Title: A Double Blind, Randomized, Placebo Controlled, Parallel Group Study to Simultaneously Qualify a Biomarker Algorithm for Prognosis of Risk of Developing Mild Cognitive Impairment due to Alzheimer's Disease (MCI due to AD) and to Test the Safety and Efficacy of Pioglitazone (AD-4833 SR 0.8 mg QD) to Delay the Onset of MCI due to AD in Cognitively Normal Subjects

NCT Number: NCT01931566

SAP Approve Date: 30 November 2017

Certain information within this Statistical Analysis Plan has been redacted (ie, specific content is masked irreversibly from view with a black/blue bar) to protect either personally identifiable (PPD) information or company confidential information (CCI).

This may include, but is not limited to, redaction of the following:

- Named persons or organizations associated with the study.
- Proprietary information, such as scales or coding systems, which are considered confidential information under prior agreements with license holder.
- Other information as needed to protect confidentiality of Takeda or partners, personal information, or to otherwise protect the integrity of the clinical study.

TAKEDA DEVELOPMENT CENTER

STATISTICAL ANALYSIS PLAN

STUDY NUMBER: AD-4833/TOMM40_301

A Double Blind, Randomized, Placebo Controlled, Parallel Group Study to Simultaneously Qualify a Biomarker Algorithm for Prognosis of Risk of Developing Mild Cognitive Impairment due to Alzheimer's Disease (MCI due to AD) and to Test the Safety and Efficacy of Pioglitazone (AD-4833 SR 0.8 mg QD) to Delay the Onset of MCI due to AD in Cognitively Normal Subjects

PHASE 3

Version**:** Amendment 1

Date**:** 30 November 2017

Prepared by:

PPD

1.1 APPROVAL SIGNATURES

Study Title:

A Double Blind, Randomized, Placebo Controlled, Parallel Group Study to Simultaneously Qualify a Biomarker Algorithm for Prognosis of Risk of Developing Mild Cognitive Impairment due to Alzheimer's Disease (MCI due to AD) and to Test the Safety and Efficacy of Pioglitazone (AD-4833 SR 0.8 mg QD) to Delay the Onset of MCI due to AD in Cognitively Normal Subjects

TDC Approvals

PPD

Date

Date

Date

Date

2.0 TABLE OF CONTENTS

LIST OF IN-TEXT TABLES

LIST OF IN-TEXT FIGURES

LIST OF APPENDICES

3.0 LIST OF ABBREVIATIONS

AA	Alzheimer's Association
AD	Alzheimer's Disease
ADAS-cog	Alzheimer's Disease Assessment Scale-Cognitive Subscale
ADCS ADL-PI	Alzheimer's Disease Cooperative Study Activities of Daily Living - Prevention Instrument
ADCS-CGIC-MCI	Alzheimer's Disease Cooperative Study – Clinical Global Impression of Change – Mild Cognitive Impairment
ADRC	Alzheimer's Disease Research Center
AE	adverse event
AESI	adverse event of special interest
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
ANOVA	analysis of variance
APOE	apolipoprotein E
ARIA	amyloid –related imaging abnormalities
AST	aspartate aminotransferase
BMI	body mass index
BPM	beats per minute
BUN	blood urea nitrogen
BVMT-R	Brief Visuospatial Memory Test - Revised
CDR	Clinical Dementia Rating
CDR-SB	Clinical Dementia Rating – Sum of Boxes
CDT	Clock Drawing Test
CGIC	Clinical Global Impression of Change
CI	confidence interval
CIAC	Cognitive Impairment Adjudication Committee
CMFV	Comprehensive Medical Follow-up Visit
CPH	Cox proportional hazards
CSR	Clinical Study Report
C-SSRS	Columbia–Suicide Severity Rating Scale
CVLT-II	California Verbal Learning Test – 2nd Edition
DNA	deoxyribonucleic acid
DSMB	Data Safety Monitoring Board
DSM-IV-TR	Diagnostic & Statistical Manual of Mental Disorders, 4th Edition - Text Revision
ECG	Electrocardiogram
(e)CRF	case report form (electronic)
EoS	end of study
EQ-5D	European Quality of life (includes single item measures of: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression)
EW	early withdrawal
FAS	full analysis set

GDS	Geriatric Depression Scale
GGT	γ-glutamyl transferase
HbA1c	glycosylated hemoglobin
ICH	International Conference on Harmonisation
IMP	investigational medicinal product
IQCODE	Informant Questionnaire on Cognitive Decline in the Elderly
IWRS	Interactive Web Response System
L	long
LDH	lactate dehydrogenase
LLN	lower limit of normal
MAR	missing at random
MCI	mild cognitive impairment
MedDRA	Medical Dictionary for Regulatory Activities
MINT	Multilingual Naming Test
mm Hg	millimeters of mercury
MMRM	mixed model for repeated measures
MMSE	Mini-Mental State Examination
NIA	National Institute on Aging
NPI-Q	Neuropsychiatric Inventory Questionnaire
NPS	neuropsychiatric symptoms
NPV	negative predictive value
PD	pharmacodynamics
PGx	pharmacogenomics
PK	pharmacokinetics
PPAR-γ	peroxisome proliferator-activated receptor-gamma
PPS	per-protocol analysis set
PPV	positive predictive value
PRO	patient-reported outcome
QD	once daily
RU	Resource Utilization
S	short
SAE	serious adverse event
SAP	statistical analysis plan
SDTM	Study Data Tabulation Model
SF-36	Short Form-36
SI	Systeme International
SOC	system organ class
TEAE	treatment-emergent adverse event
TOMM40	Translocase of the Outer Mitochondrial Membrane 40 homolog
ULN	upper limit of normal
VL	very long

vMRI	volumetric magnetic resonance imaging
WHO	World Health Organization

4.0 SUMMARY OF CHANGES

This section describes the changes in reference to the SAP Incorporating Amendment No. 1.

The primary purpose of this amendment is to revise the definition of the event time of a Cognitive Impairment Adjudication Committee (CIAC)CIAC confirmed event of MCI due to AD and provide additional analyses for the efficacy futility analysis. The following is a summary of the changes made in the amendment. Detailed description of amendments to the text is presented in Appendix E.

1. The definition of the event time of a CIAC-adjudicated event of MCI due to AD has been revised as the time from a subject's randomization date to the date of the first **scheduled visit** of the two consecutive **scheduled visits** at which that subject was assessed with a clinical diagnosis and determined as an primary endpoint event by CIAC using the data from scheduled visit and the triggered unscheduled comprehensive medical follow up visits.

   Justification: This change will reflect the thinking that, in any subject dossier reviewed by the CIAC, the first scheduled visit is the first of the 4 visits considered by CIAC in their adjudication of primary endpoint event, and it represents the first time point of these 4 visits where clinical impairment or cognitive decline is first noted. This revised definition better reflects the concept of "**time to onset**" of the MCI due to AD – "onset" meaning start of symptoms that are confirmed as MCI due to AD (or AD dementia) by the CIAC.

2. The data collected at the EOS/EW visit is allowed to be used to determine the primary efficacy events of MCI due to AD.

   Justification: the data collected at the EOS/EW is similar to those planned at the unscheduled CMFV.

3. The memory domain score derivation algorithm is revised so that at least one of the CVLT-II episodic memory tests (CVLT-II Short-delay free recall correct or CVLT-II Long-delay free recall correct) is required to be non-missing. The requirement of BMVT-R being non-missing is removed.

   Justification: The BVMT-R can be missing for many reasons other than cognition (arthritis, motor issues, examiner error). The requirement of both BVMT and at least one CVLT-II to be non-missing may cause valid data points to be omitted (such as subjects who complete CVLT only). The composite measure becomes uninformative without the memory domain; however, the CVLT alone can be used to calculate a memory domain.

4. Additional summaries and analyses are added to evaluate the biomarker qualification performance. Key secondary endpoints and additional variables (only at the domain and composite score levels) will be summarized and analyzed using the data from placebo subjects who are classified into low and high risk groups.

   Justification: These analyses will be helpful for the evaluation of the biomarker qualification performance and will be presented at the end of study.

5. At the time of the efficacy futility analysis, if the conditional power is <30%, the following subgroup analyses will be performed on the primary efficacy endpoint: gender (male vs. female), APOE risk-status (low vs. high) and baseline MMSE (<=27 vs. >=28). A particular subgroup analysis (e.g. female) will be performed only when there are at least 20% of the total primary events in the subgroup.

   Justification: This addition is expected to be informative in the efficacy futility assessment when the conditional power based on the primary event is <30%.

6. At the time of the efficacy futility analysis, if the conditional power is <30%, the following subgroup summaries and MMRM analyses will be performed on the change from baseline in the composite score, CVLT-II Long and Short Delay Free Recall Correct, BVMT-R Delayed Recall and MMSE total score: gender (male vs. female), APOE risk-status (low vs. high) and baseline MMSE (<=27 vs. >=28). The subgroup analysis will be performed only when there are at least 400 subjects in a subgroup.

   Justification: This addition is expected to be informative in the efficacy futility assessment when the conditional power based on the primary event is <30%.

7. At the time of the efficacy futility analysis, if the conditional power is <30%, and the treatment effect on the change from baseline by visit in the composite score is statistically significant (p-value<0.05), a joint model of the time-to-event data and the longitudinal composite score data will be fitted to estimate the association between the composite score and hazard rate of MCI due to AD over time during the treatment period. Details of this model are presented in Section 9.16.

   Justification: This addition is expected to be informative in the efficacy futility assessment when the conditional power based on the primary event is <30%.

8. At the time of the efficacy futility analysis, if the conditional power is between 10% (>10%) and 30% (<30%), the "progression" of the conditional power in terms of an increasing number of the events vs. total needed will be provided for the following information fractions: 10% (20 events), 20% (41 events) and 34% (69 events). The events will be selected based on the time of the occurrence of the event relative to the first dose.

   Justification: When the evidence against the null hypothesis is accumulating, the conditional power should increase as the information fraction increases. However, if the CP does not change as the information fraction increases, it may be unlikely evidence against the null hypothesis is being accumulated. Essentially if CP remains flat across 10%, 20%, 30% of events, it makes it less likely we would reject the null hypothesis at end of study.

5.0 INTRODUCTION

This document describes the statistical analyses to be performed and data presentations to be produced for this phase 3, randomized, double-blind, parallel-group, placebo-controlled study to simultaneously qualify a biomarker algorithm for prognosis of risk of developing mild cognitive impairment due to Alzheimer's Disease (MCI due to AD) and to test the safety and efficacy of pioglitazone (AD-4833 SR 0.8 mg QD) to delay the onset of MCI due to AD in cognitively normal subjects.

The purpose of this statistical analysis plan (SAP) is to ensure the credibility of the study findings by specifying the statistical approaches to the analysis of the double-blind data prior to database lock. This SAP was developed based on the International Conference on Harmonisation (ICH) E3 and E9 Guidelines and in reference to the following documents:

- Protocol AD-4833/TOMM40_301 Amendment #1 dated 09 April 2013.
- Protocol AD-4833/TOMM40_301 Amendment #2 dated 19 March 2014.
- Protocol AD-4833/TOMM40_301 Amendment #3 dated 10 November 2014.
- Protocol AD-4833/TOMM40_301 Amendment #4 (Local-Russia) dated 15 April 2015.
- Protocol AD-4833/TOMM40_301 Amendment #5 dated 24 July 2015.

Any deviations during the analysis and reporting process from the current statistical analysis plan will be described and justified in the final report. Analysis issues that suggest changes to the principal features stated in the protocol will be documented in a protocol amendment. Otherwise, the statistical analysis plan will be updated through an amendment with the changes in the analyses documented in the amendment.

6.0 OBJECTIVES AND STUDY DESIGN

6.1 Primary Objectives

For biomarker risk algorithm qualification:

- To qualify the biomarker risk algorithm composed of TOMM40 rs10524523 genotype, APOE genotype, and age for prognosis of the risk of developing MCI due to AD within 5 years.

For efficacy evaluation of pioglitazone:

- To evaluate the efficacy of pioglitazone compared with placebo to delay the onset of MCI due to AD in cognitively-normal subjects who are at high-risk, as identified by the biomarker risk algorithm at enrollment, for developing MCI due to AD within 5 years.

6.2 Key Secondary Objectives

- To evaluate the effect of pioglitazone compared with placebo on the progression of cognitive decline.
- To evaluate the effect of pioglitazone compared with placebo on functional decline and instrumental activities of daily living.

6.3 Additional Objectives

6.3.1 Safety Objective

- To evaluate long-term safety and tolerability of pioglitazone compared with placebo during the course of the treatment.

6.3.2 Exploratory Objectives

- CCI
- [redacted]

- CCI
-

6.4 Study Design

This phase 3 study is a multicenter, randomized, double-blind, placebo-controlled, parallel-group study designed to pursue two primary objectives independently yet simultaneously: (1) to qualify a biomarker risk algorithm composed of TOMM40 rs10524523 genotype, APOE genotype, and age for prognosis of the risk of developing MCI due to AD and (2) to evaluate the efficacy of pioglitazone (AD-4833 SR 0.8 mg once daily [QD]) to delay the onset of MCI due to AD in cognitively normal subjects who are prognosed to be at high risk of developing MCI due to AD within 5 years.

The study duration will be the time needed to accumulate a total of 202 conversions in non-Hispanic/Latino Caucasian subjects within the high-risk stratum from normal cognition to a diagnosis of MCI due to AD, currently anticipated to be a minimum of 4 years. For the purposes of this study, a Caucasian subject will be defined as a person having origins in any of the original peoples of Europe, the Middle East, or North Africa. Hispanic/Latino will be defined as a person of Cuban, Mexican, Puerto Rican, South or Central American origin, regardless of race. The study will primarily recruit from large community-based populations.

The subjects must be cognitively normal as assessed first by performance on the MMSE at Screening, and then on the CDR scale and the Cognitive Test Battery (Sections 9.1.15.1 and 9.1.15.2 of the protocol) at Baseline (Visit 2). The subject must be between the ages of 65 and 83 years, inclusive, at time of the Screening Visit. The subjects must have a project partner able to complete an Acknowledgement Form on his or her own behalf and take part in the study to provide information on the cognitive, functional, and behavioral status of the subject and to assist with monitoring of study medication, if needed, for as long as the subject is in the study.

Based on statistical requirement, a minimum of 2,793 subjects will be stratified into high- and low-risk (for the development of MCI due to AD over the next 5 years) groups based upon application of the biomarker risk algorithm. Subjects will be randomized via interactive web response system (IWRS) into this study as follows:

- Non-Hispanic/Latino Caucasian subjects. A minimum of 2,660 subjects further divided into:
  - Approximately 2,346 subjects in the high-risk group (pioglitazone, n=1,173; placebo, n=1,173).
  - Approximately 314 subjects in the low-risk group: All subjects stratified to the low-risk stratum will also be randomized to maintain the stratification blind, but all subjects in this stratum will receive placebo.

- Non-Caucasian subjects and Hispanic/Latino Caucasian subjects. Subjects will be allocated to risk strata using the same algorithm as for the non-Hispanic/Latino Caucasian subjects.

Since all subjects screened at the time of Amendment 5 will be allowed to randomize if they continue to meet all eligibility criteria, it is estimated that approximately 3,500 will be randomized with the following groups:

- Non-Hispanic/Latino Caucasian subjects. Estimated 3,330 subjects further divided into:
  - Approximately 2,940 subjects in the high-risk group (pioglitazone, n=1,470; placebo, n=1,470).
  - Approximately 394 subjects in the low-risk group: All subjects stratified to the low-risk stratum will also be randomized to maintain the stratification blind, but all subjects in this stratum will receive placebo.
- Non-Caucasian subjects and Hispanic/Latino Caucasian subjects. Subjects will be allocated to risk strata using the same algorithm as for the non-Hispanic/Latino Caucasian subjects (n=167).

Because the data used to develop the biomarker risk algorithm was collected in non-Hispanic/Latino Caucasian subjects only, biomarker qualification will be based upon non-Hispanic/Latino Caucasian subjects in the placebo arm of the high-risk stratum compared to data from non-Hispanic/Latino Caucasian subjects in the low-risk stratum. Likewise, the primary analyses to evaluate the effects of pioglitazone versus placebo will be within the non-Hispanic/Latino Caucasian high-risk sub-group. CCI [redacted].

An interactive response system will be used for central randomization. Stratification status and treatment assignment of all study participants will be double-blind. The randomization will be stratified based on participation in the vMRI substudy (approximately 300 subjects).

For subjects who are not participating in the vMRI substudy, subjects in the high-risk stratum will be randomized to pioglitazone and placebo in a 1:1 ratio. Subjects in the low-risk stratum will be randomized only to placebo.

For subjects who are participating in the vMRI substudy, subjects in the high-risk stratum will be randomized to pioglitazone and placebo in a 5:4 ratio. Subjects in the low-risk stratum will be randomized only to placebo.

It is estimated that at least 60% of subjects who present for screening will be low-risk as determined by the biomarker risk algorithm. Because the study requires many more high-risk subjects than low-risk subjects to be randomized into the study, a random rejection algorithm for low-risk subjects was implemented in the IWRS. The low-risk rejection algorithm is designed to lengthen the enrollment period for low-risk subjects in such a way that the ratio of high-risk to low-risk subjects randomized into the study is relatively stable over the entire enrollment period.

This will facilitate maintaining double-blinding of risk assignment for all subjects throughout the entire course of the study. Details of the low-risk random rejection algorithm are provided in the AD-4833/TOMM40_301 Study System Requirements document which was authored by the IWRS vendor, Bracket.

Subjects will be instructed to take the first dose of study medication in the morning on the day following the randomization visit.

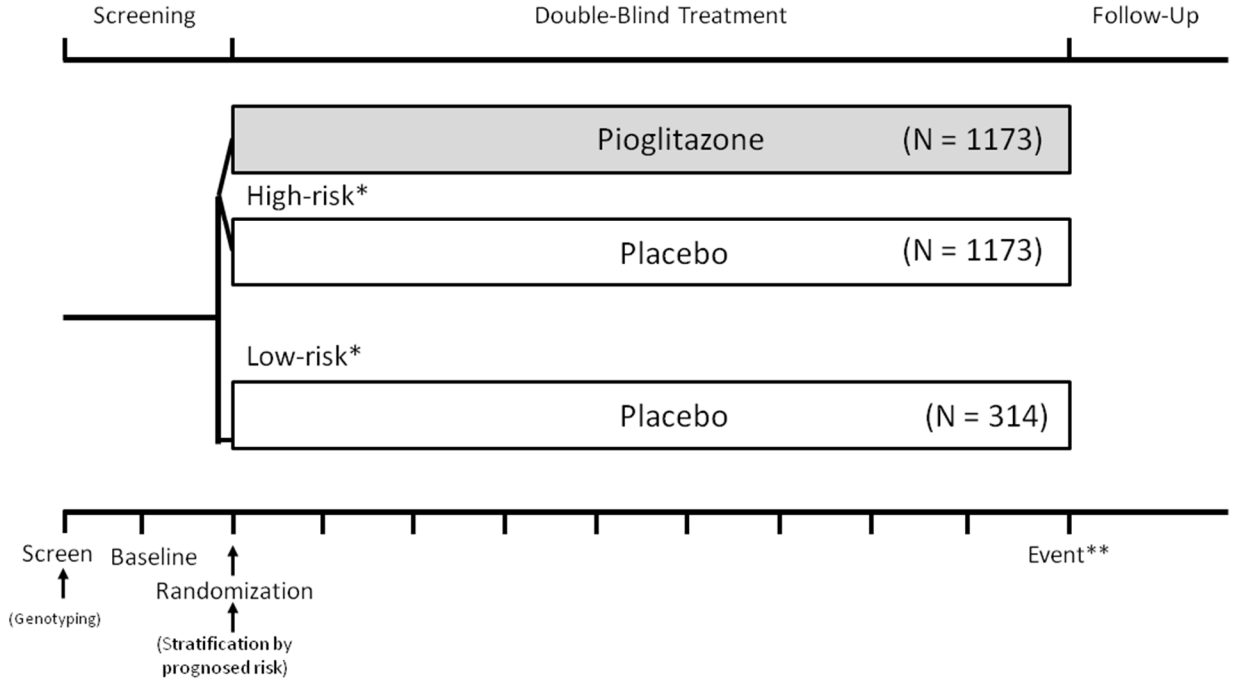
A project partner will be required to participate with the subject in the study visits, including eligibility at Baseline (Visit 2) and follow-up. The project partner is a spouse, adult child, or other person familiar with the participant's health and daily functioning for a minimum of two years prior to the Baseline Visit. Subjects and their project partners are expected to attend study visits every 6 months after randomization, for assessments of safety, efficacy, and treatment compliance. In addition, the subject will be contacted for telephone-based safety checks between visits. In-person participation of the project partner is mandatory at Baseline (Visit 2), comprehensive medical follow-up, and End-of-Study Visits; while in-person participation is also strongly encouraged at all other study visits, telephone assessments for project partners will be acceptable in cases when in-person participation is not possible. If a comprehensive medical follow-up evaluation is needed (see Section 9.1.15 of the protocol) or if a safety issue is suspected, further evaluation of the subject may be required, which may include an unscheduled clinic visit. The project partner should attend unscheduled visits, whenever possible. If the original project partner is not able to continue participation in the study for any reason, he/she may be replaced by a different individual who also meets the criteria described above. The new project partner must sign the acknowledgement form prior to any of the project partner protocol directed procedures being performed.

Two sets of blood samples for deoxyribonucleic acid (DNA) will be collected during the course of this study:

1. CCI
2. [redacted]

A schematic of the study design is included as Figure 6.a. A schedule of assessments is listed in the protocol, Appendix A.

Figure 6.a Schematic of Study Design



* Non-Hispanic/Latino Caucasian, Hispanic/Latino Caucasian, and non-Caucasian subjects alike will be stratified into high- and low-risk strata using the same biomarker risk algorithm; however, target sample size for primary endpoint analyses refers to non-Hispanic/Latino Caucasian subjects only.
** As this is an event-driven study, double-blind treatment duration will depend on the time it takes for a total of 202 conversions from cognitively normal status to MCI due to AD dementia to occur in the high-risk non-Hispanic/Latino Caucasian population; conversions from cognitively normal status to AD will also be included in this total. Study visits to the testing center will take place every 6 months after the Randomization Visit.

7.0 ANALYSIS ENDPOINTS

7.1 Primary Endpoints

For biomarker risk algorithm qualification:

Time to diagnosis of MCI due to AD for placebo-treated, high-risk, non-Hispanic/Latino Caucasian subjects versus placebo-treated, low-risk, non-Hispanic/Latino Caucasian subjects.

For efficacy evaluation of pioglitazone:

Time to diagnosis of MCI due to AD for pioglitazone-treated, non-Hispanic/Latino Caucasian subjects versus placebo-treated non-Hispanic/Latino, Caucasian subjects in the high-risk stratum.

7.2 Key Secondary Endpoints

- Change from baseline to Year 4 for cognitive decline on composite score on the cognitive battery for pioglitazone-treated subjects versus placebo-treated subjects in the high-risk stratum.
- Changes from baseline to Year 4 in instrumental activities of daily living (Alzheimer's Disease Cooperative Study Activities of Daily Living - Prevention Instrument [ADCS ADL-PI]) between pioglitazone-treated and placebo-treated groups of the high-risk stratum.

7.2.1 Additional Endpoints

7.2.1.1 Safety Assessments

- Safety and tolerability: adverse events (including adverse events of special interest), vital signs, body weight, electrocardiogram (ECG), clinical laboratory data, and physical exam findings.

7.2.1.2 Additional Endpoints

- Change from baseline to all scheduled visits where the following is assessed (6-month visits and End-of-Study (EoS) / Early Withdrawal (EW)): cognitive decline on composite score on the cognitive battery for pioglitazone-treated subjects versus placebo-treated subjects in the high-risk stratum.

Change from baseline to all scheduled visits where the following is assessed (6-month visits and End-of-Study (EoS) / Early Withdrawal (EW)): cognitive decline on tests and domain scores within the cognitive battery (see Table 9.b) for pioglitazone-treated subjects versus placebo-treated subjects in the high-risk stratum.

- Changes from baseline to all scheduled visits where the following is assessed (6-month visits and EoS/EW): instrumental activities of daily living (Alzheimer's Disease Cooperative Study Activities of Daily Living - Prevention Instrument [ADCS ADL-PI]) between pioglitazone-treated and placebo-treated groups of the high-risk stratum.

- CCI
- [redacted]

- CCI [redacted]

Although change-from-baseline endpoints will be assessed at all visits, the primary time frame for purposes of efficacy analysis is the change from baseline to Month 48 (Year 4).

8.0 DETERMINATION OF SAMPLE SIZE

Background Information for Determining Sample Sizes

Using 73 years as the age at entry (+5 years = age 78), a conversion rate of around 18% over 5 years can be expected (see Table 8.a). This is based on:

- The TOMM40 rs10524523 allele frequencies from the Duke Alzheimer's Disease Research Center (ADRC), Cache County, and Wisconsin cohorts.
- The assessment of peer-reviewed published incident MCI in normal individuals.
- The effect of APOE on the incidence of AD.

Table 8.a Conversion Rate Estimates Based on Analysis of 3 Cohorts, Cognitive Impairment Incidence, and Observed TOMM40 rs10524523 Allele Frequencies

TOMM40 rs10524523 Genotype	Duke Bryan ADRC Frequency	Wisconsin Frequency	Cache Frequency	Cohort Total Count	Age 78 (Proportion Converted)	
					Proportion	Number of Converted Subjects
L/L	0.05	0.04	0.02	189	0.53	100
VL/L	0.14	0.15	0.13	543	0.34	185
S/L	0.17	0.20	0.14	669	0.22	147
S/S	0.23	0.16	0.17	929	0.14	130
S/VL	0.31	0.31	0.36	1236	0.10	124
VL/VL	0.11	0.14	0.18	433	0.05	22
			Total count = 3999		Total converted	708
					Overall conversion rate	0.177

Taking into account the age ranges that will be enrolled in the phase 3 study, a 5-year conversion rate of 15% will be assumed for non-Hispanic/Latino Caucasian placebo subjects within the high-risk stratum identified by the enrichment process. Based on the expected 5-year conversion rate of 15%, the time-to-event was modeled using an exponential mathematical model for planning the study sample size.

Similarly, a 5-year conversion rate of 4% will be assumed for non-Hispanic/Latino Caucasian placebo subjects within the low-risk stratum. Thus, it is important to recognize that this summary of the data demonstrates nearly a four-fold enrichment of events over the population rate using the risk algorithm.

Determination of Sample Size for the Phase 3 Study Based on the High-Risk Stratum

The sample size for this study is based on the event-driven analysis that is planned for comparing treatment groups within the high-risk stratum. For the purposes of this study, it is assumed that the placebo-treated non-Hispanic/Latino Caucasian subjects in the high-risk stratum will have an event rate of 15% over a 5-year period. This is more conservative than the 18% cited in the table, to assure adequate sample size. With a clinically meaningful improvement of a reduction in event rate of at least 40%, the target event rate for the active treatment group will be 9% over 5 years. Furthermore, the following specifications are used: alpha=0.01, a 2-sided statistical test based on survival analysis, power=90%, annual dropout rate of 12% (or cumulative dropout rate of 47% over 5 years). In order to detect a difference of the specified magnitude (≥40% reduction in conversion rates over 5 years), at least 202 events must be observed during the course of the study in non-Hispanic/Latino Caucasian subjects (primary efficacy analysis). Based on these assumptions and the fact that only non-Hispanic/Latino Caucasian subjects will be used for the primary analyses, a total of 2,346 non-Hispanic/Latino Caucasian subjects (1,173 on pioglitazone/1,173 on placebo) should be enrolled from the high-risk stratum for randomization.

Sample Size for the Phase 3 Study Based On Qualifying the Test Used to Stratify Subjects

While using data from the results of risk stratification and not a comparison between randomized groups, the same analysis as specified for comparing the treatment groups within the high-risk stratum will also be used to compare the two placebo groups (high-risk placebo and low-risk placebo). Thus, assuming that the placebo-treated subjects in the low-risk stratum will have a conversion rate of 4% over a 5-year follow-up period, then approximately 314 non-Hispanic/ Latino Caucasian subjects identified as low-risk for the event should be enrolled in that stratum. This number of subjects enrolled in the low-risk stratum will provide very high power (99% with alpha=0.01) to validate the enrichment algorithm with the assumed 5-year conversion rates (ie, 15% vs. 4%).

Sample Size for the vMRI Substudy

This substudy is not designed or powered to assess any quantitative changes in the brain. Based on clinical judgment, a sample size of 300 subjects is believed to be sufficient to evaluate possible qualitative changes in the brain via vMRI.

9.0 METHODS OF ANALYSIS AND PRESENTATION

9.1 General Considerations

9.1.1 Statistical Software

All statistical analyses will be performed using SAS® Version 9.1.3 or higher.

9.1.2 Summary Statistics and Precision

All tabulations of analysis results will include summaries for all three treatment groups: high-risk placebo, high-risk pioglitazone, and low-risk placebo.

Except when noted as otherwise, all confidence intervals (CIs), statistical tests, and resulting p-values will be reported as nominal 2-sided and will be assessed at the 5% significance level. No adjustments will be made for multiplicity except for primary efficacy and key secondary endpoints.

For continuous variables, descriptive statistics will include the number of subjects (n), mean, standard deviation (SD) or standard error (SE) as appropriate, minimum, median, and maximum. The number of decimal places displayed for each statistic will be determined as follows:

- Mean and median: 1 more than the number of reported decimal places. (For electronic case report form (eCRF) data, the number of reported decimal places will be will be equal to the number of decimal places allotted in the eCRF.)
- SD and SE: 2 more than the number of reported decimal places allotted in the eCRF.
- Minimum and maximum: equal to the number of reported decimal places allotted in the eCRF.
- Confidence intervals will be presented using the same number of reported decimal places as the parameters (eg, mean).

For categorical data, frequency counts and percentages will be presented. Percentages will be reported to 1 decimal place.

The data summaries will be accompanied by individual subject data listings sorted by treatment (including low-risk placebo), study center and subject identifier. All data available from eCRFs will be listed. The actual day relative to the start of treatment will be determined and included in the listings.

Derived analysis datasets will be produced from SDTM data and laboratory data. This allows for convenient reviewing of the data as well as any necessary supplemental analyses. All data from the raw datasets will be included in the derived datasets. Derived dataset specifications will be developed to include the names and definitions of derived variables in the derived SAS datasets.

9.1.3 Definition of Study Day and Study Visit Windows

A windowing convention will be used to determine the analysis value for a given study visit that applies to observed data.

Study Day –1 corresponds to the randomization date. Other study days are defined relative to Study Day 1, the date of first double-blind study drug dose. Relative day is calculated as (date of interest – date of first dose +1) for study days on or after the date of first dose and as (date of interest – date of first dose) for study days prior to the first dose date.

For each visit, a window will be defined such that the lower and upper bounds of each window is generally the midpoint between 2 consecutive study visits. The visit windows and applicable study day ranges are presented below in Table 9.a.

Table 9.a Visit Windows

Nominal Visit Month	Nominal Visit Day	vMRI	Height, ECG, Chemistry and Hematology, PK, RU, DNA/RNA, PE	All Other Variables (c)
Baseline	-36 to -1 (a)	≤ -1	≤ -1 (b)	≤ -1
6	183			2—273
12	365		2—547	274—456
18	548			457—638
24	731	639—821	548—912	639—821
30	913			822—1003
36	1096		913—1277	1004—1186
42	1278			1187—1369
48	1461		1278—1643	1370—1551
54	1644			1552—1734
60	1826		1644—1917	1735—1917
66/Final (d)	2009	≥1370	≥1918	≥1918

(a) Consistent with SDTM dataset standards, date of first dose of double-blind study drug is defined as Day 1, and the day before is defined as Day –1. Per the protocol, date of first dose of dose of double-blind study drug is scheduled for the day after randomization. Exemptions were granted on a case-by-case basis for subjects who had a screening-to-randomization time frame greater than the 35-day window planned per the protocol.
(b) No baseline assessment for PK samples.
(c) CDR and NPI-Q are collected only at Baseline (Visit 2), unscheduled visits (eg, comprehensive medical follow-up visits) and End of Study / Early Withdrawal.
(d) The planned study duration of 5.5 years is approximate, since this is an event-driven trial.

For safety variables, visit windows are exhaustive and the lower and upper bounds of each window are the midpoints between the scheduled visit days for the current visit and the adjacent visits.

More than 1 result for a parameter may be obtained in a visit window. In such an event, the result with the date closest to the scheduled visit day will be used. In the event of 2 observations equidistant to the scheduled visit day, the later of the observations will be used.

In general, the baseline value for a variable is defined as the last observation prior to the first dose of double-blind study medication (visit date ≤ first dose date), including the screening value, if necessary.

Adverse events that start more than 30 days after the last dose of double-blind study medication (start date – last dose date >30) will be listed, but excluded from the summaries and analyses. For efficacy and other safety data, data that are obtained more than 7 days after the last dose of double-blind study medication (visit date – last dose date > 7) will be listed, but excluded from summaries and analyses. (Note: This efficacy data window only applies to determining which data are included/excluded from the summary statistics and formal statistical analysis tables for a given efficacy endpoint. All events of MCI due to AD (or AD dementia) as determined by the adjudication committee will be included in the time-to-event analyses, without regard to the date of last dose of double-blind study medication.)

If the date of first dose of double-blind study drug collected in the eCRF is missing, then for summary purposes the day after the first dispense date-will be used as an estimate for the date of first dose. However, if all dispensed study drug is returned, then the subject is assumed to have not taken any study drug, and the first dose date will not be imputed.

If the date of last double-blind study drug dose collected in the eCRF is missing, then the earliest of the following dates will be used for the last dose date for analysis and summary purposes: date of death, date of last visit (recorded in the eCRF), last double-blind study drug return date, and the last double-blind study drug dispense date (following the last drug return date) + 200 days (ie, the double-blind study drug dispensing interval corresponding to the number of study drug tablets dispensed).

The study window convention will not be applied to the eCRF data listings. The data listings for eCRF data will display the raw data as collected and entered in the eCRF. For efficacy listings, windowed visits will also be shown on the listings.

9.1.4 Grouping of Centers

Pooling of small sites will be done (if necessary) to form centers for purposes of analysis.

Before unblinding the study, sites that are considered small (<40 subjects) will be pooled with geographically similar sites to minimize artifacts in the statistical analyses from imbalances in subject counts within the sites.

Starting with the smallest center, center(s) with fewer than 40 subjects randomized will be pooled into the nearest center. The pooling will start first within each country; if it is still small, other countries within the same geographical region (for example, US, Europe, Rest of World) will also be considered. The pooled centers within a region which still have fewer than 40 subjects will be kept as they are and will not be pooled with centers in other regions.

The pooling of the centers will be reviewed and approved by the clinical and statistical team for agreement, and will then be used in the appropriate analyses once the database has been locked and unblinded.

Quality issues were detected at site 5055; therefore, data collected exclusively at that site will be excluded for purposes of efficacy analyses.

Because of the smaller number of events available at the time of the efficacy futility analysis, pooling will be done for the efficacy futility analysis using a smaller number of pooled centers, based on country/sovereign state, as shown below.

- US
- UK
- Australia
- Switzerland/Germany

9.2 Major Protocol Violations

All subjects with major protocol violations will be identified prior to unblinding, and will be listed by study center and subject number.

Subjects with the following major protocol violations will be excluded from the per-protocol set (PPS) defined in Section 9.3:

- Age outside of the range 65 to 83, inclusive, at the screening visit.
- CDR Global Score > 0 at Baseline.
- MMSE $\leq$ 24 at the screening visit (after adjusting for education and age as described in Appendix G of the protocol).
- Subject does not have at least one memory test above -1.5 SD of the demographically corrected normative mean at Baseline.
- At Baseline, the subject has a current diagnosis or history of any type of cognitive impairment or dementia, or has a current diagnosis or history of neurological/psychiatric disorder or any other diagnosis that significantly affects cognitive performance (eg, mental retardation, organic mental disorder).
- At Baseline, the subject has a current diagnosis of significant psychiatric illness, per Diagnostic & Statistical Manual of Mental Disorders, 4th Edition - Text Revision (DSM-IV-TR) (or DSM-V when published) (including but not limited to major depressive disorder, anxiety disorders) and is in an acute phase/episode, or the subject has a current diagnosis or history of schizophrenia or bipolar disorder.
- The subject has a glycosylated hemoglobin (HbA1c) >8.0% at the time of baseline or requires treatment with insulin, triple oral antidiabetic therapy or a PPAR-γ agonist. The subject should be on a stable antidiabetic regimen for at least 3 months prior to enrollment.
- The subject has a history of drug abuse (defined as any illicit drug use) or a history of alcohol abuse/dependence within 2 years prior to the Screening Visit.

- Low study drug compliance (<80%).
- Double-blind study medication exposure less than 6 months ([last dose date – first dose date +1/30.5] <6).
- Subject switched treatment during study.
- Randomized treatment different from actual treatment received.
- Incorrect risk stratum assignment used for randomization.
- Subject unblinded prematurely.

The above list may be revised and any additional major protocol violations (eg, related to prohibited concomitant medications) will be identified by the study team prior to final study unblinding.

9.3 Analysis Sets

The safety analysis set will include all subjects who were randomized and received at least 1 dose of double-blind study medication. In safety summaries, subjects will be analyzed according to the treatment they received. In the event that a subject receives more than one treatment, the actual treatment will be defined as the one which is used most frequently. If the two most common treatments are used with equal frequency, then the randomized treatment will be used as the actual treatment.

The full analysis set (FAS) will include all subjects who were randomized, received at least one dose of study drug, and had at least one valid post-baseline value for assessment of primary efficacy. In FAS efficacy summaries, subjects will be analyzed by the treatment to which they were randomized.

The per-protocol analysis set (PPS) will include all FAS subjects who had no major protocol violations. In PPS efficacy summaries, subjects will be analyzed by the treatment to which they were randomized.

9.4 Disposition of Subjects

A subject disposition summary presented by treatment group and overall will be provided. The categories will include all subjects who were randomized, subjects who were not treated, subjects who discontinued from the study categorized by reason, and subjects who completed the study. Post-randomization discontinuation reasons include pretreatment event or adverse event, major protocol deviations, lost to follow-up, lack of efficacy, voluntary withdrawal, study termination, pregnancy, and other. A listing will be presented to describe study treatment, date of first dose, date of last dose, date of completion or early withdrawal, and the reason for early discontinuation for each subject.

A summary of screening failures and listings of inclusion/exclusion criteria responses for subjects with violations will also be provided.

9.5 Demographic and Baseline Characteristics

In countries that do not allow collection of date of birth, a subject's age will be collected at each post-Baseline Study Visit, Unscheduled Visit (as needed), and at the End of Study / Early Withdrawal Visit.

Demographic and baseline characteristics including date of birth or age, gender, Hispanic/Latino ethnicity, race as identified by the subject, non-Hispanic/Latino Caucasian determination, years of education, primary language, ability to communicate in primary language, bilingualism, years lived in the country/region, occupation, smoking status, drinking habit of the subject at Screening Visit, height, weight, body mass index (BMI), family history of dementia and related conditions, APOE genotype, TOMM40 genotype, risk status as identified by the biomarker risk algorithm, diabetic status, and baseline statin use will be listed and summarized for each treatment group and overall. The demographic and baseline characteristics summary will be based on all randomized subjects.

Baseline values for selected efficacy parameters (cognitive battery composite score, ADCS ADL-PI) will also be presented for each treatment group and overall based on all randomized subjects.

Height and weight values will be presented in metric units (cm and kg respectively). BMI is calculated as [weight (kg)/height (m) ^{2}], using the weight and height collected Visit 2 (Baseline).

Race is classified into White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander. In addition, ethnicity (Hispanic or Latino) is also captured.

For continuous variables, the number of non-missing values and the mean, median, SD, minimum and maximum based on the raw scores uncorrected for demographic modifiers (ie, age, gender, education) will be tabulated by treatment group and overall. For the categorical variables, the count and percentages of each possible value will be tabulated by treatment group and overall.

All individual demographic and baseline data will be listed by treatment, study center and subject number.

9.6 Medical History and Concurrent Medical Conditions

Medical history refers to significant conditions/diseases that stopped at or prior to Screening (time of informed consent). Concurrent medical conditions are those significant ongoing conditions/diseases present at Screening (time of informed consent). Family history of dementia and related conditions will be collected at the first 6 month visit and at the End of Study/Early Withdrawal visit.

Medical history and concurrent medical conditions will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) latest version and will be summarized by treatment group and overall using System Organ Class (SOC) and MedDRA preferred term. The table will include number and percentages of subjects and will be sorted in alphabetical order by system

organ class and sorted by decreasing frequency order by preferred term. A subject will only be counted once within a particular class even if he/she has multiple conditions/symptoms. Summaries will be based on all randomized subjects.

All medical history and concurrent medical condition data will be listed by treatment, study center and subject number. The listing will contain subject identifier, whether there was any significant medical history or concurrent condition, and include system organ class, preferred term and details of the medical history or condition.

9.7 Medication History and Concomitant Medications

The medication history and concomitant medications are defined as follows:

- Medication history refers to the medication that the study subjects stopped taking within 3 months prior to the Screening Visit (ie, stop date < first screening visit date).
- Concomitant medication is defined as medication that the study subjects continued taking or took from Screening through end of study:
  - Concomitant medication that started and stopped prior to baseline (ie, stop date ≥ first screening visit date, and stop date < first dose date).
  - Concomitant medication that started prior to and was ongoing at baseline (ie, start date < first dose date, and stop date > first dose date).
  - Concomitant medication that started after baseline but before or at last dose (ie, start date ≥ first dose date, and start date ≤ last dose date).
  - Concomitant medication taken during the study (ie, start date ≤ last dose date, and stop date ≥ first dose date.
  - If start date and stop date are missing, medication will be assumed to occur both prior and concomitantly.

Medication history and concomitant medications will be coded using the latest version of the World Health Organization (WHO) Drug Dictionary and summarized by giving the number and percentage of subjects by preferred term within each therapeutic class, with therapeutic class and medications in each class sorted in order of decreasing frequency. The number of subjects with medications in each reported therapeutic class will also be presented. If a subject reports taking 2 drugs belonging to the same class, he/she will only be counted once within that class. Summaries of medication history and concomitant medication will be based on all randomized subjects.

All prior and concomitant medications will be listed by treatment, study center and subject number. The listings will contain subject identifier, WHO Drug preferred term and reported term, dose, unit, frequency, route, start date, end date, whether the medication was ongoing, whether the medication as given to treat a pretreatment event/adverse event, and reason for use.

9.8 Study Drug Exposure and Compliance

The summary of study drug exposure and compliance will be based on the safety analysis set.

Duration of exposure to double-blind study medication is defined as (date of last dose – date of first dose +1).

Treatment duration will be summarized by duration category in months and the number of subjects in each duration category by treatment group.

- <6 months (1 to 161 days).
- 6 to <12 months (162 to 344 days).
- 12 to <18 months (345 to 526 days).
- 18 to <24 months (527 to 709 days).
- 24 to <30 months (710 to 892 days).
- 30 to <36 months (893 to 1074 days).
- 36 to <42 months (1075 to 1257 days).
- 42 to <48 months (1258 to 1440 days).
- 48 to <54 months (1441 to 1622 days).
- 54 to <60 months (1623 to 1805 days).
- 60 to <66 months (1806 to 1987 days).
- 66 months ≥1988 days.

Treatment duration (months) will also be summarized using descriptive statistics (n, mean, SD, median, minimum, and maximum).

Percent of study drug compliance is defined as {(number of tablets dispensed – number of tablets returned)/(date of last dose – date of first dose +1)}x100%. If a value for the number of returned tablets is missing or the return date is missing, then 100% compliance will be assigned for each day up to the number of tablets dispensed, up to the date of return, or up to the date of the last visit, whichever is earliest.

For each treatment group, study medication compliance will be summarized by compliance category (<80%, 80 to 120%, and >120%) and the number of subjects in each compliance category. Study medication compliance will also be summarized as a continuous variable using descriptive statistics (n, mean, SD, median, minimum, and maximum) for each treatment group.

All study drug administration and accountability data will be listed by treatment, study site, and subject number. The following variables will be listed: subject identifier, visit, first and last dose dates, medication identification number, drug dispensed and drug returned dates, number of tablets dispensed and returned, and percent compliance.

9.9 Efficacy Analyses

Except as otherwise noted, the analyses and summaries will be based on the FAS.

All primary and key secondary efficacy analyses and summaries will be presented separately for the following groups of subjects:

- Non-Hispanic/Latino Caucasian subjects.
- Hispanic/Latino and/or non-Caucasian subjects.
- All subjects.

All primary, secondary, additional, and other outcomes variables listed below will be presented by treatment group using descriptive statistics by visit (including change-from-baseline for continuous variables) and presented in efficacy listings.

9.9.1 Overview of Primary, Secondary, and Additional Variables

The primary, secondary and additional variables for this study are presented in Table 9.b. While most of the variables in the table are technically discrete, the variable type column indicates how the variables will be considered for purposes of analysis.

Table 9.b Primary, Secondary and Additional Variables

Parameter	Description	Range	Variable Type
MCI_TIME	Time to event of MCI due to AD	N/A	Time-to-event
MMSE	Total of all items correct	0-32	Continuous
$MMSE_i$	MMSE single item i	0-5	Categorical
BVMT_T1	# of elements recalled, Trial 1 (0-12 possible)	0-12	Continuous
BVMT_T2	# of elements recalled, Trial 2 (0-12 possible)	0-12	Continuous
BVMT_T3	# of elements recalled, Trial 3 (0-12 possible)	0-12	Continuous
BVMT_TOT	Total # of elements recalled, Trials 1-3	0-36	Continuous
BVMT_DEL	Total # of elements recalled after a delay (25-minute delay)	0-12	Continuous
BVMT_HIT	Total # of target elements correctly recognized	0-6	Continuous
BVMT_FP	Total # of foils incorrectly endorsed	0-6	Continuous
BVMTDISC	Total target elements correctly recognized relative to Total false positives	-6 to 6	Continuous
BVMTCOPY	Total # of elements copied correctly	0-12	Continuous
DIGIT_F	Total items correct, Digit Span Forward	0-16	Continuous
DIGIT_B	Total items correct, Digit Span Backward	0-14	Continuous
DIGITTOT	Total Forward + Back	0-30	Continuous
TRAILSAS	# of seconds to complete all items, Trails A	0-300	Continuous
TRAILSAE	Errors (number of errors committed during the test)	0-5	Continuous
TRAILSBS	# of seconds to complete all items, Trails B	0-300	Continuous
TRAILSBE	Errors (number of errors committed during the test)	0-15	Continuous
MINT	Total correct, MiNT	0-32	Continuous
SEMFLU	Number of exemplars generated in 1 minute (up to max=100)	0-100	Continuous
LEXFLU	Number of exemplars generated total for F, A, and S (up to max 267)	0-267	Continuous
LEX_F	Number of exemplars generated in 1 minute (up to max=89)	0-89	Continuous
LEX_A	Number of exemplars generated in 1 minute (up to max=89)	0-89	Continuous
LEX_S	Number of exemplars generated in 1 minute (up to max=89)	0-89	Continuous
CLOCK	Number of elements correct (eg, clock face, hands, number placement), Clock Drawing Test	0-10	Continuous
CVLT_T1	# of words (out of 16), recalled 1st trial	0-16	Continuous
CVLT_T2	# of words (out of 16), recalled 2nd trial	0-16	Continuous
CVLT_T3	# of words (out of 16), recalled 3rd trial	0-16	Continuous
CVLT_T4	# of words (out of 16), recalled 4th trial	0-16	Continuous
CVLT_T5	# of words (out of 16), recalled 5th trial	0-16	Continuous
CVLT_TOT	# of words recalled (16/trial), trials 1-5	0-80	Continuous

Table 9.b Primary, Secondary and Additional Variables (continued)

Parameter	Description	Range	Variable Type
CVLT_TB	# of words (out of 16), recalled, intrusion trial B	0-16	Continuous
CVLTSDEL	# of words (out of 16), recalled, short delay	0-16	Continuous
CVLTLDEL	# of words (out of 16), recalled, long (20 min) delay	0-16	Continuous
CVLTINTR	False positive, words recalled	0-50	Continuous
CVLT_REP	Total words repeated	0-28	Continuous
CVLTHITS	Total target words correctly recognized from recognition list	0-16	Continuous
CVLT_FP	Total foil words incorrectly recognized form recognition list	0-32	Continuous
OVERALL	Composite Score (mean of cognitive battery domain scores)	N/A	Continuous
EPISODIC	Mean of cognitive battery Memory domain scores	N/A	Continuous
EXECUTIV	Mean of cognitive battery Executive Function domain scores	N/A	Continuous
LANGUAGE	Mean of cognitive battery Language domain scores	N/A	Continuous
ATTENTN	Mean of cognitive battery Attention domain scores	N/A	Continuous
GDSSC	GDS Short-Form total score	0-15	Continuous
GDS_i	GDS single item i	0-1	Binary
ADCSTOT	ADCS MCFSI total score	0-14	Continuous
$ADCS_i$	ADCS MCFSI single item i	N/A	Categorical
EQ5DINDX	EQ-5D index utility score	0-1	Continuous
EQ5D_VAS	EQ-5D VAS score	0-100	Continuous
$EQ5D_i$	EQ-5D single item i	1-3	Categorical
SF36_PCS	SF-36 Physical Component Summary score	0-100	Continuous
SF36_PCT	SF-36 Physical Component Summary T score	N/A	Continuous
SF36_PCR	SF-36 Physical Component Summary Reliability Consistency Index score	0-100	Continuous
SF36_MCS	SF-36 Mental Component Summary score	0-100	Continuous
SF36_MCT	SF-36 Mental Component Summary T score	N/A	Continuous
SF36_MCR	SF-36 Mental Component Summary Reliability Consistency Index score	0-100	Continuous
SF36_V	SF-36 Vitality domain T score	0-100	Continuous
SF36_P	SF-36 Physical Functioning domain T score	0-100	Continuous
SF36_B	SF-36 Bodily Pain domain T score	0-100	Continuous
SF36_G	SF-36 General Health Perceptions domain T score	0-100	Continuous
SF36_PRF	SF-36 Physical Role Functioning domain T score	0-100	Continuous
SF36_E	SF-36 Emotional Role Functioning domain T score	0-100	Continuous
SF36_S	SF-36 Social Role Functioning domain T score	0-100	Continuous
SF36_M	SF-36 Mental Health domain T score	0-100	Continuous
$SF36_i$	SF-36 single item i	1-5	Categorical
IQTOT	IQCODE overall mean score	1-5	Continuous
IQ_i	IQCODE single item i	1-5	Categorical
ADLTOT	ADCS ADL-PI ADL total score	0-45	Continuous
PHYTOT	ADCS ADL-PI Physical Function total score	0-5	Continuous

Table 9.b Primary, Secondary and Additional Variables (continued)

Parameter	Description	Range	Variable Type
$ADLQ_i$	ADCS ADL single item i	0-3	Categorical
NPIQ_SEV	NPI-Q Total Severity score	0-36	Continuous
NPIQ_CD	NPI-Q Total Caregiver Distress score	0-60	Continuous
$NPIQ_i$	NPI-Q single item i	0-5	Categorical
CGIC_TOT	ADCS-CGIC-MCI total score	1-7	Categorical
CGIC_COG	ADCS-CGIC-MCI cognition score	1-7	Categorical
CGIC_BHV	ADCS-CGIC-MCI behavior score	1-7	Categorical
CGICFUNC	ADCS-CGIC-MCI functional abilities score	1-7	Categorical
CDR01GS	CDR Global Score	0-3	Continuous
CDR01SB	CDR Sum of Boxes	0-18	Continuous
CDR0102	CDR orientation domain score	0-3	Continuous
CDR0101	CDR memory domain score	0-3	Continuous
CDR0103	CDR judgment and problem-solving domain score	0-3	Continuous
CDR0105	CDR home and hobbies domain	0-3	Continuous
CDR0104	CDR community affairs domain score	0-3	Continuous
CDR0106	CDR personal care domain score	0-3	Continuous
$C\text{-}SSRS_i$	C-SSRS single item i	N/A	Categorical
LHIP_ATR	Left Hippocampal Atrophy	N/A	Continuous
RHIP_ATR	Right Hippocampal Atrophy	N/A	Continuous
WHLB_ATR	Whole Brain Atrophy	N/A	Continuous
VNTENLRG	Ventricle Enlargement	N/A	Continuous

9.9.2 Task Completion, Irregular Procedure, and Missing Items on Rating Scales

The rules for handling missing items from each of the scales are designed so that as much data as possible can be used in the analyses, yet with reliable information, when calculating total scores.

Missing data may occur in the course of the study or may be due to drop-out. With regard to neuropsychological data collection within the study, missing data can occur under several scenarios. In an interview setting administration of some portions of the test battery may be precluded by tester error or factors attributable to the patient (illness, refusal, or other problems). Neuropsychological technicians will be trained to avoid missing data whenever possible. In the event that an item or a test is missing, technicians will be trained to record the reason for missing data. When tests are scored, the reason for missing data will be coded in a uniform, standardized manner across sites.

For some reasons, subjects may not complete the tasks / tests as requested, or subjects used irregular procedure and/or questionable effort. In such cases, the data may be considered unreliable and/or invalid. For baseline and follow-up assessments, the scoring of the following cognitive test instruments, including the assignment of missing value codes, will rely on the NeuroCog Trials expert judgment, and will be clearly indicated in the dataset: MMSE, CVLT-II, BVMT-R, Digit Span tests, Trail Making Tests, MiNT, Semantic and Lexical Fluency tests, and Clock Drawing Test.

The criteria for triggering possible cases for adjudication are provided in the appendix of the CIAC Charter. If values for a particular test are: '**999'** (not administered due to Cognitive Impairment) or **'992'** (Refusal), then the test is considered to be failed. This failure will be treated in the same manner as any other 'failed test' in regards to the triggering algorithm. If values for a particular test are: '**991'** (is missing due to physical/sensory reason) or **'994'** ('other' reason), then the test is treated as missing and as a non-failure.

The following tests which are collected at an item level in the eCRF will be prorated whenever a sufficient percentage of data is non-missing: GDS Short-Form total score, ADCS MCFSI total score, IQCODE Overall mean score, ADCS ADL-PI ADL total score, ADCS ADL-PI Physical Function total score, NPI-Q Total Severity score, and NPI-Q Total Caregiver Distress score.

The general rule when individual items are missing from the multiple-item assessments specified in Table 9.c below is as follows: The total score will be calculated using a SAS function ROUND, as ROUND[(sum of nonmissing items)×(total number of items)/(number of nonmissing items)]. (The IQCODE Overall mean score is a mean of the item-level values and so will remain unchanged when proration is valid and will be set to missing when prorating is not valid.) If more than 20% of the items are missing, the total score will be set to missing. The implications of this rule are indicated in Table 9.c for each rating scale. The resulting calculated total scores will be used in all analyses.

Table 9.c Number of Missing Items on Scales with a Missing Value for Total Score

Parameter	Number of Missing Items
GDS Short-Form total score	≥4
ADCS MCFSI total score	≥3
IQCODE Overall mean score	≥6
ADCS ADL-PI ADL total score	≥4
ADCS ADL-PI Physical Function total score	≥2
NPI-Q Total Severity score	≥3
NPI-Q Total Caregiver Distress score	≥3

9.9.3 Analysis of Primary Variable

There are 2 independent null hypotheses in this protocol: (1) assumes that there is no difference in conversion rate between the placebo-treated high-risk and low-risk groups as defined by the prognostic biomarker risk algorithm and (2) assumes that there is no difference in conversion rate between the high-risk active- and placebo-treated subjects. The alternate hypotheses are: (1) there is a difference between the high- and low-risk groups treated with placebo, and (2) there is a difference between placebo and active treatment in the high-risk group.

A futility analysis will be conducted for this study (see Section 9.15).

9.9.3.1 Primary biomarker risk assignment algorithm analysis

In order to test the first hypothesis, a test from a Cox proportional hazards survival model will be used to compare data from non-Hispanic/Latino Caucasian subjects in the (placebo-treated) low-risk group (ie, the combined placebo groups) with data from non-Hispanic/Latino Caucasian subjects in the placebo-treated subjects in the high-risk group. The event in this analysis will be an adjudicated event of MCI due to AD (see Section 9.9.4 below). If this test is significant at the 0.01 level, it can be concluded that the diagnostic prognostic test has differentiated between subjects at low and high risk of being diagnosed with MCI due to AD within 5 years of taking the test.

An exact partial likelihood function will be constructed to fit the Cox proportional hazards (CPH) model to accommodate events with the same times to the event. The robust sandwich estimate for the covariance matrix from the CPH model will be used. The analysis of the primary biomarker comparison will be conducted using the FAS and a CPH model of MCI due to AD events with a factor for biomarker risk assignment and including center, gender, and education as covariates. (Except as otherwise noted, education will be treated as a continuous variable and measured in years for purposes of statistical modeling.)

Parameter estimates, 2-sided 99% CIs, and appropriate p-values for the hazard ratios obtained from the CPH model will be computed to test the null hypothesis of equality of the hazard functions between high-risk placebo and low-risk placebo.

Assessment of the proportional hazards assumption for risk assignment category will be assessed. Plots of Schoenfeld residuals obtained from the CPH models that fit the primary MCI-to-AD endpoint versus ordered time of events along with a smoothing spline will be used to evaluate the proportional hazards assumption for risk assignment category. Systematic departures from a horizontal line are indicative of meaningfully non-proportional hazards. Diagnostic checks will be presented in CSR Appendix 16.1.9.

9.9.3.2 Primary efficacy analysis

A similar analysis process as described above for the primary biomarker risk algorithm analysis will be used to compare the placebo- and active-treated non-Hispanic/Latino Caucasian subjects within the high-risk group (ie, using a similar CPH model with covariates of center, gender, and education and an additional covariate of age but using a factor of high-risk treatment group instead of biomarker risk assignment). If this test is significant at the 0.01 level, it can be concluded that active treatment has delayed the development of mild cognitive impairment due to AD.

An exact partial likelihood function will be constructed to fit the Cox proportional hazards (CPH) model to accommodate events with the same times to the event. The robust sandwich estimate for the covariance matrix from the CPH model will be used. The analysis of the primary efficacy comparison will be conducted using the FAS and a CPH model of MCI due to AD events with a factor for treatment and including center, gender, education, and age as covariates.

Parameter estimates, 2-sided 99% CIs, and appropriate p-values for the hazard ratios obtained from the CPH model will be computed to test the null hypothesis of equality of the hazard functions between the high-risk pioglitazone and high-risk placebo treatment groups.

Assessment of the proportional hazards assumption for treatment group will be assessed. Plots of Schoenfeld residuals obtained from the CPH models that fit the primary MCI-to-AD endpoint versus ordered time of events along with a smoothing spline will be used to evaluate the proportional hazards assumption for treatment group. Systematic departures from a horizontal line are indicative of meaningfully non-proportional hazards. Diagnostic checks will be presented in CSR Appendix 16.1.9.

9.9.3.3 Supportive Nonparametric Interval-Censoring Analyses

Alternative methods of analysis may be considered should some of the underlying model assumptions seem inappropriate (eg, non-proportional hazards). The reasons for any changes to the planned approach and methods will be fully documented.

Most survival analysis methods, including the CPH model, work best when the true event date is known for each and every event which occurs during a study (eg, right censoring). For complex events such as MCI due to AD, generally the true event date is unknown and is only detected at a subsequent visit (ie, interval censoring).

Primary high-risk placebo versus low-risk placebo biomarker analysis

As a sensitivity analysis to the above primary biomarker risk assignment analysis, a nonparametric analysis which accounts for interval censoring by using a weighted generalized log-rank test, stratified by gender, will be done using SAS PROC ICLIFETEST with the option WEIGHT=FINKELSTEIN (which corresponds to a proportional hazards assumption).

Primary high-risk pioglitazone versus placebo efficacy analysis

As above, alternative methods of analysis may be considered if some of the underlying CPH model assumptions are violated; and a similar nonparametric sensitivity analysis using SAS PROC ICLIFETEST will be done to account for interval censoring and its possible effect on the results of the primary efficacy analysis.

9.9.3.4 Supportive Log-rank Sensitivity Analyses

Primary high-risk placebo versus low-risk placebo biomarker analysis

As a sensitivity analysis to the above primary biomarker risk assignment analysis, a log-rank test, stratified by gender, will be done using SAS PROC LIFETEST.

Primary high-risk pioglitazone versus placebo efficacy analysis

A similar log-rank sensitivity analysis as described above for the primary biomarker risk assignment analysis will be used to compare the placebo- and active-treated non-Hispanic/Latino Caucasian subjects within the high-risk group (ie, stratified by gender, but using a factor of high-risk treatment group instead of biomarker risk assignment category).

9.9.3.5 Supportive PPS and subgroup analyses

If more than 5% of total subjects in the FAS high-risk, non-Hispanic/Latino Caucasian subjects have major protocol violations, the primary efficacy analysis will also be performed based on the PPS.

Similarly, if more than 5% of total subjects in the FAS placebo (both high-risk and low-risk), non-Hispanic/Latino Caucasian subjects have major protocol violations, the primary biomarker analysis will also be performed based on the PPS.

For the primary efficacy variable (for both the primary efficacy comparison and for the primary biomarker comparison), subgroup analyses by:

- age at randomization (<75, ≥75)
- gender (female, male)
- baseline diabetic status (non-diabetic: baseline HbA1c <6.5%, diabetic: baseline HbA1c ≥6.5%)
- race/ethnicity (Hispanic/Latino and/or non-Caucasian, all subjects)
- baseline NPI-Q Apathy/Indifference Score (no symptoms; mild, moderate, or severe symptoms)
- regulatory geographic region (North America: US; Europe: UK, Germany, Switzerland; Rest of World: Australia)
- geographic region based on language (English, German)
- baseline statin use (based on the Standardized Drug Groupings maintained by Uppsala Monitoring Centre, sponsors of the WHO Drug dictionaries)

and other variables, if necessary, will be performed. Analysis results for subgroups containing less than 20% of the total subjects in the primary efficacy analysis population should be interpreted especially cautiously. The treatment groups will be compared within each subgroup.
CCI [redacted]

9.9.4 MCI Due to AD Event Description

In clinical practice, the diagnosis of MCI due to AD, like AD dementia, cannot be made solely by laboratory or other tests, but requires the judgment of a clinician who takes into account clinical, cognitive, and functional criteria that define the syndrome. The clinical diagnosis, as defined by the National Institute on Aging (NIA) and Alzheimer's Association (AA) in Albert et al., 2011 (reference given in protocol):

- Establishes clinical and cognitive criteria:
  - Cognitive concern reflecting a change in cognition reported by subject or informant or clinician (ie, historical or observed evidence of decline over time).

  - Objective evidence of impairment in one or more cognitive domains, typically including memory (ie, formal or bedside testing to establish level of cognitive function in multiple domains).
  - Preservation of independence in functional abilities.
  - Not demented.
- Examines etiology of MCI consistent with AD pathophysiological process:
  - Rule out vascular, traumatic, medical causes of cognitive decline, where possible.
  - Provide evidence of longitudinal decline in cognition, when feasible.
  - Report history consistent with AD genetic factors, were relevant.

A cognitive test battery has been specifically constructed to help study Investigators identify individuals who are beginning to decline in their cognitive status and who can be accurately diagnosed with MCI due to AD. This test battery is summarized in the Appendix D, and the individual instruments that comprise the battery are described in detail in Section 9.1.15 of the protocol.

Study Investigators use this cognitive test battery to assess subjects' cognitive status and consider the results of this battery, in combination with the findings from other subject- and project partner-reported instruments of functional change and laboratory findings, to determine diagnosis and further subject flow within the study (ie, send case for adjudication or not).

For the purposes of this clinical trial, the clinical framework described by the NIA/AA publication was operationalized through the incorporation of objective cutoff points to provide the Investigator's guidance for making the clinical diagnosis of MCI due to AD, as

Follows (Table 4.a in the protocol):

Operationalized Criteria for MCI due to AD

Establish clinical and cognitive criteria

- Clinical Dementia Rating (CDR) scale score of 0.5 **AND** one or more of the following:
  - Fails at least one of the two memory tests in the cognitive test battery (failure defined as an individual test score at or below -1.5 SD of the demographically corrected normative mean (a), and the score reflects a change from baseline).
  - Fails 2 or more of the 12 measures in the cognitive test battery, representing separate cognitive domains.
  - (Failure is defined as an individual test score at or below -1.3 SD (10th percentile) of the demographically corrected normative mean, and the score reflects a change from baseline).

Examine etiology of MCI consistent with AD pathophysiological process

- Rule out vascular, traumatic, psychiatric, and other proximal medical causes of cognitive decline.
- Continued evidence of cognitive impairment or continued decline on 6 month follow-up (ie, 2 consecutive study visits showing impairment). (b)

(a) Normative data will be available per language.
(b) It is not required that the two consecutive visits show an identical profile (ie, meeting any definition of MCI due to AD on two consecutive visits is acceptable for a subject to be forwarded for adjudication). Note that the two consecutive visits must contain non-missing determinations of MCI due to AD status in order to contribute to the determination of an adjudicated event of MCI due to AD.

The criteria for triggering possible cases for adjudication are provided in the appendix of the CIAC Charter. For purposes of triggering possible cases, the choice of tests to assess the cognitive domains differs slightly from the table in Appendix D. In particular, the Executive Function domain is assessed using Trail Making Test Part B and digit span total (rather than digit span backwards). And the attention domain consists only of Trail Making Test Part A (and does not include digit span forward).

The CVLT-II memory test consists of 3 measures, two of which are important for purposes of triggering and for calculation of the composite score: Short-Delay Free Recall and Long-Delay Free Recall. Of those two CVLT-II measures, only Long-Delay Free Recall is used for inclusion/exclusion. For post-baseline visits, failure on either of the two measures indicates failure on the CVLT-II test.

As described in the protocol, at each scheduled visit after randomization (ie, every 6 months), outcomes indicating possible MCI or AD will trigger further evaluation at an unscheduled comprehensive medical follow-up visit which should take place within one month of the routine scheduled visit when one of the triggering outcomes was met. The CDR, which is not taken at regularly scheduled post-randomization visits, is one of the procedures which is completed at the unscheduled comprehensive medical follow-up visit.

The primary endpoint is met only on the second consecutive occurrence when the diagnosis of MCI due to AD is confirmed by the adjudication committee. The adjudicators will look at all the collective information sent by the investigative sites in order to confirm (or not) the diagnosis of MCI due to AD. Subjects who convert directly to an AD dementia diagnosis by the NIA/AA criteria (reference given in protocol) will also be considered toward the numerical count of events that determines the duration of the study.

Note that data collected at the EoS/EW visit will also contribute to the determination of primary efficacy events of MCI due to AD. Further details about the definition of an event of MCI Due to AD and the adjudication process are provided in the protocol and the CIAC Charter.

Event Times and Censoring Times

For purposes of analysis, the **event time** of a CIAC confirmed event of MCI due to AD is the time from the randomization date to the **date of the first scheduled visit of the two consecutive scheduled visits at which a subject was assessed with a clinical diagnosis and determined as an primary endpoint event by CIAC using the data from scheduled visit and the triggered unscheduled comprehensive medical follow up visits.**

For subjects whose first and only clinical assessment of MCI due to AD (or AD itself) corresponds to their last visit during the study (e.g., due to subject dropout) and determined by CIAC as an primary endpoint event, the **event time** of a CIAC confirmed event of MCI due to AD is the time from the randomization date to the **date of the scheduled visit.**

A subject who does not have an event of MCI due to AD (including AD itself) will be censored for that event at the date of last visit at which a case of MCI due to AD could have been assessed (i.e., the last regularly scheduled visit at which the cognitive battery was assessed or the last unscheduled comprehensive medical follow-up visit at which the CDR was assessed, whichever is later). For subjects who discontinue from the study prior to their first assessment of the cognitive battery or CDR, the date of last contact will be used as the censoring time. At the time of the futility analysis, a subject who has not discontinued from the study and has not had an event of MCI due to AD (or AD) will be censored for that event at the date of the last visit at which a case of MCI due to AD could have been assessed prior to the futility analysis clinical database cut date (ie, the last regularly scheduled visit at which the cognitive battery was assessed or the last unscheduled comprehensive medical follow-up visit at which the CDR was assessed prior to the futility analysis clinical database cut date, whichever is later).

For subjects who die during the trial, the date of death is considered a censoring time for the event of MCI due to AD (rather than as a component of the event of MCI due to AD itself).

9.9.5 Sensitivity Analyses Using Alternative Event Definitions

As sensitivity analyses for the primary biomarker and primary efficacy analyses, analogous analyses as described in Sections 9.9.3.1 and 9.9.3.2 will be done using the following revised definitions related to an event of MCI due to AD:

- Revised Event Time Definition: For purposes of analysis, the revised **event time** of a CIAC confirmed event of MCI due to AD is the time from either:
  - (1) the randomization date to the **date of the first unscheduled comprehensive medical follow-up visit** which was triggered by the first of the two consecutive scheduled visits at which a subject was assessed with a diagnosis of MCI due to AD (or AD itself) by the CIAC; or
  - (2) the randomization date to the **date of the unscheduled comprehensive medical follow-up visit** which was triggered by the subject's last scheduled visit during the study because the subject was assessed (for the first and only time) with a diagnosis of MCI due to AD (or AD itself) by the CIAC.

The revised event time definition above changes the event time so that it is based on the date of the first unscheduled visit date at which a case of MCI due to AD was determined by CIAC (since CDR data isn't collected at scheduled visits).

9.9.6 Analysis of Key Secondary Variables

The change from baseline in cognitive battery composite score will be analyzed as a continuous variable by study visit using MMRM (mixed model for repeated measures) analysis of covariance with treatment, gender, center, visit, and treatment-by-visit interaction as fixed, categorical effects and education, age, baseline composite score, and baseline composite score-by-visit as continuous, fixed covariates. The effect at each time point for each treatment is allowed to vary freely, and an unstructured covariance matrix will be used to model the within-subject errors. The Kenward-Roger approximation will be used to estimate denominator degrees of freedom. Based on a Missing at Random (MAR), this analysis will be performed using observed case data only.

The SAS code for the MMRM analysis will be as follows:

Proc mixed data=all;

- class visit treat center gender subject;
- model Change = Baseline Baseline*visit treat center gender education age visit visit*treat /solution ddfm=KenwardRoger;
- repeated visit/subject=subject type=UN;
- lsmeans visit*treat / cl pdiff.

In the above model, Baseline is the baseline value and Change is the change from baseline value.

Similarly, the change from baseline in ADCS ADL-PI ADL total score will be analyzed as a continuous variable by study visit using MMRM, where the ADCS ADL-PI ADL total score Baseline, gender, education, and age will be used as the covariate adjustments in the MMRM.

Key secondary endpoints will be analyzed (including reporting mean differences, CIs, and p-values) using the FAS subjects in the primary efficacy population (high-risk, non-Hispanic/Latino subjects) using a 0.01 significance level.

Although change-from-baseline endpoints will be assessed at all visits, the primary time frame for purposes of efficacy analysis is the change from baseline to Month 48 (Year 4).

Missing data from this study will be reviewed carefully to evaluate any patterns in this study. Also, as a sensitivity analysis, pattern mixture models using standard SAS STAT procedures will be performed for the key secondary endpoints to address the possibility of missing data not at random. This method uses sequential regression and multiple imputation methodology to impute missing values after subjects' discontinuation from the study. The missing values from both control and experimental treatment arms will be imputed based on the available data from control subjects and using PROC MI methodology for imputation of monotone missing data

patterns to impute the outcome variables at consecutive visits in a sequential (chain) manner. The number of imputed datasets will be 100.

Alternative methods of analysis (eg, using a transformation of the response variable or by using generalized estimating equations in SAS PROC GLIMMIX or PROC GENMOD) may be considered prior to unblinding should some of the underlying model assumptions (eg, normality) seem severely violated. The reasons for any changes to the planned approach and methods will be fully documented.

9.9.7 Derivation of Cognitive Battery Composite Score

The cognitive test battery by domain is described in Appendix D. To form the composite score, z-scores will be calculated for each test listed below.

Episodic Memory domain	CVLT-II Short-delay free recall correct CVLT-II Long-delay free recall correct BVMT-R Delayed Recall
Executive Function domain	Trail Making (Part B) – Total Seconds WAIS-III Digit Span – Backward
Language domain	Semantic fluency (animals) Lexical/phonemic fluency (letters “F-A-S”)
Attention	WAIS-III Digit Span – Forward Trail Making Test (Part A) – Total Seconds

The z-scores for measures in each domain will be averaged to create the four domain scores, and all four domain scores will be averaged to form the cognitive battery composite score. Because there are two tests for each domain, the domain score can still be calculated if one test is missing. One exception is the case of memory domain, where at least one subtest from the CVLT-II episodic memory tests is required to calculate the composite (i.e., CVLT-II Short-delay free recall correct or CVLT-II Long-delay free recall correct must be non-missing). For the calculation of the memory domain score, the following 3 measures are used (and given equal weight): CVLT-II Short-delay free recall correct, CVLT-II Long-delay free recall correct, and BVMT-R Delayed Recall. If any of the domain scores is missing, then the composite score is set to missing.

For any visit, the z-score for a measure is calculated by taking the value for the measure at that visit, subtracting off the baseline mean for that measure, and then dividing by the baseline standard deviation for that measure. The baseline mean and baseline standard deviation for a measure will be calculated using all of the non-missing baseline values for subjects in the primary efficacy analysis population of high-risk, non-Hispanic/Latino Caucasian subjects.

The z-scores will only be used for the analyses and summaries of endpoints which are derived using multiple components (eg, the composite score and domain scores). All other endpoints will be analyzed and summarized using raw scores.

9.9.8 Controlling Type I Error

To control the two-sided type I error over all the efficacy endpoints that are intended to support potential claims, the primary and key efficacy endpoints will be tested in the following sequential order at significance level 0.01; as soon as an efficacy endpoint test is non-significant at alpha=0.01, the testing procedure stops for all subsequent endpoints:

- Time to diagnosis of MCI due to AD for pioglitazone-treated, non-Hispanic/Latino Caucasian subjects versus placebo-treated non-Hispanic/Latino, Caucasian subjects in the high-risk stratum.
- Change from baseline to Year 4 for cognitive decline on composite score on the cognitive battery for pioglitazone-treated, non-Hispanic/Latino Caucasian subjects versus placebo-treated non-Hispanic/Latino, Caucasian subjects in the high-risk stratum.
- Changes from baseline to Year 4 in instrumental activities of daily living (Alzheimer's Disease Cooperative Study Activities of Daily Living - Prevention Instrument [ADCS-ADL-PI]) for pioglitazone-treated, non-Hispanic/Latino Caucasian subjects versus placebo-treated non-Hispanic/Latino, Caucasian subjects in the high-risk stratum.

9.9.9 Analysis of Additional Variables

Except where otherwise noted, summaries and analyses for additional variables will be done using the FAS subjects in primary efficacy analysis population (high-risk, non-Hispanic/Latino subjects). Except for analyses involving the primary and key secondary endpoints, all other analyses (ie, analyses of additional variables) will be done using a significance level of 0.05.

The incidence of MCI due to AD will also be compared using a chi-squared test, and the difference in incidence will be reported along with 95% confidence intervals.

Relatively few cases of AD are anticipated. The incidence AD will be compared using a chi-squared test, and the difference in incidence will be reported along with 95% confidence intervals. Cases of AD will be described in the listings.

For continuous endpoints which are collected at multiple scheduled visits, change from baseline will be analyzed by study visit using MMRM analysis of covariance.

Although change-from-baseline endpoints will be assessed at all visits, the primary time frame for purposes of efficacy analysis is the change from baseline to Month 48 (Year 4).

The change from baseline variables for the BVMT-R test (BVMT_T1, BVMT_T2, BVMT_T3, BVMT_TOT, BVMT_DEL, and BVMT_COPY) will each be analyzed as a continuous variable by study visit using an MMRM model similar to the ones described above for the key secondary variables.

The change from baseline variables for the Trail Making Tests (Trails A time and Trails B time) and WAIS-III digit span tests (forward span, backward span, and total) will each be analyzed as a continuous variable by study visit using an MMRM model similar to the ones described above for the key secondary variables.

For the variables BVMT_HIT, BVMT_FP, and BVMTDISC, TRAILSAE, and TRAILSBE, counts and percentages of each possible value will be tabulated and summary statistics (n, mean, SD, median, minimum, and maximum) will be given by treatment group (including low-risk placebo) for the actual values and change-from-baseline values for each visit and EoS/EW.

The change from baseline variables for the MINT, semantic and lexical fluency test variables (SEMFLU, LEXFLU, LEX_F, LEX_A, and LEX_S), and Clock Drawing Test will each be analyzed as a continuous variable by study visit using an MMRM model similar to the ones described above for the key secondary variables.

The change from baseline variables for the CVLT-II test (CVLT_T1, CVLT_T2, CVLT_T3, CVLT_T4, CVLT_T5, CVLT_TOT, CVLT_TB, CVLTSDEL, CVLTLDEL, CVLTINTR, CVLT_REP, CVLTHITS, and CVLT_FP) will each be analyzed as a continuous variable by study visit using an MMRM model similar to the ones described above for the key secondary variables.

The change from baseline variables for the cognitive battery domain scores (memory, executive function, language, and attention) will each be analyzed as a continuous variable by study visit using an MMRM model similar to the ones described above for the key secondary variables.

The change from baseline variables for GDS Total score, ADCS MCFSI total score, IQCODE Overall mean score, and ADCS ADL-PI Physical Function total score will each be analyzed as a continuous variable by study visit using an MMRM model similar to the ones described above for the key secondary variables.

The ADCS-CGIC-MCI total, cognition, behavior, and functional ability scores will each be analyzed as continuous variables by study visit using a MMRM models similar to the ones described above for the key secondary variables (using the appropriate corresponding ADCS-CGIC-MCI baseline score). In addition, each of these 4 ADCS-CGIC-MCI variables will be analyzed at the last visit using the Cochran-Mantel-Haenszel test and stratifying by center.

The MMSE total score is scheduled to be collected at 6-month visits and EoS/EW. Summary statistics will be used to present MMSE total score and change from baseline in MMSE total score at EoS/EW.

The CDR and NPI-Q are scheduled to be collected at baseline and EoS/EW. The CDR Global Score and domain scores will be presented using shift tables from baseline to EoS/EW. Summary statistics will be used to present CDR-SB and change from baseline in CDR-SB at EoS/EW. In addition, each of the 6 CDR domain scores at Baseline and at each time point when MCI due to AD was diagnosed will be included in the listings.

For each of the NPI-Q variables Total Severity Score and Caregiver Distress Score, the change from baseline in NPI-Q from baseline to EoS/EW will be analyzed using an ANCOVA model analogous to the model for the key secondary endpoints. The following SAS code will be used for the ANCOVA analysis:

- Proc mixed data=all;
- class treat center gender;

- model Change= Baseline treat center gender education age;
- lsmeans treat / cl pdiff;

In the above model, baseline is the baseline value and Change is the change from baseline value.

Volumetric MRI (vMRI) variables were collected in a substudy of approximately 300 subjects. These volumetric MRI atrophy and enlargement endpoints are calculated with sophisticated algorithms that derive the change in the brain by registering each of the follow-up time points against the baseline and calculating algorithmically the change (ie, this is not determined by calculating the volume at baseline and then independently calculating the volume at a given post-baseline visit and then subtracting).The vMRI variables left hippocampal atrophy, right hippocampal atrophy, whole brain atrophy, and ventricular enlargement will each be analyzed as continuous variables by study visit using a ANCOVA models similar to the ones described above for the NPI-Q variables. The appropriate baseline variable (ie, left hippocampal volume, right hippocampal volume, whole brain volume, or ventricular volume), as well as intracranial volume, will be used as covariates in each these models.

For each efficacy variable for which MMRM analysis of covariance analyses will be done, summary statistics (n, mean, SD, median, minimum, and maximum) by treatment group (including low-risk placebo) will be given for the actual values and change-from-baseline values for each visit and EoS/EW.

For the single-item variables $MMSE_i$, GDS_i, $ADCS_i$, $EQ5D_i$, $SF36_i$, IQ_i, $ADLQ_i$, and $NPIQ_i$, the counts and percentages of each possible value will be tabulated by treatment group (including low-risk placebo) for each visit and EoS/EW. For single-item variables which have a meaningful numerical score, summary statistics (n, mean, SD, median, minimum, and maximum) by treatment group (including low-risk placebo) will be given for the actual values and change-from-baseline values for each visit and EoS/EW.

MRIs will be done as part of the comprehensive medical follow-up evaluation to rule out other causes of dementia and confirm the clinical diagnosis of MCI due to AD (see Section 6.2.4 of the protocol). For subjects who have contraindications for the MRI procedure (such as unremovable ferromagnetic dental work, cardiac pacemakers, etc) a CT scan can be done instead. The possible categories for the interpretation results are: Within Normal Limits; Abnormal, Not Clinical Significant; and Abnormal, Clinical Significant. The last available MRI/CT scan interpretation results for each subject will be summarized by treatment group, and all MRI/CT scan interpretation results will be presented in the listings.

Except for time-to-event analyses and the summary of MRI interpretation results, unscheduled visit values will be excluded from the summaries and analyses but will be included in the end-of-text CSR listings.

9.9.10 Additional Biomarker Qualification Analyses

At the end of this study, data from the analyses comparing high and low risk placebo groups will provide the data necessary to qualify the biomarker risk algorithm as a prognostic test and an important new tool for therapeutic decision making. Specifically, the high risk placebo group and

low risk placebo group will provide the data needed to estimate the positive predictive value (PPV) and negative predictive value (NPV) of the biomarker risk algorithm. Kaplan-Meier curves by treatment group (including low-risk placebo) will be generated for TOMM40 rs10524523 and APOE genotypes and compared to previous results.

Positive and negative predictive values will be reported for the performance of the biomarker risk algorithm. Specifically, calculations will be made of the PPV and NPV at Years 1, 2, 3, 4 and End of Study / Early Withdrawal (for subjects who were in the study at least 4 years), where PPV(t) is the proportion of subjects who are predicted to develop MCI due to AD during the 5 years who do develop MCI due to AD by time t in the high-risk group, and NPV(t) is the proportion of subjects predicted to be free of MCI due to AD at 5 years who are free of MCI due to AD at time t in the low-risk group.

Comparison of Performance Characteristics

The following additional analyses will be done to assess the contribution of age, APOE genotype, and TOMM40 rs10524523 genotype to the biomarker risk algorithm.

Because the biomarker risk algorithm uses multiple criteria for assigning the risk status (ie, APOE genotype, TOMM40 '523 genotype, and Age), a number of analyses will be conducted to delineate which characteristics of the algorithm contributed to the success of the risk assignment. These analyses will evaluate the efficacy of the risk assignment for different aspects of the biomarker risk algorithm. Parts (a) and (b) below both use CPH time-to-event model analogous to the primary biomarker high-risk pioglitazone versus low-risk pioglitazone analysis model described in Section 9.9.3.

a. Time-to-event analyses using alternative possible risk-assignment methods:
   - MCI Biomarker Risk Assignment Algorithm using APOE genotype, TOMM40 '523 genotype, and Age (ie, using the primary biomarker CPH analysis described in Section 9.9.3).
   - Use Age alone with the following cutoff value:
     - ≥76 years (This is the TOMM40 S/VL genotype age threshold for high-risk from the biomarker risk algorithm.)

     This analysis could demonstrate the value of the MCI biomarker risk algorithm over a doctor using age only to assess risk for cognitively normal patients.
   - Use Age and APOE as follows:
     - High-risk: Age ≥ 76 years or carriage of the APOE ε4 allele.
     - Low-risk: Age less than 76 years and non-carriage of APOE ε4 allele.

b. Time-to-event analyses using alternative forms of the biomarker risk algorithm variables as follows:
   - Age (continuous).

- APOE risk-status (Low, High – where High denotes carriage of at least one APOE ε4 allele).
- TOMM40 '523 risk status (Low, Indeterminate, High – where Low corresponds to VL/VL; Indeterminate corresponds to S/VL, S/S, S/L; and High corresponds to L/VL or L/L).

The following sets of the above variables will be added to the primary biomarker comparison CPH model described in Section 9.9.3 so that assessment of the contribution of each variable to assigning a risk category relative to the biomarker risk algorithm risk classification can be made.

- Age.
- APOE risk status.
- TOMM40 '523 risk status.
- Age, APOE risk status.
- Age, TOMM40 '523 risk status.
- APOE risk status, TOMM40 '523 risk status.
- Age, APOE risk status, TOMM40 '523 risk status.

Analysis of Key Secondary2ndary Endpoints and Additional Variables in Placebo Subjects

The summaries and analyses described for the key secondary endpoints and additional continuous variables (at the domain and composite score levels only) (Section 9.9.6 and Section 9.9.9.) will also be performed using high and low risk placebo groups. The summaries will be presented by high and low risk placebo groups. The analyses will include the risk factor as a fixed factor (high vs. low risk). Other covariates (age, baseline, pooled center, gender and years of education) will be included.

9.10 Analyses of Columbia-Suicide Severity Rating Scale (C-SSRS)

The following summaries will be presented for C-SSRS scale:

- Descriptive statistics by study visit.
- Number of subjects with positive reports at baseline and during treatment.
- A shift-table to demonstrate changes from baseline in C-SSRS scores during treatment.

A subject will be summarized as having a positive report at baseline if the subject reported any of the following suicidal ideation or behavior:

- Active suicidal ideation with some intent to act, without specific plan.
- Active suicidal ideation with specific plan and intent.
- Any actual suicide attempt.

- Any interrupted suicide attempt.
- Any aborted suicide attempt.
- Any preparatory acts or behavior.

A subject will be summarized as having a positive report during treatment if the subject reported any of the following suicidal ideation or behavior:

- Active suicidal ideation with some intent to act, without specific plan.
- Active suicidal ideation with specific plan and intent.
- Any actual suicide attempt.
- Any interrupted suicide attempt.
- Any aborted suicide attempt.
- Any preparatory acts or behavior.
- Completed suicide.

9.11 Pharmacokinetic/Pharmacodynamic Analyses

A Population Pharmacokinetic Analysis Plan will be prepared prior to the population PK analyses. This plan will describe the underlying hypotheses, the population PK analysis set, and the statistical methods to be used.

9.12 Pharmacogenomic Analyses

A separate pharmacogenomic analysis plan will be prepared prior to unblinding of the pharmacogenomic data and the subsequent exploratory pharmacogenomic analysis. CCI [redacted]

9.13 Other Outcomes

The change from baseline variables for the EQ-5D index utility score and VAS scores, SF-36 Physical Component and Mental Component summary T scores, and the 8 SF-36 domain T scores will each be analyzed as a continuous variable by study visit using an MMRM model similar to the ones described above for the key secondary variables. (The Physical Component and Mental Component summary scores and reliability consistency index scores will be included in the derived datasets, but analyses for these variables will not be presented in the CSR; they may be provided in the future in a separate report.)

For the following endpoints (corresponding to the analysis method for the key secondary endpoints), subgroup/responder analyses by MCI event status (ie, subject had confirmed a MCI due to AD event during the study, subject did not have confirmed MCI due to AD event during the study) will be done (using the corresponding measurements at the time of the MCI due to AD

event or using the last non-missing measurements for subject who did not have an MCI due to AD event during the study:

- Change in HRQoL (European Quality of life scale [Euro-Qol EQ-5D] health utility and VAS scores, and Short Form-36 [SF-36] Mental and Physical component scores).

Responder analyses involve subgroups which are defined using post-baseline data and so need to be interpreted cautiously. Analysis results for subgroups containing less than 20% of the total subjects in the primary efficacy analysis population should be interpreted especially cautiously. The treatment groups will be compared within each subgroup. Additional exploratory analyses examining subgroup effect and treatment-by-subgroup interaction may be performed.

Data will be collected to allow evaluation of the impact of pioglitazone on health care resource utilization (RU) in individuals at high risk of developing MCI due to AD, to be used in future pharmacoeconomic evaluation of pioglitazone. (These analyses will not be presented in the CSR, but may be provided in the future in a separate report.)

9.14 Safety Analysis

Safety summaries will be based on the safety set. Safety summaries will include descriptive statistics for values, changes, and incidence of events for all treatment groups combined in addition to summary by treatment group. Conventions for the definition of baseline values and visit windowing are provided in Section 9.1.3. Missing safety data will not be imputed.

9.14.1 Adverse Events

All adverse events will be coded using the latest version of MedDRA. In this dictionary, each verbatim term is coded to a lower level term, and then mapped to a preferred MedDRA term, which is then mapped to an SOC. All adverse events will be included in the data listings but only treatment-emergent adverse events will be included in the summary tables.

A treatment-emergent adverse event (TEAE) is defined as an adverse event with an onset that occurs after receiving study drug (AE start date ≥ first dose date) and within 30 days after receiving the last dose of study drug (AE start date – last dose date ≤30). A TEAE may also be a pretreatment adverse event or a concurrent medical condition diagnosed prior to the date of first dose of study drug that increases in severity after the start of dosing. Adverse events data with onset occurring more than 30 days after last dose of study drug (AE start date – last dose date >30) will be listed, but not included in the summary tables. Adverse events with missing onset dates will be summarized regardless of severity and relationship to study medication.

Serious adverse events (SAEs) with onset that occurs after receiving study drug (AE start date ≥ first dose date) and within 30 days after receiving the last dose of study drug (AE start date - last dose date ≤30) will be summarized.

At the adverse event level, the summary tables will present the number of subjects reporting each of these MedDRA events, ie, the number of subjects reporting one or more events which map to the given MedDRA term.

At the SOC level, the summary tables will present the number of subjects reporting one or more events that map to the given SOC. That is, the number of subjects reported at the SOC level will be less than or equal to the sum of the subject counts across all adverse events within that SOC.

In selected summaries (TEAEs overview, and TEAEs by SOC and preferred term), adverse events will be summarized by the number of events reported in addition to the number and percentage of subjects with events.

For the summary of TEAEs by SOC, preferred term and maximum intensity, if a subject experiences more than one episode of a particular coded adverse event, the subject will be counted only once by the maximum intensity of the episode (preferred term). Similarly, if a subject has more than one adverse event within an SOC, the subject will be counted only once by the maximum intensity in that SOC. Adverse events with missing severity will be classified as having the highest severity.

TEAEs classified in the eCRF as related to the study medication will also be summarized by preferred term and SOC. If a subject experiences more than one episode of a particular coded adverse event, the subject will be counted only once by the most related report for the preferred term. Similarly, if a subject has more than one adverse event within an SOC, the subject will be counted only once by the most related report in that SOC. Adverse events with missing relationship will be classified as being related to study drug.

The following summaries will be presented:

- Overview of TEAEs during the study - number and percentage of subjects, number of events.
- TEAEs by SOC and preferred term - number and percentage of subjects, number of events.
- TEAEs by SOC and preferred term by gender (male and female) - number and percentage of subjects, number of events.
- TEAEs by SOC and preferred term by age group (<75, ≥75) - number and percentage of subjects, number of events.
- TEAEs by SOC - number and percentage of subjects.
- TEAEs by preferred term - number and percentage of subjects.
- Intensity of TEAEs by SOC and preferred term - number and percentage of subjects.
- Drug-related TEAEs by SOC and preferred term - number and percentage of subjects, number of events.
- TEAEs leading to study discontinuation by SOC and preferred term - number and percentage of subjects, number of events.
- Treatment-emergent SAEs by SOC and preferred term - number and percentage of subjects, number of events.

- Most frequent TEAEs and non-serious TEAEs (>=5% based on total number of safety set subjects in any treatment group) by SOC and preferred term - number and percentage of subjects, number of events.
- Adverse Events of Special Interest using the following categories defined in the protocol:
  - Congestive Heart Failure.
  - Macular Edema.
  - Hepatic Effects.
  - Bone Fractures.
  - Bladder Cancer.
  - Hypoglycemic Events.

SOCs will be sorted by alphabetical order. Within an SOC, adverse events will be sorted in descending order of total number of subjects with preferred term among all the treatment groups.

Subjects mappings will be provided for TEAEs, TEAEs leading to study drug discontinuation, and serious TEAEs.

All adverse events will be listed by treatment, study center, subject number and onset date of the adverse event. The listing will contain: subject identifier, age, gender, body weight, race, adverse event (preferred term and reported term), SOC, onset date, end date or whether the event was ongoing, duration, type (pretreatment event, adverse event) frequency (intermittent, continuous), intensity (mild, moderate or severe), related to study procedure (yes/no; if yes, specify), relationship to study drug (not related, related), action taken concerning study drug (drug withdrawn, dose not changed, unknown, not applicable, or dose interrupted), the outcome (recovered/resolved, recovering/resolving, not recovered/not resolved, recovered with sequelae, fatal, or unknown), increased liver function tests (yes, no), whether the adverse event was an SAE, and whether the event was an adverse event of special interest (AESI).

Listings for TEAEs leading to study discontinuation, SAEs, deaths, and AEs of special interest will also be presented. A description of the AESIs for this study are provided in Section 10.2.1.3 of the protocol and consist of the following categories: congestive heart failure, macular edema, hepatic effects, bone fractures, bladder cancer, and hypoglycemia.

9.14.2 Clinical Laboratory Evaluations

The following clinical laboratory parameters will be summarized:

- Serum chemistry including alanine aminotransferase (ALT), alkaline phosphatase, aspartate aminotransferase (AST), total bilirubin, GGT, total protein, albumin, creatinine, blood urea nitrogen (BUN), potassium, sodium, glucose, calcium, vitamin B12, folate, parathyroid hormone (PTH), thyroid stimulating hormone (TSH), and free thyroxine (free T4). (Direct bilirubin will be assessed only if total bilirubin ≥2.0 mg/dL. Vitamin B12, folate, and RPR is

collected only at Baseline and as part of the comprehensive medical follow-up evaluation to rule out other causes of dementia).

- Hematology including red blood cells, white blood cells, neutrophils, eosinophils, basophils, lymphocytes, monocytes, hemoglobin, hematocrit, platelets, and HbA1c. (HbA1c is collected only at Baseline.)
- Urinalysis including pH, specific gravity, glucose, and nitrite. (Protein and blood values will not be summarized but will be included in the listings.)

For each laboratory parameter, the following will be displayed for each scheduled time point (each visit and end of study).

- Summary statistics (n, mean, SD, median, minimum, and maximum) by treatment group and overall for the actual values and change-from-Baseline values.
- Shift tables for the change from Baseline to each post-baseline time point will be presented. For these tables each subject will be categorized as low, normal, or high for the baseline value and low, normal, or high for each post-Baseline time point, according to the central laboratory reference ranges. The number of subjects in each of the combinations of shifts will be presented.
- Markedly abnormal values for laboratory parameters, as defined in Appendix A, will be summarized by treatment group and overall. The number and percentage of subjects with MAV values observed post-Baseline in each of the applicable laboratory parameters will be presented.
- Number and percentage of subjects with elevated liver enzyme lab values will be presented using the criteria described in the protocol.

Subject mappings will be provided for markedly abnormal lab values and elevated liver enzyme lab values.

A listing of all laboratory data will be provided. Laboratory data outside of the normal reference range will be flagged on the listing along with values meeting MAV criteria. The listing will also include the age (at consent) and gender of the subject. Listings of MAV laboratory values will also be presented. Direct bilirubin and rapid plasma regain (RPR) will not be summarized but will be listed.

The following listings of information from the liver function test eCRF pages will be presented: Study drug interruption, signs and symptoms, event history, test results and reports, and additional comments.

When lab values are recorded in the form of "<x" or ">y", "x" will be used for "<x" and "y" will be used for ">y" in the MAV summary tables. However, these values will be displayed as is when the individual subject data listings are presented.

Summaries and listings of laboratory data will be presented, as appropriate, in Systeme International (SI) and conventional units.

9.14.3 Vital Signs

Vital signs and weight at scheduled visits and their changes from Baseline will be summarized for each treatment group and overall using descriptive statistics by visit and end of study. The number and percentage of subjects with at least one post-Baseline MAV vital sign value during the double-blind treatment period will be presented for each variable over all visits. A listing of MAV vital signs values will also be presented. The criteria for identification of MAV vital signs values are provided in Appendix B.

9.14.4 Physical Examinations

All physical examination findings will be listed by treatment, study center and subject number. The following variables will be listed: subject identifier, age, gender, study visit, visit date, body system, whether there was an abnormality, and, a description of the abnormality.

9.14.5 12-Lead ECGs

ECG variables at scheduled visits and their changes from Baseline will be summarized for each treatment group and overall using descriptive statistics by study visit and end of study. A shift table for the investigator's ECG interpretation will provide the number of subjects in each of the appropriate categories (Normal, Abnormal but not clinically significant, or Abnormal and clinically significant) at the scheduled visit relative to the Baseline status. The number and percentage of subjects with at least one post-Baseline MAV ECG value during the double-blind treatment period will be presented for each variable over all visits. A listing of MAV ECG values will also be presented.

The criteria for identification of MAV ECG values are provided in Appendix C.

9.14.6 Other Observations Related to Safety

Not applicable.

9.15 Operational Futility Analysis

Tracking of confirmed events of MCI due to AD will be done on an ongoing basis, in a blinded fashion. A confirmed event of MCI due to AD requires two consecutive cases of MCI due to AD corresponding to two consecutive 6-month visits as adjudicated by the CIAC. The study will be declared to be "operationally futile" if the accrual of confirmed events is too slow to allow completion of the study in a feasible time frame.

Under the assumptions of the study, a projection has been made of when the target of 202 confirmed events (in the primary efficacy analysis population) of high-risk, non-Hispanic/Latino Caucasian subjects) will be achieved, as well as a corresponding last-subject-out (LSO) date.

The first assessment of operational futility will be done in April 2017. A maximum delay of up to 1 year in LSO is acceptable at the time of the first operational futility. If the projected LSO delay is 1 year or less at the time of operational futility, then the study will be considered as non-futile with respect to operational futility and will continue unchanged.

If the projected delay in LSO is greater than 1 year at the time of the first operational futility assessment, then a review of subjects with one case of MCI due to AD will be done to see whether or not it is feasible that within the next 6 months enough confirmed events of MCI due to AD are likely to accrue such that the delay in projected LSO date is less than or equal to 1 year: If not, then the study will be declared to be "operationally futile".

If the first operational futility assessment in April 2017 indicates a projected delay in LSO of greater than 1 year but a review of subjects currently being tracked as part of the adjudication process indicates the study is not operationally futile, then a second assessment of operational futility will be done in October 2017, using decision rules analogous to those used for the first operational futility analysis. If the second assessment of operational futility is passed successfully (ie, is non-futile), no further assessments of operational futility are planned.

If operational futility is declared, then an early efficacy futility analysis (ie, prior to reaching 33% of planned events as described in Section 9.16 below) will be done as soon as practicable (ie, after all pertinent data for the futility analyses has been 100% cleaned and locked in the clinical database) and will be provided to the DSMB.

In the event of an early efficacy futility analysis triggered by a declaration of operational futility, the calculated value of conditional power (described below) will be provided to the DSMB: If the conditional power value of the primary efficacy futility analysis is ≤30%, then the DSMB will recommend that the study is futile with respect to efficacy; if the conditional power value is >30%, then the DSMB may consider recommending that the trial is non-futile with respect to efficacy. The DSMB will then present their conclusions to the Executive Committee, who will decide whether or not the trial will be stopped for futility. (Note that any interim analyses will only be used to assess futility: No claim of efficacy will be made based on any interim analysis results in this study.)

9.16 Efficacy Futility Analysis

An analysis for futility for efficacy will be conducted once approximately 33% (66 out of 202) of the anticipated events (ie, conversions from cognitively normal status to adjudicated diagnosis of MCI due to AD, or to AD dementia) have occurred in the high-risk, Caucasian, non-Hispanic/Latino group. However, since the biomarker risk stratum information is blinded, the time of the futility analysis will be estimated using all subjects (both high-risk and low-risk) in the Caucasian, non-Hispanic/Latino group. Therefore, the efficacy futility analysis will occur when approximately 69 events have been confirmed through adjudication in the Caucasian, non-Hispanic/Latino group (including both high-risk and low-risk subjects). Note that no adjustment of the alpha level is required for an analysis for futility, and that for this study the results of the futility analysis will be considered to be non-binding with respect to efficacy for purposes of decision making around study continuation.

Because a review of unblinded data is involved, the DSMB will oversee this process.

Once the total number of target events for the futility analysis has been reached, the DSMB will be unblinded for a review of actual treatment groups with respect to the primary efficacy

endpoint (and potentially other analyses described below, as applicable); this review could take place at the next regularly-scheduled DSMB meeting, or at an ad hoc DSMB meeting.

The primary futility analysis will be conducted by the CCI Unblinded DSMB Statistician, based on the treatment effect within the high-risk, Caucasian, non-Hispanic/Latino group. The calculation is a test from a Cox proportional hazards survival model to evaluate a difference between the treatment groups (pioglitazone versus placebo) for the high-risk comparison at the time of the interim futility analysis. Additionally, Kaplan-Meier curves by treatment group (including low-risk placebo) will be provided for the primary analysis population of non-Hispanic/Latino Caucasian subjects. If the primary futility analysis indicates futility, then the CCI Unblinded DSMB Statistician will provide an additional set of secondary futility analyses (described below). The DSMB could then recommend study continuation to completion for the targeted number of events or recommend (based on pre-specified criteria documented in the SAP) that the study be considered futile with respect to the primary efficacy endpoint, high-risk comparison.

Immediately after the futility analysis data review meeting takes place, the DSMB recommendation regarding the continuation of the study will be promptly communicated to the DSMB Liaison. If the data indicate that study termination should be considered, the DSMB Liaison will promptly convene an ad-hoc meeting within the Takeda/Zinfandel alliance senior management (ie, the Executive Committee) to further discuss the results of the futility analyses and make the final decision as to whether or not to terminate the study. In a similar manner to the DSMB, the Executive Committee will be unblinded to the results of the futility analyses, will keep all futility analysis results and discussions strictly confidential, and will have no involvement with the daily project activities of the study or study team.

If the alliance Executive Committee decides to terminate the study as a result of the futility analyses, this decision will be promptly communicated to all the stakeholders including the AD-4833/TOMM40 project team, investigators, ethics committees, and regulatory agencies.

Specifics of the ublinded analyses to the DSMB (and to the Executive Committee, if applicable) are described below. The futility analysis will have two parts, to be assessed in a hierarchical manner:

- First, an unblinded analysis to determine futility based on lack of pioglitazone efficacy with respect to the primary efficacy endpoint will be provided to the DSMB.
- Second, if the conditional power value is >30% (see Section 9.16.1), then the DSMB will make a recommendation to the Executive Committee that the study is non-futile with respect to efficacy. In this case, the study will continue unchanged, and the Executive Committee will remain blinded.

– If the conditional power value is ≤30%, then the DSMB will receive the additional planned unblinded analyses (described below) from the CCI Unblinded Statistician.

- MMRM analyses by visit (including all visits for which at least 50% of the subjects have non-missing values) based on the composite score key secondary endpoint related to changes from baseline in cognition.
- MMRM analyses by visit (including all visits for which at least 50% of the subjects have non-missing values) based the following components of the composite score key secondary endpoint.
  - CVLT-II Long-Delay Free Recall
  - CVLT-II Short-Delay Free Recall
  - BVMT-R Delayed Recall
- MMRM analyses by visit (including all visits for which at least 50% of the subjects have non-missing values) for MMSE total score.
- Summary statistics by visit for all of the above-described CLVT-II, BVMT-R, CDR-SB, and MMSE total score continuous endpoints for all visits. For CDR-SB, visit will be derived using visit window defined in Section9.1.3.
- CCI

- For the longitudinal model, a linear trajectory model (SPM1L) will be used to model the change in the continuous efficacy endpoint over time. This model is chosen because the futility analysis is likely at an early stage of the disease development and treatment effect. Treatment and treatment*time will be included to evaluate the effect of treatment on the progression of the continuous efficacy endpoint. Values of the continuous endpoint will be used as the response in the model. Baseline value will be coded as the response at time 0. Other covariates (age, years of education, gender, and pooled center) that are included in the MMRM model for the change from baseline in the composite score will also be included in the longitudinal model.
- For the hazard rate model, number of intervals for the piecewise constant baseline hazard function will be selected using $AIC_{Surv|Long}$ [Ref. 2]. Treatment effect on the hazard function will be evaluated by including treatment as a fixed factor. Other covariates (age, years of education, gender, and pooled center) included in the primary CPH model will also be included. Conditional power will be calculated using the value of the t statistic for testing the treatment effect on the hazard rate.

- Following subgroup analyses on the primary efficacy endpoint will be performed when there are at least 20% of the total primary efficacy events in the futility analysis in a subgroup:
  - Gender (male vs. female)
  - APOE risk-status (Low, High – where High denotes carriage of at least one APOE ε4 allele)
  - Baseline MMSE (<=27 vs. >=28)
- Following subgroup summaries and MMRM analyses on the change from baseline in the composite score, CVLT-II Long and Short Delay Free Recall, BVMT-R Delayed Recall and MMSE total score will be performed when there are at least 400 subjects in a subgroup:
  - Gender (male vs. female)
  - APOE risk-status (Low, High – where High denotes carriage of at least one APOE ε4 allele)
  - Baseline MMSE (<=27 vs. >=28)
- If the conditional power value is >10% but <30%, "progression" of the conditional power in terms of an increasing number of the events vs. total needed will be provided for the following information fractions:10% (20 events), 20% (41 events) and 34% (69 events) will be provided. The events will be selected based on the time of the occurrence of the event relative to the first dose. When the evidence against the null hypothesis is accumulating, the conditional power should increase as the information fraction increases; however, if the CP does not change as the information fraction increases, it may be unlikely evidence against the null hypothesis is being accumulated. Essentially if CP remains flat across 10%, 20%, 30% of events, it makes it less likely that the null hypothesis would be rejected the null hypothesis at end of study.
- For each of the conditional power calculation in this analysis, at each cut-off day, subjects with confirmed event or censored on or prior to the cut-off day, time-to-event definition will follow the time to the primary event definition. Subjects with confirmed event after the cut-off day will be censored at the last scheduled or unscheduled visit date prior to the cut-off day. Subjects censored after the cut-off day, time-to-event is censored on the cut-off day.
- Any additional ad hoc analyses requested by the Executive Committee (which will be generated and provided by the CCI Unblinded Statistician).

The above list of additional planned unblinded analyses may be revised by the Executive Committee.

The DSMB will meet, discuss, and decide on a formal recommendation to the Executive Committee regarding continuation of the study with respect to efficacy. Subsequently, the Executive Committee will decide whether or not to stop the trial with respect to efficacy.

In all possible scenarios, the DSMB will clearly document for the Executive Committee whether or not the study appears to be futile, or non-futile, with respect to the conditional power threshold (described below) specified for the primary efficacy analysis.

9.16.1 Conditional Power for Futility for Efficacy

Any conditional power (CP) is the probability of rejecting a tested null hypothesis at the final analysis, given the interim data and a particular assumption about the population treatment effect. For the futility analysis to be performed in this study after 66 events of MCI due to AD in the high-risk, non-Hispanic/Latino Caucasian study population, rejecting the null hypothesis of no treatment effect with respect to the primary efficacy endpoint at the final analysis is equivalent to the event $\{Z(1) \leq -2.576\}$, and the assumption about the treatment effect is that the population hazard ratio equals the sample hazard ratio observed at the time of this futility analysis. The CP is therefore calculated as

$$CP(t) = \Phi[-(Z(t)/t^{1/2} + 2.576)/(1-t)^{1/2}]$$

where $Z(t)$ is the Wald statistic at information time t computed from the CPH model and Φ is the standard normal cumulative distribution function. If the CP(t) ≤30%, the study may be deemed futile for demonstrating the superiority of pioglitazone to placebo for the primary efficacy endpoint of time to MCI due to AD at the final analysis after 202 adjudicated events in the high-risk, non-Hispanic/Latino Caucasian study population.

In the above, a z-value of -2.576 corresponds to a one-sided significance level of 0.005; and t will be equal to the number of events in the high-risk, non-Hispanic/Latino Caucasian primary efficacy study population divided by 202.

9.17 Changes in the Statistical Analysis Plan from the Protocol Analysis Plan

The study protocol states that academic degrees at baseline will be listed and summarized by treatment group and overall. Information about academic degrees is not collected directly in the eCRF and so won't be presented in the CSR.

10.0 REFERENCES

1. A Double Blind, Randomized, Placebo Controlled, Parallel Group Study to Simultaneously Qualify a Biomarker Algorithm for Prognosis of Risk of Developing Mild Cognitive Impairment due to Alzheimer's Disease (MCI due to AD) and to Test the Safety and Efficacy of Pioglitazone (AD-4833 SR 0.8 mg QD) to Delay the Onset of MCI due to AD in Cognitively Normal Subjects, Takeda Development Center Inc., Protocol No. AD-4833/TOMM40_301, Amendment #3, dated 10 November 2014.
2. Zhang D, et al., JMFit: A SAS Macro for Joint Models of Longitudinal and Survivial Data. Journal of Statistical Software, 2016 July; 71(3)

Appendix A Criteria for Identification of Markedly Abnormal Laboratory Values

Hematology—Criteria for Markedly Abnormal Values

Parameter	Low Abnormal	High Abnormal
Red blood cells	<0.8 x LLN	>1.2 x ULN
White blood cells	<0.5 x LLN	>1.5 x ULN
Neutrophils	<0.5 x LLN	>1.5 x ULN
Eosinophils	--	>2 x ULN
Basophils	--	>3 x ULN
Lymphocytes	<0.5 x LLN	>1.5 x ULN
Monocytes	--	>2 x ULN
Hemoglobin	<0.8 x LLN	>1.2 x ULN
Hematocrit	<0.8 x LLN	>1.2 x ULN
Platelets (conventional)	$<75 \times 10^3/\mu L$	$>600 \times 10^3/\mu L$
Platelets (SI)	$<75 \times 10^9/L$	$>600 \times 10^9/L$
HbA1c (conventional)	--	>7%
HbA1c (SI)	--	>0.07

Chemistry—Criteria for Markedly Abnormal Values

Parameter	Low Abnormal	High Abnormal
ALT	--	>3 x ULN
Alkaline phosphatase	--	>3 x ULN
AST	--	>3 x ULN
Total bilirubin (conventional)	--	>2.0 mg/dL
Total bilirubin (SI)	--	>34.2 μmol/L
Direct bilirubin	--	>2 x ULN
GGT	--	>3 x ULN
Total protein	<0.8 x LLN	>1.2 x ULN
Albumin (conventional)	<2.5 g/dL	--
Albumin (SI)	<25 g/L	--
Creatinine (conventional)	--	>2 mg/dL
Creatinine (SI)	--	>177 μmol/L
Blood urea nitrogen (conventional)	--	>30 mg/dL
Blood urea nitrogen (SI)	--	>10.7 mmol/L
Potassium (conventional)	<3.0 mEq/L	>6.0 mEq/L
Potassium (SI)	<3.0 mmol/L	>6.0 mmol/L
Sodium (conventional)	<130 mEq/L	>150 mEq/L
Sodium (SI)	<130 mmol/L	>150 mmol/L
Glucose (conventional)	<50 mg/dL	>350 mg/dL
Glucose (SI)	<2.8 mmol/L	>19.4 mmol/L
Calcium (conventional)	<7.0 mg/dL	>11.5 mg/dL
Calcium (SI)	<1.75 mmol/L	>2.88 mmol/L
Thyroid Stimulating Hormone (TSH)	<0.8 x LLN	>2.0 x ULN
Vitamin B12 (conventional)	<125 pg/mL	--
Vitamin B12 (SI)	<92 pmol/L	--
Folate (conventional)	<2.2 pg/dL	>17.5 pg/dL
Folate (SI)	<5.0 nmol/L	>39.7 nmol/L
Parathyroid hormone (PTH)	<0.8 x LLN	>2.0 x ULN
Free thyroxine (Free T4)	<0.8 x LLN	>2.0 x ULN
Rapid plasma reagin (RPR)	positive	positive

ALT=alanine aminotransferase, AST=aspartate aminotransferase, GGT=γ-glutamyl transferase, LLN=lower limit of normal, N/A=not applicable, ULN=upper limit of normal.

Urinalysis—Criteria for Markedly Abnormal Values

Parameter	Low Abnormal	High Abnormal
pH	N/A	N/A
Specific Gravity	N/A	N/A
Protein	N/A	N/A
Glucose	N/A	N/A
Blood	N/A	N/A
Nitrite	N/A	N/A

Appendix B Criteria for Abnormal Changes from Baseline of Vital Signs

Vital Sign	Criterion Value	Change Relative to Baseline
Systolic arterial blood pressure	>180 mm Hg	Increase of ≥20 mm Hg
	<85 mm Hg	Decrease of ≥20 mm Hg
Diastolic arterial blood pressure	>110 mm Hg	Increase of ≥15 mm Hg
	<50 mm Hg	Decrease of ≥15 mm Hg
Pulse	>120 bpm	
	<50 bpm	
Body Temperature	>37.7 degrees Celsius	
	<35.6 degrees Celsius	
Weight		Change of greater than 7% body weight

Both the criterion value and the change from Baseline must be met.

Appendix C Criteria for Out-of-Range Values for the 12-Lead ECG Parameters

ECG Parameter	Criteria	
	Lower Criteria	Upper Criteria
Heart Rate	<50 bpm	>120 bpm
QT Interval	<280 msec	≥460 msec
QTc Interval	<340 msec	≥500 msec OR ≥450 msec and ≥30 msec change from baseline

Appendix D Cognitive Test Battery

Cognitive Domain	Tests*
Episodic Memory	California Verbal Learning Test – 2nd Edition (CVLT-II) Brief Visuospatial Memory Test – Revised (BVMT-R)
Executive Function	Trail Making Test (Part B) WAIS-III Digit Span Test – backwards span
Language	Multilingual Naming Test (MINT)** Semantic Fluency (animals) Lexical/phonemic Fluency (letters "F-A-S")
Attention	WAIS-III Digit Span Test – forward span Trail Making Test (Part A)
Visuospatial	Clock Drawing Test (CDT)** Copy of BVMT-R figures**

* Note: There are 12 measures derived from 8 neuropsychological tests in the battery: The CVLT-II test involves 2 primary measures (ie, short delay recall, and long delay recall); BVMT-R has 2 measures (ie, copy and recall); Digit Span has 2 measures (ie, forward and backward span); and Trails has 2 measures (ie, Parts A and B). There is one total score for each of the remaining tests: clock drawing, MINT, animal fluency and lexical fluency.
**Clock Drawing, BVMT-Copy, and the MINT, which do not allow generation of standard z scores will only be used for diagnostic purposes and will be excluded from the calculation of the composite score.

Appendix E Detailed Description of Amendments to Text

This section describes changes in reference to SAP incorporating Amendment No. 1.

Page 2 Section 1.1 Approval Signatures

Existing Text

TDC Approvals

PPD

Date

Date

Date

Date

Revised Text

PPD

Date

Date

PPD

Date

Date

Rational for Amendment

SAP signatories updated in accordance with personnel changes.

Page 37, Section 8.9.4, MCI Due to AD Event Description

Existing Text

- Note that data collected at the EoS/EW visit will *not contribute to the determine* of primary efficacy events of MCI due to AD.
  - **Revised Text**
- Note that data collected at the EoS/EW visit will **also contribute to the determination** of primary efficacy events of MCI due to AD.
  - **Rational for Amendment**

The data collected at the EOS/EW is similar to those planned at unscheduled CMFV visit.

Page 38, Section 8.9.4, MCI Due to AD Event Description

Existing Text

- For purposes of analysis, the **event time** of a CIAC confirmed event of MCI due to AD is the time from the randomization date to the *date of the first unscheduled comprehensive medical follow-up visit (i.e., corresponding to when the CDR was collected) which was triggered by the first of the two consecutive scheduled visits at which a subject was assessed with a diagnosis of MCI due to AD (or AD itself) by the CIAC*.
  - **Revised Text**
- For purposes of analysis, the **event time** of a CIAC confirmed event of MCI due to AD is the time from the randomization date to the **date of the first scheduled visit of the two consecutive scheduled visits at which a subject was assessed with a clinical diagnosis and determined as an primary endpoint event by CIAC using the data from scheduled visit and the triggered unscheduled comprehensive medical follow up visits.**
- **For subjects whose first and only clinical assessment of MCI due to AD (or AD itself) corresponds to their last visit during the study (e.g., due to subject dropout) and determined by CIAC as an primary endpoint event, the event time of a CIAC confirmed event of MCI due to AD is the time from the randomization date to the date of the scheduled visit.**

- **Rational for Amendment**
- This change will reflect the thinking that, in any subject dossier reviewed by the CIAC, the first scheduled visit is the first of the 4 visits considered by CIAC in their adjudication of primary endpoint event, and it represents the first time point of these 4 visits where clinical impairment or cognitive decline is first noted. This revised definition better reflects the concept of "time to onset" of the MCI due to AD – "onset" meaning start of symptoms that are confirmed as MCI due to AD (or AD dementia) by the CIAC.

Page 38, Section 8.9.5, Sensitivity Analyses Using Alternative Event Definitions

Existing Text

- Revised Event Definition: For purposes of analysis, the revised **event time** of a CIAC confirmed event of MCI due to AD is the time from either:
  - the randomization date to the **date of the first unscheduled comprehensive medical follow-up visit** which was triggered by the first of the two consecutive scheduled visits at which a subject was assessed with a diagnosis of MCI due to AD (or AD itself) by the CIAC; or
  - the randomization date to the **date of the unscheduled comprehensive medical follow-up visit** which was triggered by the subject's last scheduled visit during the study because the subject was assessed (for the first and only time) with a diagnosis of MCI due to AD (or AD itself) by the CIAC.

  The revised event definition above expands the definition of MCI due to AD to include subjects whose first diagnosis of MCI due to AD corresponds to their last visit during the study (eg, due to subject dropout).

- *Revised Event Time: For purposes of analysis, the revised **event time** of a CIAC confirmed event of MCI due to AD is the time from either:*
  - *the randomization date to the **date of the first scheduled visit** of the two consecutive scheduled visits at which a subject was assessed with a diagnosis of MCI due to AD (or AD itself) by the CIAC; or*
  - *the randomization date to the **date of the subject's last scheduled visit** during the study because the subject as assessed (for the first and only time) with a diagnosis of MCI due to AD by the CIAC.*

  The revised event time definition above changes the event time so that it is based on the date of the first scheduled visit date at which a case of MCI due to AD was suspected (since CDR data isn't collected at scheduled visits).

  - **Revised Text**

- Revised Event **Time** Definition: For purposes of analysis, the revised **event time** of a CIAC confirmed event of MCI due to AD is the time from either:

- the randomization date to the **date of the first unscheduled comprehensive medical follow-up visit** which was triggered by the first of the two consecutive scheduled visits at which a subject was assessed with a diagnosis of MCI due to AD (or AD itself) by the CIAC; or
- the randomization date to the **date of the unscheduled comprehensive medical follow-up visit** which was triggered by the subject's last scheduled visit during the study because the subject was assessed (for the first and only time) with a diagnosis of MCI due to AD (or AD itself) by the CIAC.

  The revised event time definition above changes the event time so that it is based on the date of the first unscheduled visit date at which a case of MCI due to AD was confirmed by CIAC (since CDR data isn't collected at scheduled visits).

– **Rational for Amendment**

– This change will keep the alternative definition of time to MCI due to AD onset as from the time of randomization to the 1st unscheduled CMFV visit in the sensitivity analyses.

Page 40, Section 8.9.7, Derivation of Cognitive Battery Composite Score

Existing Text

- One exception is the case of memory domain, where at least one *subtest from each of the episodic memory tests (CVLT-II and BMVT-R) is required to calculate the composite (ie, BVMT-R Delayed Recall must be non-missing; and CVLT-II Short-delay free recall correct and/or CVLT-II Long-delay free recall correct must be non-missing).*

  – **Revised Text**

- One exception is the case of memory domain, where at least one **subtest from the CVLT-II episodic memory tests is required to calculate the composite (i.e., CVLT-II Short-delay free recall correct or CVLT-II Long-delay free recall correct must be non-missing).**

  – **Rational for Amendment**

  – The BVMT-R can be missing for many reasons other than cognition (arthritis, motor issues, examiner error). The requirement of both BVMT and at least one CVLT-II to be non-missing may cause valid data points to be omitted (such as subjects who complete CVLT only). The composite measure becomes uninformative without the memory domain; however, the CVLT alone can be used to calculate a memory domain.

  –

Page 44, Section 8.9.10, Additional Biomarker Qualification Analyses

Existing Text

- N/A

  – **Revised Text**

- **Analysis of Key Secondary Endpoints and Additional Variables in Placebo Subjects**

  The summaries and analyses described for the key secondary endpoints and additional continuous variables (at the domain and composite score levels only) (Section 9.9.6 and Section 9.9.9) will also be performed using high and low risk placebo groups. The summaries will be presented by high and low risk placebo groups. The analyses will include the risk factor as a fixed factor (high vs. low risk). Other covariates (age, baseline, pooled center, gender and years of education) will be included.

  – **Rational for Amendment**

  – It is expected that these analyses will be helpful for the evaluation of the biomarker qualification performance. These analyses will be presented at the end of study.

  –

Page 54, Section 8.16, Efficacy Futility Analysis

Existing Text

- Summary statistics by visit for all of the above-described CLVT-II, BVMT-R, CDR-SB, and MMSE total score continuous endpoints for all visits.

  – **Revised Text**

- Summary statistics by visit for all of the above-described CLVT-II, BVMT-R, CDR-SB, and MMSE total score continuous endpoints for all visits. **For CDR-SB, visit will be derived using visit window defined in Section 9.1.3.**
- CCI

  [redacted]

  For the longitudinal model, a linear trajectory model (SPM1L) will be used to model the change in the continuous efficacy endpoint over time. This model is chosen because the futility analysis is likely at an early stage of the disease development and treatment effect. Treatment and treatment*time will be included to evaluate the effect of treatment on the progression of the continuous efficacy endpoint. Values of the continuous endpoint will be used as the response in the model. Baseline value will be coded as the response at time 0. Other covariates (age, years of education, gender, and pooled center) that are included in the MMRM model for the change from baseline in the composite score will also be included in the longitudinal model.

  For the hazard rate model, number of intervals for the piecewise constant baseline hazard function will be selected using $AIC_{Surv|Long}$ [Ref. 2]. Treatment effect on the hazard function will be evaluated by including treatment as a fixed factor. Other

covariates (age, years of education, gender, and pooled center) included in the primary CPH model will also be included. Conditional power will be calculated using the value of the t statistic for testing the treatment effect on the hazard rate.

- **Following subgroup analyses on the primary efficacy endpoint will be performed when there are at least 20% of the total primary efficacy events in the futility analysis in a subgroup:**
  - **Gender (male vs. female)**
  - **APOE risk-status (Low, High – where High denotes carriage of at least one APOE ε4 allele)**
  - **Baseline MMSE (<=27 vs. >=28)**
- **Following subgroup summaries and MMRM analyses on the change from baseline in the composite score, CVLT-II Long and Short Delay Free Recall, BVMT-R Delayed Recall and MMSE total score will be performed when there are at least 400 subjects in a subgroup:**
  - **Gender (male vs. female)**
  - **APOE risk-status (Low, High – where High denotes carriage of at least one APOE ε4 allele)**
  - **Baseline MMSE (<=27 vs. >=28)**
- **If the conditional power value is >10% but <30%, "progression" of the conditional power in terms of an increasing number of the events vs. total needed will be provided for the following information fractions:10% (20 events), 20% (41 events) and 34% (69 events) will be provided. The events will be selected based on the time of the occurrence of the event relative to the first dose. When the evidence against the null hypothesis is accumulating, the conditional power should increase as the information fraction increases; however, if the CP does not change as the information fraction increases, it may be unlikely evidence against the null hypothesis is being accumulated. Essentially if CP remains flat across 10%, 20%, 30% of events, it makes it less likely that the null hypothesis would be rejected the null hypothesis at end of study.**

  For each of the conditional power calculation in this analysis, at each cut-off day, subjects with confirmed event or censored on or prior to the cut-off day, time-to-event definition will follow the time to the primary event definition. Subjects with confirmed event after the cut-off day will be censored at the last scheduled or unscheduled visit date prior to the cut-off day. Subjects censored after the cut-off day, time-to-event is censored on the cut-off day.

  - **Rational for Amendment**
  - These additional analyses may be informative in the efficacy futility assessment when the conditional power based on the primary event is <30%.

ELECTRONIC SIGNATURES

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy HH:mm 'UTC')
PPD	Biostatistics Approval	30-Nov-2017 21:00 UTC
	Biostatistics Approval	30-Nov-2017 21:11 UTC
	Pharmacovigilance Approval	01-Dec-2017 13:33 UTC
	Clinical Approval	01-Dec-2017 15:24 UTC