I5T-MC-AACG Statistical Analysis Plan Version 2

Assessment of Safety, Tolerability, and Efficacy of LY3002813 in Early Symptomatic Alzheimer's Disease

NCT03367403

Approval Date: 07-Dec-2020

1. Statistical Analysis Plan: I5T-MC-AACG: Assessment of Safety, Tolerability, and Efficacy of LY3002813 in Early Symptomatic Alzheimer's Disease

Confidential Information

The information contained in this document is confidential and the information contained within it may not be reproduced or otherwise disseminated without the approval of Eli Lilly and Company or its subsidiaries.

Note to Regulatory Authorities: This document may contain protected personal data and/or commercially confidential information exempt from public disclosure. Eli Lilly and Company requests consultation regarding release/redaction prior to any public release. In the United States, this document is subject to Freedom of Information Act (FOIA) Exemption 4 and may not be reproduced or otherwise disseminated without the written approval of Eli Lilly and Company or its subsidiaries.

LY3002813 for Early Symptomatic Alzheimer's Disease

Multicenter, randomized, double-blind, placebo-controlled, Phase 2 study comparing up to 1400 mg of LY3002813 with placebo over 76 weeks in approximately 250 patients with early symptomatic AD.

Eli Lilly and Company
Indianapolis, Indiana USA 46285

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

SAP electronically signed and approved by Lilly on date provided below.

Approval Date: 07-Dec-2020 GMT

2. Table of Contents

Section **Page**

3. Revision History

SAP Version 1 was approved on April 20, 2020 prior to unblinding any efficacy data to Lilly senior management or Lilly study team members. SAP Version 2 was approved prior to unblinding of the placebo-controlled period in January 2021 and included the following revisions:

- Added summary of the treatment arm LY-C (6.1.2)
- Added COVID-19 summaries to general considerations (6.1.3)
- Added pooled LY population combining LY-M and LY-C (6.6)
- Revised the multiplicity graph (6.11.2)
- Added summary of the pooled LY population (6.11.5)
- Dropped health outcomes analysis of categorically defined function levels based on the ADCS-ADL
- Added description of hypersensitivity and infusion-related reactions programming (6.13.3)
- Added blinded dose reduction summary (6.13.1)
- Revised clinical lab evaluation analyses (6.13.4)
- Revised analyses of vital signs and weight (6.13.5)
- Revised ECG analyses (6.13.6)

4. Study Objectives

4.1. Primary Objective

Error! Reference source not found.. shows the primary objective and endpoint for the study.

Table AACG.4.1. Primary Objective and Endpoint

Primary Objective	Primary Endpoint
To test the hypothesis that LY3002813 administered for up to 72 weeks will decrease the cognitive and/or functional decline in patients with early symptomatic AD	Change in cognition and function as measured by the change in integrated Alzheimer's Disease Rating Scale (iADRS) score from baseline to 18 months

4.2. Secondary Objectives

Error! Reference source not found.. shows the secondary objectives and endpoints for the study.

Table AACG.4.2. Secondary Objectives and Endpoints

Secondary Objectives	Secondary Endpoints
To assess the effect of LY3002813 vs. placebo on clinical progression in patients with early symptomatic AD	Change from baseline to 18 months as measured by: • change in Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-Cog13) score • change in Clinical Dementia Rating Scale Sum of Boxes (CDR-SB) score • change in Mini Mental State Examination (MMSE) score • change in Alzheimer's Disease Cooperative Study-instrumental Activities of Daily Living scale (ADCS-iADL) score
To assess the effect of LY3002813 vs. placebo on brain amyloid deposition	Change in brain amyloid plaque deposition from baseline through 18 months as measured by florbetapir F18 PET scan
To assess the effect of LY3002813 vs. placebo on brain tau deposition	Change in brain tau deposition from baseline to 18 months as measured by flortaucipir F18

	PET scan
To assess the effect of LY3002813 vs. placebo on brain volume measures	Change in volumetric MRI measures from baseline to 18 months
To evaluate safety and tolerability of LY3002813	Standard safety assessments: • spontaneously reported adverse events (AEs) • clinical laboratory tests • vital sign and body weight measurements • 12-lead ECGs • physical and neurological examinations • MRI (ARIA and emergent radiological findings) • Columbia Suicide Severity Rating Scale (C-SSRS)

4.3. Exploratory Objectives

Error! Reference source not found.. shows the exploratory objectives and endpoints for the study.

Table AACG.4.3. Exploratory Objectives and Endpoints

Exploratory Objectives	Exploratory Endpoints
To assess the effect of LY3002813 vs. placebo on clinical progression in patients with early symptomatic AD	change in dependence level derived from ADCS-ADL scale scores
To assess peripheral PK and presence of anti-LY3002813 antibodies over 72 weeks	• Plasma Pharmacokinetics of LY3002813 • Anti-drug-antibodies (ADA) against LY3002813 including treatment-emergent ADA and neutralizing antibodies.

5. Study Design

5.1. Summary of Study Design

Study I5T-MC-AACG (AACG) is a Phase 2, double-blind, placebo-controlled, study to evaluate the safety and efficacy of N3pG antibody (LY3002813) in patients with early symptomatic AD (prodromal AD and mild dementia due to AD). Study AACG will assess whether removal of existing amyloid plaque can slow the progression of disease as assessed by clinical measures and biomarkers of disease pathology and neurodegeneration over up to 76 weeks of treatment.

Multiple biomarkers of disease progression will also be evaluated. The biomarker florbetapir F18 is a PET ligand that binds to fibrillar amyloid plaque. This biomarker can provide a qualitative and quantitative measurement of brain plaque load in AD patients. Implementation of baseline florbetapir F18 inclusion criteria will provide a screening tool for entry into the clinical trial and provide a confirmation of amyloid pathology. Florbetapir F18 PET also provides quantitative assessment of fibrillar amyloid plaque in the brain and can assess amyloid plaque reductions from the brain by LY3002813.

Alzheimer's disease progression is also associated with cerebral tauopathy, which will be assessed by flortaucipir F18 PET scans, a cortical marker of paired helical filaments. Brain atrophy, as an indicator of neurodegeneration, will be assessed by volumetric Magnetic Resonance Imaging (vMRI).

The patient population for the clinical trial is selected to be AD patients with early symptomatic disease defined both clinically (prodromal to mild AD dementia) and by biomarkers (low to medium tau burden, plus amyloid plaque presence). The 2018 NIA-AA draft AD research framework (Alzheimer's Association 2017a) proposes that AD is defined by its underlying pathophysiologic processes which can be documented in vivo by biomarkers (such as tau and amyloid PET). This framework will enable a more precise approach to therapeutic intervention trials where specific pathways can be targeted at specific points in the disease process and to the appropriate people (Alzheimer's Association 2017b). Clinical-pathological correlations also strongly suggest that baseline imaging that allows staging on the basis of neurofibrillary tangles could substantially improve the power of clinical trials aimed at changing the rate of progression of the disease (Qian et al. 2017). An early AD population defined clinically and pathologically is anticipated to be more homogeneous than populations defined without tau PET, and will be sufficiently early to respond to treatments prior to more advanced irreversible neuronal loss.

5.2. Determination of Sample Size

Approximately 250 subjects will be enrolled and randomized in a 1:1 ratio to the 2 treatment arms (placebo and LY3002813). It is expected that approximately 200 subjects will complete the double-blind treatment period of the study (approximately 100 per treatment arm). This sample size will provide approximately 84% power to demonstrate that the active treatment arm has a $\geq$ 0.6 posterior probability of slowing down iADRS progression over placebo by at least 3 points. The assumption for power calculation is that mean progression levels in the placebo and LY3002813 arms are approximately 12 and 6 points (50% slowing) over 18 months,

respectively, with common standard deviation of 17. 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 6%. 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

Patients who meet all criteria for enrollment will be assigned a study (patient) number at Visit 1 and randomized to double-blind treatment at Visit 2. Patients will be randomized to LY3002813 or Placebo in a 1:1 ratio. For between-group comparability for site factor, patient randomization will be stratified by investigative site. Assignment to treatment groups will be determined by a computer-generated random sequence using an interactive web response system (IWRS).

The IWRS will be used to assign a dosing regimen to each patient. Site personnel will confirm that they have located the correct packages by entering a confirmation number found on the label into the IWRS.

5.4. Blinding

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. The independent external DMC will potentially be unblinded for safety evaluations and dose reduction decisions related to safety. A separate blinding/unblinding plan contains a more detailed description of blinding.

6. A Priori Statistical Methods

6.1. General Considerations

6.1.1. Overall

As Study AACG is a Phase 2 study, the appropriate estimand is a de-jure estimand where efficacy of LY3303560 is assessed under the paradigm of all patients taking study drug as intended. Intercurrent events for AACG are defined to be when patients discontinue the study prior to completing the 76 weeks of treatment or stop taking study drug temporarily during the study. Currently, there is no regulatory-approved, disease-modifying standard of care. Accordingly, initiation 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 iADRS 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 6.11.3.2.

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.

6.1.2. LY3002813 in combination with LY3202626

Study AACG was originally designed as a three arm study comparing LY3002813 alone (LY-M) and in combination with LY3202626 (LY-C) versus placebo (PL). Study AACG was amended to drop the LY-C arm after a small number of patients had been enrolled to one of those three arms. Patients randomized to LY-C were allowed to continue in the study getting monthly infusions of

LY3002813 without the co-administration of the oral LY3202626. For this clinical study report, all analyses of the Randomized Population will include all three arms (LY-M, LY-C, and PL). The remainder of this SAP describe analyses comparing LY-M and PL. The following summaries will be made of LY-C alone:

- Mean change summary of efficacy scales (iADRS, CDR-SB, ADAS-Cog13, iADL, and MMSE), amyloid PET, tau PET, and vMRI
- Exposure summary
- Adverse events (TEAEs, SAEs, and DCAEs) summary
- Labs' boxplots
- Vitals' boxplots
- ECGs' boxplots
- Safety MRI summary
- Immunogenicity summary
- CSSRS summary

6.1.3. COVID-19

Study AACG was fully enrolled at the time the COVID-19 pandemic impacted the countries participating in AACG (US and Canada). Based on the FDA's guidance document regarding the conduct of clinical trials during the COVID-19 pandemic, a patient listing will be created detailing all of the trial participants impacted by COVID-19.

In the primary analysis, only scales administered on-site will be evaluated. Any scale administered via telephone, video conference or at the participant's home will be included in a sensitivity analysis of the primary endpoint. Additionally, if a scale at the final timepoint is administered more than 28 days past the planned timepoint, it will not be included in the primary analysis but will be included in the sensitivity analysis.

As a sensitivity analysis for the primary analysis, the ICH E9 "while on treatment strategy" will be assessed. For this strategy, response to treatment prior to the occurrence of an intercurrent event is of interest. Temporary discontinuation of study drug, including stoppages resulting from the COVID-19 pandemic, is considered the intercurrent event. The primary MMRM analysis will be repeated on all data up to temporary discontinuation of study drug. In other words, patients' data after a temporary discontinuation of study drug will be censored.

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 Patient 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 Totals

If any of the individual items for 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 ADAS-Cog_{13}, 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-Cog_{13} 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 ADCS-iADL and the ADCS-ADL, an imputed total score for one but not the other, or both total scores 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.

If either ADAS-Cog_{13} or ADCS-iADL is missing, iADRS score will be considered missing. The iADRS score is calculated as follows:

iADRS score = [−1(ADAS - Cog_{13}) + 85] + ADCS-iADL (Wessels et al. 2015).

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

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 α=0.05 (Bretz et al. 2009, 2011). The graph is detailed in section 6.11.2.

6.6. Analysis Populations

For purposes of analysis, populations are defined in **Error! Reference source not found.**. and Table 6.6.2. These tables also list the study measures that will be summarized and/or analyzed for each population.

Table AACG.6.1. Analysis Populations

Population	Description
Entered	All participants who sign informed consent
Randomized	All entered patients who are randomized to study treatment
Full Analysis Set	All randomized patients to LY-M or placebo with a baseline and at least one post-baseline iADRS result

Full Analysis Set (Pooled LY)	All randomized patients to LY-M, LY-C or placebo with a baseline and at least one post-baseline iADRS result
Safety	All randomized participants to LY-M or placebo who take at least 1 dose of double-blind study treatment. Participants will be analyzed according to the treatment group to which they were randomized
Safety (Combination Arm)	All randomized participants LY-C who take at least 1 dose of double-blind study treatment.
Per-Protocol	All subjects in the Full Analysis Set who also: • signed the inform consent form • had an assessment of the primary endpoint at each scheduled visit completed • had no violations of inclusion/exclusion criteria • had no study dosing algorithm violation (such as if subjects randomized to treatment A were given treatment B or subjects randomized to treatment A never received the assigned study drug) • had no unqualified raters and no raters with substantial scoring errors for the primary measure • were not considered non-compliant with regard to study drug
Completers	All Full Analysis Set 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
Full Analysis Set	Tables, listings, and figures of the following: iADRS, CDR-SB, ADAS-Cog_{13}, ADCS-ADL (basic, instrumental, and total), MMSE, plasma P-tau217, amyloid PET, tau PET, volumetric

	MRI measurements, and concomitant medications
Full Analysis Set (Pooled LY)	Tables, listings, and figures of the following: iADRS, CDR-SB, ADAS-Cog_{13}, ADCS-ADL (basic, instrumental, and total), MMSE, plasma P-tau217, amyloid PET, tau PET, and volumetric MRI measurements
Safety	Tables, listings, and figures of the following: compliance, adverse events, laboratory results, vital signs, weight, ECG, safety MRIs, C-SSRS
Safety (Combination arm)	Tables, listings, and figures of the following: iADRS, CDR-SB, ADAS-Cog_{13}, ADCS-ADL (basic, instrumental, and total), MMSE, amyloid PET, tau PET, volumetric MRI measurements, 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, plasma P-tau217, amyloid PET, tau PET, 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 summarized. Summaries will be done for the overall percentage of patients who withdraw and also for each specific reason for withdrawal.

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. 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), and MMSE will be summarized. 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. 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

As stated in section 6.1, the appropriate estimand is a de-jure estimand of treatment effect where efficacy of LY3303560 is assessed under the paradigm of all patients taking study drug as intended. The primary analysis is considered a treatment policy strategy as outlined by ICH E9 guidelines. 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 the composite measure iADRS compared with placebo in patients with early symptomatic AD. This will be assessed using an MMRM analysis comparing LY-M vs. PL.

The change from baseline score on the iADRS 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:

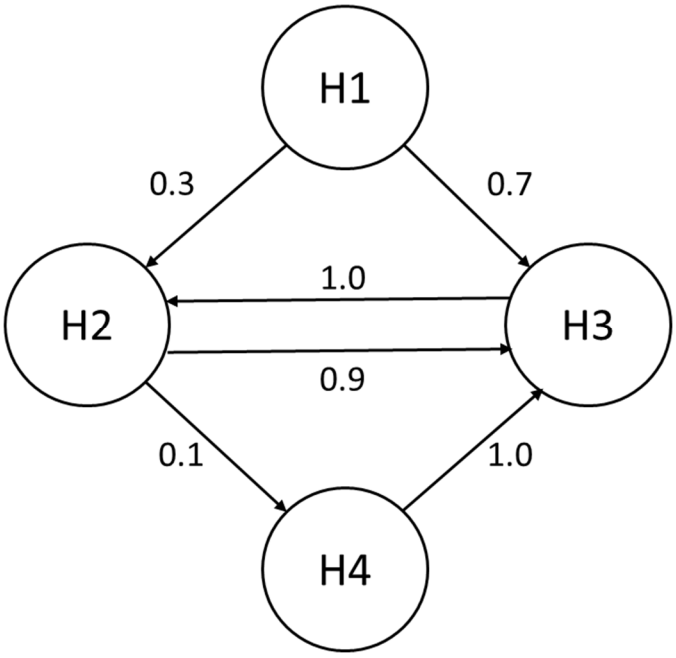
- heterogeneous Toeplitz covariance structure
- heterogeneous autoregressive covariance structure
- heterogeneous compound symmetry covariance structure
- compound symmetry covariance structure

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 a non-informative prior.

6.11.2. Gated Secondary Efficacy Analyses

Bretz's graphical approach will 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). Assuming the primary analysis is statistically significant, the MMRM analyses described for the primary analysis will be conducted on the CDR-SB, ADAS-Cog13, iADL and MMSE scores and statistical significance will be determined based on the following multiplicity graph of hypotheses regarding the IV infusion of donanemab slowing the cognitive and/or functional decline of AD:



H1. MMRM analyisis of CDR-SB at Week 76

H2. MMRM analyisis of ADAS-Cog13 total score at Week 76

H3. MMRM analyisis of iADL subscore at Week 76

H4. MMRM analyisis of MMSE at Week 76

6.11.3. Additional Analyses of the Primary Outcome Comparing LY-M vs. PL

6.11.3.1. ICH E9: While on Treatment Strategy

For this strategy, response to treatment prior to the occurrence of an intercurrent event is of interest. Temporary discontinuation of study drug, including stoppages resulting from the COVID-19 pandemic, is considered the intercurrent event. The primary MMRM analysis will be repeated on all data up to temporary discontinuation of study drug. In other words, patients' data after a temporary discontinuation of study drug will be censored.

6.11.3.2. 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 δ

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 δ 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 iADRS 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 iADRS 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.

6.11.3.3. 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 iADRS analyses with placebo data from studies 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).

6.11.3.4. Disease Progression Model

A Bayesian disease progression mixed model for repeated measures with a proportional treatment effect and noninformative prior will be used to assess statistical differences in the rate of decline of the iADRS 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.

6.11.3.5. Random Slopes Analysis

Slopes of the iADRS 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).

6.11.3.6. 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") 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, concomitant AChEI use at baseline (yes/no), pooled investigator, and the continuous effects of baseline iADRS 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.

6.11.3.7. Per Protocol Analysis

The primary efficacy outcome, iADRS, 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 Comparing LY-M vs. PL

The additional clinical and outcome measurements listed below will be analyzed using an MMRM 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. The outcomes that will be analyzed are:

- Change from baseline in CDR-SB
- 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

6.11.5. Efficacy Analyses of Pooled LY3002813 (Monotherapy plus Combination)

The primary analysis described in 6.11.1 on the iADRS13, CDR-SB, ADAS-Cog13, ADCS-ADL, ADCS-iADL, ADCS-bADL, and MMSE will be repeated on the Full Analysis Set (Pooled LY). Treatment comparisons will be made between LY (pooled LY-M and LY-C) and placebo.

6.12. Bioanalytical and Pharmacokinetic/Pharmacodynamic Methods

6.12.1. Analysis of Amyloid PET

At baseline, standardized uptake value ratio (SUVr) and centiloid value will be calculated 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 (SUVRCAA). The SUVRCAA was also used at visits 8 and 15 in order to determine if the amyloid level had been reduced enough in LY-treated patients to reduce the dose of LY.

To evaluate the amyloid PET longitudinal change from baseline (SUVRCLAA), 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 the donanemab and placebo equals zero.

Annualized change in the composite summary SUVRCLAA 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 fixed effect of treatment as well as continuous effects of baseline amyloid PET and age at baseline. The null hypothesis is that the difference in LSM between donanemab and placebo equals zero.

Spearman's rank correlation coefficient will be calculated to assess the relationship of change in amyloid PET at Week 76 to change from baseline to Week 76 in iADRS, CDR-SB, ADCS-iADL and MMSE by treatment. Partial correlation analyses will be conducted using only patients who have the amyloid PET and scale result at Week 76 and adjusted for APOE4 carrier (Y/N), age and sex. Additionally, a linear regression analysis of change in iADRS versus change in amyloid PET endpoint in centiloids (SUVRCLAA) will be graphed. The graph will include the

regression line, p-value for the explanatory variable (amyloid PET), and color to differentiate patients who are 'amyloid negative' (amyloid PET endpoint < 24.1 centiloids).

Change and annualized change analyses described above will be repeated on the Full Analysis Set (Pooled LY). Treatment comparisons will be made between LY (pooled LY-M and LY-C) and placebo.

6.12.2. Analysis of Tau PET

To evaluate the change from baseline in tau imaging parameters, an MMRM analysis will be used to compare change from baseline in global tau load measurement 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 global tau load measurement 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. The same MMRM analysis will be conducted on the tau PET MUBADA.

Change from baseline and annualized change from baseline analyses will be conducted on global tau load measurement computed from the Tau IQ algorithm. 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 global tau PET 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 by treatment, Spearman's rank correlation coefficient will be obtained on change from baseline to Week 76 for the global tau load computed from the Tau IQ algorithm and with change from baseline to Week 76 for CDR-SB, iADRS, ADAS-Cog13, ADCS-iADL, and MMSE. Partial correlation analyses by treatment group will be conducted using only patients who have the clinical outcome and global tau load measurement at Week 76; adjusted for APOE4 carrier (Y/N), age and sex.

Change and annualized change analyses described above will be repeated on the Full Analysis Set (Pooled LY). Treatment comparisons will be made between LY (pooled LY-M and LY-C) and placebo.

6.12.3. Analysis of Volumetric MRI

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

- Hippocampal volume (cm^3)
- Entorhinal cortex (cm^3)
- Inferior parietal lobe (cm^3)
- Isthmus cingulate (cm^3)
- Lateral parietal lobe (cm^3)
- Medial temporal lobe (cm^3)

- Precuneus (cm^3)
- Prefrontal lobe (cm^3)
- Superior temporal lobe (cm^3)
- Cortical (cm^3)
- Whole temporal lobe (cm^3)
- Atrophy of total whole brain volume (cm^3)
- Enlargement of Ventricular volume (cm^3)

All of the above volumes will be corrected for intracranial volume (ICV) in analyses by including ICV as an explanatory variable in the modeling. To evaluate the atrophy in the volumes, an MMRM analysis will be used 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 volume of interest, baseline ICV, and age at baseline. Visit will be considered a categorical variable with values equal to the visit numbers at which the MRI is assessed. The null hypothesis is that the difference in LSM between donanemab and placebo equals zero.

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, baseline ICV, 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 by 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. Partial correlation analyses will be conducted using only patients who have the clinical outcome and vMRI result at Week 76; adjusted for APOE4 carrier (Y/N), age, baseline ICV and sex; and include patients from both treatment groups.

Spearman's rank correlation coefficient will be calculated to assess the relationship of change in vMRI to week 76 with change in amyloid PET to Week 76 by treatment. Partial correlation analyses will be conducted using only patients who have the PETSCAN and vMRI result at Week 76 and adjusted for APOE4 carrier (Y/N), age, baseline ICV and sex.

Spearman's rank correlation coefficient will be calculated to assess the relationship of change in vMRI to week 76 with change in tau PET to week 76 by treatment. Partial correlation analyses will be conducted using only patients who have the PETSCAN and vMRI result at Week 76 and adjusted for APOE4 carrier (Y/N), age, baseline ICV and sex.

Change and annualized change analyses described above will be repeated on the Full Analysis Set (Pooled LY). Treatment comparisons will be made between LY (pooled LY-M and LY-C) and placebo.

6.13. Safety Analyses

6.13.1. 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. Blinded dose reduction was implemented based on the amount of amyloid measured by the PET scans at weeks 24 and 52. The amounts of amyloid PET in centiloids will be categorized (>= 25, >= 11 and < 25, and < 11) and tabulated at weeks 24, 52, and 76 for patients randomized to LY.

Study drug treatment assignment will be listed.

6.13.2. 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.

TEAEs by decreasing frequency of PT within SOC will be summarized. Summaries of adverse events by decreasing frequency 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 AEs. Treatment comparisons will be carried out using Fisher's Exact Test.

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

6.13.3. 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.

Hypersensitivity and Infusion-Related Reactions (IRR) will be summarized and compared between treatment groups using Fisher's Exact test. Hypersensitivity and IRR will be broken out

between Potential Immediate (defined as TEAEs occurring on the date of infusions) and Potential Non-Immediate (defined as TEAEs not occurring on the date of infusions but prior to the administration of a subsequent infusion). The following will be used to identify such TEAEs:

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

The number and percentage of patients who experienced a TEAE for the following will be analyzed for each of the two time periods:

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

Within query, individual PTs that satisfy the queries will be summarized. For Infusion related reaction (PT), the individual Lower Level Terms (LLTs) will be summarized.

The Anaphylactic reaction SMQ algorithm will be run only for Potential Immediate TEAEs. The SMQ defines a category (A, B, C, D) for each SMQ PT. All Narrow terms have category A, and the occurrence of a Narrow term satisfies the algorithm. Additionally, a pair of PTs *following the same drug administration* satisfies the algorithm if the two events are from different categories (i.e., B&C, or B&D, or C&D). Both contributing events must begin on the same infusion date. Tables will summarize (the number of patients experiencing) each PT that contributes to such an algorithmic pair, and include such terms in the combined narrow and algorithmic search. Broad events that do not contribute to the algorithm will be summarized in a distinct portion of the table.

6.13.4. Clinical Laboratory Evaluation

Laboratory measurements will be analyzed using continuous data (change from baseline) and categorical or ordinal data (proportion of treatment-emergent abnormalities). If there are multiple records of laboratory measurements at baseline or postbaseline visit, the last record will be used. Summaries and analyses of continuous data (change from baseline) will be performed using both conventional and International System of Units (SI units).

Measures of central tendency for planned lab analytes' raw measurements and change from baseline (in CN and SI units) will be summarized with boxplots. Boxplots will display results semi-annually (visits 8, 15, and 21) and for the last visit (LOCF) and will include summary tables of N, mean, median, quartiles, min, max, standard deviation, and p-value (for change scores). If there are considerable missing visits, the measures of central tendency may be based on MMRM analyses.

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) semi-annually (visits 8, 15, and 21) 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 protocol-specified 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.

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.

6.13.5. Vital Signs and Other Physical Findings

Vital sign measurements and weight will be analyzed using continuous data (change from baseline) and categorical data (proportion of potentially clinically significant changes) 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 semi-annual postbaseline visit. Systolic and diastolic blood pressure and pulse (collected in sitting position), orthostatic diastolic and orthostatic systolic blood pressures and orthostatic pulse, temperature, and weight by treatment group for all patients in the safety population will be summarized.

Measures of central tendency for vital sign or weight raw measurements and change from baseline will be summarized with boxplots. Boxplots will display results semi-annually (visits 8, 15, and 21) and for the last visit (LOCF) and will include summary tables of N, mean, median, quartiles, min, max, standard deviation, and p-value (for change scores). If there are considerable missing visits, the measures of central tendency may be based on MMRM analyses.

In order to assess outliers and potentially clinically significant changes from baseline, the number and percent of patients meeting criteria for treatment-emergent abnormalities in vital signs and weight at any time during study will be summarized. Treatment group comparisons will be performed using Fisher's exact test. Baseline is defined as the entire screening period (visits 1 and 2). Vital signs and weight limits, presented in Appendix 1, will be used to define low and high. 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) semi-annually (visits 8, 15, and 21).

A listing of patients with vital signs or weight exceeding limits will be generated.

6.13.6. 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.

Measures of central tendency for ECG raw measurements and change from baseline will be summarized with boxplots. Boxplots will display results semi-annually (visits 8, 15, and 21) and for the last visit (LOCF) and will include summary tables of N, mean, median, quartiles, min, max, standard deviation, and p-value (for change scores). If there are considerable missing visits, the measures of central tendency may be based on MMRM analyses.

In order to assess outliers and potentially clinically significant changes from baseline, the number and percent of patients meeting criteria for treatment-emergent abnormalities in ECGs will be summarized. Treatment group comparisons will be performed using Fisher's exact test. Baseline is defined as the entire screening period (visits 1 and 2). Incidence of treatment-emergent abnormal ECGs will be assessed by comparisons at (1) anytime and (2) semi-annually (visits 8, 15, 22, and 28).

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.

6.13.7. Safety MRIs

The frequency and percentage of subjects with any ARIA (ARIA-E or ARIA-H), ARIA-E, and ARIA-H will be compared between treatments. The frequency and percentages of ARIA-E will be further broken out by asymptomatic vs. symptomatic and by ApoE genotype. A shift table of ARIA-E from baseline by visit will be created, and a shift table of ARIA-H from baseline by visit will be created.

The frequency and percentage of subjects with ARIA-H microhemorrhage, ARIA-H superficial siderosis, and ARIA-H macrohemorrhage will be compared separately between treatments. Changes in ARIA-H microhemorrhages will be further categorized from less than or equal to 4 to greater than 4 by visit.

6.13.8. 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. If the number of subjects experiencing TE-ADA is sufficiently high, further exploratory analyses will be performed using Spotfire to characterize their impact on exposure and clinical outcomes.

6.13.9. 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.

6.14. Subgroup Analyses

To assess the effects of baseline characteristics on treatment outcome, subgroup analyses for the change in iADRS, CDR-SB, and amyloid PET 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 (defined as $27 \leq MMSE \leq 30$) or mild AD (defined as $20 \leq MMSE \leq 26$)
- Tau PET level at baseline – no tau ($SUVR < 1.10$), low tau ($1.10 \leq SUVR < 1.23$), and medium tau ($SUVR \geq 1.23$)

The 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. If there are insufficient numbers of patients across the three tau PET levels at baseline, the no tau and low tau groups will be combined.

6.15. 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 ADAS-Cog or ADL 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 ADAS-Cog or ADL.

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.

6.16. 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 or separate interim Statistical Analysis Plan. 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.

The DMC will meet quarterly to review unblinded safety data. At least 1 interim analysis may be conducted for Study AACG. Operational details and a quantitative framework to provide

information for decisions resulting from the interima analyses will be documented in the interim analysis Statistical Analysis Plan.

6.17. Planned Exploratory Analyses

6.17.1. 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 by 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-Cog13, ADCS-iADL, and MMSE. Partial correlation analyses will be conducted using only patients who have the clinical outcome and biomarker result at Week 76; adjusted for APOE4 carrier (Y/N), age and sex; and include patients from both treatment groups.

6.17.2. Analysis of Plasma P-tau217

To evaluate the change in plasma P-tau analyte (P-tau217) after treatment, an MMRM will be used to compare change from baseline to 76 weeks. This analysis will be run 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 values for P-tau217 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 P-tau217. 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 P-tau217 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 plasma tau with cognition and function by 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-Cog13, ADCS-iADL,

and MMSE. Partial correlation analyses will be conducted using only patients who have the clinical outcome and biomarker result at Week 76; adjusted for APOE4 carrier (Y/N), age and sex; and include patients from both treatment groups.

6.17.3. 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.).

6.18. 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.

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.

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 (supine to standing) (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
QTcF Delta Changes		
i)		> 30 msec
ii)		> 60 msec
iii)		> 75 msec

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