Statistical Analysis Plan I5T-MC-AACQ (3.0)

Investigating the Effect of Different Donanemab Dosing Regimens on ARIA-E and Amyloid Lowering in Adults with Early Symptomatic Alzheimer's Disease

NCT05738486

Approval Date: 29-Jul-2024

Title Page

Statistical Analysis Plan (I5T-MC-AACQ): Investigating the Effect of Different Donanemab Dosing Regimens on ARIA-E and Amyloid Lowering in Adults with Early Symptomatic Alzheimer's Disease

Protocol Title: Investigating the Effect of Different Donanemab Dosing Regimens on ARIA-E and Amyloid Lowering in Adults with Early Symptomatic Alzheimer's Disease

Protocol Number: I5T-MC-AACQ

Compound Number: LY3002813

Short Title: Exploration of Different Donanemab Dosing Regimens

Acronym: TRAILBLAZER-ALZ 6

Sponsor Name: Eli Lilly and Company

Legal Registered Address: Indianapolis, Indiana USA 46285

Regulatory Agency Identifier Number(s)

Ind: 109157

Approved on 29 Jul 2024 GMT

Table of Contents

Version history

This SAP for Study I5T-MC-AACQ is based on the protocol dated 28OCT2022.

SAP Version	Approval Date	Change	Rationale
1.0	27APR2023	Not Applicable	Original version
2.0		• Change reduction in proportion to difference in proportion in Section 1.1.	• Such a change is more aligned with the language in primary objective.
2.0		• In Section 2, define p_1-p_4.	• Needs such definitions.
2.0		• Introduce relative risk in addition to relative risk reduction.	• Highlight the fact a relative risk <1 implies a positive relative risk reduction.
2.0		• Add posterior probability being more than 0.8 to be considered as a significant result.	• SAP (V1.0) did not mention benchmark value of posterior probability to be considered significant.
2.0		• Replace amyloid PET analysis set with evaluable efficacy analysis set.	• Current EES includes any randomized patient with baseline and at least 1 post-baseline measure of interest.
2.0		• Add per-protocol analysis set.	• SAP (V1.0) did not include this analysis set.
2.0		• For the Bayesian model specification, baseline PET values are divided into tertiles.	• In previous SAP, baseline PET in centiloids was a continuous covariate in the logistic regression model
2.0		• Change threshold value to 0.8.	• SAP (V1.0) did not cite this number correctly.
2.0		• Analysis of PET data at 24, 52, and 76 weeks.	• SAP (V1.0) used around 24, 52, and 74 weeks.
2.0		• Minor editing.	• Increase readability.

2.0		• Add PET analysis based on per-protocol analysis set.	• This is a sensitivity analysis for PET results using randomized population.
2.0		• Add an alternative method of CMH test to analyze severity of ARIA-E and ARIA-H.	• Previously, only Fisher's exact test was proposed. CMH test can treat severity as an ordinal measurement.
2.0		• Add statistical analysis plan for exploratory MRI sequence data	• When SAP (V1.0) was approved, no DTS was ready for exploratory data.
2.0		• Add baseline plasma biomarker data in Section 4.5.2.	• Baseline plasma biomarker data need to be considered for analysis itself.
2.0		• Add a list of plasma biomarker measurements collected in AACQ.	• This list clarifies what should be included in plasma biomarker analysis.
2.0		• Add "change from baseline" in the 2nd paragraph of Section 4.5.2.	• "change from baseline" is the parameter to estimate in MMRM analysis for plasma biomarker.
2.0		• For a logistic regression model to predict ARIA rate from change of NfL, baseline amyloid PET, mean arterial pressure, and baseline weight will be added as predictors.	• In the previous version of SAP, such a logistic regression model did not include the following predictors: baseline amyloid PET, mean arterial pressure, and baseline weight.
2.0		• Rewrote methods for exposure analysis (Section 4.6.1).	• Previous version was out of date and did not fit exposure analysis output.
2.0		• Add an analysis regarding overview of ARIA incidence rates using per-protocol analysis set.	• Including this analysis will show robustness of ARIA numbers to protocol deviations.

2.0		• Remove "The anaphylactic reaction SMQ algorithm will be run only for potential immediate TEAEs" from 2nd paragraph in Section 4.6.7.	• The removed sentence was not accurate as the anaphylactic reaction SMQ algorithm will be run on both immediate and non-immediate TEAEs of IRR/hypersensitivity.
2.0		• Power is recalculated assuming discontinuation rate at 10% by Week 24.	• SAP (V1.0) assumed every subject would reach PO lock at Week 24.
3.0		• Focus of baseline models for exploratory endpoints at PO lock will be ARIA-E prediction.	• SAP (V2.0) said both ARIA-E and ARIA-H would be modeled.
3.0		• PVS score from centrum semi ovale will be used.	• Earlier version said using total PVS score.
3.0		• Outline 4 types of baseline exploratory endpoints models.	• Previous SAP version suggested only 1 type of modelling.
3.0		• For exploratory endpoints analysis as well as biomarker analysis, baseline anti-hypertensives will be treated as one predictor. Baseline age and weight will no longer be considered as predictors.	• The models in SAP (V2.0) excluded baseline anti-hypertensives but included age and weight. This was not the exact model outlined by Greenburg et al. (2024).
3.0		• Will use ROC curve and AUC to determine how well exploratory endpoint model classifies subjects with ARIA-E and those without ARIA-E.	• It was not mentioned in previous SAP versions.
3.0		• In Section 6.1, ethnicity will be reported on both	• Previously, ethnicity was required only for US population.

		US and GB populations randomized.	
3.0		• A reference related to baseline risk factors of ARIA-E was added to Section 7.	• It was not published before finalization of previous SAP.

Abbreviations: ARIA = amyloid related imaging abnormalities; AUC = area under the curve; CMH test = Cochran-Mantel-Haenszel test; DTS = data transfer specification; EES = evaluable efficacy set; IRR = infusion related reactions; MMRM = mixed model repeated measures; MRI = magnetic resonance imaging; PET = positron emission tomography; PVS = perivascular space; ROC curve = receiver operating characteristics curve; SMQ = standardized MedDRA query; PO = primary outcome; TEAE = treatment-emergent adverse event; SAP = statistical analysis plan.

1. Introduction

Donanemab is a humanized immunoglobulin G1 antibody developed to remove existing amyloid plaques through microglial-mediated clearance. Amyloid-related imaging abnormalities (ARIA) have been observed with amyloid plaque-targeted therapies, including donanemab. Study I5T-MC-AACQ (AACQ) is a multicenter, randomized, double-blind, Phase 3b study in adults with early symptomatic Alzheimer's Disease (AD) that refers to the combination of 2 stages, mild cognitive impairment and mild AD dementia. Study AACQ will investigate different donanemab dosing regimens and their effect on the frequency and severity of ARIA-E in adults with early symptomatic AD and explore patient characteristics that might predict risk of ARIA.

1.1. Objectives, Endpoints, and Estimands

Objectives	Endpoints
Primary	
To assess the effect of alternative donanemab dosing regimens versus the standard donanemab dosing regimen on ARIA-E frequency	Proportion of participants with any occurrence of ARIA-E by Week 24
Secondary	
To assess the effect of alternative donanemab dosing regimens versus the standard donanemab dosing regimen on ARIA-E frequency	Proportion of participants with any occurrence of ARIA-E by Week 52
To assess the effect of alternative donanemab dosing regimens versus standard donanemab dosing regimen on brain amyloid deposition	Change from baseline in brain amyloid plaque deposition as measured by amyloid PET scan through • Week 24 • Week 52 • Week 76
To evaluate the effect of alternative donanemab dosing regimens on ARIA events	• Proportion of participants with any occurrence of ARIA-H by o Week 24 o Week 52 • ARIA-E and ARIA-H by severity

To assess peripheral PK and presence of anti-donanemab antibodies	• Serum PK of donanemab • ADAs against donanemab, including TE-ADAs and neutralizing antibodies
Tertiary/Exploratory	

To explore characteristics that predict risk of ARIA-E and ARIA-H	• Association between exploratory MRI imaging sequences and the occurrence and severity of ARIA events
	• Association between blood-based biomarkers and the frequency and severity of ARIA events
	  ○ NfL
	  ○ pTau
	  ○ Aß levels (40 and 42)
	  ○ GFAP
	  ○ sTREM2
	  ○ hsCRP

To evaluate the effect of alternative donanemab dosing regimens on IRR	IRR events frequency and severity

Abbreviations: Aß = amyloid-beta peptide; ADA = antidrug antibody; ARIA = amyloid-related imaging abnormalities; GFAP = glial fibrillary acidic protein; hsCRP = high-sensitivity C-reactive protein; IRR = infusion related reaction; NfL = neurofilament light chain; PET = positron emission tomography; PK = pharmacokinetic; pTau = phosphorylated tau; sTREM2=soluble triggering receptor expressed on myeloid cells 2; TE = treatment-emergent.

Primary estimand

The primary clinical question of interest is:

> *What is the frequency of ARIA-E events at Week 24 in participants with early symptomatic AD treated with an alternative donanemab dosing regimen and is there a difference compared to the standard donanemab dosing regimen, regardless of discontinuation or interruption of donanemab dosing for any reason and regardless of change in background AD medication?*

The estimand is described by the following attributes:

- Population - Participants with early symptomatic AD
- Endpoint - Occurrence of ARIA-E by Week 24
- Treatment condition - 4 randomized treatment arms of donanemab with varying dosing schemes

- Population-level summary - The relative difference in proportion of participants with any occurrence ARIA-E events in the alternative donanemab regimen compared to the standard donanemab regimen.
- Intercurrent events - "Discontinuation or interruption of donanemab dosing for any reason" and "change in background AD medication." No other intercurrent events are considered.

The estimand strategy is treatment policy. Data will be collected and analyzed regardless of post randomization events in the first 24 weeks of the trial. This strategy will be implemented by a Bayesian logistic regression model. Rationale for this estimand is to evaluate the effect of alternative donanemab dosing regimens versus the standard donanemab dosing regimen on ARIA-E rate, as the study primary objective stated.

1.2. Study Design

Study AACQ is a multicenter, randomized, double-blind, Phase 3b study of donanemab in participants with early symptomatic AD. There will be 4 treatment arms in this study. All participants will receive a dosing regimen that includes donanemab, but at different dose levels and frequency of dosing. There is no placebo arm in this study, however, placebo is given at specific visits to preserve the blind for different dosing regimens.

The following table describes the planned dosing regimens for the first 16 weeks. At or after Week 16, all participants are targeted to receive 1400 mg of donanemab monthly until they meet the dose stopping criteria or until the end of the study. Before Week 16, total amount of donanemab dosage equals 3500 mg among all 4 treatment arms.

Visit Number	2	3	4	5	6	7	8	9	10
Study Week	0	2	4	6	8	10	12	14	16
Treatment Arm	**Dose (mg)**								
1-standard regimen	700	PBO	700	PBO	700	PBO	1400	PBO	1400
2-dose skipping	700	PBO	PBO	PBO	1400	PBO	1400	PBO	1400
3-titration	350	PBO	700	PBO	1050	PBO	1400	PBO	1400
4-Cmax	350	350	350	350	350	350	700	700	1400

Abbreviations: Cmax = concentration maximum; PBO = placebo.

If a participant meets amyloid plaque reduction criteria defined by the sponsor, as measured by amyloid PET scan, at Visit 12 (Week 24) or Visit 19 (Week 52), the participant will discontinue donanemab and stay in the study for the remaining study visits.

AACQ study duration will be up to approximately 91 weeks, which includes a screening period (7 weeks), a double-blind treatment period (76 weeks), and a follow-up visit that occurs 8 weeks after the last dose of study intervention.

2. Statistical Hypotheses

Suppose p_1, p_2, p_3, and p_4 are ARIA-E rates at Week 24 in the standard dosing arm, the dosing skipping arm, the titration arm and the Cmax arm, respectively. There are 3 parameters associated with the primary objective of Study AACQ: 1) the relative risk reduction in ARIA-E in the 2nd treatment arm (dose skipping regimen) versus the 1st treatment arm (standard dosing regimen) at Week 24, denoted by $RD_{21} = \frac{p_1 - p_2}{p_1} = 1 - \frac{p_2}{p_1}$, where $\frac{p_2}{p_1}$ represents a relative risk between the 2nd treatment arm and 1st treatment arm. A positive relative risk reduction means the relative risk is smaller than 1; 2) the relative risk reduction in ARIA-E in the 3nd treatment arm (titration regimen) versus the 1st treatment arm at Week 24, denoted by $RD_{31} = 1 - \frac{p_3}{p_1}$, where $\frac{p_3}{p_1}$ represents a relative risk between the 3rd treatment arm and 1st treatment arm; 3) the relative risk reduction in ARIA-E in the 4th treatment arm (Cmax regimen) versus the 1st treatment arm at Week 24, denoted by $RD_{41} = 1 - \frac{p_4}{p_1}$, where $\frac{p_4}{p_1}$ represents a relative risk between the 4th treatment arm and 1st treatment arm.

A relative risk reduction in ARIA-E greater than 0 (which implies relative risk smaller than 1) favors alternative treatment arm. The posterior probability of obtaining at least 20% relative risk reduction of ARIA-E comparing the alternative dosing regimen versus the standard dosing regimen by Week 24 is considered as a positive primary outcome. Bayesian logistic regression will yield 3 posterior distributions of RD_{21}, RD_{31}, and RD_{41}. Afterwards, following 3 posterior probabilities will be derived:

$$\Pr(RD_{21} \geq 0.2 \mid \text{24-week ARIA-E data}); \quad (1)$$

$$\Pr(RD_{31} \geq 0.2 \mid \text{24-week ARIA-E data}); \quad (2)$$

$$\Pr(RD_{41} \geq 0.2 \mid \text{24-week ARIA-E data}). \quad (3)$$

Conditional on 24-week ARIA-E data, the computed posterior probability from (1), (2), or (3) will evaluate how likely it is to achieve at least 20% relative risk reduction in ARIA-E comparing the alternative treatment arm with the standard one. A posterior probability >80% will lead to a conclusion that at least 1 alternative dosing regimen reduces ARIA-E risk by at least 20% compared to standard dosing regimen.

2.1. Multiplicity Adjustment

Adjustments for multiplicity will not be applicable for this study.

3. Analysis Sets

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

Participant Analysis Set	Description
Entered	• All participants who sign the informed consent form.
Randomized	• All participants who are randomly assigned to 1 of 4 treatment arms.
Safety analysis set	• All randomized participants who are exposed to at least 1 dose of donanemab. This safety analysis set is used to analyze all the safety endpoints during the treatment period plus safety follow-up.
Evaluable efficacy analysis set	• Randomized participants with baseline and at least 1 post-baseline measure of interest.
Per-Protocol analysis set	All subjects in the Randomized Analysis Set population who also: • signed the inform consent form • 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) • adhered to the dosing titration scheme required in the first 16 weeks of this study • were not considered non-compliant with regard to study drug

4. Statistical Analyses

4.1. General Considerations

As 4 donanemab dosing regimens including the standard one are available for Study AACQ, any safety or biomarker analysis will involve 3 comparisons: 1) dose skipping vs. standard dosing; 2) titration vs. standard dosing; 3) Cmax (concentration maximum) vs. standard dosing.

Safety analyses will include subjects who are randomized to study treatment and who take at least 1 dose of treatment. Subjects will be analyzed according to the treatment arm to which they were assigned.

Unless otherwise specified, analyses for ARIA events are based on MRI findings or AE (adverse events) reporting. ARIA safety analyses will cover both ARIA-E and ARIA-H events by Week 24 as well as by Week 52. Bayesian logistic regression method is proposed to obtain a posterior distribution regarding relative risk reduction in ARIA-E or ARIA-H at Week 24 or 52 for each of 3 comparisons mentioned above. The details of logistic regression model, prior elicitation, and computation for the posterior distribution will be presented in Section 4.3.2. From each derived posterior distribution of RD_{21}, RD_{31}, and RD_{41}, summary statistics will be reported which include posterior mean, posterior median, posterior mode, a 2-sided equal-tailed 95% posterior interval (often cited as credible interval in Bayesian literature), and a posterior probability of achieving a minimum of 20% relative risk reduction in ARIA-E.

Frequentist methods will be used for the rest of the study objectives. All frequentist tests 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.

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.

To allow for evaluation of the primary and secondary hypotheses at 24 and 52 weeks, database locks and the assessments of corresponding primary and secondary objectives are expected to occur after all participants in the Safety Analysis Set have had the opportunity to complete visits 12 and 19 including PET and MRI assessments. The sponsor may choose not to have a separate database lock after visit 19 and wait to perform the analyses for 52-week data until the final database lock.

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 (CSR).

4.2. Participant Disposition

Reason(s) for treatment and study discontinuation will be summarized by treatment arm for all randomized participants. The percentage of participants discontinuing from each treatment arm

will be compared between arms using Fisher's exact test. The comparisons will be done for the overall percentage of participants who discontinue and for select specific reasons for discontinuation.

4.3. Primary Endpoint Analysis

4.3.1. Definition of endpoint(s)

The primary endpoint will be the occurrence of ARIA-E as diagnosed by the scheduled or unscheduled MRIs or reported as treatment-emergent cluster up to and including Visit 12 (Week 24). The scheduled MRIs for the primary endpoint occur at weeks 4, 12, and 24.

4.3.2. Main analytical approach

The number of ARIA-E events and the percentage of participants by treatment arm with any occurrence of ARIA-E events will be tabulated. The percentage of any occurrence of ARIA-E events is equivalent to ARIA-E incidence proportion. The primary analysis will use a Bayesian logistic regression model to estimate relative risk reduction in ARIA-E between alternative donanemab dosing regimen and standard donanemab dosing regimen given ARIA-E data up to Week 24.

The logistic regression model will include categorical effects of treatment arm, APOE e4 genotype, baseline presence of microhemorrhage, baseline presence of superficial siderosis, and baseline amyloid level. Treatment arm is a multinomial variable with 4 levels. APOE e4 genotype status has 3 levels: heterozygous carrier, homozygous carrier, and non-carrier. As participants with ≤4 microhemorrhages at baseline are eligible to enroll into this study, the model needs to adjust for baseline presence of microhemorrhage which is a binary predictor. The proposed model will also assess the impact of baseline presence of superficial siderosis because the study allows enrollment of eligible subjects with 1 area of cortical superficial siderosis. Baseline amyloid values are divided into terciles.

An integral part of Bayesian analysis is the formulation of prior distributions for regression coefficients. A diffuse prior will be assigned to all slope parameters to reflect the impact of the alternative dose regime and the effect of baseline covariates solely based on the data collected from the study. For the intercept parameter, a mixture prior will be using information on ARIA-E rates of the standard donanemab dosing regimen by 24 weeks from AACI – Addendum 9 study. The mixture prior will have a diffuse and an informative component. However, the informative component will have much less variability than the diffuse component. Both prior means are equal to log odds of the ARIA-E incidence in AACI-Addendum 9 study at Week 24. One normal prior has much less variability compared to another normal prior that is relatively flat. 50% weight will be assigned to each component of the mixture a priori.

Through an iterative MCMC (Markov chain Monte Carlo) sampling method, a posterior distribution of ARIA-E incidence in Arm 1 (p_1), Arm 2 (p_2), Arm 3 (p_2), or Arm 4 (p_4) at Week 24 will be simulated knowing the likelihood function of the logistic regression and prior density of individual regression coefficient. Afterwards, a posterior distribution of relative risk reduction in ARIA-E such as RD_{21}, RD_{31}, or RD_{41} as defined in Section 2 will be generated as well.

Posterior summary statistics, like posterior mean, posterior median, 2-sided equal-tailed 95% of posterior interval of RD_{21}, RD_{31}, or RD_{41} may be reported. The posterior probability of RD_{21}, RD_{31}, or RD_{41} being no smaller than 20% will be presented as well. The Bayesian logistic regression model can be implemented via both MCMC procedure in SAS and rjags package in R.

4.3.3. Sensitivity Analyses

One type of sensitivity analysis that will be conducted is to assess how a posterior probability described in (1), (2), or (3) in Section 2 will be changed if the margin or threshold is changed for relative risk reduction in ARIA-E. The calculated posterior probability regarding effect of interests will be computed.

One such analysis will be to compute a posterior probability of RD_{21}, RD_{31}, or RD_{41} being greater than 0. This is analogous to testing any positive relative risk reduction when a frequentist logistic regression was applied. Suppose from a posterior distribution of RD_{21},

$$\Pr(RD_{21} > 0 \mid \text{24-week ARIA-E data}) > 0.8,$$

the conclusion would be compared to the standard dosing arm, the dosing skip arm had a significant positive relative risk reduction of ARIA-E.

Another type of sensitivity analysis will be to test how robust the yielded posterior probabilities will be if a neutral prior is assigned to the intercept coefficient rather than the mixture prior stated in Section 4.3.2.

4.4. Secondary Endpoints Analysis

4.4.1. 12-month and 18-month incidence of ARIA-E

For the analysis of 52-week ARIA-E incidence, all 4 scheduled MRIs (weeks 4, 12, 24, 52) and unscheduled MRI data as well as treatment-emergent cluster data will be used first to obtain descriptive statistics including total number and proportion of patients with ARIA-E in 1-year in AACQ and by treatment arm.

Like the method for analyzing the primary endpoint, a Bayesian logistic regression will be applied before answering whether relative risk reduction of ARIA-E in 52 weeks is significant comparing the dosing skip arm with standard dosing arm, or comparing the titration arm with standard dosing arm, or comparing the Cmax arm with standard dosing arm. A diffuse prior will be used for each slope parameter in this Bayesian logistic regression model and a mixture prior for the intercept parameter will be elicited from 52-Week ARIA-E outcome in AACI-Addendum 9 study. The probability that one alternative dosing arm has a minimum of 20% risk reduction in ARIA-E in one year compared to the standard dosing arm will be derived from each of 3 simulated posterior distributions in relative risk reductions. In addition, the sensitivity analyses proposed in Section 4.3.3 will be used for 52-week analysis of ARIA-E.

All scheduled postbaseline MRIs (weeks 4, 12, 24, 52, and ED) and unscheduled MRI data as well as treatment-emergent cluster data will be analyzed for ARIA-E events up to Week 76 using the framework in Section 4.3.2 or following the descriptions for the analysis of 1-year incidence of ARIA-E.

4.4.2. Amyloid plaque deposit

Participants' brain amyloid deposition will be measured by amyloid PET imaging at visits of screening, 24, 52 and 76 weeks. The frequency and percentage of subjects who achieved amyloid PET clearance (< 24.1 centiloids) in AACQ will be reported. And the count and percentage of subjects who meet dose stopping criteria either due to amyloid < 11 centiloids on one occasion or amyloid between 11 and 25 centiloids on two consecutive measurements will be presented at Week 24, Week 52, or Week 76.

The change from baseline to the post-baseline visit of the amyloid imaging centiloid will be evaluated using a MMRM (mixed model repeated measures) model which includes the fixed effects of treatment arm, time, treatment arm-by-time interaction, baseline age, baseline amyloid level, and baseline amyloid-by-time interaction. Treatment arm, time, as well as treatment arm-by-time interaction are the categorical effects and baseline age, baseline amyloid level, as well as baseline amyloid-by-time interaction are continuous effects. The within-patient covariance matrix of the MMRM is an unstructured one with no assumption made about the variances and covariances. 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 difference in mean change in the amyloid plaque level between treatment arms at 24, 52, and 76 weeks from baseline will be obtained through the least squares (LS) method. Its associated SE (standard error), 95% of confidence interval, and *p*-value derived from *t* test statistic will be reported as well.

The amyloid PET analysis described above will be based on the EES. A sensitivity PET analysis will be conducted using the per-protocol analysis set.

4.4.3. ARIA-H incidence and severity of ARIA events

ARIA-H includes microhemorrhage and superficial siderosis. Total number of participants with microhemorrhage, superficial siderosis, as well as overall ARIA-H events will be reported at Week 24 and Week 52 by treatment arm. Incidence proportion of microhemorrhage, superficial siderosis, and overall ARIA-H per arm will be estimated from both time points. A posterior distribution of relative risk reduction in ARIA-H between one alternative dosing regimen and the standard dosing regimen by Week 24 or by Week 52 will be derived through a Bayesian logistic regression using 24-week or 52-week ARIA-H data coupled with priors constructed in a similar way as mentioned in Section 4.3.2 or Section 4.4.1. Then a posterior probability of at least 20% relative risk reduction in ARIA-H in 24 weeks or 52 weeks comparing one alternative treatment

arm versus the standard arm will be calculated. Sensitivity analyses will be similar to approaches as mentioned in Section 4.3.3.

For outputs including radiographic severity, below table contains the definitions used for 3-point ARIA-E and ARIA-H categorizations. A 5-point ARIA-E categorization has mild+ and moderate+ levels besides mild, moderate, and severe levels defined by the 3-point ARIA-E categorizations. Among participants with at least 1 ARIA-E or ARIA-H event during the study, number and percentage of participants with their maximum radiographic severity value in the categories of mild, moderate, and severe will be presented by treatment arm from 24-week and 52-week ARIA data when a 3-point rating of ARIA severity is used. Among participants with at least 1 ARIA-E event during the study, number and percentage of participants with their maximum radiographic severity value using the 5-point severity scale will be reported by treatment arm for both times points as well. At Week 24 and Week 52, Fisher's exact tests will be applied to determine if severe ARIA-E or ARIA-H events are distributed evenly while comparing the dosing skip regimen vs. standard dosing regimen, the titration regimen vs. standard dosing regimen, or the Cmax regimen vs. standard dosing regimen.

An additional approach to Fisher's exact test will be used to investigate the association between treatment arm and ARIA (-E or -H) severity with a Cochran-Mantel-Haenszel (CMH) test. Fisher's exact test treats severity as a nominal variable whereas CMH can handle severity as an ordinal variable.

Table 4.1. Radiographic severity of ARIA-E and ARIA-H.

ARIA Type	Radiographic Severity	Definition
ARIA-E	Mild	FLAIR hyperintensity confined to sulcus and/or cortex/subcortex white matter in one location <5 cm.
	Mild+	Mild presentation in more than one site of involvement
	Moderate	FLAIR hyperintensity 5 to 10 cm in single greatest dimension, or more than 1 site of involvement, each measuring < 10 cm.
	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.
	Severe	FLAIR hypersensitivity > 10 cm with associated gyral swelling and sulcal effacement. One or more separate/independent sites of involvement may be noted.

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

Abbreviations: ARIA-E = amyloid-related imaging abnormalities- edema/effusions; ARIA-H = amyloid-related imaging abnormalities- hemorrhage/hemosiderin deposition; FLAIR = fluid attenuated inversion recovery.

4.4.4. PK/PD analysis

PK/PD analyses will be covered in a separate PK/PD statistical analysis plan.

4.4.5. 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 by treatment arm. 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 to characterize their impact on exposure and clinical outcomes.

.

4.5. Tertiary/Exploratory Endpoints Analysis

4.5.1. Exploratory MRI sequences

The table below lists exploratory MRI sequences, and pre-specified endpoints for exploratory analyses, as well as scheduled visits for exploratory endpoints. Unscheduled MRIs will be

included for analyses of exploratory endpoints. Information about four dosing regimens will be combined before performing any analysis here.

Table 4.2. Prespecified endpoints for exploratory MRI sequences.

MRI Sequences	Scheduled Visit	Endpoints
3D FLAIR	2, 4, 8, 12, 19	• ARIA-E presence • ARIA-E severity • White matter disease burden
3D T2	2, 4, 8, 12, 19	• PVS (perivascular spaces) score in centrum semi ovale
SWI	2, 4, 8, 12, 19	• Number of new and total ARIA-H microhemorrhages • Number of total ARIA-H superficial siderosis
DTI	2, 4, 8, 12, 19	• Whole cortex MD (mean diffusivity), and FW (free water fraction) • Hippocampal MD and FW • Whole white matter FA (fractional anisotropy)
fMRI	2	• Peak activation in occipital cortex (top 200 voxels based on z-scored activation maps)
vMRI	2, 4, 8, 12, 19	• Hippocampal volume, cortical volume, ventricle volume, whole brain volume (all normalized for intracranial volume) • Cortical thickness
FLAIR 3D, 3DT2, SWI	2	• Vascular risk score for CAA based on Boston 2.0 Criteria

Abbreviations: ARIA = amyloid-related imaging abnormalities; CAA = cerebral amyloid angiopathy; DTI = diffusion tensor imaging; FLAIR = fluid attenuated inversion recovery; FW = free water fraction; fMRI = functional magnet resonance imaging; MD = mean diffusivity; PVS = perivascular spaces; SWI = susceptibility weighted imaging; vMRI = volumetric magnet resonance imaging.

Prediction of ARIA

Logistic regression analyses will be used to assess four types of baseline models to predict ARIA-E incidence at Week 24:

1) The model will include baseline risk factors of ARIA-E: APOE e4 genotype, number of microhemorrhages at baseline (determined by standard safety sequences), presence of superficial siderosis at baseline (determined by standard safety sequences), amyloid centiloid at baseline, mean arterial pressure, and anti-hypertensives.

2) The model will include all baseline exploratory MRI endpoints discussed before. The following markers will also be considered for inclusion to the model: white matter disease burden from 3D FLAIR, number of microhemorrhages at baseline form SWI, cortical superficial siderosis from SWI, PVS in centrum semi ovale, whole cortex MD, whole cortex FW, hippocampal MD and FW, whole white matter FA, fMRI peak activation in occipital cortex, hippocampal volume, cortical volume, ventricle volume, whole brain volume, cortical thickness, and CAA.

3) Exploratory MRI endpoints listed above will be added individually to the baseline ARIA-E risk model described in 1.

4) The model will include all baseline risk factors described in 1) and all exploratory endpoints described in 2).

AIC (Akaike information criteria) as well as BIC (Bayesian information criteria) statistics will be generated for each proposed model above. While both the AIC and BIC will be generated, the BIC will be used to evaluate the model. In the case of SWI endpoints, the counts of microhemorrhages and superficial siderosis will come from baseline SWI instead of its standard safety MRI equivalent. In the case of 3D FLAIR, white matter burden score will come from 3D FLAIR instead of the standard safety sequence MRI equivalent. An ROC curve (receiver operating characteristic curve) can be plotted from individual ARIA-E risk prediction model and AUC (area under the curve) will be computed to tell how each model is capable of distinguishing between ARIA-E subjects and ARIA-E free subjects. A possible AUC value is between 0.5 and 1, and the larger the AUC, the better the discriminatory ability.

Comparison between standard and exploratory MRI sequences

To assess the agreement of exploratory versus standard MRI sequences at baseline, a contingency table assessing agreement on white matter disease burden, number of microhemorrhages, and presence of cortical superficial siderosis will be evaluated separately using a Bowker's test of symmetry.

For scans with both SWI and T2*-GRE data, Spearman's rank correlations between the number of microhemorrhages identified by SWI and that identified by T2*-GRE will be provided for baseline scans and all scans. For scans with both 3D FLAIR and 2D FLAIR data in which ARIA-E was detected in either 3D FLAIR or 2D FLAIR, a Spearman's rank correlation will be computed from the severity of ARIA-E assessed by 3D FLAIR versus that assessed by 2D FLAIR (severity will be assigned to zero if there is no presence of ARIA-E).

To assess the agreement of exploratory versus standard MRI sequences in monitoring ARIA, a contingency table assessing agreement on maximum severity, will be constructed for ARIA-E and ARIA-H separately using a Bowker's test of symmetry.

To compare temporal dynamics between exploratory versus standard MRI sequences, mean and standard deviation of number of days to 1st ARIA event will be reported for 2D FLAIR versus 3D FLAIR (ARIA-E) and for GRE (gradient recalled echo) versus SWI (ARIA-H). Percentage of exploratory scans still showing ARIA-E or ARIA-H when standard scan said ARIA-E or ARIA-H was resolved.

4.5.2. Blood-based biomarkers

Blood-based biomarkers including the ones shown below are measured at baseline and weeks 4, 8, 12, 24, 36, 52, 64, and 76.

- Neurofilament Light chain (NfL)
- Phosphorylated tau (P-tau181 and P-tau217)
- Glial fibrillary acidic protein (GFAP)
- Aβ levels (Aβ 1-42 and 1-40)
- Soluble triggering receptor expressed on myeloid cells 2 (sTREM2)
- High sensitivity C-reactive protein (hsCRP)
- Other blood biomarkers when results become available.

To evaluate the change from baseline difference by treatment groups, an MMRM analysis will be used to compare change from baseline at each post-baseline timepoint for each of these blood-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. Visit will be considered a categorical variable with values equal to the planned visit numbers at which the blood-based biomarker is assessed. The null hypothesis is that the difference in LS mean change between the alternative treatment regimen and the standard treatment regimen equals zero. The values for these biomarkers may be log transformed to fit the normality assumption of the model.

To assess the relationship of baseline levels of these biomarkers with ARIA-E detected by standard safety MRI, after pooling all 4 dosing regimens, logistic regression model will be used with baseline amyloid centiloid, number of microhemorrhages at baseline, presence of superficial siderosis at baseline, mean arterial pressure, anti-hypertensives being predictors of such a model other than baseline biomarker values.

A logistic regression model will be used to analyze the association between biomarker changes and ARIA-E. This model will include percent change of NfL from baseline in 24 weeks, APOE e4 genotype, number of microhemorrhages at baseline, presence of superficial siderosis at baseline, baseline amyloid centiloid, mean arterial pressure, as well as anti-hypertensives. Similar analyses will be conducted to assess the association between ARIA-E incidence at Week 52 from mean percent change of NfL from baseline in the first 24 weeks. Whether or not percent change of NfL from baseline in 52 weeks can predict incidence of ARIA-E at Week 52 will be confirmed through a logistic regression model as well.

Only participants with at least 1 occurrence of ARIA-E will be included for analyses to correlate severity of ARIA-E with change of NfL scores from baseline. For instance, to correlate severity of ARIA- E at Week 24, a proportional odds model will be used. Such a model will include treatment arm, APOE e4 genotype status, number of microhemorrhages at baseline, baseline presence of superficial siderosis, baseline amyloid centiloid, mean arterial pressure, anti-hypertensives, and percent change of NfL from baseline in 24 weeks.

The association between other plasma-based biomarkers, like pTau, Aβ levels (40 and 42), GFAP, sTREM2, or hsCRP and the frequency and severity of ARIA-E will follow the same analysis plan outlined above.

4.5.3. IRR events

Here, infusion related reactions will be analyzed through safety analysis set. IRR events can be divided into potential immediate, which refers to IRRs occurring within 24 hours of donanemab administration, and potential non-immediate, which refers to IRRs occurring after 24 hours of donanemab administration. A Fisher's exact test will be employed to infer if potential immediate or potential non-immediate IRR incidence proportion is identical or not between one alternative treatment arm and the standard treatment arm. Whether the severity of immediate or non-immediate IRR events (severe vs. non-severe) is the same or not between one alternative dosing regimen and the standard dosing regimen will be confirmed through a Fisher's exact test as well.

4.6. Safety Analyses

Baseline for AACQ safety analyses refers to all adverse events that started before first dose of donanemab. Postbaseline for safety reviews includes data obtained during treatment period plus safety follow-up period.

4.6.1. Extent of exposure

Descriptive statistics, like mean, median, interquartile range, and range will be used to summarize the number of infusions per treatment arm. In each dosing regimen, the number and percentage of subjects with the number of infusions will be reported as well.

4.6.2. Adverse events

Treatment-emergent adverse events (TEAEs) in the safety analysis set will be defined as events that first occurred or worsened on or after the day of first donanemab dose. The MedDRA lower-level term (LLT) will be used in the treatment-emergent computation and the LLT of each baseline event will be used as baseline. Should there be insufficient data for AE start date, stop date, and time to make this comparison, imputation method will be used for partially missing dates. Missing dates of adverse events or concomitant medicine will be handled such that incomplete start dates of the current disease episode will be imputed in a manner resulting in the earliest possible onset. In order to calculate the durations in years from the onset of AD symptoms from the diagnosis of AD, the start date of the onset of AD symptoms and AD diagnosis will be defined in the following manner:

- The date itself if it is available as day, month and year.
- The date of the first day of the month, if month and year are available but the day is missing. For example, ??FEB2022 will be completed as 01FEB2022.
- The date of the first day of the year, if year is available but day and month are missing. For example, ?????2022 will be completed as 01JAN2022.

- If the date is completely missing, the duration will not be calculated and will form part of a "Missing" category.

The imputed values of incomplete data will not be presented in the patient data listings and will only be used to derive the durations described above.

An overview of AEs by treatment arm, including the number and percentage of patients who died, experienced serious adverse events (SAEs), discontinuations from study treatment due to an AE and TEAE will be displayed. Besides descriptive statistics, odds ratios with any alternative dosing regimen as the numerator and standard dosing regimen as the denominator, and *p*-values derived from Fisher's exact test may be reported for AE overview.

Summaries of AEs by decreasing frequency of preferred term (PT) within system organ class (SOC) will be provided per treatment arm for the following:

- TEAEs
- TEAE with 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
- Adverse events reported as reason for study discontinuation

4.6.3. Clinical laboratory evaluation

Laboratory measurements will be analyzed using continuous data (change from baseline) and categorical/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).

In AACQ safety analysis set, mean and SD in change from baseline to post-baseline visit by treatment arm at which laboratory measurements are taken will be provided. An ANCOVA (analysis of covariance) method will be used to assess if 4 treatment means of a post-baseline lab measurement are significantly different after adjusting baseline lab measurement.

The proportion of patients with treatment-emergent high or treatment-emergent low or treatment-emergent abnormal laboratory values will be summarized by treatment arm. 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. Fisher's exact test will be able to tell whether compared to the standard dosing regimen, one alternative dosing regimen has a significantly lower rate of treatment-emergent high, treatment-emergent low, or treatment-emergent abnormal laboratory values.

The proportion of patients by treatment arm with treatment-emergent clinically meaningful changes from a low value or normal value at all baselines at any time in ALT, AST, ALP, direct bilirubin, total bilirubin, and GGT will be summarized through safety analysis set. Clinically significant changes of interest at any time are: ALT ≥1 x, 3 x, 5 x, 10 x, 20 x upper limit of

normal (ULN), AST ≥1 x, 3 x, 5 x, 10 x, 20 x ULN, ALP ≥2 x, 3 x ULN, direct bilirubin ≥2 x, 5 x ULN, total bilirubin ≥2 x, 5 x, 8 x ULN and GGT ≥2 x ULN. TEAE analysis of potentially drug-related hepatic disorders will be provided by treatment arm. In addition, a plot will graphically display maximum of ALT (x ULN) or AST (x ULN) versus maximum total bilirubin of every patient included in the safety analysis set.

4.6.4. 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 all safety data set. Vital sign measurements include 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 5 minutes in the supine position minus that after at least 3 minutes in the standing position), and temperature.

If there are multiple records of vital sign or weight measurements at baseline or postbaseline visit, the last record will be used. For the analysis of vital signs and weights, summary statistics like mean and SD per treatment arm will be presented for observed values at baseline and for change from baseline results at each scheduled postbaseline visit. The application of ANCOVA will be used to assess whether 4 treatment means of a post-baseline vital sign or weight are different or not after adjusting baseline vital sign or weight.

The incidence of treatment-emergent abnormal high or low vital signs and weight will be presented using safety analysis set. Treatment-emergent vital sign evaluations are defined for evaluations collected after the initiation of donanemab. Abnormal criteria for post-baseline vital signs and weight are presented in Table 4.1. Any vital sign or weight meeting the criteria will be considered abnormal. The proportion of patients with treatment-emergent abnormal high or low vital signs and weight by treatment arm will be provided regardless of their baseline results. And an odds ratio of having treatment-emergent abnormal high or low vital signs or weight in each alternative dosing regimen compared to the standard dosing regimen can be computed. A *p*-value generated from Fisher's exact test will confirm if proportions of treatment-emergent abnormal high or low vital signs or weight are statistically different between the alternative and standard treatment arms.

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 summarized in each treatment arm using safety analysis set.

Table 4.1. Potentially Clinically Significant Changes in Vital Sign 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., standing minus supine ≤–20)	
Orthostatic diastolic blood pressure (mmHg)	≥10 mmHg decrease in diastolic blood pressure (supine to standing) (i.e., standing minus supine ≤ -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; mmHg = millimeters of mercury.

4.6.5. Columbia-Suicide Severity Rating Scale

Suicide-related thoughts and behaviors occurring through the end of the study which includes safety follow-up period will be summarized based on the Columbia-Suicide Severity Rating Scale (C-SSRS). The composite measure is determined at each assessment by the "yes" or "no" responses in C-SSRS categories by study patient.

- 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.

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. The proportion of patients with suicidal ideation, suicidal behavior, or either suicidal ideation or behavior per treatment arm will be tabulated through safety analysis set.

Number of patients with suicide-related treatment-emergent events based on the C-SSRS may be reported for safety analysis set:

- 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

Emergence of suicidal behavior compared to all prior history "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.

In case the following suicide-related events occur, the counts and percentages of patients of the event by treatment arm will be summarized from safety analysis set:

- died by 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
- nonsuicidal, self-injurious behavior

4.6.6. Amyloid-Related Imaging Abnormalities (ARIA)

ARIA incidence was discussed in Section 4.3, Section 4.4.1, and Section 4.4.3. Severity of ARIA events was covered in Section 4.4.3.

Unless otherwise specified, safety analyses of ARIA events are based on centrally read MRI findings from both scheduled and unscheduled visits. Where applicable, ARIA-E summaries will be presented for symptomatic and asymptomatic ARIA-E. ARIA-H summaries will be presented by ARIA-H microhemorrhage and ARIA-H superficial siderosis. Macrohaemorrhage will be described separately and will not be included in the ARIA-H category.

An overview of ARIA incidence by treatment arm from safety analysis set will be presented using frequency and percentage of patients with any ARIA (ARIA-E or ARIA-H), ARIA-E, and ARIA-H as defined by safety MRIs or treatment-emergent AE clusters. The proportion of the patients with serious adverse events of ARIA, ARIA-E, and ARIA-H with or without ARIA-E will be summarized via treatment arm. The frequency and percentage of ARIA-E will be further

broken out by asymptomatic vs. symptomatic and by APOE genotype. The frequency and percentage of subjects with ARIA-H microhemorrhage, ARIA-H superficial siderosis, and macrohemorrhage will be further broken out by APOE genotype. A sensitivity analysis of overview of ARIA incidence will be performed utilizing per-protocol analysis set.

A severity shift table of ARIA-E, ARIA-H microhemorrhage, and ARIA-H superficial siderosis severity from baseline to Week 76 by visit and by treatment arm will be presented.

To investigate the onset of the first ARIA-E or ARIA-H event among different dosing regimens, Kaplan-Meier plots will be presented using safety analysis set. Besides Kaplan-Meier plots, histograms of the onset of the first ARIA-E or ARIA-H based on safety MRI may be provided to visually inspect timing of the first ARIA event across treatment arms.

From safety analysis set, mean and median duration before the resolution of ARIA-E or the mean and median duration of ongoing ARIA-E events for each dosing regimen will be derived.

4.6.7. Hypersensitivity/Infusion Related Reactions

Section 4.5.3 described ways to explore whether alternative dosing regimen will impact IRR frequency and severity. In this subsection, both hypersensitivity and IRRs will be analyzed and summarized from safety analysis set. Hypersensitivity and IRRs will be broken out between potential immediate (defined as TEAEs occurring on the date of infusions, or for which a hypersensitivity follow-up form indicates the event occurred within 24 hours of study drug administration) and potential non-immediate (defined as TEAEs that are not immediate but occur prior to the administration of a subsequent infusion).

The AE database will be searched using predefined SMQs to identify preferred terms (PTs) that map hypersensitivity events or IRRs in both potential immediate and potential non-immediate analysis periods. For each treatment arm, the number and percentage of patients who experienced a TEAE of hypersensitivity or IRR will be reported.

The following summary tables may be provided about hypersensitivity/IRRS:

- treatment-emergent serious hypersensitivity, anaphylactic and IRRs
- potential immediate hypersensitivity, anaphylactic and IRRs by maximum severity
- potential non-immediate hypersensitivity, anaphylactic and IRRs by maximum severity
- symptoms for potential immediate hypersensitivity, anaphylactic and IRRs
- symptoms for potential non-immediate hypersensitivity, anaphylactic and IRRs.

4.6.8. Safety MRIs

Treatment-emergent white matter disease and other abnormality findings will be summarized as incidence in safety analysis set.

To evaluate white matter changes over time, a shift table from baseline to selected visits via treatment arm will be created from the following categories:

- 0 = No lesions
- 1 = Focal lesions
- 2 = Beginning confluence of lesions

- 3 = Diffuse involvement of entire region

4.7. Interim Analysis

An external Data Monitoring Committee (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. The DMC will consist of a minimum of 3 members, including a physician with expertise in AD and a statistician. 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:

- 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
- 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
- C-SSRS data

At least 1 interim safety analysis may be conducted for the entire AACQ Study. Operational details and a quantitative framework to provide information for these decisions will be documented in interim analysis Statistical Analysis Plan.

5. Sample Size Determination

Approximately 800 participants will be randomly assigned to 1 of 4 treatment arms in a 1:1:1:1 ratio. It is hypothesized here that in each treatment arm, 10% of participants will not reach primary outcome lock. Below we showed that 720 participants who completed visit 12 can provide more than 80% power to demonstrate that there is over 80% probability that at least one of the alternative dosing regimens will reduce relative risk of ARIA-E by at least 20% compared to the standard dosing regimen.

The Bayesian power calculation was performed through simulations in both data generation and model fitting steps. In the data simulation stage, 1000 ARIA-E data sets were generated for every treatment arm. It was assumed that ARIA-E rate by Week 24 in the standard dosing regimen was 18.5% based on observations from AACI Addendum 9 data. In each alternative dosing regimen, ARIA-E rate was set to be equal to 11.1%, which translates to 40% relative risk reduction in ARIA-E compared to standard dosing regimen.

In the model fitting stage, a normal prior, $N(0, 4)$ was assigned to individual slope parameter of Arm 2, 3, and 4, respectively in the Bayesian logistic regression model. The intercept parameter was given a normal mixture prior, $0.5\times N(-1.48, 4)+ 0.5\times N(-1.48, 30.19)$. The component prior mean (-1.48) in the mixture prior is equal to the natural log odds ratio of ARIA-E rate at 18.5% observed from historical study. For any simulated ARIA-E data with a complete sample size of 720, posterior samplers of RD_{21}, RD_{31}, and RD_{41} were simulated from 3 Markov chains with each chain composed of 10000 iterations after 10000 burn-ins. Among 1000 datasets generated, 927 satisfied at least one of the following probability statements:

$$\Pr(RD_{21} \geq 0.2 \mid \text{simulated 24-week ARIA-E data}) > 0.8;$$

$$\Pr(RD_{31} \geq 0.2 \mid \text{simulated 24-week ARIA-E data}) > 0.8;$$

$$\Pr(RD_{41} \geq 0.2 \mid \text{simulated 24-week ARIA-E data}) > 0.8.$$

Hence, the reported power was 92.7% based on simulations.

In the null scenario where all the dosing regimens have the same true ARIA-E rate of 18.5% by Week 24, the false positive rate was controlled at 5% one-sided, or equivalently 10% two-sided.

The power and false-positive rate of the study were determined by simulations using R 4.1.2. Particularly, posterior samplers were generated through rjags package in R.

6. Supporting Documentation

6.1. Appendix 1: Demographic and Baseline Characteristics

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

- age
- gender
- race
- ethnicity
- country
- years of education
- work status
- height
- body weight
- body mass index (weight (kg) / [height (m)]^2)
- amyloid PET burden
- APOE4 carrier status (carrier [ε2/ε4, ε3/ε4, ε4/ε4], noncarrier [ε3/ε3, ε2/ε2, ε3/ε2])
- APOE4 genotype (ε2/ε4, ε3/ε4, ε4/ε4, no ε4)
- AChEI and/or memantine use at baseline
- time since onset of first AD symptoms
- time since AD diagnosis
- having one or more first degree relatives with AD
- tobacco use
- alcohol use
- caffeine use
- MMSE

6.2. Appendix 2: Treatment Compliance

Because dosing occurs at study visits, patients who attend all visits and successfully receive donanemab infusions are automatically compliant with this treatment. Any infusion where 75% (approximately 60 mL, where the starting volume is 80 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 arm in the treatment period.

6.3. Appendix 3: 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 prior medications being

taken in any time during the treatment period and follow-up period. A summary of concomitant medications that were used will be presented through frequencies and percentages for safety analysis set. If the start or stop dates of therapies are missing or partial to the degree that determination cannot be made of whether the therapy is prior or concomitant, the therapy will be deemed concomitant.

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

6.4. Appendix 4: Multicenter Studies

This study will be conducted by multiple investigators at multiple sites internationally. In case 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. Appendix 5: Handling Missing Data from Participant Dropouts

Because AACQ study follows a patient population that is elderly with multiple comorbidities for 76 weeks, 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.

A likelihood-based mixed effects model for repeated measures will be used to handle missing data for biomarker analyses. The model parameters are simultaneously estimated using restricted likelihood estimation incorporating all 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 early from the study, there may be safety or biomarker measurements at visits where the variables were not scheduled to be collected. This data will be used in all other analyses.

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.

Christensen R, Johnson W, Branscum A, et al. Bayesian Ideas and Data Analysis: An introduction for scientists and statisticians. 2010. CRC Press.

Charidmou A, Boulouis G, Frosch M, et al. The Boston criteria V2.0 for cerebral amyloid angiopathy: A multicenter MRI neuropathology diagnostic accuracy study. *Lancet Neurol*. 2023; 21(8):714-725.

Greenburg S, Battioui C, Lu M, et al. ARIA insights from the Donanemab Trails. *Neurology*. April 9, 2024 issue, 102 (17_supplement_1), https://doi.org/10.1212/WNL.0000000000205071.

Integrated Statistical Analysis Plan (LY3002813): Assessment of safety of LY3002813 in early symptomatic Alzheimer's disease. 2023.

Protocol (I5T-MC-AACQ): Investigating the effect of different donanemab dosing regimens on ARIA-E and amyloid lowering in adults with early symptomatic Alzheimer's disease. 2022.

Wang F and Gelfand EG. A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. *Stat Sci.* 2002;17(2):193-208.

Signature Page for VV-CLIN-160476 v1.0

Approval	PPD 29-Jul-2024 19:37:24 GMT+0000

Signature Page for VV-CLIN-160476 v1.0