AD-signature region weighted SUVr and regional tau will be measured at pre-specified region of interest (ROI) including frontal, parietal, and posterior lateral temporal. All SUVr values will be referenced to cerebellar crusteneous region. To evaluate donanemab treatment effect on brain tau accumulation, the change from baseline in tau imaging parameters (including global and regional tau SUVr) will be assessed by an ANCOVA analysis in the Evaluable Efficacy Set (EES). The model will be adjusted by baseline tau SUVr, and age at baseline. Baseline tau category will also be included as a fixed effect to the model applied to overall population.

To assess the relationship of biomarker with cognition and function with treatment, Spearman's rank correlation coefficient will be obtained on change from baseline to Week 76 for the SUVr with change from baseline to Week 76 for iADRS, CDR-SB, ADAS-$Cog_{13}$, ADCS-iADL, and MMSE. Correlation analyses will be conducted using only patients who have the clinical outcome and SUVr result at Week 76 and include patients from both treatment groups, as well as by treatment groups.

#### 4.4.3.3. Analysis of Volumetric MRI

Analyses of the following volumetric MRI (vMRI) parameters will be conducted:

- Bilateral hippocampal volume ($mm^3$)
- Atrophy of total whole brain volume ($cm^3$)
- Enlargement of Ventricular volume ($cm^3$)

To evaluate the changes in vMRI data after treatment, an MMRM model will be used to compare change from baseline to 76 weeks in the EES dataset. The change from baseline to the endpoint visit will be the dependent variable. The model will include the fixed, categorical effect of treatment as well as the continuous effects of baseline vMRI value and age at baseline. Baseline tau category will also be included as a fixed effect to the model applied to overall population.

The null hypothesis is that the difference in LS means between donanemab and placebo equal zero.

## 4.5. Tertiary/Exploratory Endpoints Analysis

### 4.5.1. Analysis of Plasma-based Biomarkers

Donanemab treatment effect will be evaluated with these plasma-based biomarkers:

- Neurofilament Light chain (NfL)
- Glial fibrillary acidic protein (GFAP)
- Phosphorylated tau (P-tau181 and P-tau217)
- Other plasma biomarkers when results become available. These include but not limited to Aβ levels (Aβ 1-42/1-40 ratio) and high sensitivity C-reactive protein (hsCRP)

To evaluate the change from baseline difference by treatment groups, an MMRM analysis will be used to compare change from baseline at 76 weeks in the EES for each of these plasma-based biomarkers. The model will include the fixed, categorical effects of treatment, visit, and treatment-by-visit interaction, as well as baseline value, baseline value-by-visit interaction and age at baseline. Baseline tau category will also be included as a fixed effect to the model applied to overall population. Visit will be considered a categorical variable with values equal to the planned visit numbers at which the plasma-based biomarker is assessed. The null hypothesis is that the difference in LS mean change between donanemab and placebo equals zero. The values for these biomarkers may be log transformed to fit the normality assumption of the model.

To assess the relationship of these biomarkers with cognition and function with treatment, Spearman's rank correlation coefficient will be obtained on change from baseline to Week 76 for these biomarker values and with change from baseline to Week 76 for iADRS, CDR-SB, ADAS-Cog13, ADCS-iADL, and MMSE. Correlation analyses will be conducted using only patients who have the clinical outcome and plasma-based biomarker results at Week 76 and include patients from both treatment groups.

### 4.5.2. Analysis of DSST – Medicines Version

To evaluate the changes in DSST-Medicine version data after treatment, an MMRM model will be used to compare change from baseline to 76 weeks by treatment groups in the EES. The same MMRM analysis as described in Section 4.3.3.1 will be conducted using DSST data, following ITT rule.

### 4.5.3. Analysis of Time to Substantial Decline

For this analysis, the change in CDR scores, both CDR global and CDR-SB, and iADRS as described below will be considered as meeting the criteria of time to substantial decline (MCID, Andrews et al. 2019; Wessels et al. 2022; Lansdall et al. 2023):

1. Any increase in CDR-global score from baseline.
2. 1 point or more increase in CDR-SB from baseline for participants with baseline clinical status as mild cognitive impairment (MCI), or 2 points increase from baseline for participants with baseline clinical status as mild AD.

3. 5 points decrease in iADRS from baseline for participants with baseline clinical status as MCI, or 9 points decrease from baseline for participants with baseline clinical status as mild AD.

The definitions of MCI and Mild for 2) and 3) will be based on the MMSE value at screening. The MCI definition will be a score of 27-30 and the Mild AD definition will be a score of 20-26.

For each of the clinical endpoints as detailed above, a clinical worsening event is defined as meeting the criteria at 2 consecutive visits during the double blinded phase. A Cox proportional hazard (CPH) model will be fit to the EES data to evaluate the hazards of progressing to the defined clinical worsening events by treatment arms. The analysis will be modeling as time to first occurrence of the event as determined above, and adjusting for baseline age, score, and concomitant AChEI and/or memantine use at baseline (yes/no). The model will be stratified by pooled investigator sites. The analyses will be conducted for both overall and baseline intermediate tau level populations. For the analysis of overall population, the model will also be stratified by the Baseline tau category. The ties will be handled using discrete method. The hazard ratio (HR) for donanemab treated group versus placebo group, 95% CI and associated p-value will be provided.

### 4.5.4. "Responder" Type of Analyses

#### 4.5.4.1. Probability of Non-Progressing Post Treatment

To further evaluate the treatment benefit of donanemab, participants' status will be classified as "non progressing" if their CDR-SB change from baseline is less than or equal to 0, which will be calculated as a binary outcome at each of the scheduled visits. A generalized linear mixed model (GLMM) will be applied to assess the difference in probability of "non progressing" by treatment arm. The GLMM model will use the dichotomized "non progressing" status (Yes or No) as dependent variable with a binary distribution option. The model will include these fixed effects: baseline score, baseline score-by-visit interaction, treatment, visit, treatment-by-visit interaction, concomitant AChEI and/or memantine use at baseline (yes/no), and age at baseline. The baseline score-by-visit interaction may be excluded from the model if this term causes a model convergence issue. Baseline tau category will also be included as a fixed effect to the model applied to overall population. Visit will be considered a categorical variable. 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 order 1 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.

The probability of "non progressing" by treatment groups will be compared at each of the scheduled follow up visits. The treatment group contrast in LS mean estimates and its associated p-value and 95% CI will be calculated.

#### 4.5.4.2. Probability of Meeting Prespecified Disease Slowing Criteria by Treatment

A set of criteria will be applied to each patient that will classify whether or not they met a pre-specified percentage of disease slowing at month 12 and 18. The analysis will be conducted for the iADRS and CDR-SB, for both the intermediate tau population and the overall population. The percentage slowing calculation is based on the estimated placebo decline using the NCS2 model for the iADRS and the MMRM model for the CDR-SB, for both the intermediate tau and overall population, respectively. The analysis will be conducted for disease slowing percentages of 50%, 70%, 80%, 90%, 100%, and potentially other percentages.

An example is now provided to identify the patients who had at least 70% slowing. Suppose the NCS2 placebo decline estimate from AACI is 7.3 points at month 12 and 10.5 points at month 18.

- If a patient declines by 2 points or less from baseline on the iADRS at month 12, the patient would have an estimated disease slowing of 100*(1-2/7.3) = 72.6%, which would meet the criterion for at least 70% slowing.
- If a patient declines by 3 points or less from baseline on the iADRS at month 18, the patient would have an estimated disease slowing of 100*(1-3/10.5) = 71.4%, which would meet the criterion for at least 70% slowing.

The same logic will be applied to identify the thresholds of change from baseline for the other disease slowing percentages, and similarly for the CDR-SB.

The probability of meeting the disease slowing percentage criterion at month 12 and 18 by treatment will be compared using a GLIM model as described in Section 4.5.4.1, respectively. The model will be fit separately for each disease slowing percentage.

### 4.5.5. Analysis of PET and Plasma-based Biomarkers by Amyloid Clearance Status

Donanemab antibody targets removal of deposited amyloid plaque. To evaluate the downstream impact of amyloid plaque removal to other AD related biomarkers, including tau PET and plasma-based biomarkers, the study participants will be divided into groups as below, according to treatment and amyloid clearance status by amyloid PET scan results at Week 24:

1. LY-EC (early amyloid complete clearance): donanemab treated and amyloid centiloid value <24.1 by week 24 amyloid PET scan;
2. LY-nEC (not early amyloid complete clearance): donanemab treated and amyloid centiloid value ≥24.1 by week 24 amyloid PET scan;
3. Placebo

ANCOVA analysis with tau PET SUVr as described in Section 4.4.3.2, MMRM analysis as described in Section 4.5.1 with plasma-based biomarkers including P-tau, GFAP, NfL and Aβ

level will be repeated by replacing treatment variable with this treatment/amyloid clearance variable as defined above.

### 4.5.6. Complete Amyloid Clearance

As described in Section 4.5.5, amyloid complete clearance is defined as amyloid centiloid value <24.1. The percent of subjects who meet this complete clearance criteria at each of the scheduled post treatment PET visit will be calculated. A 95% CI for this percentage will be calculated using Wilson score method. In addition, a binomial test will be applied to test whether this percentage equals to 0.

### 4.5.7. Amyloid Reaccumulation Assessment

Donanemab treated participants could switch to placebo during the trial if they meet these criteria: 1) any scheduled posttreatment amyloid PET scan has centiloid <11 or 2) two consecutive scheduled posttreatment amyloid PET scans have centilod value <25. Proportion of participants who meet each of the criteria at the scheduled amyloid PET visits will be summarized. Donanemab-treated subjects who meet these criteria will also be included to assess the amyloid re-accumulation posttreatment switch with MMRM analysis. Amyloid centiloid change from baseline values will be used as the dependent variable, the fixed effect variables will include baseline centiloid value, age, and visits. Baseline tau category will also be included as a fixed effect to the model applied to overall population. The LS mean change estimates at Visit 15 (Week 52) and 21 (Week 76) will be compared to Visit 8 (Week 24) to evaluate the amyloid reaccumulations throughout the study. An unstructured variance-covariance structure matrix will be used to within-subject variance-covariance errors. If the unstructured variance-covariance structure matrix results in a lack of convergence, the following structures will be used in sequence:

- Heterogeneous Toeplitz covariance structure
- Heterogeneous autoregressive order 1 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.

## 4.6. Safety Analyses

### 4.6.1. Extent of Exposure

Summary statistics will be provided for the total number of infusions received per participants. Study drug treatment assignment will be listed.

### 4.6.2. Adverse Events

Treatment-emergent adverse events (TEAEs) will be defined as events that first occurred or worsened after the treatment initiation date. Since participants will continue to long term extension phase (LTE) of Study AACI after Visit 21, the TEAEs for double-blinded phase are

defined as events happened up to either the first visit date of LTE -1 day or end of treatment period in double blinded phase + 57 days, whichever occurs first. 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 LLT during the baseline period will be used as baseline.

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

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

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

SAEs and discontinuations due to AEs will be listed.

### 4.6.3. Deaths, Other Serious Adverse Events, and Adverse Events of Special Interest

An overview of AEs, including the number and percentage of patients who died or experienced SAEs during the study, discontinued due to AEs and who experienced TEAEs, will be provided. Comparison between treatments will be performed using Fisher's Exact Test.

#### 4.6.3.1. Amyloid-Related Imaging Abnormalities (ARIA)

An overview of ARIA incidence will be presented using frequency and percentage of subjects with any ARIA (ARIA-E or ARIA-H), ARIA-E, and ARIA-H as defined by safety MRIs or treatment emergent AE clusters. ARIA-H includes microhaemorrhage and superficial siderosis; macrohaemorrhage will be described separately and not included in the ARIA-H category. The respective TEAE clusters are defined as below:

- ARIA-E: amyloid-related imaging abnormality-oedema/effusion, brain oedema, and vasogenic cerebral oedema.
- ARIA-H: amyloid related imaging abnormality-microhaemorrhages and haemosiderin deposits, brain stem microhaemorrhage, cerebellar microhaemorrhage, cerebral haemosiderin deposit, cerebral microhaemorrhage, and superficial siderosis of central nervous system.
- Macrohemorrhage: cerebral haemorrhage and haemorrhagic stroke.

The incidences will be compared between treatments using Fisher's exact test. The frequency and percentages of ARIA-E will be further broken out by asymptomatic versus symptomatic and by APOE genotype. The frequency and percetage of subjects with ARIA-H microhemorrhage, ARIA-H superficial siderosis, and macrohemorrhage, and co-existing ARIA-E and ARIA-H will be compared separately between treatments and will be further broken out by APOE genotype.

Serious ARIA events will be based on TEAE cluster reported events and MRI although the latter will not be comprehensive as the need to have central MRIs linked to these events may limit the analyses.

The radiographic severity of ARIA-E and ARIA-H is defined according to Table AACI.4.1 and Table AACI.4.2. ARIA events will be summarized by maximum radiographic severity level.

**Table AACI.4.1. ARIA-E Radiographic Severity Classifications**

| Radiographic Severity | ARIA-E Extent |
|---|---|
| 0 (no ARIA-E) | Absence of FLAIR hyperintensity suggestive of ARIA-E |
| 1 (mild) | Mild FLAIR hyperintensity confined to sulcus and/or cortex/subcortex white matter (with or without gyral swelling and sulcal effacement), which affects an area of less than 5 cm in a single greatest dimension. Only a single region of involvement detected. |
| 2 (mild+) | Mild presentation (see 1) in more than one site of involvement |
| 3 (moderate) | Moderate involvement (area of FLAIR hyperintensity measuring 5-10 cm in single greatest dimensions). Only a single region of involvement detected. |
| 4 (moderate+) | Moderate involvement (area of FLAIR hyperintensity measuring 5-10 cm in single greatest dimensions) in more than one site of involvement, each measuring less than 10 cm in a single greatest dimension. |
| 5 (severe) | Severe involvement (area of FLAIR hyperintensity measuring greater than 10 cm in single greatest dimension (white matter and/or sulcal involvement with associated gyral swelling and sulcal effacement)). One or more separate/independent sites of involvement may be noted. |

Abbreviations: ARIA-E = Amyloid Related Imaging Abnormality – Edema; FLAIR = Fluid Attenuated Inversion Recovery.

**Table AACI.4.2. ARIA-H Radiographic Severity Classifications**

| ARIA-H Type | Radiographic Severity | | |
|---|---|---|---|
| | Mild | Moderate | Severe |
| ARIA-H microhaemorrhage | ≤4 treatment-emergent total microhaemorrhages and new incident microhaemorrhages | 5-9 treatment-emergent total microhaemorrhages or new incident microhaemorrhages, whichever is greater | ≥10 treatment-emergent total microhaemorrhages or new incident microhaemorrhages, whichever is greater |
| ARIA-H superficial siderosis | 1 new or increased focal area of superficial siderosis | 2 new or increased focal areas of superficial siderosis | >2 new or increased focal areas of superficial siderosis |

Abbreviations: ARIA-H = Amyloid Related Imaging Abnormality – Haemosiderin.

Shift tables of ARIA-H (microhemorrhage and superficial siderosis) and ARIA-E from baseline by visit will be presented. In addition, a summary and a listing of patients with clinical symptoms associated with ARIA-E will be provided.

Kaplan-Meier plots to describe the onset of the first ARIA-E reported identified by MRI and ARIA-H identified by MRI will be presented in the donanemab treatment group compared with the placebo.

Concomitant antithrombotic drug use was also summarized for participants with and without ARIA. For outputs including antithrombotic drugs, antithrombotic includes all subcategories (aspirin, nonaspirin antiplatelets, anticoagulants, and thrombolytics).

- Aspirin includes platelet aggregation inhibitors excluding heparin. Include those with drug name containing the following in drug name:
  - Acetylsalicylic acid;
  - Acetylsalicylate.
- Nonaspirin antiplatelets include medications with the following Anatomical Therapeutic Chemical (ATC) code: ATC B01AC. However, exclude medications in the Aspirin group.
- Anticoagulants include medications with the following ATC codes: ATC B01AA, ATC B01AB, ATC B01AE, and ATC B01AF
- Thrombolytics include Thrombolytic drugs (Enzymes). Thrombolytics include medications with the following ATC code: ATC B01AD.

ARIA and macrohemorrhage events will also be summarized by sex, age group, and baseline MRI findings. Treatment emergent SAEs, death, discontinuations, and symptomatic events for ARIA and macrohemorrhage will also be summarized. In addition, participants who experienced multiple episodes of ARIA-E, who have resolution of symptoms related to ARIA-E, and first ARIA-E event by donanemab infusion numbers, time to ARIA-E resolution or ongoing ARIA-E based on MRI findings will be summarized. Other treatment-emergent new or worsened MRI findings will be summarized accordingly.

#### 4.6.3.2. Hypersensitivity/Infusion-Related Reactions

Hypersensitivity and Infusion-Related Reactions 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 event occurring either on the same day of drug administration per the AE database or has an associated Hypersensitivity, Anaphylactic, and Infusion Related Reactions Follow-up (HAIRRFU) form that indicates an event within 24 hours of drug administration) 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)

The algorithm portion of the search applies only for the Immediate analysis period. The number and percentage of patients who experienced a TEAE for the following will be analyzed for each of the 2 time periods:

- Any narrow or algorithmic term from any 1 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)

### 4.6.4. Additional Safety Assessments (if applicable)

#### 4.6.4.1. 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 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 ANCOVA model adjusting for baseline value. This analysis will be done separately for each laboratory analyte.

Treatment differences in the proportion of patients 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 patients 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 patient with treatment-emergent high. Similarly, only patients 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 patient with treatment-emergent low. In addition, treatment differences in the proportion of patients who have normal baselines with a change to abnormal high or abnormal low values at any post-baseline visits will be summarized.

For urinalysis parameters, baseline to post-baseline shifts will be summarized at each visit. Fisher's exact test with Monte Carlo estimates 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 patients 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 patients and stratified by low, normal, or high at baseline.

The proportion of patients 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 patients 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.

#### 4.6.4.2. 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) using the safety set.

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 at least 3 minutes in the standing position minus that after at least 5 minutes in the supine position), temperature, and weight by treatment group for all patients in the safety set will be summarized.

Change from baseline to each post-baseline visit at which vital signs are taken will be assessed using an ANCOVA model adjusting for baseline value.

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 post-baseline vital signs and weight are presented in Table AACI.4.3. Any vital sign or weight meeting the criteria will be considered abnormal. Treatment differences in the proportion of patients 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.

**Table AACI.4.3. Potentially Clinically Significant Changes in Vital Signs and Weight**

| Vital Sign Parameter (Unit) | Postbaseline Low Criteria | Postbaseline High Criteria |
|---|---|---|
| Sitting systolic blood pressure (mmHg) | Absolute value ≤90 and ≥20 decrease from baseline | Absolute value ≥160 and ≥20 increase from baseline |
| Sitting diastolic blood pressure (mmHg) | Absolute value ≤50 and ≥10 decrease from baseline | Absolute value ≥100 and ≥10 increase from baseline |
| Sitting pulse (bpm) | Absolute value <50 and ≥15 decrease from baseline | Absolute value >100 and ≥15 increase from baseline |
| Weight | ≥7% decrease | ≥7% increase |
| | | |
| **Vital Sign Parameter (Unit)** | **Postbaseline Criteria for Abnormality** | |
| Orthostatic systolic blood pressure (mmHg) | ≥20 mmHg decrease in systolic blood pressure (supine to standing) (i.e., supine minus standing ≥20) | |
| Orthostatic diastolic blood pressure (mmHg) | ≥10 mmHg decrease in diastolic blood pressure (supine to standing) (i.e., supine minus standing ≥10 mm Hg) | |

| Vital Sign Parameter (Unit) | Postbaseline Low Criteria | Postbaseline High Criteria |
|---|---|---|
| Orthostatic pulse (bpm) | ≥30 increase in bpm (standing to supine) (i.e., standing minus supine ≥30) | |
| Temperature | Absolute value ≥38.3°C and ≥1.1°C increase from baseline<br>(Absolute value ≥101°F and ≥2°F increase from baseline) | |

Abbreviation: bpm = beats per minute.

For each vital sign at each post-baseline visit, only patients 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 patients with treatment-emergent high, low, or abnormal values.

Summary and analyses of change from baseline in weight will be provided. The proportion of patients 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.

#### 4.6.4.3. Electrocardiograms

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

The ECG measurements are derived from three 10 second readings taken every 30 seconds. These 3 readings are to be averaged prior to analysis. Additionally, whenever ECG is measured in triplicate, the average of these readings will be used in the analysis. 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 will be assessed using an ANCOVA model, adjust for baseline ECG value. 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 Table AACI.4.4.

**Table AACI.4.4. Potentially Clinically Significant Changes in ECGs**

| ECG Parameter | Low Criteria | High Criteria |
|---|---|---|
| Heart Rate | <50 bpm | >100 bpm |
| PR Interval | <120 msec | ≥220 msec |
| QRS Duration | <60 msec | ≥120 msec |
| QTcF Interval | | |
| Males | <330 msec | ≥450 msec |
| Females | <340 msec | ≥470 msec |
| Males and females | | >500 msec |

Abbreviations: bpm = beats per minute; ECG = electrocardiogram; QTcF = Fridericia-corrected QT interval.

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

#### 4.6.4.4. Safety MRI

Besides ARIA findings as described in Section 4.6.3.1, treatment-emergent white matter disease and other abnormality findings will be summarized as incidence by treatment assignment, and the incidences between treatment arms will be compared using Fisher's exact test.

#### 4.6.4.5. Immunogenicity

Analyses of immunogenicity data will be covered in a separate immunogenicity statistical analysis plan.

#### 4.6.4.6. 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 patients 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 patients with non-suicidal self-injurious behavior occurring during the treatment period will also be summarized by treatment.

In addition, the number and percent of patients 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 patients who experienced at least one of various composite 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 patient and is used for determining treatment emergence.

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

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 (that is, 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.

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

## 4.7. Other Analyses

This trial is conducted during the Coronavirus (COVID-19) pandemic. The impact of COVID-19 to study will be assessed with the follow analyses:

1. Summary of treatment emergent COVID-19 adverse events, including the discontinuation due to COVID-19;
2. Summary of missed visits due to COVID-19.

### 4.7.1. Subgroup Analyses

To assess the consistency of treatment effects across various demographic and baseline characteristics, the following subgroup analyses may be conducted for the primary and

secondary efficacy endpoint including iADRS, CDR-SB, ADAS-Cog13, ADCS-iADL, MMSE, amyloid centiloid, tau PET, and plasma-based biomarker assessments.

- Age group: <65, 65-74 versus ≥75 years
- Sex: female vs male
- Race: White, black or African American, or Asian
- Ethnicity: Hispanic or Latino versus not Hispanic or Latino
- APOE4 Carrier Status: Carrier defined as E2/E4, E3/E4, or E4/E4 genotype; Non-Carrier defined as all other genotypes
- Number of APOE 4 alleles: 0, 1, or 2 E4 alleles
- Clinical staging at screening – MCI or mild AD
- Baseline brain tau burden category: intermediate vs. high tau
- Baseline tau SUVr tercile groups as defined by screening MUBADA SUVr for overall population: subjects with MUBADA SUVr <33% percentile, MUBADA SUVr within 33%-67% percentiles, and MUBADA SUVr >67% percentile.
- Baseline tau SUVr tercile groups as defined by screening MUBADA SUVr for intermediate tau level population: subjects with MUBADA SUVr <33% percentile, MUBADA SUVr within 33%-67% percentiles, and MUBADA SUVr >67% percentile.
- BMI: <25, 25- <30, ≥ 30

NCS2 analyses will be conducted to assess the subgroup effect for iADRS, CDR-SB, ADAS-Cog13, ADCS-iADL, and MMSE. MMRM analyses will be conducted for amyloid centiloid and plasma-based biomarkers, and ANCOVA analysis will be conducted for tau PET endpoints. The model setup and included covariates are similar to what is described in the corresponding Sections 4.3.2, 4.3.3.1, 4.4.3.1, 4.4.3.2, and 4.5.1, with these additional covariates to include to the NCS models: subgroup by treatment, subgroup-by-basis expansion terms, and subgroup-by-basis expansion-by-treatment interactions. For MMRM models, these additional covariates will be included: subgroup by treatment, subgroup by visit, and treatment by visit by subgroup. The analyses will be done with overall population and with intermediate population separately. The efficacy by subgroups will also be displayed using a forest plot.

In addition, the analyses on iADRS and CDRSB as described in Section 4.3.2 and 4.4.1.2 will be conducted using a subset of participants with intermediate tau and MCI at screening.

## 4.8. Interim Analyses

If any interim analysis is planned, operational details and a quantitative framework to provide information for these decisions will be documented in a later version of this Clinical Trial Statistical Analysis Plan.

### 4.8.1. Data Monitoring Committee

An external DMC is authorized to evaluate results from unblinded interim analyses for the assessment of safety and futility and to recommend any modifications to the study (including stopping the study). Operational details and the decision rules are provided in the DMC charter.

The DMC will have the responsibility to review accumulating unblinded study data and make recommendations to protect the safety of patients. Each member of the DMC is a recognized expert in the fields of Alzheimer's Disease, neurology, cardiology, immunology or biostatistics. All members will be external to the Sponsor. The approved DMC charter enumerates the roles of the DMC members, the frequency with which it meets, and the structure of their meetings. Study sites will receive information about interim results ONLY if relevant for the safety of their patients.

For safety reviews, the DMC will receive data monitoring results that will include at least the following:

- Rates of enrollment and patient discontinuations, including reasons for discontinuation
- Demographic characteristics of enrolled subjects
- Adherence to assigned treatment regimen
- Serious adverse events (SAEs)
- Non-serious adverse events
- Adverse events necessitating unblinding at the site or by the sponsor
- Vital signs data
- Electrocardiographic data
- Central lab data
- Safety MRI data
    - Number of patients with significant treatment-emergent MRI findings, especially Amyloid Related Imaging Abnormalities (ARIA) events such as vasogenic edema or microhemorrhage
    - Listing of all significant treatment-emergent MRI findings
    - For patients with ARIA events, standard listings of medical history, concomitant medications, adverse events, baseline demographics
- CSSRS data
- Immunogenicity/anti-drug antibody data

## 4.9. Changes to Protocol-Planned Analyses

In protocol, the multiplicity of statistical hypothesis testing was planned to be controlled using a chain procedure method (Millen and Dmitrienko 2011). To provide a strong control of overall study-wise type I error rate at 2-sided 0.05 level, a graphical control approach (Bretz et al. 2009, 2011) was developed and described in detail in Section 2.1.

# 5. Sample Size Determination

Approximately 1800 participants will be randomized in the trial. It is anticipated that approximately two-thirds of participants have low-medium tau and approximately one-third of participants have high tau pathology.

The powering and sample size determination of the trial is based on the intermediate tau pathology population. The assumptions for the power calculation were based on the results of the Study AACG data. The mean progression levels in the placebo and donanemab arms from the MMRM analysis on iADRS were -10.06 and -6.86 points (approximately 32% slowing) over 18 months, respectively, with a standard deviation of 11.06. The assumed discontinuation rate of AACI is 30%. Multiple longitudinal data sets were simulated, and the NCS model with 2 degrees of freedom was fit to each sample to determine the power. With a sample size of approximately 1000 randomized participants in the intermediate tau pathology population, the NCS model with 2 degrees of freedom provides greater than 95% power to achieve statistical significance at a 2-sided 0.05 level for the treatment difference relative to placebo, as measured by iADRS at month 18. If both treatment arms are placebo-like with no efficacy, the 2-sided Type I error is 5%.

# 6. Supporting Documentation

## 6.1. Appendix 1: Demographic and Baseline 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. 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. Patient characteristics to be presented include:

- age
- Sex
- race
- Country
- ethnicity (for US and Puerto Rico participants only)
- height
- body weight
- body mass index (weight (kg) / [height (m)]2)
- tobacco use
- alcohol use
- years of education
- work status
- Caffeine use
- time since onset of first AD symptoms
- tau PET burden (MUBADA)
- amyloid PET burden (centiloid)
- time since diagnosis
- 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 CDR-SB, CDR-Global, ADAS-$Cog_{13}$, ADCS-ADL total score and instrumental (ADCS-iADL) and basic subscores (ADCS-bADL), MMSE, and DSST (medicines version).
- Screening MMSE, and the disease stage as determined by the screening MMSE (<20: moderate AD; 20-26: mild AD; 27-28: MCI)

## 6.2. Appendix 2: Treatment Compliance

Because dosing occurs at study visits, patients who attend all visits and successfully receive donanemab or placebo 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.

## 6.3. Appendix 3: 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 patients/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.

# 7. References

Andersen SW, Millen BA. On the practical application of mixed effects models for repeated measures to clinical trial data. *Pharm Stat*. 2013;12(1):7-16. https://doi.org/10.1002/pst.1548

Andrews JS, Desai U, Kirson NY, et al. Disease severity and minimal clinically important differences in clinical outcome assessments for Alzheimer's disease clinical trials. *Alzheimers Dement (N Y)*. 20195(1):354-363. https://doi.org/10.1016/j.trci.2019.06.005

Beyene J, Moineddin R. Methods for confidence interval estimation of a ratio parameter with application to location quotients. *BMC Med Res Methodol*. 2005;5:32. https://doi.org/10.1186/1471-2288-5-32

Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. *Stat Med*. 2009;28(4):586-604. https://doi.org/10.1002/sim.3495

Bretz F, Posch M, Glimm E, et al. Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. *Biom J*. 2011;53(6):894-913. https://doi.org/10.1002/bimj.201000239

Carpenter JR, Kenward MG. Multiple Imputation and its Application. 1st ed. John Wiley & Sons Ltd; 2013.

Donohue MC, Langford O, Insel PS, et al. Natural cubic splines for the analysis of Alzheimer's clinical trials [published online January 10, 2023]. *Pharm Stat*. https://doi.org/10.1002/pst.2285

Kahle-Wrobleski K, Andrews JS, Belger M, et al. Clinical and economic characteristics of milestones along the continuum of Alzheimer's disease: transforming functional scores into levels of dependence. *J Prev Alzheimers Dis*. 2015;2(2):115-120. https://doi.org/10.14283/jpad.2015.53

Kahle-Wrobleski K, Andrews JS, Belger M, et al. Dependence levels as interim clinical milestones along the continuum of Alzheimer's disease: 18-month results from the GERAS Observational Study. *J Prev Alzheimers Dis*. 2017;4(2):72-80. https://doi.org/10.14283/jpad.2017.2

Lansdall CJ, McDougall F, Butler LM, et al. Establishing clinically meaningful change on outcome assessments frequently used in trials of mild cognitive impairment due to Alzheimer's Disease. *J Prev Alzheimers Dis*. 2023;10(1):9-18. https://doi.org/10.14283/jpad.2022.102

Marcus R, Peritz E, Gabriel KR. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*. 1976;63(3):655-660. https://doi.org/10.1093/biomet/63.3.655

McLaughlin T, Feldman H, Fillit H, et al. Dependence as a unifying construct in defining Alzheimer's disease severity. *Alzheimers Dement*. 2010;6(6):482-493. https://doi.org/10.1016/j.jalz.2009.09.004

Millen BA, Dmitrienko A. Chain procedures: a class of flexible closed testing procedures with clinical trial applications. *Stat Biopharm Res*. 2011;3(1):14-30. https://doi.org/10.1198/sbr.2010.09014

Raket LL. Progression models for repeated measures: estimating novel treatment effects in progressive diseases. *Stat Med*. 2022;41(28):5537-5557. https://doi.org/10.1002/sim.9581

Spackman DE, Kadiyala S, Neumann PJ, et al. The validity of dependence as a health outcome measure in Alzheimer's disease. *Am J Alzheimers Dis Other Demen.* 2013;28(3):245-252. https://doi.org/10.1177/1533317513481092

Viele K, Berry S, Neuenschwander B, et al. Use of historical control data for assessing treatment effects in clinical trials. *Pharm Stat*. 2014;13(1):41-54. https://doi.org/10.1002/pst.1589

Wessels AM, Rentz DM, Case M, et al. Integrated Alzheimer's Disease Rating Scale: clinically meaningful change estimates. *Alzheimers Dement (N Y)*. 2022;8(1):e12312. https://doi.org/10.1002/trc2.12312

Wessels AM, Siemers ER, Yu P, et al. A combined measure of cognition and function for clinical trials: the Integrated Alzheimer's Disease Rating Scale (iADRS). *J Prev Alzheimers Dis*. 2015;2(4):227-241. https://doi.org/10.14283/jpad.2015.82

Signature Page for VV-CLIN-075800 v1.0

| | |
|---|---|
| Approval | PPD<br>20-Apr-2023 17:44:34 GMT+0000 |

Signature Page for VV-CLIN-075800 v1.0

# 1. Statistical Analysis Plan: I5T-MC-AACI: Assessment of Safety, Tolerability, and Efficacy of Donanemab in Early Symptomatic Alzheimer's Disease



Generic Name (LY3002813) Early Symptomatic Alzheimer's Disease

Study I5T-MC-AACI (AACI) is a Phase 2, double-blind, placebo-controlled study to evaluate the safety and efficacy of N3pG antibody (donanemab) in patients with early symptomatic AD (prodromal AD and mild dementia due to AD) with the presence of brain tau burden.


Eli Lilly and Company
Indianapolis, Indiana USA 46285


[Protocol I5T-MC-AACI]
[Phase 2]

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

Approval Date: 29-Mar-2020 GMT

# 2. Table of Contents

**Section** **Page**

## Table of Contents

**Appendix** **Page**

# 3. Revision History

SAP Version 1 was approved prior to the first visit when a subject receives study drug.

# 4. Study Objectives

## 4.1. Primary Objective

The primary objective of study AACI is to assess the effect of donanemab versus placebo on clinical progression in participants with early symptomatic Alzheimer's Disease (AD) with demonstrated presence of tau pathology. Clinical progression will be assessed as the change from baseline to Week 76 in the Clinical Dementia Rating (CDR) scale's Sum of Boxes summary score (CDR-SB). Study success will be based on showing a statistically significant difference in the mixed model-repeated measures (MMRM) analyses of changes from baseline to Week 76 in donanemab versus placebo in the overall population and/or the low-medium tau PET subpopulation.

## 4.2. Secondary Objectives

The secondary objectives of AAIC are to assess the effect of donanemab versus placebo on clinical progression, on brain aggregated tau deposition, on brain aggregated amyloid deposition, peripheral PK and presence of anti-donanemab antibodies, and on attenuating downstream markers of the neurodegenerative process in patients with early symptomatic AD and evaluate the safety and tolerability of donanemab.

Clinical progression will be assessed by change from baseline to 76 weeks in cognition and/or function as measured by the following:

- Mini-Mental State Examination (MMSE) score
- Alzheimer's Diasease Assessment Scale—Cognitive subscale (ADAS-Cog13) score
- integrated Alzheimer's Disease Rating Scale (iADRS) score
- Alzheimer's Disease Cooperative Study—instrumental Activities of Daily Living scale (ADCS-iADL) score

Brain aggregated amyloid deposition will be assessed by change from baseline through 76 weeks as measured by florbetapir F 18 PET scan.

Brain aggregated tau deposition will be assessed by change from baseline through 76 weeks as measured by flortaucipir F 18 PET scan.

Attenuation of downstream markers of the neurodegenerative process in AD will be assessed by change from baseline through 76 weeks as measured by volumetric magnetic resonance imaging (vMRI).

Peripheral PK and presence of anti-LY3002813 antibodies (ADA) through 76 weeks will be assessed by summarizing the maximum serum concentration of LY3002813 at steady state (Cmax,ss) and summarizing treatment-emergent ADA against LY3002813 and neutralizing antibodies.

Safety assessments used to evaluate the safety and tolerability of donanemab include the following:

- Spontaneously reported adverse events
- Clinical laboratory tests
- Vital signs and body weigh measurements
- 12-lead ECGs
- Physical and neurological exams
- Safety MRIs
- Infusion-related reactions
- Columbia Suicide Severity Rating Scale (C-SSRS)

## 4.3. Exploratory Objectives

The exploratory objectives described in the protocol for study AAIC are in the following table:

| **Exploratory Objectives** | **Exploratory Endpoints** |
|---|---|
| To assess the effect of donanemab versus placebo on blood-based biomarkers | Plasma<br>• NfL<br>• P-tau<br>• Aβ levels |
| To assess the effect of donanemab versus placebo on cognition | Change in DSST from baseline to Week 76.<br>Change in DSST - Medicines Version from baseline to Week 76 |
| To assess the efficacy of donanemab to prolong time in the current disease state | CDR global score<br>CDR-SB |

# 5. Study Design

## 5.1. Summary of Study Design

Study AACI is a multicenter, randomized, double-blind placebo-controlled, Phase 2 study of donanemab in participants with early symptomatic AD (where early symptomatic AD refers to the combination of 2 stages: MCI-AD and mild AD dementia), MMSE 20-28, and with cerebral tau burden that is elevated, as measured by flortaucipir. The study is intended to further characterize the benefits and risks of treatment with donanemab versus placebo in patients with early symptomatic AD. Participants who meet entry criteria will be randomized in a 1:1 ratio to 1 of the following treatment groups:

- Donanemab: 1400 mg IV Q4W
- Placebo

The duration of the double-blind period of the study is 76 weeks and includes up to 72 weeks of treatment with endpoint measures at the end of the double-blind period (Week 76), to assess the safety, tolerability and efficacy of donanemab versus placebo. In addition to clinical outcomes, imaging biomarkers will also be measured to assess the direct effect of donanemab 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. Amyloid pathology is theorized to be a mediator for clinical decline and therefore it is hypothesized that the removal of amyloid may slow clinical decline.

## 5.2. Determination of Sample Size

Approximately 500 participants will be enrolled and randomized in a 1:1 ratio to the 2 treatment arms (placebo and donanemab). It is expected that approximately 350 subjects will complete the double-blind treatment period of the study (approximately 175 per treatment arm).

This sample size will provide approximately 93% power to demonstrate that the active treatment arm has a $\geq 0.6$ posterior probability of slowing down CDR-SB progression over placebo by at least 0.71 points (i.e., 25% slowing).

The assumption for power calculation is that mean progression levels in the placebo and donanemab arms are approximately 2.85 and 1.71 points (40% slowing) over 18 months, respectively, with common standard deviation of 2.3. These estimates are from the Low/Medium/High tau (defined as baseline composite neocortical tau SUVr >1.1) subset of patients in study AZES. If the active treatment arm is placebo-like with no efficacy, the probability of passing the efficacy criterion specified above (i.e., false positive) is approximately 0.2%. The simulation for the power calculation and sample size determination was carried out in FACTS Version 6.0.

## 5.3. Method of Assignment to Treatment

This is a double-blind study, with design to maintain blinding to treatment. To preserve the blinding of the study, a minimal number of Lilly personnel will see the randomization table and treatment assignments before the study is complete.

Donanemab will be prepared by an unblinded pharmacist or other qualified unblinded personnel, and will be administered by a blinded nurse or other qualified blinded personnel, as described in the pharmacy manual. Site personnel will confirm that they have located the correct packages by entering a confirmation number found on the label into the interactive web-response system (IWRS).

Patients who meet all criteria for enrollment will be assigned a study (patient) number at Visit 601 or Visit 1 and randomized to double-blind treatment at Visit 2. Patients will be randomized to donanemab or Placebo in a 1:1 ratio. For between-group comparability, patient randomization will be stratified by investigative site and tau burden (low-medium versus high). Assignment to treatment groups will be determined by a computer-generated random sequence using an IWRS. Randomization into 1 stratum may be discontinued at the discretion of the sponsor.

# 6. A Priori Statistical Methods

## 6.1. General Considerations

The protocol calls for a Data Monitoring Committee (DMC) charged with making decisions regarding patient safety and study futility. This analysis plan describes analyses for the final clinical study report and all interim analyses for the DMC.

As Study AAIC is a Phase 2 study, the appropriate estimand is a de-jure estimand where efficacy of donanemab is assessed under the paradigm of all patients taking study drug as intended. Intercurrent events for AAIC are defined to be when patients discontinue the study prior to completing the 76 weeks of treatment. After randomization, the protocol does not allow patients to initiate any therapeutic interventions that have demonstrated to be effective in treating AD. Because of this restriction, intiation of standard of care is not considered an intercurrent event. The primary analysis is to use a mixed-model repeated measures (MMRM) analysis of the CDR-SB to compare the cognitive and functional decline between treatment groups at 76 weeks. This MMRM analysis assumes the intercurrent events lead to data that is missing at random (MAR). The MAR assumption will be assessed by the Tipping Point Delta Adjustment analysis described in section 6.12.5.

All analyses will follow the intention-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. Unless otherwise noted, all pairwise tests of treatment effects will be conducted at a 2-sided alpha level of 0.05; 2-sided confidence intervals (CIs) will be displayed with a 95% confidence level. All tests of interactions between treatment and other factors will be conducted at an alpha level of 0.05.

Unless otherwise noted baseline is defined as the last measurement prior to dosing. When change from baseline is assessed, subjects will only contribute to the analysis if both a baseline and a post-baseline measurement are available. Endpoint is the last non-missing post-baseline measurement within the time period for the given analysis. For mixed-effect model for repeated measures (MMRM) models, observations collected at nonscheduled visits will not be included in the analyses (Andersen and Millen 2013). For analyses using last observation carried forward (LOCF), the last nonmissing post-baseline observation (scheduled or unscheduled) will be used to calculate change from baseline.

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

## 6.2. Adjustments for Covariates

The repeated measures models will include the fixed, categorical effects of baseline score, pooled investigator, treatment, visit, treatment-by-visit interaction, baseline-by-visit interaction, concomitant AChEI and/or memantine use at baseline (yes/no), 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, APOE4 status, and appropriate baseline value included as a covariate. When an ANCOVA model is used to analyze a continuous safety variable, the model will contain the main effects of treatment, age, and appropriate baseline value included as a covariate.

## 6.3. Handling of Dropouts or Missing Data

### *6.3.1. Handling Missing Data from Participant 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.

Repeated measures analyses will only use data from visits where the data was scheduled to be collected (Andersen and Millen 2013). When patients discontinue from the study early, there may be efficacy or safety data measurements at visits where the variables were not scheduled to be collected. This data will be used in all other analyses.

### *6.3.2. Handling Missing Items in Calculating Total Scores*

If any of the individual items for the CDR, ADAS-Cog or ADCS-ADL are missing or unknown, every effort will be made to obtain the score for the missing item or items.

For the CDR-SB score, if only 1 box (of 6) of the CDR is missing, the sum of the boxes will be imputed by prorating the sum from the other 5 boxes. If the score from more than 1 box is not available, the CDR-SB at that visit will be considered missing.

For ADAS-Cog13, if 4 or fewer of a total of 13 items are missing, the total score (maximum =85) will be imputed as follows: the total from the remaining items will be multiplied by a factor that includes the maximum score for the missing items. For example, if the first item, "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 to 5 (maximum = 5), is missing, then the multiplication factor = 85/(85 - [10 + 5]) = 85/70 = 1.21. Thus, the total score for this example will be the sum of the remaining 11 items multiplied by 1.21. The imputed number will be rounded up to the nearest integer. If more than 4 items are missing, the total score for ADAS-Cog13 at that visit will be considered missing.

For the ADCS-iADL, if <30% of the items are missing, the total score will be imputed. The sum of the nonmissing items will be prorated to the sum of total items. The imputed number will be rounded up to the nearest integer. If the nearest integer is greater than the maximum possible

score, the imputed score will be equal to the maximum score. If >30% of the items are missing, the total score for ADCS-iADL at that visit will be considered missing. The same imputation technique will be applied to the ADCS-ADL total score. Note that, depending on the specific item responses that are missing, it is possible to have an imputed total score for both the ADCSiADL and the ADCS-ADL, an imputed total score for one but not the other, or both total scores missing.

The iADRS score is calculated as follows: iADRS score = [−1(ADAS - Cog13) + 85] + ADCS-iADL (Wessels et al. 2015). If either ADAS-Cog13 or ADCS-iADL is missing, iADRS score will be considered missing.

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

## 6.4. Multicenter Studies

This study will be conducted by multiple investigators at multiple sites internationally. In the event that any investigator has an inadequate number of subjects (defined as 1 or 0 randomized subjects per treatment group) for the planned analyses, the following strategy will be implemented. Data from all such investigators will be pooled. The pooling will be done first within a country. If the resulting pool within a country is still inadequate (1 or 0 randomized subjects to 1 or more treatment arms), no further pooling will be performed. A listing including country, investigator site with address, number of patients enrolled (randomized) by each site, and unique patient IDs will be presented.

## 6.5. Multiple Comparisons/Multiplicity

The overall population will include all early symptomatic AD patients with elevated tau. This study includes high tau patients, a population not studied in Study AACG.

The primary efficacy objective of Study AACI is to demonstrate donanemab slows the cognitive and/or functional decline in AD versus placebo as measured by the CDR-SB after up to 76 weeks in at least 1 of the following populations: the overall population or the patients with low-medium tau burden.

The testing scheme that will be used to determine statistical significance of the primary analysis and major secondary analyses is described below. All p-values and alpha level thresholds are 1-sided.

The 1-sided p-values and associated test statistics are defined as follows:

- $p_1$ is the low-medium tau burden population p-value; $z_1$ is the associated test statistic
- $p_2$ is the overall population p-value; $z_2$ is the associated test statistic

The hypotheses are defined as follows:

- $H_{low\text{-}med}$ is the null hypothesis corresponding to the low-medium tau burden population
- $H_{overall}$ is the null hypothesis corresponding to the overall population

The low-medium tau burden population hypothesis $H_{low\text{-}med}$ will be rejected if:

- $p_1<0.010$ OR ($p_1<0.025$ AND $p_2<0.0226$)

The overall population hypothesis Hoverall will be rejected if:

- $p_2<0.0226$ OR ($p_2<0.025$ AND $p_1<0.010$)

If either the low-medium tau burden population test or the overall population test is declared significant, the trial will have been considered to have met its primary endpoint.

The preceding testing scheme provides strong control of Type-I error for the study at a 1-sided 0.025 level, based on the closed testing principle (Marcus et al. 1976). The testing scheme is parametric where the known correlation is accounted for between the 2 test statistics $z_1$ and $z_2$.

The initial alpha level for the overall population test is set to 0.0226 and the low-medium tau population alpha will be calculated to control the study Type I error rate at 0.025. In the specific case of 400 low-medium tau burden patients and 100 high tau burden patients, the initial alpha for the low-medium subpopulation will be 0.010. If the overall population test is rejected and the low-medium tau population test is not, or vice versa, the remaining alpha is recycled and used for subsequent testing of the population that was not significant (Millen and Dmitrienko 2011). The alpha levels are based on the assumption that 80% of the overall population are patients with low-medium tau burden.

The correlation between the 2 test statistics defined above representing the low-medium tau only and overall populations is now described and is the basis of the parametric testing procedure. Let $Z_1$ and $Z_2$ denote random variables whose joint distribution is identical to that of $z_1$, $z_2$. Under the null hypothesis of no treatment effect, that is, $Z_1$ and $Z_2$ follow the standard bivariate normal

distribution with the following known correlation:

$$\text{corr}(Z_1 , Z_2) = \sqrt{\frac{400}{500}} = \sqrt{0.8}$$

Using the standard bivariate normal distribution and the defined correlation, the critical values $c_1$ and $c_2$ can be defined to solve the equation $\Pr(Z_1 > c_1 \text{ or } Z_2 > c_2) \leq 0.025$. The alpha levels corresponding to the critical values $c_1$ and $c_2$, are 0.010 and 0.0226, as given above. If the study enrolls a different percentage other than the assumed 80% of low-medium tau patients relative to the overall population, all alpha levels will be adjusted accordingly. If the alpha levels are adjusted due to a different percentage of the low-medium tau only population, then the initial alpha level for the overall population will remain at 0.0226 and the low-medium tau population alpha will be recalculated to control the study Type I error rate at 0.025.

Major secondary hypotheses tested in a similar manner are that donanemab is superior to placebo with regards to:

- Clinical progression in participants with early symptomatic AD at Week 76, as measured by MMSE, ADAS-Cog13, iADRS, and ADCS-iADL

- Sensitivity analyses of the primary outcome
- Health outcomes analyses
- Biomarker analyses

All other hypotheses will compare donanemab to placebo in the overall population.

A graphical strategy may be used for testing key secondary hypotheses to protect against Type I error of falsely rejecting a null hypothesis (Section 6.11.2.). The use of a prespecified analysis plan that employs Bretz' graphical approach will provide strong control of the study-wise Type I error rate for the primary and key secondary hypotheses at level $\alpha$=0.05 (Bretz et al. 2009, 2011). The graph will be added to a future version of the SAP prior to the first unblinding of efficacy data (Interim Analysis 2).

## 6.6. Use of an "Efficacy Subset" of Patients

For purposes of analysis, populations are defined in Table 6.6.1 and Table 6.6.2. These tables also list the study measures that will be summarized and/or analyzed for each population.

**Table 6.6.1. Analysis Populations for Study I5T-MC-AACI**

| Population | Description |
|---|---|
| Entered | All participants who sign informed consent |
| Randomized | All entered participants who are randomized to study treatment |
| Evaluable Efficacy | All randomized participants with a baseline and at least one post-baseline efficacy scale |
| Safety | All randomized participants who are exposed to study drug. Participants will be summarized according to the treatment group to which they were randomized |
| Per-Protocol | All subjects in the Evaluable Efficacy population who also:<br>• signed the inform consent form<br>• had an assessment of the primary endpoint at each scheduled visit completed<br>• had no violations of inclusion/exclusion criteria<br>• 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)<br>• had no unqualified raters and no raters with substantial scoring errors for the primary measure<br>• were not considered non-compliant with regard to study drug |
| Completers | All randomized subjects who have disposition status of 'complete' or have at least 2 weeks exposure in visit interval 21 |

**Table 6.6.2. Efficacy and Safety Measures Summarized and/or Analyzed per Analysis Population**

| **Population** | **Variables Assessed** |
|---|---|
| Entered | Listings |
| Randomized | Tables and listings for patient characteristics, baseline severity, and patient disposition |
| Evaluable Efficacy | Tables, listings, and figures of the following: CDR-SB, ADAS-$Cog_{13}$, ADCS-ADL (basic, instrumental, and total), MMSE, Digit Symbol Substitution Test (DSST), DSST (medicines version), plasma total tau, plasma p-tau, florbetapir parameters, flortaucipir paramters, volumetric MRI measurements, and concomitant medications |
| Safety | Tables, listings, and figures of the following: compliance, adverse events, laboratory results, vital signs, weight, ECG, safety MRIs, C-SSRS |
| Per-Protocol | Tables, listings, and figures of the following: CDR-SB, ADAS-$Cog_{13}$, ADCS-ADL (basic, instrumental, and total), MMSE |
| Completers | Tables, listings, and figures of the following: CDR-SB, ADAS-$Cog_{13}$, ADCS-ADL (basic, instrumental, and total), MMSE, DSST, DSST |

| | |
|---|---|
| | (medicines version), plasma total tau, plasma p-tau, florbetapir parameters, flortaucipir paramters, and volumetric MRI measurements |

## 6.7. Patient Disposition

Because this is a long-term study in a patient population that is elderly with multiple comorbidities, patient withdrawal is of particular concern. Additional efforts will be undertaken to reduce patient withdrawals and to obtain information on patients who are initially categorized as lost to follow-up.

From the randomized population, the percentage of patients withdrawing from each treatment group will be summarized. From the safety population, the percentage of patients withdrawing from each treatment group will be compared between groups using Fisher's exact test. Comparisons using Fisher's exact test will be done for the overall percentage of patients who withdraw and also for each specific reason for withdrawal.

The median time to discontinuation will also be compared between treatment groups using the Kaplan-Meier product limit estimator. For any-cause discontinuation as well as discontinuation due to adverse event or death, comparisons of time-to-discontinuation will be conducted using the Kaplan-Meier product limit estimator and the associated log-rank test.

## 6.8. 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. 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. Patient characteristics to be presented include:

- age
- gender
- race
- ethnicity
- height
- body weight
- body mass index (weight (kg) / [height (m)]2)
- tobacco use
- alcohol use
- years of education
- work status
- time since onset of first AD symptoms
- tau PET burden (various measures)
- time since diagnosis

- 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 CDR-SB, ADAS-$Cog_{13}$, ADCS-ADL total score and instrumental (ADCS-iADL) and basic subscores (ADCS-bADL), MMSE, DSST, and DSST (medicines version). Baseline characteristics and baseline severity will also be listed.

## 6.9. Treatment Compliance

Because dosing occurs at study visits, patients who attend all visits and successfully receive donanemab or placebo 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.

## 6.10. 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. A summary table will also be provided for concomitant AChEI/memantine medications. Medications will be coded using the World Health Organization (WHO) drug dictionary. Concomitant medications will be listed.

## 6.11. Efficacy Analyses

### *6.11.1. Primary Outcome and Methodology*

The primary objective of this study is to test the hypothesis that IV infusion of donanemab will slow the cognitive and/or functional decline of AD as measured by CDR-SB score compared with placebo in patients with early symptomatic AD. This will be assessed using an MMRM analysis of the overall population and a separate MMRM analysis of the low to medium tau subpopulation. The study will be deemed to be positive if at least 1 of these analyses is statistically significant.

The change from baseline score on the CDR-SB at each scheduled postbaseline visit (according to the SoA) during the treatment period will be the dependent variable. The model for the fixed

effects will include the following terms: baseline score, pooled investigator, treatment, visit, treatment-by-visit interaction, baseline-by-visit interaction, concomitant AChEI and/or memantine use at baseline (yes/no), and age at baseline. Visit will be considered a categorical variable. The null hypothesis is that the contrast between the donanemab group versus placebo at the last visit equals 0. 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
- compound symmetry covariance structure

The Kenward-Roger approximation will be used to estimate the denominator degrees of freedom.

The primary time point for treatment comparison will be at Week 76. The treatment group contrast in least-squares mean progression and its associated p-value and 95% CI will be calculated for the treatment comparison of donanemab versus placebo using the MMRM model specified above. In addition, Bayesian posterior probability of the active treatment arm being superior to placebo by at least a margin of interest (25% slowing of placebo progression) will also be calculated using AACG results as the prior.

## *6.11.2. Gated Secondary Efficacy Analyses*

Bretz's graphical approach may be utilized to provide strong control of the study-wise type I error rate for the primary and key secondary hypotheses at alpha level of 0.05 (Bretz et al. 2009, 2011). Details on the final graphical approach and testing strategy will be specified in a future version of the statistical analysis plan (SAP) prior to unblinding efficacy data at an interim analysis assessing efficacy data.

## *6.11.3. Additional Analyses of the Primary Outcome*

All of the analyses described in this section will be performed on both the overall population and the low-medium tau subpopulation.

### 6.11.3.1. Delta Adjustment Tipping Point Analysis

Sensitivity to departures from the missing-at-random (MAR) assumption will be investigated using a tipping point analysis (Carpenter and Kenward 2013). This method is a sensitivity analysis in multiple imputation under the missing-not-at-random (MNAR) assumption that searches for a tipping point that reverses the study conclusion. Departures from MAR in the donanemab treatment group will be assessed assuming that patients who discontinue the study have, on average, efficacy outcomes after discontinuation that are worse by some amount $\delta$ compared to other similar patients with observed data (ie, compared to a value which would have been assumed under an MAR model). A series of analyses will be performed with increasing values of $\delta$ until the analysis conclusion of a statistically significant treatment effect no longer

holds. The value of δ that overturns the primary results will represent a tipping point. An interpretation of clinical plausibility of the assumption underlying the tipping point will be provided.

Mean changes from baseline in CDR-SB scores will be analyzed based on data observed while the patient remains on study as well as data imputed using multiple imputation (MI) methodology for time points at which no value is observed. Imputed values in the donanemab treatment group will first be sampled from an MAR-based multiple imputation model and then δ-adjusted as described below.

Missing-at-random-based imputations will be generated for CDR-SB scores at each time point, and then a value of $\delta = \{\Delta\}$ will be added to all imputed values in the donanemab treatment group prior to analyzing multiply imputed data. This approach assumes that the marginal mean of imputed patient measurements is worse by δ at each time point after discontinuation compared to the marginal mean of patients with observed data at the same time point. Analyses will be conducted with values of δ starting from 0 with increments of 0.10 until the null hypothesis can no longer be rejected.

#### 6.11.3.2. Bayesian Analysis of Shared Control

Sensitivity of comparative inference for slowing the cognitive and/or functional decline of AD based on the shared control may be accomplished via Bayesian mixture modeling. Supplementing the CDR-SB analyses with placebo data from studies I5T-MC-AACG, I8G-MC-LMDC, I8D-MC-AZES, and H8A-MC-LZAX will be explored and potentially based on matching baseline tau PET scan results (Viele, 2014).

#### 6.11.3.3. Disease Progression Model

A disease progression mixed model for repeated measures with a proportional treatment effect will be used to assess statistical differences in the rate of decline of the CDR-SB between the donanemab group and the placebo group. The analysis is testing the hypothesis that the disease cognitive progression ratio (CPR), defined as the rate of decline of the donanemab arm to the rate of decline of the placebo arm, is less than 1.

$$H_0: CPR = 1$$

$$H_1: CPR < 1$$

To test the hypothesis of a cognitive disease progression benefit we calculate the posterior probability of the alternative hypothesis and if it is greater than a pre-specified threshold then the claim of superiority (cognitive disease progression slowing) will be made. A 95% credible interval (from the 2.5th to 97.5th percentiles) and posterior mean and median cognitive disease progression ratio will be presented.

#### 6.11.3.4. Random Slopes Analysis

Slopes of the CDR-SB will be assessed using an MMRM analysis. The change from baseline score at each post-baseline visit during the treatment period will be the dependent variable. The model will include the fixed, categorical effects of treatment, APOE4 status (carrier versus non-

carrier), concomitant AChEI or memantine use at baseline (yes/no), pooled investigator, and continuous effects of baseline score, time, time-by-treatment interaction, and age at baseline. Time will be assumed to be a continuous variable calculated as number of days between baseline and each postbaseline visit (ie, [visit-baseline]+1) during the treatment period. The actual visit dates will be used to calculate number of days (time). The null hypothesis is that the contrasts of slopes of donanemab versus placebo equal zero.

A quadratic slopes model may also be fit to these same scales. The quadratic model would include the linear component of time (TIME) and a quadratic component of time (TIME*TIME), the linear component of time and treatment interaction (TIME*TREATMENT) and quadratic component of time and treatment interaction (TIME*TIME*TREATMENT).

#### 6.11.3.5. Completer Analysis

The primary efficacy outcome, CDR-SB, from the dataset of those patients who remained in the study and on treatment through Week 76 ("completers") will be analyzed using an ANCOVA. The change from baseline at Week 76 will be the dependent variable. The model will include the fixed, categorical effects of treatment, APOE4 status (carrier versus non-carrier), concomitant AChEI use at baseline (yes/no), pooled investigator, and the continuous effects of baseline CDR-SB score and age at baseline. The null hypothesis is that the differences in least-squares means between donanemab and placebo at Week 76 equals zero.

#### 6.11.3.6. Per Protocol Analysis

The primary efficacy outcome, CDR-SB, from the per-protocol dataset will be analyzed using the MMRM analysis from the primary analysis. The change from baseline at each scheduled postbaseline visit will be the dependent variable. The model for the fixed effects will include the following terms (same as primary efficacy analysis): baseline score, pooled investigator, treatment, visit, treatment-by-visit interaction, baseline-by-visit interaction, concomitant AChEI and/or memantine use at baseline (yes/no), and age at baseline. The null hypothesis is that the differences in least-squares means between donanemab and placebo at Week 76 equals zero.

### *6.11.4. Other Secondary Efficacy Analyses*

The additional clinical and outcome measurements listed below will be analyzed separately using an MMRM analysis on both the overall population and the low-medium tau subpopulation. The change from baseline at each scheduled postbaseline visit will be the dependent variable. The model for the fixed effects will include the following terms (same as primary efficacy analysis): baseline score, pooled investigator, treatment, visit, treatment-by-visit interaction, baseline-by-visit interaction, concomitant AChEI and/or memantine use at baseline (yes/no), and age at baseline. The null hypothesis is that the differences in least-squares means between donanemab and placebo at Week 76 equals zero. The outcomes that will be analyzed are:

- Change from baseline in iADRS
- Change from baseline ADAS-$Cog_{13}$ total score
- Change from baseline in ADCS-ADL total score
- Change from baseline in ADCS-iADL score

- Change from baseline in ADCS-bADL score
- Change from baseline in MMSE
- Change from baseline in DSST
- Change from baseline in DSST (medicine version)

## 6.12. Health Outcomes/Quality-of-Life Analyses

Dependence, or the level of assistance required by a patient, has been suggested as a construct for assessing the effect of AD treatment. The process of increasing dependence on others is intended as a complementary measure to existing clinical measures in order to help explain the impact of AD on economic issues such as the risk of institutionalization and caregiver burden (McLaughlin et al. 2010; Spackman et al. 2013). Recently, the ADCS-ADL scores were used to map individuals into 1 of 6 dependence levels (0 to 5): Level 0 – No iADL/bADL impairment; Level 1 – Some supervision needed on isolated iADLs; Level 2 – Supervision on multiple iADLs or loss of at least 1 household activity; Level 3 – Supervision on all types of iADLs or homebound; Level 4 – Supervision on some bADLs; and Level 5 – Impaired transfer or complete incontinence (Kahle-Wrobleski et al. 2015). An approach to transforming continuous functional scale scores into discrete levels of dependence was examined previously in a longitudinal observational study, with preliminary results suggesting acceptable validity and progression in dependence level over time (Kahle-Wrobleski et al. 2017). At baseline, 49.6% of those with mild AD dementia had dependence level 2 and 42.7% were at levels 3 or 4. At 18 months, the proportion of patients at level 2 declined to 31.2% whereas those at levels 3 and 4 rose to 58.8%.

Analyses will be conducted to examine changes in dependence levels across the trial population as well as potential differences on dependence level by treatment group assignment. Treatment differences in dependence levels will be assessed using logistic ordinal regression analysis. The logistic ordinal regression model will include independent variables for baseline dependency level and treatment. The null hypothesis is that the contrast of donanemab versus placebo equals zero. This analysis will be run on the entire population and the low-medium tau subpopulation.

## 6.13. Bioanalytical and Pharmacokinetic/Pharmacodynamic Methods

All of the analyses described in this section will be performed on both the overall population and the low-medium tau subpopulation.

### *6.13.1. Analysis of Florbetapir F 18 PET scan (AV-45)*

At baseline, standard uptake value ratio (SUVr) and centiloid will be calculated using as a ratio of the composite summary region that is an average of 6 different cortical regions (anterior cingulate, posterior cingulate, medial orbital frontal, lateral temporal, lateral parietal, precuneus) with whole cerebellum as a reference region. However, post-baseline SUVr values will be calculated using 2 different reference regions whole cerebellum and a correction factor using atlas based white matter (AWM). The SUVr with whole cerebellum will be calculated as a ratio of composite summary region to whole cerebellum as a reference region, similar to the calculation at baseline. The SUVr values using AWM correction factors will be calculated by

dividing the composite summary ratio by an AWM correction factor. This correction factor is a ratio of SUV values of AWM to whole cerebellum from baseline to post-baseline.

The change from baseline to the post-baseline visit of the SUVr of AV-45 (amyloid imaging) normalized (based on AVID guidelines) will be done using an analysis of covariance (ANCOVA) model with fixed effects of baseline AV-45 result, and treatment. The null hypothesis is that the difference in LSM between donanemab versus placebo equals zero. This analysis will be repeated with centiloid values.

Annualized change in the composite summary SUVr of AV-45 for each patient will be calculated using the change at the last post-baseline visit. The annualized change will be compared between the treatment groups with an ANCOVA. The ANCOVA model will include the following independent variables: baseline AV-45 value and treatment. The null hypothesis is that the difference in LSM between donanemab and placebo equals zero.

To assess the relationship of biomarker with cognition and function with treatment, Spearman's rank correlation coefficient will be obtained on change from baseline to Week 76 for the SUVr normalized to bimodal white matter and with change from baseline to Week 76 for CDR-SB, iADRS, ADAS-$Cog_{13}$, ADCS-ADL, and MMSE. Correlation analyses will be conducted using only patients who have the clinical outcome and SUVr result at Week 76 and include patients from both treatment groups.

### *6.13.2. Analysis of Flortaucipir F 18 PET scan (AV-1451)*

To evaluate the change from baseline in tau imaging parameters, an MMRM analysis will be used to compare change from baseline in SUVr at 76 weeks in the Evaluable Efficacy dataset. The model will include the fixed, categorical effects of treatment, visit, and treatment-by-visit interaction, as well as continuous effects of baseline SUVr and age at baseline. Visit will be considered a categorical variable with values equal to the visit numbers at which tau imaging is assessed. The null hypothesis is that the difference in LSM between donanemab and placebo equals zero.

Change from baseline and annualized change from baseline analyses will be conducted on SUVrs computed from the MUBADA region with the bimodal white matter serving as the reference region. The annualized change will be compared between the treatment groups with an ANCOVA on the full efficacy dataset. The ANCOVA model will include the fixed effect of treatment as well as continuous effects of baseline AV-1451 value and age at baseline. The null hypothesis is that the difference in LSM between donanemab and placebo equal zero.

To assess the relationship of biomarker with cognition and function with treatment, Spearman's rank correlation coefficient will be obtained on change from baseline to Week 76 for the SUVr normalized to bimodal white matter and with change from baseline to Week 76 for CDR-SB, iADRS, ADAS-$Cog_{13}$, ADCS-ADL, and MMSE. Correlation analyses will be conducted using only patients who have the clinical outcome and SUVr result at Week 76 and include patients from both treatment groups.

### *6.13.3. Analysis of Volumetric MRI*

Analyses of the following volumetric MRI (vMRI) parameters will be conducted (right + left for all but whole brain volume and ventricular volume):

- Hippocampal volume ($mm^3$)
- Entorhinal cortex ($mm^3$)
- Inferior parietal lobe ($mm^3$)
- Isthmus cingulate ($mm^3$)
- Lateral parietal lobe ($mm^3$)
- Medial temporal lobe ($mm^3$)
- Precuneus ($mm^3$)
- Prefrontal lobe ($mm^3$)
- Superior temporal lobe ($mm^3$)
- Atrophy of total whole brain volume ($cm^3$)
- Enlargement of Ventricular volume ($cm^3$)

All of the above volumes are corrected for intracranial volume. To evaluate the changes in vMRI data after treatment, an ANCOVA model will be used to compare change from baseline to 76 weeks in the Evaluable Efficacy dataset. The change from baseline to the endpoint visit will be the dependent variable. The model will include the fixed, categorical effect of treatment as well as the continuous effects of baseline vMRI value and age at baseline. The null hypothesis is that the difference in LSM between donanemab and placebo equal zero. A similar analysis will be performed for completers.

Annualized change in vMRI for each patient will be calculated using the change in vMRI at the last post-baseline visit. The annualized change will be compared between the treatment groups with an ANCOVA model on the full efficacy dataset. The ANCOVA model will the include fixed, categorical effect of treatment as well as the continuous effects of baseline vMRI value, and age at baseline. The null hypothesis is that the difference in LSM between donanemab and placebo equals zero.

To assess the relationship of vMRI with cognition and function with treatment, Spearman's rank correlation coefficient will be obtained on change from baseline to Week 76 for vMRI parameters with change from baseline to Week 76 for iADRS, ADAS-$Cog_{13}$, ADCS-ADL, MMSE, and CDR-SB; this will be performed using all patients who have the clinical outcome and vMRI result at Week 76.

## 6.14. Safety Analyses

### *6.14.1. Extent of Exposure*

Days of exposure will be calculated for each patient (date of last dose – date of first dose + 28). Summary statistics will be provided for the total number of days and patient-years of exposure by treatment. Study drug treatment assignment will be listed.

### 6.14.2. 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 patients 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 patients by PT
- Serious adverse events
- Adverse events reported as reason for study treatment discontinuation

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

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

### 6.14.3. Deaths, Other Serious Adverse Events, and Other Notable Adverse Events

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

### 6.14.4. 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 post-baseline visit at which laboratory measurements are taken will be compared between treatment groups using an MMRM model on the Safety Dataset. For each lab analyte, the rank-transformation will be applied to the change from baseline for all patients and all visits prior to analysis. Similarly, an independent rank-transformation will be applied to the baseline values prior to analysis. 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 patients 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 patients 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 patient with treatment-emergent high. Similarly, only patients 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 patient with treatment-emergent low. In addition, treatment differences in the proportion of patients who have normal baselines with a change to abnormal high or abnormal low values at any post-baseline visits 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 patients with notable changes (ie, increases or decreases of a prespecified amount unique to each analyte) from baseline to each postbaseline visit were also summarized for all patients and stratified by low, normal, or high at baseline.

The proportion of patients 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 patients 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 patient'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.

### *6.14.5. 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) using the Safety Dataset.

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 2 minutes and 5 minutes in the standing position), temperature, and weight by treatment group for all patients in the safety population will be summarized.

With the large number of visits at which vital signs are scheduled to be measured, the MMRM model is not suitable for the change from baseline comparison of treatments due to computational challenges. Change from baseline to each post-baseline visit at which vital signs are taken will be assessed using an ANCOVA model with treatment as an independent factor 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 post-baseline vital signs and weight are presented in Appendix 1. Any vital sign or weight meeting the criteria will be considered abnormal. Treatment differences in the proportion of patients 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 patients 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 patients with treatment-emergent high, low, or abnormal values.

Summary and analyses of change from baseline in weight will be provided. The proportion of patients 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.

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

### *6.14.6. Electrocardiograms*

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

The ECG measurements are derived from three 10 second readings taken every 30 seconds. These 3 readings are to be averaged prior to analysis. Additionally, whenever ECG is measured in triplicate, the average of these readings will be used in the analysis. 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 will be assessed using an MMRM model. The model will include the fixed effects of

treatment, visit, and treatment-by-visit interaction as well as continuous effects of baseline ECG score and age at baseline. 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 2.

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

### 6.14.7. Safety MRIs

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.

### 6.14.8. Immunogenicity

The frequency and percentage of subjects 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 patients) will also be summarized.

### 6.14.9. 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 patients 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 patients with non-suicidal self-injurious behavior occurring during the treatment period will also be summarized by treatment.

In addition, the number and percent of patients 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 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.

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 patient and is used for determining treatment emergence.

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

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.

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

## 6.15. Subgroup Analyses

Similar to the primary analysis, each of the secondary efficacy outcomes will be assessed using an MMRM analysis for the overall population and an MMRM analysis for the low-medium tau subpopulation. These secondary efficacy outcomes include ADAS-$Cog_{13}$, ADCS-iADL, and MMSE. For each secondary efficacy measure, the change from baseline score at each scheduled postbaseline visit (according to the SoA) during the treatment period will be analyzed using the same MMRM model described for the primary analysis.

To assess the effects of various demographic and baseline characteristics on treatment outcome, subgroup analyses for the primary endpoint, CDR-SB, will be conducted:

- 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 or mild AD

The primary outcome measure will be modeled using a MMRM approach. This general model will include terms for baseline, treatment, pooled investigator, visit, concomitant AChEI/memantine use at baseline (yes/no), baseline age, treatment by visit, subgroup by treatment, subgroup by visit, and treatment by visit by subgroup. Redundant terms will be dropped from the model in those cases where the subgroup of interest is overlapping with this general model. In order to run these analyses, at least 20 patients are required in each strata-treatment combination.

## 6.16. Protocol Violations

Listings of patients with significant protocol violations will be provided for the Randomized population. 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 an assessment of either the CDR-SB at any of the visits at which the scales were scheduled to be assessed.
- 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 (such as if patients randomized to treatment A were given treatment B or patients randomized to treatment A never received the assigned study drug.)
- Unqualified raters for the CDR.

Other protocol violations 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.

## 6.17. Interim Analyses and Data Monitoring

An external DMC is authorized to evaluate results from unblinded interim analyses for the assessment of safety and futility and to recommend any modifications to the study (including stopping the study). Operational details and the decision rules will be provided in the DMC charter. The DMC will have the responsibility to review accumulating unblinded study data and make recommendations to protect the safety of patients. Each member of the DMC is a recognized expert in the fields of Alzheimer's Disease, neurology, cardiology, or biostatistics. All members will be external to the Sponsor. The approved DMC charter enumerates the roles of the DMC members, the frequency with which it meets, and the structure of their meetings. Study sites will receive information about interim results ONLY if relevant for the safety of their patients.

For safety reviews, the DMC will receive data monitoring results that will include at least the following:

- Rates of enrollment and patient discontinuations, including reasons for discontinuation
- Demographic characteristics of enrolled subjects
- Adherence to assigned treatment regimen
- Serious adverse events (SAEs)
- Non-serious adverse events
- Adverse events necessitating unblinding at the site or by the sponsor
- Vital signs data
- Electrocardiographic data
- Central lab data
- Safety MRI data
    - Number of patients with significant treatment-emergent MRI findings, especially Amyloid Related Imaging Abnormalities (ARIA) events such as vasogenic edema or microhemorrhage
    - Listing of all significant treatment-emergent MRI findings
    - For patients with ARIA events, standard listings of medical history, concomitant medications, adverse events, baseline demographics
- CSSRS data
- Immunogenicity/anti-drug antibody data

At least 1 interim analysis may be conducted for Study AACI; for example, when 50% of randomized subjects have had a chance to complete 52 weeks of treatment (Visit 15), and data will be used to evaluate whether to stop the study for futility. Operational details and a

quantitative framework to provide information for these decisions will be documented in a later version of this Clinical Trial Statistical Analysis Plan.

## 6.18. Planned Exploratory Analyses

### *6.18.1. Exploratory Bioanalytical and PK/PD Analyses*

All of the analyses described in this section will be performed on both the overall population and the low-medium tau subpopulation.

#### 6.18.1.1. 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 at 76 weeks in the Evaluable Efficacy dataset. 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 LSM between donanemab and placebo equals zero. The values for NfL may be log transformed to fit the normality assumption of the model.

Change from baseline and annualized change from baseline analyses will be conducted on NfL. The annualized change will be compared between the treatment groups with an ANCOVA on the full efficacy dataset. The ANCOVA model will include the fixed effect of treatment as well as continuous effects of baseline NfL value and age at baseline. The null hypothesis is that the difference in LSM between donanemab and placebo equals zero.

To assess the relationship of biomarker with cognition and function with treatment, Spearman's rank correlation coefficient will be obtained on change from baseline to Week 76 for the NfL and with change from baseline to Week 76 for CDR-SB, iADRS, ADAS-$Cog_{13}$, ADCS-ADL, and MMSE. Correlation analyses will be conducted using only patients who have the clinical outcome and SUVr result at Week 76 and include patients from both treatment groups.

#### 6.18.1.2. 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 76 weeks. This analysis will be run separately for each plasma tau parameter using the Evaluable Efficacy dataset. The model will include the fixed, categorical effects of treatment, visit, and treatment-by-visit interaction, as well as the continuous effect of baseline plasma tau. 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 LSM between donanemab and placebo equals zero. A similar analysis will be performed for completers.

To assess the relationship of plasma tau with cognition and function with treatment, Spearman's rank correlation coefficient will be obtained on change in plasma tau from baseline to Week 76 and with change from baseline to Week 76 for CDR-SB, iADRS, ADAS-$Cog_{13}$, ADCS-ADL, and MMSE. Correlation analyses will be conducted using only patients who have the clinical outcome and plasma tau result at Week 76.

#### 6.18.1.3. 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 SUVr, cognitive endpoints, 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.).

## 6.19. 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 patients/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.

# 7. References

Andersen SW and Millen BA. On the practical application of mixed effects models for repeated measures to clinical trial data. *Pharm Stat*. 2013;12(1):7-16.

Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. *Stat Med*. 2009;28(4):586-604.

Bretz F, Posch M , Glimm E, Klinglmueller F, Maurer W, Rohmeyer K. Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. *Biom J*. 2011;53(6):894-913.

Carpenter JR, Kenward MG. Multiple Imputation and Its Application. New York, NY: John Wiley & Sons Ltd; 2013.

Kahle-Wrobleski K, Andrews JS, Belger M, Gauthier S, Stern Y, Rentz DM, Galasko D. Clinical and economic characteristics of milestones along the continuum of Alzheimer's disease: transforming functional scores into levels of dependence. J Prev Alzheimers Dis. 2015;2(2):115-120.

Kahle-Wrobleski K, Andrews JS, Belger M, Ye W, Gauthier S, Rentz DM, Galasko D. Dependence levels as interim clinical milestones along the continuum of Alzheimer's disease: 18-month results from the GERAS Observational Study. J Prev Alzheimers Dis. 2017;4(2):72-80.

Marcus R, Peritz E, Gabriel KR. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*. 1976;63(3):655-660.

McLaughlin T, Feldman H, Fillit H, Sano M, Schmitt F, Aisen P, Leibman C, Mucha L, Ryan JM, Sullivan SD, Spackman DE, Neumann PJ, Cohen J, Stern Y. Dependence as a unifying construct in defining Alzheimer's disease severity. *Alzheimers Dement*. 2010;6(6):482-493.

Millen BA, Dmitrienko A. Chain procedures: a class of flexible closed testing procedures with clinical trial applications. *Stat Biopharm Res*. 2011;3(1):14-30.

Spackman DE, Kadiyala S, Neumann PJ, Veenstra DL, Sullivan SD. The validity of dependence as a health outcome measure in Alzheimer's disease. *Am J Alzheimers DisOther Demen.* 2013;28(3):245-252.

Viele, Kert et al. "Use of Historical Control Data for Assessing Treatment Effects in Clinical Trials." Pharmaceutical statistics 13.1 (2014): 41–54. PMC. Web. 9 Oct. 2018.

Wessels AM, Siemers ER, Yu P, Andersen SW, Holdridge KC, Sims JR, Sundell K, Stern Y, Rentz DM, Dubois B, Jones RW, Cummings J, Aisen PS. A Combined Measure of Cognition and Function for Clinical Trials: The Integrated Alzheimer's Disease Rating Scale (iADRS). *J Prev Alzheimers Dis*. 2015;2(4):227-241.

# 8. Appendices

# Appendix 1. Potentially Clinically Significant Changes in Vital Signs and Weight

| Vital Sign Parameter (Unit) | Postbaseline Low Criteria | Postbaseline High Criteria |
|---|---|---|
| Sitting systolic blood pressure (mmHg) | Absolute value ≤90 and ≥20 decrease from baseline | Absolute value ≥160 and ≥20 increase from baseline |
| Sitting diastolic blood pressure (mmHg) | Absolute value ≤50 and ≥10 decrease from baseline | Absolute value ≥100 and ≥10 increase from baseline |
| Sitting pulse (bpm) | Absolute value <50 and ≥15 decrease from baseline | Absolute value >100 and ≥15 increase from baseline |
| Weight | ≥7% decrease | ≥7% increase |
| | | |
| **Vital Sign Parameter (Unit)** | **Postbaseline Criteria for Abnormality** | |
| Orthostatic systolic blood pressure (mmHg) | ≥20 mmHg decrease in systolic blood pressure (supine to standing) (i.e., supine minus standing ≥20) | |
| Orthostatic diastolic blood pressure (mmHg) | ≥10 mmHg decrease in diastolic blood pressure (supine to standing) (i.e., supine minus standing ≥10 mm Hg) | |
| Orthostatic pulse (bpm) | ≥30 increase in bpm (standing to supine) (i.e., standing minus supine ≥30) | |
| Temperature | Absolute value ≥38.3°C and ≥1.1°C increase from baseline (Absolute value ≥101°F and ≥2°F increase from baseline) | |

Abbreviation: bpm = beats per minute.

# Appendix 2. Potentially Clinically Significant Changes in ECGs

| ECG Parameter | Low Criteria | High Criteria |
|---|---|---|
| Heart Rate | <50 bpm | >100 bpm |
| PR Interval | <120 msec | ≥220 msec |
| QRS Duration | <60 msec | ≥120 msec |
| QTcF Interval | | |
| Males | <330 msec | ≥450 msec |
| Females | <340 msec | ≥470 msec |
| Males and females | | > 500 msec |

Abbreviations: bpm = beats per minute; ECG = electrocardiogram; QTcF = Fridericia-corrected QT interval.

Statistical Analysis Plan Amendment History

| SAP Amendments | Substantial Changes |
|---|---|
| Version 1<br>01 Nov 2021 | Original SAP. |
| Version 2<br>29 March 2022 | Study was updated to Phase 3 and additional safety analyses were added for addendum 9 cohort before the first patient visit for addendum 9. |
| Version 3<br>20 Apr 2023 | The primary analysis method (DPM) was changed to natural cubic spline with 2 degrees of freedom based on clinical study protocol amendment e. Time based analyses and final gating scheme were also added. |

Abbreviations: DPM= Disease Progression Models.