Statistical Analysis Plan I8D-MC-AZES (Version 7)

A 24-month, Multicenter, Randomized, Double-blind, Placebo-controlled, Parallel-group, Efficacy, Safety, Tolerability, Biomarker, and Pharmacokinetic Study of AZD3293 in Early Alzheimer's Disease (The AMARANTH Study)

NCT02245737

Approval Date: 17-Aug-2018

1. Statistical Analysis Plan: I8D-MC-AZES: A 24-month, Multicenter, Randomized, Double-blind, Placebo-controlled, Parallel-group, Efficacy, Safety, Tolerability, Biomarker, and Pharmacokinetic Study of AZD3293 in Early Alzheimer's Disease (The AMARANTH Study)

Confidential Information

AZD3293 (LY3314814) Early AD

Study I8D-MC-AZES is a multicenter, randomized, parallel-group, double-blind, placebo-controlled, international study of 2 fixed dose levels of AZD3293 in patients with early AD.

Eli Lilly and Company
Indianapolis, Indiana USA 46285
Protocol I8D-MC-AZES
Phase 2/3

SAP Version 7 electronically signed and approved on the date provided below:

Approval Date: 17-Aug-2018 GMT

2. Table of Contents

Section **Page**

Table of Contents

Table of Contents

Figure **Page**

Table of Contents

Appendix **Page**

3. Revision History

Statistical Analysis Plan (SAP) Version 1 was approved on 03 March 2015, prior to any unblinding. Version 2 of the SAP was approved on 20 March 2015, prior to unblinding and prior to any of the planned 3 interim analyses. Version 3 of the SAP was approved on December 1, 2016. Version 4 of the SAP was approved on May 1, 2017 prior to unblinding post baseline efficacy data for Interim Analysis 2. Version 5 of the SAP was approved on October 4, 2017 prior to unblinding efficacy data for Interim Analysis 3. Version 6 was approved on March 21, 2018 prior to unbinding efficacy data for Interim Analysis 3.

Because the futility criteria were met at the interim analysis, the lanabecestat phase 3 trials are being stopped for futility. While efficacy and safety data will be reported, the scope of the analysis will be streamlined to focus on the key analyses. Version 7 of the SAP reflects this streamlining, and will be approved prior to final unblinding for the clinical summary report and includes the following changes:

1. Reported p-values will not be adjusted for multiplicity.
2. Deleted Full Effectiveness and Per Protocol dataset definitions.
3. Dropped Kaplan-Meier analyses of disposition results.
4. Dropped prior medications listing.
5. Dropped summary of concomitant medications with anticholinergic properties affecting cognition.
6. Deleted the multiplicity graph.
7. Reduced the Subgroup analyses.
8. Dropped responder analyses.
9. Dropped slope analyses.
10. Reduced the number of explanatory variables in the logistic ordinal regression model.
11. Dropped LOCF analyses.
12. Dropped completer analyses.
13. Dropped MMRM effectiveness dataset analyses.
14. Dropped per protocol dataset analyses.
15. Dropped certain efficacy MMRM analyses.
16. Added MMRM analyses of RBANS index scores.
17. Dropped categorical analyses of NPI.
18. Dropped sensitivity analyses.
19. Dropped subgroup analyses of plasma A-Beta.
20. Added ADCS-iADL and dropped ADAS-Cog11 and ADAS-Cog12 from plasma A-Beta vs. efficacy scale correlation analysis.
21. Added ADCS-iADL and dropped ADAS-Cog11 and ADAS-Cog12 from CSF biomarkers vs. efficacy scale correlation analysis.
22. Dropped additional amyloid PET regions analyzed.
23. Dropped additional normalizations of amyloid PET analyses.
24. Added ADCS-iADL and dropped ADAS-Cog11 and ADAS-Cog12 from amyloid PET vs. efficacy scale correlation analysis.

25. Clarified the FDG-PET SUVR calculations to be summarized.
26. Added ADCS-iADL and dropped ADAS-Cog11 and ADAS-Cog12 from FDG PET vs. efficacy scale correlation analysis.
27. Added detailed calculations for AV-1451 composites and added two brain regions for analysis.
28. Added ADCS-iADL and dropped ADAS-Cog11 and ADAS-Cog12 from vMRI vs. efficacy scale correlation analysis.
29. Added listing of study treatment assignment.
30. Clarified the treatment-emergent adverse event time period to include 5 days after stopping study drug.
31. Specified treatment-emergent adverse events will be based on lower level term (LLT) processing.
32. Dropped summary of TEAE's by visit.
33. Dropped subgroup analysis of treatment-emergent adverse events.
34. Clarified Skin Disorder adverse events of special interest.
35. Added PCS labs analysis.
36. Added a listing based on the hepatic safety CRF.
37. Changed categorical analysis of weight gain or loss from 4% to 7%.
38. Dropped analysis of BMI categories for patients changing more than 7%.
39. Dropped shift analysis of BMI.
40. Replaced CSSRS analyses originally based on studies where no/few suicidal ideations/behaviors are expected to analyses where suicidal ideations/behaviors are anticipated.
41. Clarified analyses of skin exam.
42. Clarified analyses of eye exam.
43. Added ADCS-ADL to be included in subgroup analyses.
44. Reduced subgroups to disease status, age group, and region.
45. Added details to Safety follow-up summary.
46. Various typographical corrections and minor editing for clarity.

4. Study Objectives

4.1. Primary Objective

The primary objective of this study is to test the hypothesis that LY3314814 (previously known as AZD3293), administered orally at doses of 20 mg and 50 mg daily for 104 weeks, will slow the decline of Alzheimer's Disease (AD) as compared with placebo in patients with early AD dementia as measured by ADAS-Cog$_{13}$. Early AD is defined as the continuum of patients with mild cognitive impairment (MCI) due to AD (ie, prodromal AD) and patients diagnosed with mild dementia of the Alzheimer's type.

4.2. Secondary and Exploratory Objectives

Table 1 below defines the objectives and endpoints of the study.

Table 1. Objectives and Endpoints

Double-Blind Treatment Period	
Secondary Objectives	**Endpoints**
Clinical efficacy objectives: • To evaluate the efficacy of LY3314814 on functional, clinical, and cognitive outcomes in patients with early AD dementia at the end of the double-blind treatment period (Week 104)	• Functional Outcome Measures o The Alzheimer's Disease Cooperative Study Activities of Daily Living Inventory instrumental items (ADCS-iADL) score o Functional Activities Questionnaire (FAQ) score • Cognitive/Functional Outcome Measures o Integrated Alzheimer's Disease Rating Scale (iADRS) score o Clinical Dementia Rating – Sum of Boxes (CDR-SB) score • Clinical Outcome Measures o CDR-Global Score o Neuropsychiatric Inventory (NPI) score • Cognitive Outcome Measures o Mini-Mental State Examination (MMSE)
• To evaluate the relationship between treatment effect of LY3314814 and time (at points other than the end of the double-blind treatment period [Week 104], such as Week 26, Week 52, and Week 78). Specific time points will vary by instrument.	• ADAS-Cog$_{13}$, ADCS-iADL, FAQ, CDR-SB, and iADRS

Objectives and Endpoints

Double-Blind Treatment Period	
Secondary Objectives	**Endpoints**

• To evaluate the efficacy of LY3314814 to prolong time in the current disease state	• CDR global score
• To evaluate the clinical worsening and need for symptomatic treatments	• AD symptomatic treatments time of initiation
Biomarker objectives: • To evaluate the effect of LY3314814 on cerebrospinal fluid (CSF) amyloid beta peptide (Aβ) pharmacodynamics (PD) markers	• CSF $A\beta_{1-42}$ and $A\beta_{1-40}$ concentrations
• To evaluate the effect of LY3314814 on CSF markers of neurodegeneration	• CSF total tau and phosphorylated tau concentrations
• To evaluate the effect of LY3314814 on brain amyloid burden	• Florbetapir amyloid scan
• To evaluate the effect of LY3314814 on brain aggregated tau levels	• ^{18}F-AV-1451 Tau PET [separate addendum]
• To evaluate the effect of LY3314814 on brain metabolism	• Fluorodeoxyglucose (FDG) positron emission tomography (PET) [separate addendum]
• To evaluate the effect LY3314814 on brain atrophy	• Brain volumes measured by magnetic resonance imaging (MRI)
Pharmacokinetic objective: To assess the population PK of AZD3293 and metabolite AZ13569724 in patients with early AD dementia	• Apparent Oral Clearance of AZD3293 • Central Volume of Distribution of AZD3293
Safety Objective • To evaluate the safety and tolerability of LY3314814 in patients with early AD dementia	• Standard safety assessments: o spontaneously reported adverse events (AEs) o clinical laboratory tests o vital sign and body weight measurements o 12-lead electrocardiograms (ECGs) o physical examinations including neurological examinations • Additional safety assessments: o Eye examinations o Skin examinations o Serial MRI o Columbia-Suicide Severity Rating Scale (C-SSRS)
Exploratory Objectives*	**Endpoints***
• To evaluate the efficacy of LY3314814 on executive function	• Changes from baseline in the Letter Fluency, Category Fluency, and Digit Symbol-Coding test scores
• To evaluate the efficacy of LY3314814 on neuropsychological status	• Change from screening in the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) score

Objectives and Endpoints

Double-Blind Treatment Period	
Exploratory Objectives[a]	**Endpoints**[a]
• To explore the effects of LY3314814 on	• Change from baseline in the Resource

resource utilization and health status	Utilization in Dementia (RUD)-Lite and EQ-5D-5L score
• To explore the effects of LY3314814 on caregiver distress	• Change from baseline in the NPI-D score
• To evaluate the dose-exposure effect relationships for selected efficacy, safety, and biomarker endpoints	• Changes in scales, adverse events and blood, CSF and imaging biomarkers in relation to exposure
• To explore the genetic effects of APOE4 allele status (carrier or non-carrier)	• Change from baseline on selected efficacy, safety, and biomarker endpoints
• To explore the effect of LY3314814 on gene expression	• Change in levels of whole blood messenger ribonucleic acid (mRNA) and on plasma and CSF micro-ribonucleic acid (miRNA) levels and the relation to selected efficacy, safety, and biomarker endpoints
• To explore the effect of LY3314814 on peripheral and central nervous system protein expression	• Change in CSF or peripheral protein biomarkers, including CSF or plasma $A\beta_{1-42}$ and $A_{\beta 1-40}$, neurodegenerative inflammatory markers, and proteomic assessments and the relation to selected efficacy, safety, and other biomarker endpoints
• To examine LY3314814 population PK/PD	• Exposure and selected measures of efficacy and/or tolerability
• To optionally collect and store deoxyribonucleic acid (DNA) for future exploratory research into genes/genetic variation that may influence response (ie, distribution, safety, tolerability, and efficacy) to LY3314814. Genes that will be tested for heritable genetic (DNA) or expression (mRNA, miRNA) effects will be limited to those for which a hypothesis for involvement in the pharmacology of LY3314814 or in the pathogenesis of AD can be formulated, as currently understood and as directed by future scientific progress in the field.	• Association of genetic variation with selected endpoints (optional pharmacogenomics sampling)

[a] The results of analyses that address some exploratory objectives may not form part of the Clinical Study Report, but may be described in supplementary reports, as appropriate.

5. Study Design

5.1. Summary of Study Design

This is a multicenter, randomized, parallel-group, 104-week, double-blind, placebo-controlled, international study of 2 fixed dose levels of LY3314814 (20 mg or 50 mg) in patients with early AD (ie, the continuum of patients with MCI due to AD and patients diagnosed with mild dementia of the Alzheimer's type) and abnormal levels of amyloid.

The study consists of a screening period of approximately 8 weeks (Visit 1), a baseline assessment period of approximately 1 week (Visit 2), a double-blind treatment period of 104 weeks (Visit 3 through Visit 20), and a follow-up period of 4 to 6 weeks (Visit 21) after the last dose of study treatment is administered (see Figure 1). The total study duration is approximately 118 weeks. Visit 9 (Week 32), Visit 13 (Week 58), Visit 15 (Week 71), and Visit 17 (Week 84) will be conducted by telephone with the patient and/or study partner. The remaining visits will be in-clinic visits. Patients are expected to complete all visits through Week 104, even if they discontinue study treatment. After the study was stopped for futility, all ongoing patients were asked to come in for a futility-based discontinuation visit. Patients at this visit followed the schedule of events for an early discontinuation visit.

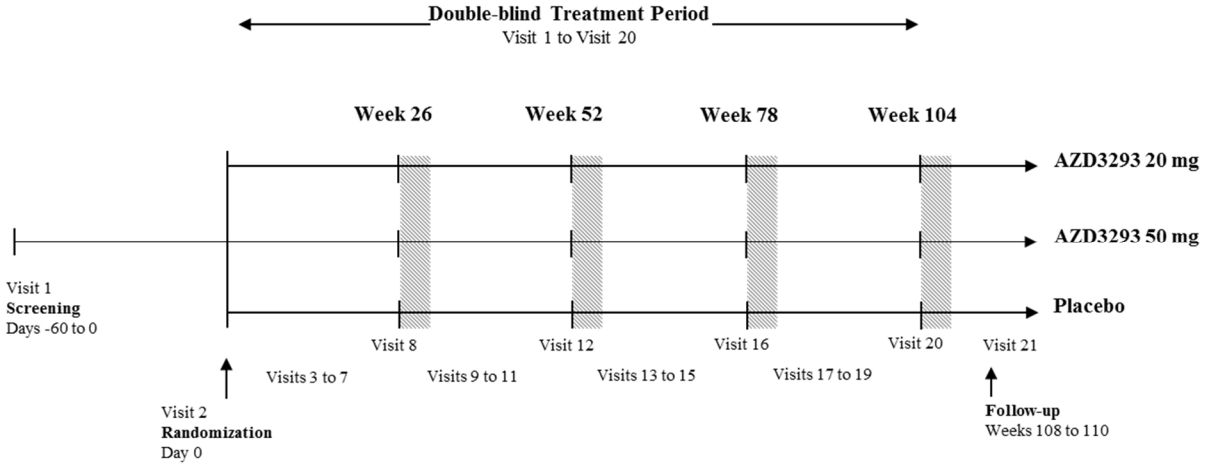

Figure 1. Study flow chart.

Male and female patients, aged 55 to 85 years, who meet all inclusion criteria and no exclusion criteria, will be randomly assigned on Day 0 to 1 of 3 treatment groups in a 1:1:1 ratio: 20 mg LY3314814, 50 mg LY3314814, or placebo (approximately 734 patients per treatment group). Study treatment will be administered once daily, in the morning, beginning on Day 1.

Randomized treatment allocation will be centrally controlled and stratified by disease status at baseline (MCI due to AD or mild AD). Randomization into either the MCI due to AD or mild AD groups may be stopped before the completion of the study to ensure an adequate distribution of both strata.

Patients who meet other study entry requirements will be required to undergo either a florbetapir F18 positron emission tomography (PET) scan or a lumbar puncture for cerebrospinal fluid (CSF) sampling (but not both) at screening to document the presence of abnormal levels of brain or CSF amyloid for study inclusion. Historical amyloid PET scans may be allowed for study eligibility with central reader confirmation of amyloid positivity.

The main study protocol also includes 2 mandatory longitudinal biomarker sub-studies: a florbetapir F18 amyloid PET sub-study and a CSF sub-study. At least 900 individuals will participate in the florbetapir F18 amyloid PET sub-study and at least 175 patients will participate in the CSF sub-study. After the required number of patients are enrolled in these 2 sub-studies, subsequent enrollment is optional and enrollment may be discontinued at any time at the discretion of the sponsor.

Each patient will have a study partner who has sufficient interaction with the patient to provide meaningful input regarding the patient and is willing to participate at every study visit.

An independent data monitoring committee (IDMC) will be established to monitor data on an ongoing basis to ensure the continuing safety of patients enrolled in this study, to ensure the integrity of the blinded nature of the study, and to oversee the planned interim analyses.

5.2. Determination of Sample Size

The sample size for this study is 734 randomized patients in each of 3 treatment groups. The sample-size calculations are based on the primary efficacy endpoint, change from baseline in ADAS-Cog_{13} total score at Week 104. An unblinded sample size re-estimation is planned for interim 2 and the sample size could increase up to 1000 randomized patients in each of the 3 treatment groups.

Assumptions about expected decline of study participants as measured by the ADAS-Cog total score were based on 18 month results extrapolated to 24 months from placebo-treated patients in a global AD Phase 3 studies database, because this data source is expected to reasonably predict decline in the current study. In this database, the least-squares mean change in ADAS-Cog total score from baseline at 80 weeks in the mild placebo-treated population of APOE4 carriers with CDR global score = 0.5 was 1.68 units with a standard deviation (SD) of 9.72. A linear extrapolation using these results coupled with results at 28 and 52 weeks resulted in 104 week estimates of 2.24 (12.96). Using these estimates to approximate a MCI population, this expected difference translates into a 0.17 effect size.

Compared with the MCI population, cognitive deterioration in patients with mild AD has been observed to be somewhat faster. The least-squares mean (LSM) change in ADAS-Cog total score from baseline at 80 weeks in the mild AD population (enriched by restricting to APOE4 carriers) was 2.87 units, with an SD of 10.16. A linear extrapolation using these results coupled with results at 28 and 52 weeks resulted in 104 week estimates of 3.83 (13.54). This expected difference translates into a 0.28 effect size.

Similar calculations were made for the key secondary endpoint ADCS-iADL. The LSM change in ADCS-iADL score from baseline at 80 weeks in the mild placebo-treated population of

APOE4 carriers with CDR global score = 0.5 was 0.31 (8.39) units. A linear extrapolation using these results coupled with results at 28 and 52 weeks resulted in 104 week estimates of 0.41 (11.19). Using these estimates to approximate an MCI population, this expected difference translates into a 0.04 effect size. The LSM change in ADCS iADL score from baseline at 80 weeks in the mild AD population (enriched by restricting to APOE4 carriers) was 2.14 (9.81) units. A linear extrapolation using these results coupled with results at 28 and 52 weeks resulted in 104 week estimates of 2.85 (13.08). This expected difference translates into a 0.22 effect size.

Not accounting for a possible sample size increase or potential arm dropping for safety or clinical worsening reasons, the study will enroll a total of 2202 patients (734 patients per group). Assuming α=0.025 and the effect size among the patients who discontinue without completing the study is 50% of the effect size of the completers, and a patient population consisting of 65% mild AD patients and 35% MCI patients, the power for the overall population for each dose is 96% for ADAS-Cog (adjusted effect size is 0.21); for iADL, 60% (adjusted effect size is 0.13). The power to show statistical significance for the Mild AD subpopulation is 93% (adjusted effect size is 0.24) and 75% (adjusted effect size is 0.19) for ADAS-Cog and iADL, respectively. The power to show statistical significance for the MCI subpopulation is 26% (adjusted effect size is 0.14) and 3% (adjusted effect size is 0.03) for ADAS-Cog and iADL, respectively.

Table 2. Effect Size and Power per Dose for Mild AD and MCI Patient Populations

Population	ADAS-Cog ES per Dose	ADAS-Cog Power per Dose	ADCS-iADL ES per Dose	ADCS-iADL Power per Dose
Overall[a]	0.21	96%	0.13	60%
Mild AD	0.24	93%	0.19	75%
MCI	0.14	26%	0.03	3%

Abbreviations: AD = Alzheimer's Disease; ADAS-Cog = Alzheimer's Disease Assessment Scale-Cognition; ADCS-iADL = The Alzheimer's Disease Cooperative Study Activities of Daily Living Inventory - instrumental items score; ES = Effect size; MCI = Mild cognitive impairment.

[a]The overall population effect sizes and power calculations use a distribution of 65% mild AD patients and 35% MCI patients.

5.3. Method of Assignment to Treatment

Eligibility will be established before randomized treatment assignment. The randomization will be stratified for disease status at baseline (MCI due to AD or mild AD).

The Integrated Voice/Web Response System (IVRS/IWRS) will allocate the treatment assigned through the stratified randomization scheme and provide a unique randomization code and the appropriate treatment kit number from those available at the site for the treatment assignment, as patients are eligible for randomization, at completion of Visit 2 procedures. Randomization codes will be assigned strictly sequentially as patients become eligible for randomization.

The stratified randomization schedule will be prepared by the CCI on behalf of the sponsor. Randomization codes will be allocated in blocks within strata to

LY3314814 20 mg, LY3314814 50 mg, or placebo in a ratio of 1:1:1. Enrollment rates in MCI due to AD and mild AD strata will be monitored.

If a patient withdraws from the study, the patient's study identifier and randomization code will not be reused, and the patient will not be allowed to re-enter the study.

6. A Priori Statistical Methods

6.1. General Considerations

As the study was stopped for futility, the planned statistical tests lose their scientific validity and multiplicity is no longer a concern. All reported p-values will not be adjusted for multiplicity. No test will be interpreted as statistically significant, and tests resulting with low p-values will only be considered as potential results of interest.

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 patients are assigned by random allocation, even if the patient does not take the assigned treatment, does not receive the correct treatment, or otherwise does not follow the protocol.

When change from baseline is assessed, patients will be included in the analysis only if both a baseline and a postbaseline measure are available. Unless otherwise defined, a baseline measure is the last nonmissing observation collected during Visit 1 and Visit 2.

6.2. Adjustments for Covariates

The repeated measures models will include the fixed, categorical effects of treatment, visit (categorical covariate), treatment-by-visit interaction, disease status at baseline (MCI due to AD or mild AD), APOE4 status (carrier versus non-carrier), concomitant AChEI use at baseline (yes/no), and country, as well as the continuous, fixed covariate of baseline score, age at baseline, and baseline score-by-visit interaction.

When an analysis of covariance (ANCOVA) model is used to analyze a continuous efficacy variable, the model will contain the main effects of treatment, disease status at baseline, 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. For skin analyses the model will also include the main effect for baseline Fitzpatrick score.

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

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

For ADAS-Cog_{13}, if <30% (4 or fewer of a total of 13 items) of the items are missing, the total score (maximum = 85) will be imputed as follows: The total from remaining items will be multiplied by a factor that includes the maximum score for the missing items. For example, if the first item, "Word-Recall Task," which ranges from a score of 0 through 10 (maximum = 10) is missing, and the second item "Commands," which ranges from a score of 0 through 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.

The same imputation technique will be applied to any other ADAS-Cog subscore, if tested.

For theADCS-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 techniques will be applied to the Functional Activities Questionnaire (FAQ). if <30% of the items are missing, the total score will be imputed as follows. 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 FAQ at that visit will be considered missing.

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

For the RBANS, if <30% of the sub-items are missing (ie, no more than 3 of the 12 sub-items), the item score will be imputed. For the missing subtest, the scaled score from the other subtest within that index will be used to impute the missing scaled score, which is then converted to a raw score. If List Recognition is missing, the scaled score mean for List Recall, Story Recall, and Figure Recall should be used to impute the missing value. If two sub items are missing within the

same index and/or if >30% of the sub-items are missing, the total score for the RBANS at that visit will be considered missing.

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

6.4. Multicenter Studies

This study will be conducted by multiple investigators at multiple sites internationally. Country is a covariate of the primary and many of the secondary efficacy analyses. In the event that any country has an insufficient number of patients (defined as less than 20), the data from these countries will be pooled with the closest geographical country.

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

As the study was stopped for futility, multiplicity is no longer a concern. All reported p-values will not be adjusted for multiplicity.

If an adaptation is made to the trial (either a sample size increase at the second interim analysis or a dose is dropped for safety concerns), the following methodology will be utilized to strongly control the Family-wise Error Rate (FWER) and applied to all hypotheses included in the multiplicity graph. If there is no adaptation to the trial, a conventional final analysis will be utilized which includes the combined data from all patients enrolled; however, sensitivity analyses will still compare the results for consistency between the first and second cohorts (those patients included in the sample size re-estimation at IA2 and those not included). Patients included in the sample size re-estimation will be those randomized on or after the date of the last randomized patient who has a 6 month ADAS-Cog13 or iADL score as of the IA2 cut-off date.

A confirmatory adaptive design should satisfy the conditional invariance principle (Brannath et al. 2007); separate test statistics will be calculated from the 2 patient cohorts and combined in a prespecified way for the final test decisions. Hence, any design modification that preserves the distributional properties of the separate test statistics under a given null hypothesis of interest will not lead to an inflation of the FWER. The p-value combination test principle combines the 2 p-values using a prespecified combination function (the weighted inverse-normal function, in this case) and by construction satisfies the invariance principle.

Lehmacher and Wassmer (1999) propose a generalization of the adaptive, weighted inverse normal test which facilitates exact K-stage tests when an exact p-value can be obtained from each cohort. Pre-assigned weights ω_1 and ω_2 are defined for k=1,2 test statistics where $Z_{winv}=(\omega_1 Z_1 + \omega_2 Z_2)$ represents the weighted inverse normal test statistic at conclusion of the study should an adaptation occur. Here, Z_k for k=1,2 are Z-statistics based on data in each cohort and $\sum\omega_i^2=1$. For normal responses with unknown variance, the process for obtaining Z_k from cohort k with n_k observations per treatment and per patient cohort is as follows: First calculate the t-statistic t_k for testing the null hypothesis, then covert this to the p-value $P_k=P(T_{2nk-2})> t_k$, and finally, obtain the

Z-value $Z_k = \phi^{-1}(1-P_k)$. P-values for the 2 patient cohorts will be presented separately, and for the 2 patient cohorts combined with and without using weighted inverse normal test.

The weights are derived from the sample sizes for each of the 2 patient cohorts based on the assumption of no sample size change and no discontinued treatment arms. Enrollment projections anticipate approximately 1000 patients will be included in the SSR analysis at IA2 and approximately 1202 patients not included in the SSR. Patients included in the sample size re-estimation will be those randomized on or after the date of the last randomized patient who has a 6 month ADAS-Cog13 or iADL score as of the IA2 cut-off date. This patient allocation corresponds to planned cumulative information fractions of 0.45 and 1; the weights for the inverse normal method would be therefore given as $\omega_1 = \sqrt{0.45}=0.67$ and $\omega_2 = \sqrt{((1-0.45))}=0.74$. These weights will remain fixed even if the sample size is increased or a treatment arm is stopped, thus protecting the Type I error rate. Due to the uncertainty in enrollment, the number of patients enrolled at the time of the interim may differ than what is currently projected. Should that occur, the stage 1 weight will be defined as the square root of the information fraction at IA2, which is the number of patients included in the SSR analysis divided by 2202, and the stage 2 weight will be the square root of 1 minus the information fraction at IA2. The weighted inverse normal method will be applied to all hypotheses included in the multiplicity graph. The same weights will be used for all hypotheses. The weighted inverse normal method will not be applied to any other hypotheses.

6.6. Descriptions of Analysis Datasets

The Full Analysis Set will group patients according to randomized treatment assignment, even if the patient does not take the assigned treatment, does not receive the correct treatment, switches to a different treatment group if the assigned treatment group is dropped at an interim analysis, or otherwise does not follow the protocol. In the event that the study is reduced to 2 arms at an interim analysis, patients originally randomized to the dropped dose will have their data analyzed separately. Additionally, when change from baseline is assessed in the Full Efficacy dataset, patients will be included in the analysis only if both a baseline and at least 1 valid post-baseline measure are available.

All patients who received at least 1 dose of randomized study treatment (LY3314814 or placebo) will be included in the safety analysis set. If an LY3314814 dose group is dropped following an interim analysis, patients switched to the remaining LY3314814 dose and remaining on treatment will be accounted for as a separate group.

For the CSF biomarker and PET (amyloid) sub-studies, the analysis sets will include all those patients whose eligibility was determined by the respective biomarker, who signed consent for the respective sub-studies, and who have a baseline and at least 1 valid post-baseline CSF assessment or PET scan.

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

The primary and secondary efficacy measures will be analyzed using the full efficacy dataset unless otherwise specified. Summaries and analyses for safety measures will be based on the safety population.

Table 3 below defines each of the analysis populations used in this study. Table 4 lists the study measures that will be summarized and/or analyzed for each population.

Table 3. Analysis Datasets for Study I8D-MC-AZES

Dataset Name	Description of Population
All Patients Entered	All patients who signed informed consent
Full Efficacy	All randomized patients using scale data captured on or prior to study treatment discontinuation date
Safety	All randomized patients with at least 1 dose of study medication

Table 4. Efficacy and Safety Measures Summarized and/or Analyzed for Each Analysis Dataset

Dataset Name	Variables Assessed
All Patients Entered	Listings
Full Efficacy	Tables, Listings, and/or Figures of the following: patient disposition, patient characteristics, pre-existing conditions, significant historical diagnoses ADAS-Cog_{11}, basic, instrumental and total ADCS-ADL,, ADAS-Cog_{13}, CDR-SB, RBANS, NPI, FAQ, RUD-Lite, EQ-5D, Letter & Category Fluency/Digit Symbol Coding Tests, MMSE, plasma Aβ parameters, vMRI parameters, CSF parameters, AV-45 parameters, FDG parameters, AV-1451 parameters and concomitant medications.
Safety	Tables, Listings, and/or Figures of the following: Compliance, adverse events, laboratory results, vital signs, weight, ECG, MRI, eye exam, skin exam, and C-SSRS.

Abbreviations: ADAS-Cog = Alzheimer's Disease Assessment Scale-Cognition; ADCS-iADL= The Alzheimer's Disease Cooperative Study Activities of Daily Living Inventory – instrumental items scores; CDR-SB = Clinical Dementia Rating - Sum of Boxes; C-SSRS = Columbia-Suicide Severity Rating Scale; ECG = electrocardiogram; FAQ = Functional Activities Questionnaire; FDG = fluorodeoxyglucose; MMSE =Mini-Mental State Examination; MRI = magnetic resonance imaging; EQ-5D = EuroQol five dimensions questionnaire; RBANS = Repeatable Battery for the Assessment of Neuropsychological Status; RUD-Lite = Resource Utilization in Dementia-Lite.

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.

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

From the randomized population, the percentage of patients withdrawing from each treatment group will be summarized. Patients discontinuing treatment due to the sponsor's decision to end the phase 3 program, following the futility analysis, will have "sponsor decision" as reason for discontinuation. From the safety population, the percentage of patients withdrawing from each treatment group will be compared between groups using Fisher's exact test. Comparisons using Fisher's exact test will be done for the overall percentage of patients who withdraw and also for each specific reason for withdrawal. Patient disposition will be listed.

6.8. Patient Characteristics

Baseline characteristics will be summarized for the randomized population by treatment group and overall. Summaries will include descriptive statistics for continuous and categorical measures. Fisher's exact test or Pearson's chi-square test will be used for treatment-group comparisons of categorical data. For continuous data, analysis of variance (ANOVA), with independent factors for treatment, will be used. Patient characteristics to be presented include:

- disease status
- age
- age group: 55 to 64, 65 to 74, and 75 to 85
- gender
- race
- ethnicity
- height
- body weight
- body mass index (weight (kg) / [height (m)]2)
- region
- tobacco use
- alcohol use
- years of education
- work status
- method of amyloid positivity determination (PET, CSF, historical PET)
- APOE4 carrier status (carrier [ε2/ε4, ε3/ε4, ε4/ε4], noncarrier [ε3/ε3, ε2/ε2, ε3/ε2])
- APOE4 genotype (ε2/ε4, ε3/ε4, ε4/ε4, no ε4)
- having 1 or more first degree relatives with AD
- AChEI and/or memantine use at baseline
- Baseline severity of impairment as measured by ADAS-Cog$_{13}$, ADAS-Cog$_{11}$, ADCS-ADL total score and instrumental (ADCS-iADL) and basic subscores (ADCS-bADL), CDR Sum of Boxes, MMSE, Letter and Category Fluency tests, Symbol Coding test, NPI, RUD-Lite, EQ-5D Proxy, QoL-AD, and FAQ.

Baseline characteristics will also be listed.

6.9. Treatment Compliance

Only patients who consume at least 80% of the prescribed daily dose during this study will be considered compliant. The percentage of compliant patients will be compared across treatment groups.

6.10. Concomitant Therapy

Prior medications are defined as those that stop before randomization (the day prior to the first administration of study drug). Concomitant medications are defined as those being taken on or after randomization (the day prior to the first administration of study drug). A summary of concomitant medications will be presented as frequencies and percentages for each treatment group. Fisher's exact test will be used to test for treatment differences between groups. If the start or stop dates of therapies are missing or partial to the degree that determination cannot be made of whether the therapy is prior or concomitant, the therapy will be deemed concomitant. A summary table will also be provided for concomitant AChEI/memantine medications. Medications will be coded using the World Health Organization (WHO) drug dictionary. Concomitant medications will be listed.

6.11. Study Partners

The protocol states every effort should be made to keep the same study partner through the duration of this trial. However, changes may be unavoidable. The percentage of patients with the same study partner will be compared across treatment groups using a Fisher's Exact test. Additionally, study partner changes will be categorized (0 changes, 1 change, and more than 1 change) and compared using Pearson's chi-square test.

6.12. Efficacy Analyses

6.12.1. Primary Outcome and Methodology

The primary objective of this study is to test the hypothesis that treatment with 20 mg or 50 mg of LY3314814 in patients with early Alzheimer's disease will slow the decline of AD as compared with placebo. Early AD is defined as the continuum of patients with MCI due to AD (ie, prodromal AD) and patients diagnosed with mild dementia of the Alzheimer's type. The hypothesis will be assessed using a mixed model repeated measures (MMRM) analysis of ADAS-Cog_{13} total score, in which the specific hypothesis is that the change from baseline at Week 104 (ie, the end of the treatment period) for LY3314814 20 mg or 50 mg will be significantly less than that for placebo.

The change from baseline score at each scheduled post-baseline visit (according to the Study Schedule - data collected at unscheduled visits, including early discontinuation visits if not matching the study schedule, will not be included in this analysis) during the treatment period will be the dependent variable. The model will include the fixed effects of treatment, visit (categorical covariate), treatment-by-visit interaction, disease status at baseline (MCI due to AD or mild AD), APOE4 status (carrier versus non-carrier), concomitant AChEI use at baseline (yes/no), and country, and the continuous effects of baseline ADAS-Cog_{13} total score

(continuous covariate), age at baseline, and baseline ADAS-Cog$_{13}$ total score-by-visit interaction. Visit will be considered a categorical variable with values equal to the visit numbers at which the scales were assessed. The null hypothesis is that both contrasts between each LY3314814 dose group versus placebo at the last visit equals zero. An unstructured covariance matrix will be used to model the within-patient variance-covariance errors.

If the model fails to converge, the covariance structures will be evaluated in the following order until model convergence is met: heterogeneous toeplitz, heterogeneous first-order autoregressive, heterogeneous compound symmetry, toeplitz, first-order autoregressive, compound symmetry. This order is specified according to a decreasing number of covariance parameters in the structure. The sandwich estimator (Diggle et al. 1994) for the covariance estimation will be used by specifying the EMPIRICAL option in SAS PROC MIXED. When sandwich estimation is used, the Kenward-Roger approximation for denominator degrees of freedom cannot be used. Instead, DDFM= BETWITHIN option will be used to estimate denominator degrees of freedom.

6.12.2. Additional Analyses of the Primary and Key Secondary Outcomes

6.12.2.1. Subgroup Analyses

The ADAS-Cog$_{13}$ and ADCS-ADL results for change from baseline at Week 104 will be assessed for consistency among subgroup strata (Section 6.16). The MMRM model will be repeated for each of these factors and will include a term for the interaction between the factor and treatment. Additionally, the MMRM analyses will be run on each subgroup strata individually.

6.12.2.2. Logistic Ordinal Regression

Treatment differences in final post-baseline CDR global scores will be assessed using logistic ordinal regression analysis. The logistic ordinal regression model will include independent variables for treatment and disease status at baseline (MCI due to AD or mild AD). The null hypothesis is that the contrast of LY3314814 groups versus placebo equals zero.

6.12.2.3. Time to Conversion Analyses

Time in the current disease state, defined as the time until progression to the next disease state as measured by the CDR global score, will be analyzed using time-to-event techniques. For patients enrolled with a CDR global score of 0.5, progression is defined as a CDR global score of 1 or greater. For patients enrolled with a CDR global score of 1, progression is defined as a CDR global score of 2 or greater. Kaplan-Meier methodology will be used to estimate median time-to-event for each treatment group and Kaplan-Meier curves by treatment group will be presented. A stratified log-rank test will be used to test the hypothesis of equal survival curves between the LY3314814 treatment groups and the placebo group. All analyses will be stratified by disease status at baseline (MCI due to AD or mild AD).

6.12.2.4. Time to Initiation or Dose Increase of Symptomatic Medications

Time to initiation or dose increase of symptomatic medications (memantine and AChEI) will be analyzed using time-to-event methods. Kaplan-Meier methodology will be used to estimate the median time-to-initiation or increase for each treatment group. Kaplan-Meier curves by treatment group will be presented. A log-rank test will be used to test the hypothesis of equal survival curves between the LY3314814 treatment groups and the placebo group.

6.12.3. Other Secondary Efficacy Analyses

The additional clinical and outcome measurements listed below will be analyzed separately 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 terms for the 7 independent effects listed previously (Section 6.12.1). The null hypothesis is that the differences in least-squares means between the LY3314814 dose groups versus placebo at Week 104 equal zero. The outcomes that will be analyzed are:

- Change from baseline as obtained from the ADAS-Cog_{11}.
- Change from baseline in ADCS-ADL total score, the instrumental ADL subscale, and the basic ADL subscale.
- Change from baseline in CDR Sum of Boxes
- Change from baseline in MMSE
- Change from baseline in FAQ
- Change from baseline in behavioral disturbance as measured by NPI (12 item) total score (frequency multiplied by severity).
- Change from baseline in NPI score (frequency multiplied by severity) for each of the 12 NPI modules.
- Change from baseline in total NPI frequency.
- Change from baseline in total NPI severity.
- Change from baseline in the Letter Fluency, Category Fluency, and Digit Symbol-Coding test scores.
- Change from baseline to Week 97 in RBANS Total, Immediate Memory, Delayed Memory, Visuospatial/Construction, Language, and Attention scores.

6.13. Health Outcomes/Quality-of-Life Analyses

Shifts from baseline in quality of life as measured by EQ-5D Proxy utility component scores and QoL-AD subscale scores will be presented. Analyses of these measures will follow the same methods as outlined in Section **Error! Reference source not found.** for additional MMRM analyses.

Resource utilization as measured by RUD-Lite (basic ADL hours/day, instrumental ADL hours/day, supervision hours/day, and the sum of basic and instrumental ADL hours/day) will be compared across treatment groups using an MMRM with fixed effects including terms for the 7 independent effects listed in Section 6.12.1 will be used. Number of patients experiencing hospitalizations will be compared across treatment groups using Fisher's Exact test. If supported by the data, a logistic regression will be run using number of patients experiencing

hospitalizations comparing treatment groups with baseline age as a covariate and treatment as fixed effects will also be conducted.

The proportion of patients who have a change in permanent living accommodation will be summarized and treatment comparisons will be conducted using Fisher's Exact Test.

Caregiver change from previous visit (yes/no) will be summarized by treatment group. Incidence of caregiver change will be compared across treatment groups using Fisher's Exact test.

6.14. Bioanalytical and Pharmacokinetic/Pharmacodynamic Methods

6.14.1. Analysis of Plasma Aβ

To evaluate the change in plasma Aβ analytes (including assayed plasma $A\beta_{1-40}$ and $A\beta_{1-42}$) after treatment, an MMRM will be used to compare change from baseline to 52 and 104 weeks. This analysis will be run separately for each plasma Aβ parameter using the full efficacy dataset. The model will include the fixed, categorical effects of treatment, visit, treatment-by-visit interaction, and disease status at baseline (MCI due to AD or mild AD) as well as the continuous effect of baseline plasma Aβ. Visit will be considered a categorical variable with values equal to the visit numbers at which plasma Aβ is assessed. The null hypothesis is that the difference in LSM between the LY3314814 dose groups and placebo equal zero. A similar analysis will be performed for completers.

To assess the relationship of plasma Aβ with cognition and function with treatment, Spearman's rank correlation coefficient will be obtained on change in plasma Aβ from baseline to Week 104 and with change from baseline to Week 104 for $ADAS\text{-}Cog_{13}$ and ADCS-ADL, MMSE, FAQ, and CDR-SB. Correlation analyses will be conducted using only patients who have the clinical outcome and plasma Aβ result at Week 104.

6.14.2. Analysis of CSF Data

Analyses of CSF biomarkers (including total CSF $A\beta_{1-40}$, total CSF $A\beta_{1-42}$, free CSF $A\beta_{1-40}$, free CSF $A\beta_{1-42}$, CSF total tau, and CSF p-tau from lumbar puncture) will be run using the full efficacy dataset. The dependent variable for each CSF parameter is its change from baseline to endpoint. The model will include the fixed, categorical effects of treatment and disease status at baseline (MCI due to AD or mild AD) as well as the continuous effects of baseline CSF and age at baseline. The null hypothesis is that the difference in LSM between the LY3314814 dose groups versus placebo equal zero. Analyses of additional CSF analytes including RBC and WBC will be done in a similar manner. Similar analyses will be performed for completers.

Annualized change in CSF biomarkers for each patient will be calculated using the change in CSF at the last post-baseline visit. The annualized change will be compared between the treatment groups with an ANCOVA. The ANCOVA model will include the following independent variables: baseline CSF value, treatment, age at baseline, and disease status at baseline (MCI due to AD or mild AD). The null hypothesis is that the difference in LSM between the LY3314814 dose groups and placebo equal zero.

To assess the relationship of CSF parameters with cognition and function with treatment, Spearman's rank correlation coefficient will be obtained on change in CSF result from baseline to Week 104 versus changes in ADAS-Cog_{13} and ADCS-ADL, MMSE, FAQ, and CDR-SB; this will be performed using all patients who have the clinical outcome and CSF result at Week 104.

6.14.3. Analyses of Amyloid PET Data

The PET images acquired in the study will be processed using previously mentioned methods (Clark et al. 2012). At baseline, standard uptake value ratio (SUVr) will be calculated using as a ratio of the composite summary region that is an average of 6 different cortical regions (anterior cingulate, posterior cingulate, medial orbital frontal, lateral temporal, lateral parietal, precuneus) with whole cerebellum as a reference region. However, post-baseline SUVr values will be calculated using 2 different reference regions whole cerebellum and a correction factor using atlas based white matter (AWM). The SUVr with whole cerebellum will be calculated as a ratio of composite summary region to whole cerebellum as a reference region, similar to the calculation at baseline. The SUVr values using AWM correction factors will be calculated by dividing the composite summary ratio by an AWM correction factor. This correction factor is a ratio of SUV values of AWM to whole cerebellum from baseline to post-baseline. The complete listing of AV45 regions is provided in Appendix 1.

Change from baseline and annualized change from baseline in the composite summary SUVr of AV-45 for each patient will be calculated using the change at the last post-baseline visit. The annualized change in the full efficacy dataset will be compared between the treatment groups with an ANCOVA. The ANCOVA model will include the fixed, categorical effects of treatment and disease status at baseline (MCI due to AD or mild AD), as well as continuous effects of baseline AV-45 value and age at baseline. The null hypothesis is that the difference in LSM between the LY3314814 dose groups and placebo equal zero. Similar analyses will be performed for completers.

To assess the relationship of AV-45 with cognition and function with treatment, Spearman's rank correlation coefficient will be obtained on change from baseline to Week 104 for the composite summary standard uptake value ratio (SUVr) normalized (based on whole cerebellum and based on atlas-based white matter), and with change from baseline to Week 104 for ADAS-Cog_{13} and ADCS-ADL, MMSE, FAQ, and CDR-SB. Correlation analyses will be conducted using only patients who have the clinical outcome and AV-45 result at Week 104 and include patients from all 3 dose groups.

6.14.4. Analyses of FDG PET Data

The primary analysis of the fluorodeoxyglucose positron emission tomography (FDG PET) scans will follow the established methods of Landau et al (2011). Composite FDG SUVr will be calculated using the following regions: the left and right parietal (angular gyrus), left and right posterior cingulate, and the left and right temporal lobes (Landau et al, 2011). Two composite summary standard uptake value ratios (SUVr) of FDG PET normalized to the pons + vermis will be assessed: (1) Composite Meta and (2) Composite Meta Automated Anatomical Labeling atlas (AAL).

An ANCOVA model will be run on the full efficacy dataset using change from baseline to the post-baseline visit of the composite SUVR as the dependent variable. The model will include the fixed, categorical effects of treatment and disease status at baseline (MCI due to AD or mild AD) as well as continuous effects of baseline biomarker result and age at baseline. The null hypothesis is that the difference in LSM between the LY3314814 dose groups versus placebo equal zero. Similar analyses will be performed for completers.

Annualized change in the composite summary SUVr for each patient will be calculated using the change at the last post-baseline visit. The annualized change will be compared between the treatment groups with an ANCOVA. The ANCOVA model will include the following independent variables: baseline biomarker value, treatment, age at baseline, and disease status at baseline (MCI due to AD or mild AD). The null hypothesis is that the difference in LSM between the LY3314814 dose groups and placebo equal zero.

To assess the relationship of biomarker with cognition and function with treatment, Spearman's rank correlation coefficient will be obtained on change from baseline to Week 104 for the composite SUVrand with change from baseline to Week 104 for ADAS-Cog_{13} and ADCS-ADL, MMSE, FAQ, and CDR-SB. Correlation analyses will be conducted using only patients who have the clinical outcome and SUVr result at Week 104 and include patients from all 3 dose groups.

6.14.5. Analysis of AV-1451 Data

To evaluate the change from baseline in tau imaging parameters, an MMRM analysis will be used to compare change from baseline in SUVr at 104 weeks in the full efficacy dataset. The model will include the fixed, categorical effects of treatment, visit, treatment-by-visit interaction and disease status at baseline (MCI due to AD or mild AD), 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 LY3314814 dose groups and placebo equals zero.

Change from baseline and annualized change from baseline analyses will be conducted on SUVrs computed from the MUBADA region with the bimodal white matter serving as the reference region.The annualized change will be compared between the treatment groups with an ANCOVA on the full efficacy dataset. The ANCOVA model will include the fixed effects of treatment and disease status at baseline (MCI due to AD or mild AD), as well as continuous effects of baseline AV-1451 value and age at baseline. The null hypothesis is that the difference in LSM between the LY3314814 dose groups and placebo equal zero.

To assess the relationship of biomarker with cognition and function with treatment, Spearman's rank correlation coefficient will be obtained on change from baseline to Week 104 for the composite summary standard uptake value ratio (SUVr) normalized to bimodal white matter and with change from baseline to Week 104 for ADAS-Cog_{13} and ADCS-ADL, MMSE, FAQ, and CDR-SB. Correlation analyses will be conducted using only patients who have the clinical outcome and SUVr result at Week 104 and include patients from all 3 dose groups.

6.14.6. Analyses of vMRI Data

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

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

All of the above volumes are corrected for intracranial volume. To evaluate the changes in vMRI data after treatment, an ANCOVA model will be used to compare change from baseline to 104 weeks in the full efficacy dataset. The change from baseline to the endpoint visit will be the dependent variable. The model will include the fixed, categorical effects of treatment and disease status at baseline (MCI due to AD or mild AD) as well as the continuous effects of baseline vMRI value and age at baseline. The null hypothesis is that the difference in LSM between the LY3314814 active dose groups and placebo equal zero. A similar analysis will be performed for completers.

Annualized change in vMRI for each patient will be calculated using the change in vMRI at the last post-baseline visit. The annualized change will be compared between the treatment groups with an ANCOVA model on the full efficacy dataset. The ANCOVA model will include fixed, categorical effects of treatment and disease status at baseline (MCI due to AD or mild AD) as well as the continuous effects of baseline vMRI value, and age at baseline. The null hypothesis is that the difference in LSM between the LY3314814 dose groups and placebo equal zero.

To assess the relationship of vMRI with cognition and function with treatment, Spearman's rank correlation coefficient will be obtained on change from baseline to Week 104 for vMRI parameters with change from baseline to Week 104 for ADAS-Cog_{13} and ADCS-ADL, MMSE, FAQ, and CDR-SB; this will be performed using all patients who have the clinical outcome and vMRI result at Week 104.

6.15. Safety Analyses

6.15.1. Extent of Exposure

Days of exposure will be calculated for each patient (date of last dose – date of first dose +1). Summary statistics will be provided for the total number of days and patient-years of exposure by treatment. Study drug treatment assignment will be listed.

6.15.2. Adverse Events

Treatment-emergent adverse events (TEAEs) will be defined as events that first occurred or worsened after the randomization date (Visit 2 date). The treatment-emergent period ends on the last day of treatment plus 5 days (these 5 days constitute at least 5 half-lives of lanabecestat). Events occurring within the study but during a period of treatment interruption will only be treated as treatment-emergent if they occur within 5 days of the last dose prior to the treatment interruption. Should there be insufficient data for AE start date, stop date, and time to make this comparison, the AE will be considered treatment-emergent. The MedDRA lower-level term (LLT) will be used in the treatment-emergent computation. The maximum severity for each lower-level term (LLT) during the baseline period will be used as baseline.

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

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

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

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

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

6.15.3. 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. A listing and summary will be provided for major adverse cardiovascular events (MACE) that are blindly evaluated by external independent consultants.

In addition, the proportion of patients within specific clusters of TEAEs will be summarized and treatment comparisons will be conducted using Fisher's Exact Test. Clusters will be created from MedDRA High Level Group Terms (Table 5) and MedDRA SMQs.

Table 5. Adverse Events of Special Interest

AE Groups of Interest (Clusters)	MedDRA HLGT
Nervous System Disorders	Neuromuscular Disorders HLGT Demyelination SMQ Peripheral Neuropathy SMQ
Eye Disorders	Retinal disorders SMQ
Skin Disorders	Sub-group A: Epidermal and Dermal Conditions HLGT (excluding sub-group B terms) Sub-group B (Hypopigmentation-related events): Hypopigmentation disorders HLT Pigmentation changes, NEC HLT Preferred terms: 'hair depigmented', 'eyelash discolouration', 'iris hypopigmentation', 'eye colour change', 'lip colour altered', 'lip discolouration', 'hair colour changes', 'achromotrichia aquired', 'poliosis'
Liver Disorders	Drug related hepatic disorders - comprehensive search SMQ
Cardiovascular-type events – Arrhythmic	Arrhythmia related investigations, signs and symptoms SMQ Cardiac arrhythmia terms (incl bradyarrhythmias and tachyarrhythmias) SMQ TdP/QT prolongation SMQ
Cardiovascular-type events – Ischemic	Ischaemic heart disease SMQ
Cardiovascular-type events – Stroke	Central nervous system vascular disorders SMQ
Cardiovascular-type events – including orthostatic hypotension	Decreased and Nonspecific Blood Pressure Disorders and Shock HLGT

Abbreviations: SMQ = standardized MedDRA Query, NEC = Not Elsewhere Classified, HLGT = High Level Group Term, TdP/QT = Torsades de pointes /QT interval.

6.15.4. Clinical Laboratory Evaluation

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

Change from baseline to post-baseline visit at which laboratory measurements are taken will be compared between treatment groups using an MMRM model on the Safety Dataset. For each lab

analyte, the rank-transformation will be applied to the change from baseline for all patients and all visits prior to analysis. Similarly, an independent rank-transformation will be applied to the baseline values prior to analysis. The model will include the fixed, categorical effects of treatment, visit, treatment-by-visit interaction, and disease status at baseline (MCI due to AD or mild AD) as well as the continuous effects of ranked baseline value and age at baseline. This analysis will be done separately for each laboratory analyte.

Treatment differences in the proportion of patients with treatment-emergent high or treatment-emergent low or treatment-emergent abnormal laboratory values at (1) anytime and (2) each post-baseline visit will be assessed using Fisher's exact test. Treatment-emergent high or low laboratory abnormality will be based on SI unit. For each laboratory analyte, only patients who were low or normal at baseline and have at least 1 post-baseline will be included in the denominator when computing the proportion of patient with treatment-emergent high. Similarly, only patients who were high or normal at baseline and have at least 1 post baseline will be included in the denominator when computing the proportion of patient with treatment-emergent low. In addition, treatment differences in the proportion of patients who have normal baselines with a change to abnormal high or abnormal low values at any post-baseline visits will be summarized.

A second categorical analysis will be conducted on laboratory analytes. This analysis is considered a PCS analysis and will use limits typically wider than the first categorical analysis. Abnormal criteria for these treatment-emergent PCS changes are presented in Appendix 3.

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

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

The proportion of patients with treatment-emergent clinically significant changes from a low value or normal value at all baselines at any time in ALT and total bilirubin will be summarized by treatment group. Clinically significant changes of interest at any time are: ALT ≥3 x upper limit of normal (ULN) and total bilirubin ≥2 x ULN, AST ≥3 x ULN, ALT ≥5 x ULN, ALT ≥10 x ULN, and total bilirubin ≥2 x ULN. Additionally, Hy's Law analysis will be conducted by comparing treatment groups with regard to the proportion of patients with (ALT ≥3 x ULN OR AST ≥3 x ULN) AND total bilirubin ≥2 x ULN at any time. Comparisons between treatment groups will be carried out using Fisher's Exact test. When criteria are met for hepatic evaluation and completion of the hepatic safety case report form (CRF), investigators are required to answer a list of questions pertaining to the patient's history, relevant pre-existing medical conditions, and other possible causes of liver injury. A listing of the information collected on the hepatic-safety CRF will be generated.

6.15.5. Vital Signs and Other Physical Findings

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

If there are multiple records of vital sign or weight measurements at baseline or postbaseline visit, the last record will be used. Summary statistics will be presented for observed values at baseline and for change from baseline results at each scheduled postbaseline visit. Systolic and diastolic blood pressure and pulse (collected in sitting position), orthostatic diastolic and orthostatic systolic blood pressures and orthostatic pulse (measurement after 5 minutes in the supine position minus that after 2 minutes and 5 minutes in the standing position), temperature, and weight by treatment group for all patients in the safety population will be summarized.

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

The incidence of treatment-emergent abnormal high or low vital signs and weight will be presented by treatment group and visit. Treatment-emergent vital sign evaluations are defined for evaluations collected after the initiation of study medication. Abnormal criteria for post-baseline vital signs and weight are presented in Appendix 4. Any vital sign or weight meeting the criteria will be considered abnormal. Treatment differences in the proportion of patients with treatment-emergent abnormal high or low vital signs and weight will be assessed between treatment groups using Fisher's exact test at (1) any time (2) post-baseline visit.

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

Summary and analyses of change from baseline in weight will be provided. The proportion of patients with a weight gain or loss of greater than or equal to 7 percent of baseline body weight will be compared between treatment groups using Fisher's Exact test at each visit and at any time.

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

6.15.6. Electrocardiograms

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

The ECG measurements are derived from three 10 second readings taken every 30 seconds. These 3 readings are to be averaged prior to analysis. Additionally, whenever ECG is measured

in triplicate, the average of these readings will be used in the analysis. If there are multiple records after averaging ECG triplicates within a visit, the last record of averages will be used.

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

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

Abnormal ECG criteria and criteria for abnormal QTcF prolongation are presented in Appendix 5.

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

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

6.15.7. Analyses of MRI Data

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

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

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

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

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

A listing of MRI data will also be presented.

6.15.8. Additional Safety Concerns

6.15.8.1. Columbia Suicide Severity Rating Scale

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

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

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

Specifically, the following outcomes are C-SSRS categories and have binary responses (yes/no). The categories have been re-ordered from the actual scale to facilitate the definitions of the composite and comparative endpoints, and to enable clarity in the presentation of the results.

Category 1 – Wish to be Dead
Category 2 – Non-specific Active Suicidal Thoughts
Category 3 – Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act

Category 4 – Active Suicidal Ideation with Some Intent to Act, without Specific Plan
Category 5 – Active Suicidal Ideation with Specific Plan and Intent
Category 6 – Preparatory Acts or Behavior
Category 7 – Aborted Attempt
Category 8 – Interrupted Attempt
Category 9 – Actual Attempt (non-fatal)
Category 10 – Completed Suicide

Self-injurious behavior without suicidal intent is also a C-SSRS outcome (although not suicide-related) and has a binary response (yes/no).

Composite endpoints based on the above categories are defined below.

- Suicidal ideation: A "yes" answer at any time during treatment to any one of the five suicidal ideation questions (Categories 1-5) on the C-SSRS.
- Suicidal behavior: A "yes" answer at any time during treatment to any one of the five suicidal behavior questions (Categories 6-10) on the C-SSRS.
- Suicidal ideation or behavior: A "yes" answer at any time during treatment to any one of the ten suicidal ideation and behavior questions (Categories 1-10) on the C-SSRS.

The following outcome is a numerical score derived from the C-SSRS categories. The score is created at each assessment for each patient and is used for determining treatment emergence.

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

Comparative endpoints of interest are defined below. "Treatment emergence" is used for outcomes that include events that first emerge or worsen. "Emergence" is used for outcomes that include events that first emerge.

- Treatment-emergent suicidal ideation compared to recent history:
  An increase in the maximum suicidal ideation score during treatment (Visits Y1-Y2) from the maximum suicidal ideation category during the screening and lead-in periods (C-SSRS scales taken at Visits X1-X2). Recent history excludes "lifetime" scores from the Baseline C-SSRS scale or Baseline/Screening C-SSRS scale.
- Treatment-emergent serious suicidal ideation compared to recent history: An increase in the maximum suicidal ideation score to 4 or 5 on the C-SSRS during treatment (Visits Y1-Y2) from not having serious suicidal ideation (scores of 0-3) during the screening and lead-in periods (C-SSRS scales taken at Visits X1-X2). Recent history excludes "lifetime" scores from the Baseline C-SSRS scale or Baseline/Screening C-SSRS scale.
- Emergence of serious suicidal ideation compared to recent history:
  An increase in the maximum suicidal ideation score to 4 or 5 on the C-SSRS during treatment (Visits Y1-Y2) from no suicidal ideation (scores of 0) during the screening and

lead-in periods (C-SSRS scales taken at Visits X1-X2). Recent history excludes "lifetime" scores from the Baseline C-SSRS scale or Baseline/Screening C-SSRS scale.

- Improvement in suicidal ideation at endpoint compared to baseline:
  A decrease in suicidal ideation score at endpoint (the last measurement during treatment; Visits Y1-Y2) from the baseline measurement (the measurement taken just prior to treatment; (Visit X2). This analysis should only be performed for a non-lifetime baseline measurement (i.e., having improvement from the worse event over a lifetime is not clinically meaningful). A specific point in time can be used instead of endpoint.
- Emergence of suicidal behavior compared to all prior history:
  The occurrence of suicidal behavior (Categories 6-10) during treatment (Visits Y1-Y2) from not having suicidal behavior (Categories 6-10) prior to treatment (Visits X1-X2). Prior to treatment includes "lifetime" and/or "screening" scores from the Baseline C-SSRS scale, Screening C-SSRS scale, or Baseline/Screening C-SSRS scale, and any "Since Last Visit" from the Since Last Visit C-SSRS scales taken prior to treatment.

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

6.15.8.2. Skin Examination

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

Any hypopigmentation will be assessed by location, percentage of body surface area involvement, degree (partial/decreased pigmentation to complete depigmentation), and other findings in or around the hypopigmentation area (eg, redness or induration). A static physician's global assessment (sPGA) will be used to determine the patient's overall hypopigmentation severity at a given time point using a visual analog scale (VAS) ranging from 0 to 100. In addition, patients noted to have evidence of hypopigmentation will be asked to record how bothersome they find the hypopigmentation to be on a VAS ranging from 0 to 100. Additionally, the percentage of patients with emergence of greater than expected hair hypopigmentation will be summarized for patients 'no' at baseline if question is available on worksheet.

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

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

6.15.8.3. Eye Examination

Frequency tables will be produced for all time points for performance of eye examination, visual acuity examination, intraocular pressure examination, and slit lamp exam status and dilated fundus exam status (normal, abnormal – clinically not significant, and abnormal – clinically significant). Clinically significant abnormalities will be displayed together with the corresponding specifications of abnormalities in separate individual data listings.

Summary statistics will be produced for the following continuous parameters: left eye total visual acuity score, right eye total visual acuity and both eyes score (scores expressed as logMAR calculated as the negative log [base 10] of the decimal scores), as well as left eye intraocular pressure and right eye intraocular pressure (both in mmHg). For visual acuity, "count fingers" will be given a decimal score of 0.01 and a logMAR of 2 (reference http://www.hicsoap.com/publications/ProperMethodforCalculating.pdf). “Light perception” and “no light perception” cannot be assigned decimal or LogMAR values and so are treated as missing in the mean change summary tables. Visual acuities of patients with these values at any time during the study will be provided in a separate listing.

In order to display any changes/deteriorations during treatment, the following will be reported: number of patients with potentially clinically significant changes (slit lamp examination or dilated fundus examination) documented during treatment that were not already present at baseline; summary statistics for the difference of maximum value during treatment minus baseline value for left eye total visual acuity score, right eye total visual acuity score, left eye intraocular pressure (mmHg), and right eye intraocular pressure (mmHg). Worst assessment of overall eye examination results during treatment with possible entries “unchanged”, “new”, “improved”, and “worsened”.

6.16. Subgroup Analyses

To assess the effects of various demographic and baseline characteristics on treatment outcome, subgroup analyses for the primary endpoint, ADAS-Cog$_{13}$, and key secondary endpoint ADL total score will be performed based on the following variables:

- Disease status at baseline – MCI due to AD or mild AD
- Age - ≥ 55 and < 72 or ≥ 72 and < 85
- APOE4 Carrier Status – Carrier defined as E2/E4, E3/E4, or E4/E4 genotype; No-Carrier defined as all other genotypes.
- Region – US or OUS
- Newly initiated concomitant AD therapy or increase in dose from baseline – No Standard of Care (StOC) treatment versus acetylcholinesterase inhibitors (AChEIs) or memantine

The primary outcome measure will be modeled using a MMRM approach. This general model will include terms for baseline, treatment, 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.

6.17. 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 at any of the visits at which the scales were scheduled to be assessed.
- Not compliant with study drug calculated as taking less than 80% or greater than 120% of study drug while the subject was expected to be on treatment.

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.
-

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.

6.18. Interim Analyses and Data Monitoring

An IDMC will have the responsibility to review accumulating unblinded study data on a periodic basis and make recommendations to protect the safety of patients. Each member of the IDMC 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 IDMC members, the frequency with which it meets, and the structure of their meetings. A SAP for analyses associated with the DMC list out the specific analyses that the IDMC will review.

Four interim analyses may be conducted during this study. The IDMC charter will specify the final details of the interim analyses. Only the IDMC is authorized to evaluate unblinded interim efficacy and safety analyses. Study sites will receive information about interim results only if they need to know for the safety of their patients.

6.18.1. Interim Analysis 1

The objective of the first interim analysis (IA1) is to assess the viability of continuing with at least 1 dose of LY3314814 with regard to safety. An IA1 will be performed when approximately 70 patients per group have 13 weeks' worth of data. No statistical adjustments will be made to account for this interim since it is a safety only interim and there is no possibility of stopping for efficacy.

6.18.2. Interim Analysis 2

The objectives of the second interim analysis (IA2) are to assess the viability of continuing with at least 1 dose of LY3314814 with regard to safety and to re-assess the total sample size. For safety reasons, the IDMC will also evaluate ADAS-Cog 13-item scores results at IA2 to confirm that clinically meaningful worsening of cognition was not present in patients treated with LY3314814. The data cut-off for IA2 will be when approximately 80% of patients have been randomized. Patients included in the sample size re-estimation will be those randomized on or after the date of the last randomized patient who has a 6 month ADAS-Cog13 or iADL score as of the IA2 cut-off date.

The sample size may be increased from 3 up to approximately 3000 randomized patients. To determine the final sample size, the Bayesian predictive probabilities for superiority on the ADAS-Cog_{13} and iADL scores will be calculated at IA2 based on the 50 mg dose (to be consistent with the multiplicity graph). Bayesian predictive probabilities of statistical significance for each dose at Week 104 will be calculated for the ADAS-Cog_{13} and iADL scores. All available ADAS-Cog and iADL scores will be included for the subset of patients with scores at week 26. The Bayesian predictive probabilities will be calculated from a Bayesian joint model of ADAS-Cog_{13} and iADL scores using diffuse priors (refer to the interim SAP for details). If the Bayesian predictive probabilities for success is "Promising" (ie, neither very low nor very high), then the final sample size may be increased up to the maximum size (ie, 3000 patients) to maintain the desired powering of the study. The sample size would be increased to the sample size necessary to achieve a predicted probability of 80%. Otherwise if the interim result is "Unfavorable" (very low Bayesian predictive probabilities; indicating that increasing the sample size is not warranted and would not sufficiently increase the likelihood success), or "Favorable" (very high Bayesian predictive probabilities; indicating that the trial is adequately powered at the planned sample size), then the approximate planned/minimum sample size will be maintained (ie, 2202 patients).

Table 6. Sample Size Re-Estimation Zone Definitions

Endpoint	Unfavorable Zone	Promising Zone	Favorable Zone
ADAS-Cog_{13}	< 55%	55 – 80%	> 80%

Abbreviations: ADAS-Cog13 = Alzheimer's Disease Assessment Scale-Cognition.

Based on the zones, the final sample size is determined from the following algorithm:

For the 50 mg dose

1. If ADAS-Cog_{13} is in the Promising Zone, then the sample size is increased
2. If ADAS-Cog_{13} is in the Unfavorable Zone, then the sample size remains at 2202
3. If ADAS-Cog_{13} is in the Favorable Zone , then the sample size is remains at 2202

If the sample size is increased as a result of the interim analysis, the weighted, inverse normal (Lehmacher and Wassmer, 1999) will be applied to each of the primary and key secondary endpoints to control the Type I error (Section 6.5).

For assessing cognitive worsening, ADAS-Cog 13-item scores will be analyzed using the primary MMRM model described in the sample size re-estimation section. All available ADAS-Cog scores will be included. Using a 1-sided $\alpha = 0.0125$ (Bonferroni adjust for 2 dose and 2 time points) for each LY3314814 dose group, cognitive worsening of an LY3314814 dose group will be declared if the LSM of that group is statistically significantly worse than the LSM of the placebo-treated group at either the 6-month or 12-month time point or both.

6.18.3. Interim Analysis 3

A futility interim analysis (IA3) may be performed if the study sample size does not increase at IA2 or there is decreased confidence in current clinical trial design based on external data.

6.18.3.1. Data Included in IA3 Futility Analysis

The Bayesian predictive probability of statistical significance for the 50 mg dose vs. placebo at Week 104 will be calculated for the ADAS-Cog13 endpoint. The Bayesian predictive probability will be calculated from a Bayesian joint model of ADAS-Cog_{13} and iADL scores using diffuse priors. The iADL data is included (in addition to the ADAS-Cog13 data) as there is a correlation between the ADAS-Cog13 and iADL data and it is anticipated that including the iADL data will enhance the prediction of ADAS-Cog13 scores.

For the patients included in the interim analysis data set, all available scores from ADAS-Cog13 and iADL will be utilized. This will potentially include data from an individual patient at weeks 0, 13, 26, 52, 78, and 104 for ADAS-Cog13, and weeks 0, 26, 52, 78, and 104 for iADL. Note that the iADL is not collected at Week 13. A response vector is formulated for the observed scores from all patients in the interim analysis set, including the week 0 baseline assessment, for both the ADAS-Cog13 and the iADL. If a patient has a missing value at a visit, or has not progressed in the trial to a visit, a missing value indicator (an NA) is included in the response vector for all missing visits. The purpose of including the 'NA' values at missing visits is that it enables the prediction of the missing values which are provided in a very convenient fashion in the Bayesian framework.

6.18.3.2. Joint Linear Model

The primary purpose of the joint model is to obtain a distribution of predicted values of ADAS-Cog13 at week 104 for patients that do not have an observed value at week 104. This joint model was the same model used at IA2. The key features of the model are that it assumes linearity for both ADAS-Cog13 and iADL, and that it includes subject specific intercepts and slopes for both ADAS-Cog13 and iADL. The linearity assumption is reasonable based on current understanding of the progression of Alzheimer's Disease. The predicted values at week 104 are obtained assuming the same trend observed at IA3 will continue for all data collected after IA3. The model is a 'joint' model since the overall error terms for ADAS-Cog13 and iADL are allowed to be correlated in a 2 by 2 variance-covariance matrix, and the subject-specific

random slopes and intercepts for ADAS-Cog13 and iADL are allowed to be correlated in a 4 by 4 variance-covariance matrix. No structure is imposed on either of the variance-covariance matrices. The OpenBugs code that will be used to fit the joint model is provided below.

```
## OpenBugs code for the Bayesian joint model
{
    for (i in 1:Ntotal) {
        Y[i, 1:2] ~ dmnorm(mu[i, 1:2], omega[, ])
        mu[i, 1] <- b[subject[i], 1] + (b[subject[i], 2] + Beta5 *
            LY1[i] + Beta6 * LY2[i]) * time[i]
        mu[i, 2] <- b[subject[i], 3] + (b[subject[i], 4] + Beta7 *
            LY1[i] + Beta8 * LY2[i]) * time[i]
    }
    for (j in 1:Nsubj) {
        b[j, 1:4] ~ dmnorm(Beta.re[j, ], omega.re[, ])
        Beta.re[j, 1] <- Beta1
        Beta.re[j, 2] <- Beta2
        Beta.re[j, 3] <- Beta3
        Beta.re[j, 4] <- Beta4
    }
    Beta1 ~ dnorm(30, 0.005)
    Beta2 ~ dnorm(8, 0.005)
    Beta3 ~ dnorm(44, 0.005)
    Beta4 ~ dnorm(-8, 0.005)
    Beta5 ~ dnorm(0, 0.005)
    Beta6 ~ dnorm(0, 0.005)
    Beta7 ~ dnorm(0, 0.005)
    Beta8 ~ dnorm(0, 0.005)
    omega[1:2, 1:2] ~ dwish(R[, ], 2)
    sig[1:2, 1:2] <- inverse(omega[, ])
    R[1, 1] <- 1
    R[1, 2] <- 0
    R[2, 1] <- 0
    R[2, 2] <- 1
    omega.re[1:4, 1:4] ~ dwish(R.re[, ], 4)
    sig.re[1:4, 1:4] <- inverse(omega.re[, ])
    R.re[1, 1] <- 10
    R.re[1, 2] <- 0
    R.re[1, 3] <- 0
    R.re[1, 4] <- 0
    R.re[2, 1] <- 0
    R.re[2, 2] <- 10
    R.re[2, 3] <- 0
    R.re[2, 4] <- 0
    R.re[3, 1] <- 0
    R.re[3, 2] <- 0
    R.re[3, 3] <- 10
    R.re[3, 4] <- 0
    R.re[4, 1] <- 0
    R.re[4, 2] <- 0
    R.re[4, 3] <- 0
    R.re[4, 4] <- 10
}
```

The model includes generally diffuse priors on all parameters. For the parameters that estimate the mean baseline value for ADAS-Cog13 and iADL (Beta1 and Beta 3, respectively), the location parameter is defined to what is anticipated as the average baseline value for each scale. For the parameters that estimate the rate of decline for placebo at the end of the trial for ADAS-Cog13 and iADL (Beta2 and Beta4, respectively), the location parameter is defined to what is anticipated as the overall rate of decline of placebo at week 104. The precision value for the prior distribution of all parameters is set to a small value; therefore, the prior distributions on all parameters have very little impact on the calculated predicted probability. The variable Ntotal has a value of 1 to 13,308 (2218 potential total patients x 6 visits). The variable Nsubj has a value of 1 to 2218. The variables LY1 and LY2 represent indicator variables for patients on the 20 mg and 50 mg doses, respectively. The variable time includes values of 0, 0.125, 0.25, 0.5, 0.75, and 1, corresponding to the data collected at weeks 0, 13, 26, 52, 78, and 104, respectively.

6.18.3.3. Calculating the Predicted Probability of ADAS-Cog13 at IA3

The joint model will be fit at IA3 and a distribution of predicted values at week 104 for ADAS-Cog13 will be obtained for patients that do not have an observed value at week 104. The single fitting of the joint model provides the predicted values at week 104 needed to create the predicted probability metric. The process used to generate the predicted probabilities is outlined in the following steps:

1. After a sufficient burn-in period, keep the collection of samples from the first iteration of the MCMC chain and obtain the week 104 predicted scores for patients that don't have an observed value.
2. Merge the week 104 predicted scores with the values from patients that have observed data at week 104 to create a complete dataset of 2218 patients.
3. Calculate the change from baseline score as: week 104 ADAS-Cog13 score – Week 0 baseline ADAS-Cog13 score.
4. Fit the following ANCOVA model: Change from baseline = baseline + 20 mg Treatment Indicator + 50 mg Treatment Indicator.
5. Store the test statistic corresponding to the contrast of the 50 mg dose vs. placebo.
6. Repeat steps 1-6 50,000 times to generate 50,000 test statistics for the 50 mg vs. placebo contrast (note that updated predicted values—where baseline or week 104 values were not observed—will be utilized for each iteration of the MCMC chain).
7. The predicted probability is calculated as the proportion of the 50,000 times that the change score of the 50 mg dose is statistically significantly less than the change score of the placebo dose (at a 1-sided 0.025 alpha level).

6.18.3.4. IA3 Futility Threshold

If the predictive probability calculated according to the steps outlined in 6.18.3.3. is less than 0.157, then the study stops at IA3 for futility. A futility threshold of 0.157 at IA3 has the following operating characteristics:

- If the assumed effect size is 0, 85% of trials that would fail if the study were to run to completion would be stopped by this futility threshold;

- If the assumed effect size is 0.11 (the minimal statistically significant effect size at the end of trial), 3.2% of trials that would succeed if the study were to run to completion would be stopped by this futility threshold.

6.18.3.5. Assessing the Joint Linear Model

The iDMC will evaluate whether the joint linear model is appropriate for the data of this interim analysis. The iDMC will examine the plot of ADAS-Cog13 least square mean change estimates from the MMRM analysis versus time for each treatment in order to assess whether the linear projection out to 24 months from the joint linear model is valid. Additionally, the iDMC will compare the estimated treatment difference at 24 months (with 95% intervals) calculated from the joint linear model versus the estimate calculated from the MMRM analysis in order to assess robustness of the joint linear model.

6.18.3.6. iDMC Futility Recommendation

The iDMC will consider the joint linear model assessment described in 6.18.3.5 as well as the weight of evidence of treatment benefit from other efficacy scales (MMRM tables of iADL, CDR-SB, and MMSE), disease severity subgroups (all patients, MCI only patients, and mild AD only patients), and both doses when determining whether the study is futile at IA3.

- If the predictive probability is less than 0.157 AND (joint linear model is deemed appropriate AND the weight of evidence does support stopping for futility) then the iDMC will recommend "Stop for futility."
- If the predictive probability is less than 0.157 AND (joint linear model is NOT deemed appropriate OR the weight of evidence does NOT support stopping for futility) then the iDMC may exercise judgement and recommend "Continue the study as planned."
- If the predictive probability is greater than 0.157 AND (joint linear model is deemed appropriate AND the weight of evidence does support continuing the study) then the iDMC will recommend "Continue the study as planned."
- If the predictive probability is greater than 0.157 AND (joint linear model is NOT deemed appropriate OR the weight of evidence does NOT support continuing the study) then the iDMC may exercise judgement and recommend "Stop for futility."

6.18.4. Interim Analysis 4

An additional interim analysis (IA4) may be performed to assess futility and for early stopping of the study for efficacy.

6.18.4.1. Data Included in IA4

Data cut-off for IA4 would occur approximately 52 weeks after the last patient is randomized; however, it may occur later or earlier than this expected date if external data warrant.

6.18.4.2. Analytical Model for Futility Analysis

The joint linear model described in 6.18.3.2 and 6.18.3.3 will be used to calculate the predictive probability that the primary endpoint will be achieved at the end of the study. The iDMC will evaluate whether this model is appropriate in the same manner as described in 6.18.3.5. If the

model is deemed inappropriate, the iDMC will use the conditional power estimate from the MMRM analysis.

6.18.4.3. IA4 Futility Threshold

If the predicted probability calculated at IA4 is less than 0.229, then the study fails the IA4 futility test. A predictive probability futility threshold of 0.229 at IA4 combined with a futility threshold of 0.157 at IA3 has the following operating characteristics:

- If the assumed effect size is 0, 92.4% of trials that would fail if the study were to run to completion would be stopped by one of the two futility thresholds (at IA3 or IA4);
- If the assumed effect size is 0.11 (the minimal statistically significant effect size at the end of trial), 5.0% of trials that would succeed if the study were to run to completion would be stopped by at least one of the futility thresholds (at IA3 or IA4).

If the iDMC deems the linear model to be inappropriate for assessing futility at this interim analysis, the iDMC will compare a conditional power (CP) statistic to a futility threshold. The CP ranges from 0 to 1 and is calculated in R as follows (IF4 denotes the information fraction at the time of IA4 and is equal to 0.58):

CP = round(pnorm(((-MMRM test statistic)/sqrt(IF4) – qnorm(0.975)) / sqrt(1-IF4)), 4)

If the conditional power calculated at IA4 is less than 0.0735, then the study fails the IA4 futility test. A conditional power futility threshold of 0.0735 at IA4 combined with a predictive probability futility threshold of 0.157 at IA3 has the following operating characteristics:

- If the assumed effect size is 0, 90.0% of trials that would fail if the study were to run to completion would be stopped by at least one of the two futility thresholds (at IA3 or IA4);
- If the assumed effect size is 0.11, 4.7% of trials that would succeed if the study were to run to completion would be stopped by one of the futility thresholds (at IA3 or IA4).

6.18.4.4. iDMC Futility Recommendation

After the iDMC decides the appropriate futility statistic (predictive probability from joint linear model or conditional power from MMRM analyses), the iDMC will compare the statistic to the futility threshold.

- If the futility statistic is less than the futility threshold, then the iDMC will recommend "Stop for futility."
- If the futility statistic is greater than the futility threshold, then the iDMC will recommend "Continue the study as planned."

6.18.4.5. IA4 Early Efficacy Stopping Rules

To control for the remote possibility that the study will be stopped for efficacy, a Haybittle-Peto adjustment (Haybittle 1971, Peto et al. 1976) will be used if efficacy is tested at IA4. A two-sided significance level of α=0.001 will be used for comparisons at the start of the multiplicity graph described in 6.12.2. The final analysis will use the full α=0.05 at the start of the

multiplicity graph for comparisons between the LY3314814 treatment groups and the placebo group.

6.19. Safety Follow-Up Visit

Patients who choose to withdraw from the study upon discontinuing study treatment and after completing the early discontinuation visit assessments, as appropriate, should be asked to return for a follow-up visit (Visit 21) within 4 to 6 weeks of discontinuing treatment. Adverse events, concomitant medications, vital signs, and ECGs will be collected at these visits. Separate summaries of adverse events, concomitant medications, PCS vital signs and PCS ECGs (where baseline is AZES baseline) will be created for these data from Visit 21.

6.20. Clinical Trial Registry Analyses

Analyses provided for the CTR requirements will be a summary of AEs, provided as a dataset which will be converted to an XML file. Both SAEs and 'Other' AEs are summarized: by treatment group, by MedDRA PT. An AE is considered 'Serious' whether or not it is a treatment emergent adverse event (TEAE). An AE is considered in the 'Other' category if it is both a TEAE and is not serious. For each SAE and 'Other' AE, for each term and treatment group, the following are provided:

- the number of participants at risk of an event
- the number of participants who experienced each event term
- the number of events experienced.

Consistent with www.ClinicalTrials.gov requirements, 'Other' AEs that occur in fewer than 5% of patients/patients in every treatment group may not be included if a 5% threshold is chosen (5% is the minimum threshold).

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.

Brannath W, Koenig F, Bauer P. Multiplicity and flexibility in clinical trials. *Pharm Stat*. 2007;6:205-216.

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.

Clark CM, Pontecorvo MJ, Beach TG, Bedell BJ, Coleman RE, Doraiswamy PM, Fleisher AS, Reiman EM, Sabbagh MN, Sadowsky CH, Schneider JA, Arora A, Carpenter AP, Flitter ML, Joshi AD, Krautkramer MJ, Lu M, Mintun MA, Skovronsky DM; AV-45-A16 Study Group. Cerebral PET with florbetapir compared with neuropathology at autopsy for detection of neuritic amyloid-β plaques: a prospective cohort study [published correction appears in *Lancet Neurol*. 2012;11(8):658]. *Lancet Neurol*. 2012;11(8):669-678.

Haybittle JL. Repeated assessment of results in clinical trials of cancer treatment. *Brit J Radiol*. 1971;44(526):793-797.

Herholz K, Salmon E, Perani D, Baron JC, Holfhoff V, Frolich L, Schonknecht P, Ito K, Mielke R, Kalbe E, Zundorf G, Delbeuck X, Pelati O, Anchisis D, Fazio F, Kerrouche N, Desgranges B, Eustache F, Beuthien-Baumann B, Menzel C, Schroder J, Kato F, Arahata Y, Henze M, Heiss WD. Discrimination between Alzheimer dementia and controls by automated analysis of multicenter FDG PET. *Neuroimage*. 2002;17(1):302-316.

Landau SM, Harvey D, Madison CM, Koeppe RA, Reiman EM, Foster NL, Weiner MW, Jaqust WJ; Alzheimer's Disease Neuroimaging Initiative. Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI. *Neurobiol Aging*. 2011;32(7):1207-1218.

Lehmacher W, Wassmer G. Adaptive sample size calculations in group sequential trials. *Biometrics*. 1999;55(4):1286-1290.

Lykestos CG, Sheppard JE, Steinberg M, Tschanz JT, Norton MC, Steffens DC, Breitner JCS. Neuropsychiatric disturbance in Alzheimer's disease clusters into three groups: the Cache County study. *Int J Geriatr Psychiatry*. 2001b.16(1):1043-1053.

Mohs RC, Doody RS, Morris JC, Ieni JR, Rogers SL, Perdomo CA, Pratt RD; "312" Study Group. A 1-year placebo-controlled preservation of function survival study of donepezil in AD patients. *Neurology*. 2001;57(3):481-488.

Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, Mantel N, McPherson K, Peto J, Smith PG. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Brit J Cancer*. 1976;34(6):585-612.

Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley and Sons, 1987.

Schafer JL. Analysis of Incomplete Multivariate Data. London: Chapman & Hall; 1997.

Tun SM, Murman DL, Long HL, Colenda CC, von Eye A. Predictive validity of neuropsychiatric subgroups on nursing home placement and survival in patients with Alzheimer Disease. *Am J Geriatr Psychiatry*. 2007;15(4):314-327.

8. Appendices

Appendix 1. Complete List of AV-45 Parameters

SUVr will be obtained for the regions listed below normalized to whole cerebellum and to patient-specific white matter:

- composite summary
- anterior cingulate
- frontal medial orbital
- parietal
- posterior cingulate
- precuneus
- temporal

The composite summary measure is an unweighted average of the 6 smaller regions listed.

Appendix 2. Complete List of FDG Pet Parameters

SUVr will be obtained for the regions listed below normalized to the pons, whole cortex and group of voxels with AD-preserved activity (Herholz et al. 2002).

composite summary	lateral temporal cortex left	pons
caudate left	lateral temporal cortex right	putamen right
caudate right	mean cerebellum gray matter	putamen left
cerebellar cortex left	mean whole cerebellum	rectus left
cerebellar cortex right	mesial temporal cortex left	rectus right
cerebellar white matter	mesial temporal cortex right	subcortical white matter
cingulum anterior left	lateral occipital cortex left	temporal cortex left[a]
cingulum anterior right	lateral occipital cortex right	temporal cortex right[a]
cingulum posterior left[a]	orbitofrontal cortex left	thalamus left
cingulum posterior right[a]	orbitofrontal cortex right	thalamus right
lateral frontal cortex left	lateral parietal cortex left	angular left[a]
lateral frontal cortex right	lateral parietal cortex right	angular right[a]
		whole cortex
		region with AD-preserved uptake
		precuneus left
		precuneus right

[a] Regions used in calculation of the composite summary SUVr.

Appendix 3. Potentially Clinically Significant Laboratory Values

Parameter	SI Unit	Low PCS Criteria	High PCS Criteria
Hematology (whole blood)			
Hemoglobin (male)	mml/L-Fe	<6.8266	>11.1708
Hemoglobin (female)	mml/L-Fe	<6.206	>10.5502
Hematocrit	Proportion of 1.0	<0.3	>0.50 (F); >0.55 (M)
Leukocyte (WBC Count)	10^9/L	≤2.8	≥15
Neutrophils	10^9/L	≤1.5	NA
Platelet Count	10^9/L	≤75	≥700
Chemistry (serum or plasma)			
ALT (SGPT)	U/L	NA	≥3 X ULN
AST (SGOT)	U/L	NA	≥3 X ULN
Total Bilirubin	umol/L	NA	≥1.5 ULN
BUN	mmol/L	NA	≥1.2 ULN
Creatinine Kinase (CK)	U/L	NA	≥3ULN
Sodium	mmol/L	≤125	≥155
Potassium	mmol/L	≤3.0	≥5.5
Calcium	mmol/L	≤0.7 ULN	≥1.2 ULN
Alkaline Phosphatase	U/L	NA	≥3ULN
Albumin	g/L	≤26	≥60
Chloride	mmol/L	≤85	≥120
Glucose (random)	mmol/L	≤0.3 ULN	≥1.5 ULN
Serum Creatinine	umol/L	NA	>1.5 ULN
TSH	mIU/L	below normal range	above normal range
Urinalysis			
Hb/RBCs/Blood		NA	≥ + 2
Protein/Albumin		NA	≥ + 2
Glucose		NA	≥ + 2

Abbreviations: ALT/SGPT = alanine aminotransferase/serum glutamic pyruvic; AST/SGOT = aspartate aminotransferase/serum glutamic oxaloacetic transaminase; BUN = blood urea nitrogen; Hb = heart beat, PCS = potentially clinically significant; RBC = red blood cells; TSH = thyroid stimulating hormone; ULN = upper limit of normal; WBC = white blood cells.

Appendix 4. Potentially Clinically Significant Vital Signs and Weight

Potentially Clinically Significant 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 decrease in systolic blood pressure (supine to standing) (ie, supine minus standing ≥20)	
Orthostatic Diastolic Blood Pressure (mmHg)	≥10 decrease in diastolic blood pressure (supine to standing) (ie, supine minus standing ≥10)	
Orthostatic Pulse (bpm)	≤ -30 decrease (supine to standing) (ie, supine minus standing ≤ -30)	
Temperature	Absolute value ≥38.3°C and ≥1.1°C increase from baseline (Absolute value ≥101°F and ≥2°F increase from baseline)	

Abbreviations: bpm=beats per minute, NA=not applicable.

Appendix 5. Potentially Clinically Significant ECGs

Potentially Clinically Significant ECGs

Parameter	Unit	Low PCS Criteria	High PCS Criteria
QRS Interval	msec	NA	≥120
PR Interval	msec	<100	≥220
Heart Rate	bpm	<45	≥120
QTcF	msec	<320	>500
QTcF interval: change from baseline	>60 msec at any time after randomization		

Abbreviation: PCS = potentially clinically significant.