Janssen Research & Development

Statistical Analysis Plan

A Phase 2b/3 Randomized, Double-blind, Placebo-Controlled, Parallel Group, Multicenter Study Investigating the Efficacy and Safety of JNJ-54861911 in Subjects who are Asymptomatic At Risk for Developing Alzheimer's Dementia

Protocol 54861911ALZ2003; Phase 2b/3

JNJ-JNJ:54861911 (atabecestat)

Status: Approved
Date: 31 October 2018
Prepared by: Janssen Research & Development, LLC
Document No.: EDMS-ERI-110315452, 2.0

Compliance: The study described in this report was performed according to the principles of Good Clinical Practice (GCP).

TABLE OF CONTENTS

AMENDMENT HISTORY

Version 1.0 of the SAP was amended to modify the planned analyses due to the cessation of screening, randomization, and dosing in the study as 17 May 2018. 'JNJ-54861911' is changed throughout to 'atabecestat'.

ABBREVIATIONS

Aβ	amyloid beta
A-IADL-Q	Amsterdam instrumental activities of daily living questionnaire
AD	Alzheimer's disease
ADCS-ADL-PI	Alzheimer's Disease Cooperative Study - Activities of Daily Living-Prevention Instrument
ADNI	Alzheimer's Disease Neuroimaging Initiative
AE	adverse event
AIBL	Australian Imaging, Biomarkers and Lifestyle Flagship Study of Aging
ANCOVA	Analysis of Covariance
APOE ε4	apolipoprotein E, ε4 allele
ARH	Autoregressive heterogeneous
ARIA-E	amyloid-related imaging abnormalities – edema or effusion
ATC	Anatomical-Therapeutic-Chemical
BMI	Body mass index
CDR	Clinical Dementia Rating scale
CDR-SB	Clinical Dementia Rating scale – sum of boxes
CFI	Cognitive Function Index
CI	confidence interval
CRF	case report form
CRO	Clinical Research Organization
CS	Homogeneous compound symmetry
CSF	cerebrospinal fluid
CSH	Compound symmetry heterogeneous
C-SSRS	Columbia Suicide Severity Rating Scale
DMC	Data Monitoring Committee
DRV	Driving scenes
ECG	electrocardiogram
eCRF	electronic case report form
EM	Expectation-maximization
EQ-5D-5L	European Quality of Life-5 Dimensions-5 Levels
FCI	Financial Capacity Instrument
FCSRT	Free and Cued Selective Reminding Test
FTP	Future Time Perspective
GCP	Good Clinical Practice
GDS	Geriatric Depression Scale
HRUQ	Healthcare Resource Utilization Questionnaire
HSI	Health Status Index
IA	interim analysis
IES	Impact of Events scale
IRB	Institutional Review Board
ITT	Intent-to-Treat
IWRS	interactive web response system
JRD	Janssen Research and Development
LFT	Liver function tests
LS	Least squares
MAR	Missing at Random
MCMC	Monte Carlo Markov chain
MedDRA	Medical Dictionary for Regulatory Activities
MED-drc	Medication instructions – delayed recall
MED-drg	Medication instructions – delayed recognition
MED-irc	Medication instructions – immediate recall
MHIS	Modified Hachinski Ischaemic Scale
MI	Multiple imputation
MMRM	mixed effect model for repeated measurement
MMSE	Mini Mental State Examination
MRI	magnetic resonance imaging

NAB-DLT	Neuropsychological Assessment Battery–Daily Living Test
NAP-drc	Name, Address, Phone number – delayed recall
NAP-drg	Name, Address, Phone number – delayed recognition
NAP-irc	Name, Address, Phone number – immediate recall
PACC	Preclinical Alzheimer's Cognitive Composite
PCI	Potentially clinically important
PD	pharmacodynamics
PET	positron emission tomography
PK	pharmacokinetics
RBANS	Repeatable Battery for the Assessment of Neuropsychological Status
RDO	Retrieved drop-outs
SAE	serious adverse event
SAP	Statistical Analysis Plan
SC	Sponsor Committee
SD	standard deviation
SF	Semantic Fluency
SF-36	Short Form-36 Scale
SSG	Statistical Support Group
SSR	Sample size re-estimation
STAI	State-Trait Anxiety Inventory
SUVr	standardized uptake value ratio
VAS	Visual analog scale
WAIS-IV	Wechsler Adult Intelligence Scale, version IV
WHO-DD	World Health Organization – Drug Dictionary
WMS	Wechsler Memory Scale

1. INTRODUCTION

This statistical analysis plan (SAP) contains definitions of analysis populations, derived variables and statistical methods for the analysis of efficacy and safety for study 54861911ALZ2003.

Screening, randomization, and dosing in this study was stopped on 17 May 2018. Randomized subjects were to continue in the double-blind treatment phase without treatment (as allowed by protocol) for approximately 3 to 6 months from cessation of dosing. This permitted continued safety monitoring of the subjects per protocol. The planned last-patient-out date is 17 September 2018, except for some subjects who may require further follow-up.

At the time of cessation of screening and randomization, 557 (34%) subjects of the planned 1650 had been randomized. The planned duration of treatment was 54 months. As of the time of cessation of dosing, no subject had more than 27 months of treatment and 447 had 12 months of treatment or less.

As a result of the significantly reduced sample size, the objectives of the study as stated in the protocol and in Section 1.1 below cannot be evaluated. Safety outcomes will be summarized with descriptive statistics. The planned interim analysis of CSF $A\beta_{1-40}$ was performed but other planned interim analyses will not be performed. Between-group comparisons of selected efficacy outcomes will be performed. Other efficacy outcomes will be summarized with descriptive statistics

1.1. Trial Objectives

1.1.1. Primary Objective

The primary objective of this study is to determine whether treatment with atabecestat slows cognitive decline compared with placebo treatment, as measured by a composite cognitive measure, Preclinical Alzheimer Cognitive Composite (PACC), in amyloid-positive subjects who are asymptomatic at risk for developing Alzheimer's dementia.

1.1.2. Secondary Objectives

The key secondary objective of this study is the following:

- To determine if atabecestat will slow the decline of cognitive function and performance of everyday activities, compared with placebo, based on the Cognitive Function Index (CFI).

The other secondary objectives of this study are the following:

- To assess the overall safety and tolerability of atabecestat versus placebo.
- To determine if a decline in activities of daily living can be detected based on the ADCS-Activities of Daily Living-Prevention Instrument (ADCS-ADL-PI) scale, and if so, to assess the effect of atabecestat compared with placebo.
- To compare changes in cognitive performance between atabecestat and placebo based on the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS).

- To determine if the Neuropsychological Assessment Battery–Daily Living Tests (NAB-DLTs) for Memory and Attention can detect decline in cognitive function, and if so, to assess the effect of atabecestat compared with placebo.
- To assess the effect of atabecestat compared with placebo on the Clinical Dementia Rating (CDR) scale, including progression from CDR 0 to CDR 0.5 or higher.
- To assess the plasma and cerebrospinal fluid (CSF) pharmacokinetics (PK) of atabecestat following chronic treatment using a population PK approach and to explore their relationship with efficacy and safety parameters (including biomarkers).
- To assess the effects of atabecestat on accumulation of cerebral fibrillar amyloid, as measured by amyloid Positron Emission Tomography (PET) imaging.
- To assess the effects of atabecestat on markers of neurodegeneration (e.g., tau peptides) in CSF compared with placebo.
- To assess the maintenance of atabecestat effects on markers of Aβ processing in CSF and plasma compared with placebo.

1.1.3. Exploratory Objectives

The exploratory objectives of this study are the following:

- To investigate the effect of atabecestat compared with placebo on brain volume as measured by volumetric Magnetic Resonance Imaging (MRI).
- To investigate the impact of atabecestat compared with placebo on markers of synaptic dysfunction on task-free functional MRI.
- To assess the effects of atabecestat on additional downstream markers of neuronal injury, neurodegeneration, or inflammation in CSF compared with placebo.
- To assess the effects of atabecestat on progression of tau spreading pathology in the brain as measured by tau PET imaging.
- To explore if baseline markers of neurodegeneration (e.g., volumetric MRI, CSF t-tau or p-tau, or tau PET) are related to cognitive decline and response to treatment with atabecestat.
- To explore the correlation between the effect of atabecestat on biomarkers (e.g., amyloid PET, plasma Aβ, CSF Aβ and t-tau or p-tau, tau PET) and clinical outcomes.
- To explore the impact of disclosure of amyloid status on questionnaires probing subject perception of amyloid imaging and concern about developing Alzheimer's disease (AD) dementia.
- To investigate the effect of atabecestat compared with placebo on cognition as measured by a computerized cognitive battery.
- To investigate the effect of atabecestat compared with placebo on function as measured by the Financial Capacity Instrument (FCI).
- To investigate the effect of atabecestat compared with placebo on medical resource utilization as measured by the Healthcare Resource Utilization Questionnaire (HRUQ) scale.

- To investigate the effect of atabecestat compared with placebo on health outcomes as measured by the Short Form-36 (SF-36) and European Quality of Life-5 Dimensions (EQ-5D-5L) scales.
- To explore the ability of the Cognitive Function Index-acute (CFI-a) to measure decline of cognitive function and performance of everyday activities.
- To explore the effect of atabecestat compared with placebo on instrumental activities of daily life (AIDL) as measured by the Amsterdam IADL questionnaire (A-IADL-Q).

1.2. Trial Design

This is a multi-center, double-blind, placebo-controlled, randomized, parallel-group study assessing the efficacy and safety of atabecestat over approximately 4.5 years of treatment in subjects (60 to 85 years of age) who are asymptomatic and at risk for developing Alzheimer's dementia due to evidence of elevated amyloid accumulation based on CSF or amyloid PET imaging.

The study will consist of 3 phases: a screening phase of approximately 90 days during which subject eligibility will be assessed; a fixed dose, double-blind treatment phase during which eligible subjects will receive randomly assigned study drug once daily for up to 4.5 years; and a follow-up phase to be conducted 7 to 28 days after the last dose of study drug. This study will be conducted in an outpatient setting. The maximum study duration for a subject will be 58 months (3-month screening phase + 54-month double-blind treatment phase + 1-month follow-up phase).

Prior to Amendment 3 of the protocol, the fixed dose, double-blind treatment phase included the following treatment groups: placebo, atabecestat 10 mg, and atabecestat 25 mg. Beginning with Amendment 3, the double-blind phase treatment groups are: placebo, atabecestat 5 mg, and atabecestat 25 mg. A total of 11 subjects were randomized prior to the implementation of Amendment 3. Subjects randomized to the atabecestat 10 mg group prior to Amendment 3 and still being dosed after implementation of Amendment 3 will be dispensed atabecestat 5 mg tablets at visits after implementation of Amendment 3. The dosing regimen for all subjects is once-daily, preferably in the morning between 7:00 am and 11:00 am.

1.3. Statistical Hypotheses for Trial Objectives

This section describes the pre-specified statistical hypotheses. However, given the cessation of screening, randomization, and dosing, no formal statistical hypothesis testing will be done.

There are two families of statistical hypotheses: a primary family in which the atabecestat dose groups are compared to the placebo group on the primary efficacy endpoint and a secondary family in which the atabecestat dose groups are compared to the placebo group on the key secondary efficacy endpoint.

The statistical null hypothesis for the primary efficacy endpoint is that there is no difference between either atabecestat dose group and placebo for cognitive decline as measured by the PACC change score at Month 54 (Year 4.5). The alternative hypothesis is that at least one atabecestat dose group is different from placebo for cognitive decline at Month 54 (Year 4.5).

The statistical null hypothesis for the key secondary efficacy endpoint is that there is no difference between either atabecestat dose group and placebo in the decline of cognitive function and performance of everyday activities as measured by the Cognitive Function Index change from baseline at Month 54 (Year 4.5). The alternative hypothesis is that at least one atabecestat dose group is different from placebo in the decline of cognitive function and performance of everyday activities at Month 54.

A gate-keeping strategy combined with a multiple comparison procedure will control overall Type I error over both families of hypotheses at 0.05 (two-sided). Details of the multiple-testing strategy are provided in Section 5.1.1.

1.4. Sample Size Justification

The decline in PACC from baseline was assessed using data from clinically normal populations in the Alzheimer's Disease Neuroimaging Initiative (ADNI); Australian Imaging, Biomarkers, and Lifestyle Flagship Study of Aging (AIBL); and ADCS Prevention Initiative studies. The average difference in PACC decline at 36 months in AIBL between subjects with and without elevated brain amyloid was 1.40 units (95% CI: 0.52 to 2.29), and the standard deviation (SD) at Week 144 (Month 36) was estimated at 2.44 units. If a similar decline in PACC at Month 54 is seen and the SD is 2.44, then 385 completers per treatment arm would provide 83% power (at the 2-sided alpha of 0.05) to detect a difference from placebo of 0.49. The SD at 54 months will likely be higher. If the SD at month 54 increases by 20% to 2.93, then 555 completers per group would be required in order to maintain the same power. The sample size is estimated based on a mixed effects repeated measures (MMRM) model assuming a constant correlation of 0.5[7]. Under the above assumptions and assuming the attrition is no more than 30%, a total of 1,650 randomized subjects (550/treatment arm) will be required at minimum, and 2,400 subjects (800/treatment arm) at maximum. The treatment difference of 0.49 represents 35% of the 1.40-unit difference, or a 35% slowing of decline in PACC in subjects with elevated brain amyloid. If the average time from amyloid positivity to dementia is 15 years[166], then, assuming linear decline, a 35% slowing of cognition may translate into approximately a 5-year delay in the onset of dementia. A 5-year delay could lead to a 57% reduction in the number of patients with dementia[18].

A total of 1,650 randomized subjects (550/treatment arm) will be required, if the attrition does not exceed 30%. This sample size may be further adjusted for increased variability or attrition rate.

1.5. Randomization and Blinding

Randomization

A stratified permuted block randomization will be used in this study. There are 2 stratification factors: country and APOE ε4 carrier status (yes/no). Subjects will be assigned to 1 of 3 treatment groups in a 1:1:1 ratio based on a computer-generated randomization schedule prepared by a CRO under the supervision of the sponsor. The randomization scheme will be implemented in an interactive web response system (IWRS). The IWRS will assign a unique

treatment code for each subject, which will dictate the treatment assignment and matching study drug kit(s) for that subject.

Blinding

Under normal circumstances, the blind should not be broken until all subjects have completed the study and the database is finalized. Otherwise, the blind should be broken only if a specific emergency treatment or course of action would be dictated by knowing the treatment status of the subject. Subjects who have had their treatment assignment unblinded should continue to return for scheduled evaluations.

Data that may potentially unblind the treatment assignment (i.e., study drug concentrations in serum or CSF; treatment allocation, longitudinal CSF/PET biomarker data, or other specific laboratory data) will be handled with special care to ensure that the integrity of the blind is maintained and the potential for bias is minimized. This can include making special provisions, such as segregating the data in question from view by the investigators, clinical team, or others as appropriate until the time of database lock and unblinding.

2. GENERAL ANALYSIS DEFINITIONS

2.1. Subject IDs

In some cases, subjects who screen fail for reasons that are subsequently addressed may be considered for re-screening after approval by the sponsor's medical monitor. In addition, subjects in screening at the time of a previous temporary pause in screening and randomization could be re-screened. In the analysis datasets, data from multiple screening periods for re-screened subjects will be mapped to a single unique subject ID. The data handling rules described below will be applied to all the data collected for that unique subject ID.

2.2. Treatment Group Labels

The following treatment group labels will be used in summary tables:

- Placebo
- Atabecestat 5 mg: This group will include subjects randomized to atabecestat 10 mg prior to Amendment 3.
- Atabecestat 25 mg

2.3. Reference Start Date and Analysis Phases

The analysis reference start date (Study Day 1) is defined as the date of the first dose of study medication. The Study Day of an event start or stop date is defined relative to the analysis reference start date as:

- Event date – analysis reference start date + 1, if the visit/event date is on or after the analysis reference start date, or

- Event date – analysis reference start date -1, if the visit/event date is before the analysis reference start date

Study Days <1 will be considered in the Screening analysis phase; Study Days ≥1 (there is no Study Day 0) will be considered in the Double-blind analysis phase.

For visit dates prior to, or on, the date of the last dose of study medication, an Analysis Day will be defined as above relative to the analysis reference start date. For visit dates after the treatment end date, an Analysis Day is defined relative to the treatment end date as: Visit date – treatment end date + 1.

- Analysis reference start date (date of first dose as above) for all timepoints after first dose (whether before or after last dose)
- For assessments after last dose, use last observation before last dose as the reference time point.

2.4. Visit Windows

As subjects do not always adhere to the protocol visit schedule, the following rules are applied to assign actual visits to analysis visits. Listed below are the visit windows and the target days for each visit. The same definitions will apply to visits before or after the treatment end date using the Analysis Day as defined in Section 2.3 above. The time period description will be appended with '(On-Trt)' for post-baseline visits on or prior to the treatment end date. The time period description will be appended with '(Off-Trt)' for post-baseline visits after the treatment end date. If a subject has 2 or more actual visits in one visit window, the visit closest to the target day will be used as the protocol visit for that visit window. The other additional visit(s) will not be used in the summaries or analyses but they can be used for determination of clinically important endpoints. If 2 actual visits are equidistant from the target day within a visit window, the later visit will be used.

All assignments will be made in chronological order. Once a visit date is assigned to a visit window, it will no longer be used for a later time point except for the endpoint. Listed below (Table 1) are the visit windows and the target days for each visit defined in the protocol.

Table 1: Visit Windows for on and off-treatment Analysis

Scheduled Visit Number	Time period (label on output)	Time Interval (Day*)	Target Time Point (Day)
7	Baseline	-120 to 1	1
8	Week 2	2 to 22	15
9	Week 4	23 to 36	29
10	Week 6	37 to 50	43
11	Week 8	51 to 64	57
12	Week 10	65 to 81	71
13	Month 3	82 to 106	92
14	Month 4	107 to 134	120
15	Month 5	135 to 165	148
16	Month 6	166 to 211	183
17	Month 8	212 to 256	239
18	Month 9	257 to 288	274
19	Month 10	289 to 333	302
20	Month 12	334 to 410	365
21	Month 15	411 to 501	456
22	Month 18	502 to 592	547
23	Month 21	593 to 683	638
24	Month 24	684 to 774	729
25	Month 27	775 to 865	820
26	Month 30	866 to 956	911
27	Month 33	957 to 1047	1002
28	Month 36	1048 to 1138	1093
29	Month 39	1139 to 1229	1184
30	Month 42	1230 to 1320	1275
31	Month 45	1321 to 1411	1366
32	Month 48	1412 to 1502	1457
33	Month 51	1503 to 1593	1548
34	Month 54	1594 to 1684	1639
	Month >54	1685 to ∞	1685

* Relative to analysis reference start date if visit date is prior to or on the treatment end date. Relative to treatment end date if visit date is after treatment end date.

'End point (On-Trt)' is defined as the last available post-baseline result among the on-treatment visits. Unscheduled visit results are included in this definition and will be considered as the endpoint value if the unscheduled visit result is the last post-baseline result available among all on-treatment visits. 'Baseline (Off-Trt)' will be defined as the same as 'End point (On-Trt)'. 'End point (Off-Trt)' is defined similarly to 'End point (On-Trt)' but based on off-treatment visits.

2.5. Pooling Algorithm for Analysis Centers

Country is a randomization stratification factor and will be used as a factor in the analysis of the primary efficacy endpoint.

Countries within a geographical region (as defined by the United Nations Statistics Division [http://unstats.un.org/unsd/methods/m49/m49regin.htm#ftnb]) with fewer than 12 subjects in the intent-to-treat (ITT) analysis set (see Section 2.6.1.1 below) across all sites within the country will be combined to form a single pooled 'country' for analysis.

If only one country within a geographical region has fewer than 12 ITT subjects, it will be combined with the next smallest country within the geographical region.

If the combination of all countries within a geographical region totals fewer than 12 ITT subjects, that geographical region will be combined with an appropriate country from a neighboring geographical region.

As of the cessation of randomization, the following countries had fewer than 12 randomized, treated subjects: Canada (n = 7), Mexico (n = 9), Germany (n = 8), Finland (n = 7), and Italy (n = 3). For analysis purposes, Canada and Mexico will be combined with USA (n = 188) and Germany, Finland, and Italy will be combined to form a single EU pooled 'country'.

Unless otherwise noted, the use of 'country' in this document refers to pooled countries as defined in this section.

2.6. Analysis Sets

2.6.1. Efficacy Analysis Sets

2.6.1.1. Primary Efficacy Analysis Set

The primary efficacy analysis set includes all randomized subjects (intent-to-treat [ITT] analysis set).

2.6.2. Safety Analysis Set

The safety analysis set includes all randomized subjects with at least one dose of study medication.

2.6.3. Pharmacokinetics Analysis Set

The pharmacokinetics (PK) analysis set is defined as subjects who have received at least one dose of study medication and have at least one valid blood sample drawn for PK analysis.

2.6.4. Biomarker Analysis Sets

The CSF Aβ analysis set includes all randomized subjects with valid CSF Aβ data at baseline and at ≥1 post-baseline visit. All CSF Aβ values will be blindly reviewed for qualification.

The amyloid PET analysis set includes all randomized subjects with valid amyloid PET data at baseline and at ≥1 post-baseline visit.

The CSF p-tau and t-tau analysis set includes all randomized subjects with valid CSF p-tau and t-tau data at baseline and at ≥1 post-baseline visit.

The tau PET analysis set includes all randomized subjects with valid tau PET data at baseline and at ≥1 post-baseline visit.

2.7. Definition of Subgroups

The primary efficacy endpoint and adverse events will be summarized based on the following subgroups:

- Sex (male, female)
- Race (White, Black or African American, Asian, or Other)
- Age Group (<65 years, 65 to 74 years, ≥75 years)
- Education level
- Country (per Section 2.5)
- Disease stage: Stage 1, 2, or 3 based on the criteria below:

	Amyloid positive	Baseline hippocampal volume < median baseline hippocampal volume	Baseline[a] RBANS Total Scale <85
Stage 1	Yes	No	No
Stage 2	Yes	Yes	No
Stage 3	Yes	Yes	Yes

[a] RBANS assessment performed 5 to 10 days before randomization.

3. INTERIM ANALYSIS AND DATA MONITORING COMMITTEE REVIEW

An independent, external Data Monitoring Committee (DMC) has been established to perform safety monitoring over the entire duration of the study and review interim analysis results for futility. An independent, external statistical support group (SSG) will perform the interim efficacy analysis and provide statistical reports to the DMC. Of the interim analyses planned in the protocol, only the futility interim analysis of $A\beta_{1-40}$ in CSF was performed because of the cessation of screening, randomization, and dosing.

Details of the DMC and its procedures are provided in the DMC Charter[1]. The names, roles, and contact information of the members of the Janssen R&D Sponsor Committee (SC) are also available in the DMC charter. Additional details about the timing of the interim analyses and the flow of information between the DMC, SSG, SC, JRD, and other involved parties is provided in an operational plan (Attachment 1). Note, the operational plan is no longer in effect due to the early cessation of the study.

3.1. Safety Monitoring

The DMC will conduct safety monitoring meetings approximately every 3 months over the entire duration of the study. The content of the safety monitoring reports is specified in the DMC Charter.

3.1.1. Interim Analyses of $A\beta_{1-40}$ in CSF

An unblinded interim analysis of $A\beta_{1-40}$ in CSF will be performed by the SSG and DMC.

The IA for CSF $A\beta_{1-40}$ will take place when at least 20 subjects per group (60 total subjects) in the CSF $A\beta_{1-40}$ analysis set (Section 2.6.4) meet the following criteria:

- The subject has a Month 12 CSF $A\beta_{1-40}$ result.
- The subject was ≥85% compliant with treatment through the Month 12 visit.
- The subject was dosed for ≥6 of the 7 days prior to the Month 12 visit.

Justification for this sample size is provided in Attachment 2.

If the proportion of subjects meeting the treatment compliance criteria is low, the IA will be repeated including all subjects with a biomarker result at the specified time point. In this case, the Sponsor Committee will take both sets of results into account to make a final decision.

There will be no Type I error adjustment for the interim analysis since superiority will not be declared.

Tables summarizing data will have columns A, B and C randomly representing the study treatments 'Placebo', 'Atabecestat 5 mg' and 'Atabecestat 25 mg'. Subjects randomized to the atabecestat 10 mg group before implementation of Amendment 3 will be included in the atabecestat 5 mg group.

Demographics and APOE ε4 carrier status will be summarized as described in Section 4.1.1 and Section 4.1.2. Subject disposition, treatment compliance, and extent of exposure will be summarized as described in Section 4.2, Section 4.3, and Section 4.4, respectively.

The objective of this IA is to assess the treatment effect on CSF $A\beta_{1-40}$. Since screening, randomization, and dosing has been stopped, the DMC will not make a recommendation to the JRD Sponsor Committee, but will report the results to the JRD Sponsor Committee.

The following endpoints will be analyzed:

- Key endpoint: Percent change from baseline to Month 12 in CSF $A\beta_{1-40}$.
- Supportive endpoint: Percent change from baseline in plasma $A\beta_{1-40}$.

Absolute values, changes from baseline and percent changes from baseline in $A\beta_{1-40}$ in CSF and plasma will be summarized by time point and treatment group using descriptive statistics (mean, standard deviation, median, range, and 95% CI).

Percent changes from baseline in $A\beta_{1-40}$ in CSF and plasma at Month 12 will be analyzed based on an analysis of covariance (ANCOVA) model that includes treatment group and the baseline value as a covariate. The Least Squares (LS) means and treatment differences versus placebo with the corresponding 95% confidence intervals will be provided.

The criterion for lack of effect on $A\beta_{1-40}$ in CSF is based on the LS mean difference (A-P) in percent reduction from baseline in $A\beta_{1-40}$ between the 25 mg dose (A) and placebo (P):

- If (A – P) < 50%, then conclude lack of effect for atabecestat
- If (A – P) ≥ 50%, then do not conclude lack of effect for atabecestat

Where P is the LS mean for percent reduction versus baseline in $A\beta_{1-40}$ in the placebo group, and A is the LS mean for percent reduction versus baseline in $A\beta_{1-40}$ in the 25 mg group.

4. SUBJECT INFORMATION

The ITT analysis set and the safety analysis set will be used throughout this document. The number of subjects screened and the number of subjects in each analysis set will be summarized by treatment group and overall. In addition, the distribution of subjects by region, country, and site will be presented. Region is defined as found at http://unstats.un.org/unsd/methods/m49/m49regin.htm.

4.1. Demographics and Baseline Characteristics

4.1.1. Demographics

Table 2 presents a list of the demographic variables that will be summarized by treatment group and overall for the ITT analysis set.

Table 2: Demographic Variables

Continuous Variables	**Summary Type**
Age, years	Descriptive statistics (N, mean, standard deviation [SD], median and range [minimum and maximum])
Weight, kg	
Height, cm	
Body Mass Index (BMI), kg/m^2	
Categorical Variables	
Age (<65years, 65 to 74 years, ≥75 years)	Frequency distribution with the number and percentage of subjects in each category
Sex (male, female)	
Race[a] (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or other Pacific Islander, White, Other, Multiple)	
Ethnicity (Hispanic or Latino, Not Hispanic or Latino, Not Reported, Unknown)	
BMI (underweight <18.5 kg/m^2, normal 18.5-<25 kg/m^2, overweight 25-<30 kg/m^2, obese ≥30 kg/m^2)	

[a] If multiple race categories are indicated, then Race is recorded as "Multiple."

4.1.2. Baseline Characteristics

Previous study participation

Subjects who had previously participated in the Janssen non-interventional studies REGISTRYALZ0001 (CHARIOT-PRO) or 54861911ALZ0001 (study to evaluate psychometric comparability of translated, culturally adapted versions of the PACC components with the US English versions) could participate in the current study. The number and percentage of subjects who previously participated in these studies will be summarized by treatment group and overall.

APOE ε4 carrier status

The APOE genotype and the number and percentage of subjects who are APOE ε4 carriers will be summarized by treatment group and overall.

Hippocampal volume

Left and right hippocampal volume will be derived from MRI. For each subject, the average of the left and right volumes will be calculated. Baseline left, right, and average hippocampal volume will be summarized with descriptive statistics by treatment group and overall.

Family history of dementia

The number and percentage of subjects with at least one first degree family member, at least one second degree family member, or either a first- or second-degree family member diagnosed with dementia will be summarized by treatment group and overall, with additional summaries for each type of dementia (mixed, Alzheimer's, vascular, Lewy body, unknown, or other).

Marital status and education level

The number and percentage of subjects in each marital status category (single, separated, married, divorced, widowed) and in each category of highest education level completed (less than upper secondary/high school, upper secondary/high school, some education after upper secondary/high school, bachelor's degree or equivalent, master's degree or equivalent or higher) will be summarized by treatment group and overall.

4.1.3. Medical History and Baseline General Physical Examination

The verbatim terms used in the CRF by investigators to identify medical history diseases or conditions will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). For each disease/condition, the number and percentage of subjects with at least 1 past occurrence and with at least 1 ongoing occurrence of the disease or condition will be summarized by treatment group and overall.

For each general physical examination body system, the number and percentage of subjects with an abnormal finding will be summarized by treatment group and overall. A listing of specific physical examination abnormalities will be provided.

A listing of pre-planned surgeries or procedures including the specific surgery/procedure, the planned date of the surgery/procedure, and the study day of the planned surgery/procedure will be provided.

4.1.4. Informant characteristics

Table 3 presents a list of the informant characteristics that will be summarized by treatment group and overall for the ITT analysis set.

Table 3: Informant Characteristics

Continuous Variable	**Summary Type**
Age, years	Descriptive statistics (N, mean, standard deviation [SD], median and range [minimum and maximum])
Categorical Variables	
Sex (male, female)	Frequency distribution with the number and percentage of subjects in each category
Relationship to subject (spouse or partner, sibling, child, cousin, friend, paid companion, other)	
Does informant live with the subject? (yes, no)	

4.1.5. Rosen-modified Hachinski Ischemic Scale (MHIS)

The MHIS is an 8-item clinical questionnaire which evaluates items and characteristics typical of vascular dementia[15]. It is collected during screening only. Abrupt onset, history of strokes, focal neurological symptoms, and focal neurological signs are scored as 2 if present and 0 if absent. Stepwise deterioration, somatic complaints, emotional incontinence, and history of hypertension are scored as 1 if present and 0 if absent. A MHIS total score is calculated as the sum of the 8 individual item scores. Item scores are recorded on an electronic device which will provide the total score.

For each item or characteristic, the number and percentage of subjects rated as having the item or characteristic (rated as 'present') will be summarized by treatment group and overall. The MHIS total score will be summarized with descriptive statistics by treatment group and overall.

4.1.6. Geriatric Depression Scale (GDS)

The GDS is a 30-item measure used to identify depression in elderly individuals[23]. The 30-item GDS is collected during screening only. Each item is answered 'yes' or 'no' and scored as 0 or 1 according to the table in Attachment 3. The GDS-30 total score is the sum of the 30 individual item scores. Item scores are recorded on an electronic device which will provide the total score.

The GDS-30 total score will be categorized as follows:

- Normal (GDS-30 total score ≤9)
- Mildly Depressed (GDS-30 total score 10 to 19)
- Severely Depressed (GDS-30 total score ≥20)

The GDS-30 total score will be summarized with descriptive statistics by treatment group and overall and the number and percentage of subject in each GDS-30 total score category will also be summarized by treatment group and overall.

4.1.7. Assessment of Psychological Well Being

The Assessment of Psychological Well Being is composed of the 15-item short version of the Geriatric Depression Scale (GDS-15) and the 6-item short version of the State-Trait Anxiety Inventory (STAI). See Section 6.4.1 for additional details. The total scores of each scale during screening will be summarized with descriptive statistics by treatment group and overall.

4.1.8. Amyloid Status

At screening, amyloid status will be evaluated in all subjects using either CSF amyloid β_{1-42} ($A\beta_{1-42}$) concentrations or brain amyloid burden by PET imaging or both. Study eligibility will be determined by a positive result showing evidence of elevated accumulation on either evaluation.

For subjects with results of both biomarkers at screening, a cross-tabulation of elevated amyloid accumulation status (positive or negative) between the 2 biomarkers will be provided to evaluate concordance.

4.1.9. Amyloid Disclosure Assessments

The Concerns about Alzheimer's Disease Dementia Scale and the Future Time Perspective Scale are conducted for all subjects during the screening phase prior to and after the amyloid status disclosure visit and at regularly scheduled visits during the double-blind phase. For subjects with a PET scan performed to assess eligibility, the Views and Perceptions of Amyloid Imaging Scale will be assessed during the screening phase prior to and after the amyloid status disclosure visit. The Impact of Events Scale will be assessed for all subjects during screening after the amyloid status disclosure visit.

Concerns about Alzheimer's Disease Dementia Scale

See Section 5.5.1.1 for details about the Concerns about Alzheimer's Disease Dementia Scale. The distribution of item scores, which range from 1 (strongly agree) to 5 (strongly disagree), for each of the 6 questions will be summarized by treatment group and overall for the pre- and post-amyloid status disclosure visits.

Future Time Perspective Scale

See Section 5.5.1.1 for details about the Future Time Perspective (FTP) Scale and calculation of the overall FTP score. Overall FTP score will be summarized with descriptive statistics by treatment group and overall for the pre- and post-amyloid status disclosure visits.

Impact of Events Scale

The Impact of Events (IES) scale is a 15-item self-report measure that assesses 2 common responses related to a specific stressful life event: intrusion and avoidance[3]. The IES has been adapted for amyloid disclosure-related distress. The scale will be administered during the screening phase within 3 days after disclosure of amyloid status.

Subjects will rate how frequently each item was true for them since disclosure of amyloid results as "not at all" (0), "rarely" (1), "sometimes" (3), or "often" (5). The Intrusion sub-scale is the sum of the scores of the 7 intrusion-related items and the Avoidance sub-scale is the sum of the scores of the 8 avoidance-related items (see Attachment 4 for a list of which items belong to which sub-scale). The total score is the sum of all 15 items. If any item is missing, the relevant sub-scale and the total score will be missing.

The total IES score and the Intrusion and Avoidance subscales will be summarized with descriptive statistics by treatment group and overall.

Views and Perceptions of Amyloid Imaging Scale

Subjects undergoing amyloid PET imaging to determine study eligibility will complete the Views and Perceptions of Amyloid Imaging scale during the screening phase prior to and after the amyloid status disclosure visit. Subjects will rate how important 9 reasons for seeking results of amyloid PET imaging are to them on a 5-point scale with 0 = "not at all important" to 5 = "extremely important". The distribution of scores for each item will be summarized by treatment group and overall for the pre- and post-amyloid status disclosure visits.

4.2. Disposition Information

Screened subjects and reason for screen failures will be summarized overall.

The number of subjects in the following disposition categories will be summarized throughout the study by treatment group and overall:

- Subjects randomized
- Subjects receiving study medication
- Subjects completing the study
- Subjects who discontinued study medication
- Reasons for discontinuation of study medication
- Subjects who terminated study prematurely
- Reasons for termination of study

4.3. Treatment Compliance

Study medication compliance will be calculated as follows: Study medication compliance (%) = 100 x number of days taking study medication/total treatment duration (including days off study medication). For subjects who continue in the double-blind phase after discontinuing study medication, compliance will be based only on the days up to the last day of study medication dosing.

The number and percentage of subjects who have at least 85% study medication compliance will be summarized by treatment group.

4.4. Extent of Exposure

The number and percentage of subjects who receive study medication will be summarized by treatment group.

Study medication duration is defined as: (date of last dose of study medication – date of first dose of study medication) +1.

Descriptive statistics for study medication duration (N, mean, SD, median, and range (minimum, maximum)) will be presented by treatment group for the safety analysis set. Subject-years of exposure are calculated as days of exposure/365.25. Subject-years will be presented by treatment group.

Duration of exposure will be summarized in the following duration categories by treatment group and presented graphically in a histogram:

- <6 months (<180 days)
- 6 months - <1 year (181 to <365 days)
- 1 year – <1.5 years (365 to <548 days)
- 1.5 years - <2 years (548 to <730 days)
- 2 years - <2.5 years (730 to <913 days)

Cumulative exposure (≥6 months, ≥1 year, ≥2 years) will also be summarized.

4.5. Protocol Deviations

In general, the following list of major protocol deviations may have the potential to impact subjects' rights, safety or well-being, or the integrity and/or result of the clinical study. Subjects with major protocol deviations will be identified prior to database lock and the subjects with major protocol deviations will be summarized by category:

- Developed withdrawal criteria but not withdrawn
- Entered but did not satisfy criteria
- Received wrong treatment or incorrect dose
- Other

A listing of major protocol deviations will be provided.

4.6. Prior and Concomitant Medications

Prior and concomitant medications will be coded using the World Health Organization Drug Dictionary (WHO-DD). Prior medications are defined as any therapy used before the day of first dose of study medication. Concomitant medications are defined as any therapy used on or after the same day as the first dose of study medication, including those that started before and continued on after the first dose of study medication. Rules for handling partial concomitant medication start and stop dates are in Attachment 5.

Summaries of concomitant medications will be presented by Anatomical-Therapeutic-Chemical (ATC) class and treatment group. The proportion of subjects who receive each concomitant medication will be summarized as well as the proportion of subjects who receive at least one concomitant medication. For subjects who continued in the double-blind phase after discontinuing study medication, separate summaries of concomitant medications taken while on study medication and after discontinuing study medication will be provided.

Prior medications will be summarized by treatment group and ATC class.

5. EFFICACY

Inferential analyses or descriptive analyses will be performed for the on-treatment period. For off-treatment period, only descriptive statistics will be presented.

5.1. Analysis Specifications

5.1.1. Level of Significance

At the time of cessation of dosing, 34% of the planned 1650 subjects had been randomized and 80% of those randomized had no more than 12 months of the planned 54 months of treatment. As a result, any statistical comparisons will be considered exploratory in nature and will be performed at the 0.05 level of significance without adjustment for multiple comparisons.

5.1.2. Data Handling Rules

Missing data handling rules will be applied to the scales that are included in the composite primary efficacy endpoint (see Section 5.2.2). For on-treatment visits, if a scale value is missing and the scale value was observed at the previously scheduled visit, the value at the previous visit will be carried forward. If a scale value is missing at baseline and the scale value was observed at the next scheduled visit, the value at that visit will be used as the baseline value. No observation will be carried forward or backward more than one visit. Missing data in the off-treatment period will not be imputed.

5.2. Primary Efficacy Endpoint

5.2.1. Primary Estimand and Analysis Set

Given that all analyses are exploratory due to the reduced sample size, no primary estimand is specified.

The primary analysis will be based on the ITT analysis set (see Section 2.6.1.1). The ITT analysis set includes all randomized subjects. The analysis of the PACC change score will only include subjects for whom the PACC change score is non-missing at ≥1 post-baseline time point.

Measurements collected after initiation of any Alzheimer adjunctive therapy will be included in the analysis. Measurements collected after treatment discontinuation will not be included in the analysis in the on-treatment analysis.

5.2.2. Definition

The primary endpoint is the PACC change score.

The PACC change score at a time point is a composite of 4 standardized changes from baseline (two episodic memory tests, one timed executive function test, and a test of global cognition) at that time point as follows:

PACC change score = Z1+Z2+Z3+Z4, where

- Z1 = (change from baseline in FCSRT score / SD of the baseline FCSRT score), where the FCRST score is twice the number of correct free recalls in 3 trials plus the number of correct cued recalls in 3 trials from the Free and Cued Selective Reminding Test (FCSRT). See Section 5.4.1.1 for details of the FCSRT.
- Z2 = (change from baseline in WMS test score / SD of the baseline WMS test score), where the WMS test score is the total number of correct recalls after a 30-minute delay from the Wechsler Memory Scale (WMS) Logical Memory subtest (delayed paragraph recall). See Section 5.4.1.2 for details of the WMS Logical Memory subtest.
- Z3 = (change from baseline in DSS test score / SD of the baseline DSS test score), where the DSS test score is the number of correctly drawn symbols completed in 120 seconds from the Wechsler Adult Intelligence Scale (WAIS)-IV Coding subtest (also known as the Digit Symbol Substitution [DSS] test). See Section 5.4.1.3 for details of the WAIS-IV Coding subtest.
- Z4 = (change from baseline in MMSE total score / SD of the baseline MMSE total score), where the MMSE total score is the total number of correctly completed items on the Mini-Mental State Examination (MMSE). See Section 5.4.1.4 for details of the MMSE.

The PACC baseline score is the sum of the 4 components' standardized baseline value (baseline value / SD of baseline values). The PACC scores will be considered missing if more than 1 of its 4 components is imputed (see Section 5.1.2). Change from the off-treatment baseline (see Section 2.4) will be calculated. For off-treatment visits, however, the components will not be imputed. The PACC scores will be considered missing if any of the components is missing.

The scales contributing to the PACC all follow the same schedule: during screening, at the randomization visit, and every 6 months after randomization through Month 54 (or early termination).

5.2.3. Analysis Methods

The primary analysis is based on the primary analysis set defined in Section 5.2.1 and uses an MMRM model. This relies on the Missing at Random (MAR) assumption: after accounting in the model for the effects of the covariates and for the available efficacy scores, the unobserved data have the same distribution as the observed data.

The PACC change score will be analyzed by an MMRM model based on observed data at on-treatment visits. The fixed terms for this model are treatment group, sex, country, APOE ε4 carrier status, education level, PACC baseline score, age at study entry, average hippocampal volume at baseline, time point (as a factor), and the interaction of treatment and time point. Subject is included as a random effect.

Comparisons between each atabecestat dose group and placebo will be performed at Months 6, 12, and 18 by using the appropriate contrasts based on this model. The Kenward-Roger approximation will be used to estimate the denominator degrees of freedom. Between-treatment differences in LS Means with the corresponding 95% confidence intervals (CIs) and p-values will be provided. If the unstructured covariance matrix results in a lack of convergence, starting values will be supplied using estimates from simpler structures. If this strategy does not result in

convergence then the following structures will be used in sequence: heterogeneous autoregressive of order 1 (ARH(1)), heterogeneous compound symmetry (CSH), and homogeneous compound symmetry (CS).

Plots of mean (±SE) PACC change score over time by treatment group will be provided for observed data and for estimates based on the MMRM.

For the off-treatment period, change of PACC and the components from off-treatment baseline will be summarized.

5.2.3.1. Subgroup Analyses

Certain subgroups of subjects, based on their baseline clinical presentation, may be more sensitive to a treatment effect than the population as a whole. PACC change score will be summarized in the following subgroups:

- APOE ε4 carriers
- APOE ε4 non-carriers

In addition, the PACC change score will also be summarized across various other subgroups of subjects (as defined in Section 2.7)

5.3. Key Secondary Endpoint

5.3.1. Definition

The key secondary efficacy endpoint is the change from baseline in the subject-reported Cognitive Function Index (CFI) total score. The CFI is a modified version of the Mail-in Cognitive Function Screening Instrument[20], a subject- and informant-reported outcome measure developed by the ADCS. The subject-reported and informant-reported CFIs are assessed on the same schedule: 5 to 10 days before randomization, then annually after randomization through Month 48, and at Month 54 (or early termination if not performed at previous visit). This assessment includes 14 questions that assess the subject's perceived ability to perform high-level functional tasks in daily life and sense of overall cognitive functional ability (yes=1; no/does not apply=0; maybe=0.5 for each question) and 1 additional question about whether the subject has seen a doctor in the last year about memory concerns. Study subjects and their informants independently rate the subject's abilities.

The CFI total score is the sum of the first 14 questions, with higher scores indicating greater impairment (range = 0 to 14). If there are fewer than 5 (<30%) missing items, an imputed total score will be calculated as a prorated total based on the observed items:

$$14 * \left(\frac{\text{Total score of non-missing items}}{\text{Total number of non-missing items}}\right)$$

Imputed total scores will be rounded to 1 decimal place. If there are 5 or more missing items, the total score will be missing. A subject-reported and an informant-reported total score will be calculated. Change from baseline in the CFI total score will be calculated at each time point. For off-treatment visits, change from the off-treatment baseline (see Section 2.4) will also be calculated.

5.3.2. Analysis Methods

Change from baseline in subject-reported CFI total score in the on-treatment period will be analyzed using the same MMRM approach as described for the primary efficacy endpoint in Section 5.2.3. For off-treatment visits, change from off-treatment baseline will be summarized. If the recall period of an off-treatment CFI assessment overlapped with the on-treatment period, then the score will be excluded from the off-treatment summary.

5.4. Other Secondary Endpoints

5.4.1. Definition

5.4.1.1. Free and Cued Selective Reminding Test (FCSRT)

The FCSRT is administered during screening, at the randomization visit, and every 6 months after randomization through Month 54 (or early termination, if the RBANS was administered at the subject's previous visit). In the FCSRT, subjects are shown pictures of 16 items in groups of 4 and the category of each item is identified (for example, "grapes" is an item in the "fruit" category). After a controlled learning period, subjects are asked to recall as many of the 16 items as they can in 90 seconds ("free recall"). For each item not freely recalled, subjects are cued by the category name and the number of items recalled in this way ("cued recall") is recorded. For remaining items not recalled, the subject is reminded of the item and category ("selective reminding"). Three trials of free and cued recall with selective reminding are conducted approximately 20 seconds apart.

Twenty to 30 minutes after the third trial of free and cued recall, the subject is asked again to recall as many of the items as they can in 90 seconds ("delayed free recall"). For each item not freely recalled, subjects are cued by the category name and the number of items recalled in this way ("delayed cued recall") is recorded.

Three versions of the FCSRT, using three different sets of items, will be administered in an alternating manner during the study.

Total free recall is the cumulative sum of free recall items over the 3 trials (range, 0-48). Total recall is the cumulative sum of free and cued recall items from the three trials (range, 0-48). Total delayed free recall is the sum of delayed free recall items (range, 0-16) and total delayed recall is the sum of delayed free recall items and delayed cued recall items (range, 0-16). Higher scores indicate better performance. Item scores are recorded on an electronic device which will provide the total scores.

The FCSRT score used in the PACC is the sum of Total recall and Total free recall (or two times the sum of free recall items plus the sum of cued recall items over the three trials).

Change from baseline at each time point will be calculated for each total score. Values and changes of each total score at each time point will also be standardized by dividing by the baseline SD. For off-treatment visits, change from the off-treatment baseline (see Section 2.4) will also be calculated.

5.4.1.2. Wechsler Memory Scale (WMS) Logical Memory Subtest

The WMS Logical Memory subtest is administered during screening, at the randomization visit, and every 6 months after randomization through Month 54 (or early termination, if the RBANS was administered at the subject's previous visit). In the WMS Logical Memory subtest, a short story is used that consists of 25 items of information. Free recall of the 25 items of information will be elicited immediately after the story is read aloud to the subject and again after approximately a 30-minute delay. The total number of items of information from the story that are recalled immediately (Immediate Paragraph Recall; range, 0-25) and after the delay interval (Delayed Paragraph Recall; range, 0-25) are recorded. Higher scores indicate better performance.

Three versions of the short story, with different items of information, will be administered in an alternating manner during the study. Item scores are recorded on an electronic device which will provide the total scores.

Change from baseline at each time point will be calculated for both the Immediate Paragraph Recall and Delayed Paragraph Recall total scores. The Delayed Paragraph Recall total score is used as the WMS test score in the PACC. Values and changes of each total score at each time point will also be standardized by dividing by the baseline SD. For off-treatment visits, change from the off-treatment baseline (see Section 2.4) will also be calculated.

5.4.1.3. Wechsler Adult Intelligence Scale – IV (WAIS-IV) Coding Subtest

The WAIS-IV Coding subtest is administered during screening, at the randomization visit, and every 6 months after randomization through Month 54 (or early termination, if the RBANS was administered at the subject's previous visit). WAIS-IV Coding is a subtest of the WAIS-IV battery and is also known as WAIS-R Digit Symbol Substitution (DSS) test. In this subtest, the subject copies symbols that are paired with numbers as quickly as possible for 120 seconds. Specifically, the top of the page has a 'key' that contains boxes with the numbers 1 to 9 paired with random symbols. The remainder of the page contains 135 boxes with numbers and blank boxes below the numbers. The participants are directed to fill in the blank boxes with the symbol corresponding to the number as quickly as possible working from left to right without skipping numbers or rows.

Three versions of the subtest, utilizing different sets of symbols, will be administered in an alternating manner during the study. Item scores are recorded on an electronic device which will provide the total scores.

The number of correctly drawn symbols completed in 90 seconds is totaled (Total Correct (90 seconds); range: 0-135) as is the number of correctly drawn symbols in 120 seconds (Total Correct (120 seconds); range: 0-135). Higher scores indicate better performance.

Change from baseline at each time point will be calculated for both total scores. The Total Correct (120 seconds) score is used as the DSS test score in the PACC. Values and changes of each total score at each time point will also be standardized by dividing by the baseline SD. For off-treatment visits, change from the off-treatment baseline (see Section 2.4) will also be calculated.

5.4.1.4. Mini-Mental State Examination (MMSE)

The MMSE is administered during screening, at the randomization visit, and every 6 months after randomization through Month 54 (or early termination, if the RBANS was administered at the subject's previous visit). It consists of 11 items (30 questions) that test 5 areas of cognitive function: orientation, registration, attention and calculation, recall, and language. Each question is rated as 1 for a correct answer or 0 otherwise. Three versions of the MMSE, which are identical except for the list of words used in the registration and recall questions, will be administered in an alternating manner during the study.

A total score is computed as the sum of individual scores. The maximum score is 30. If there are fewer than 10 (30%) missing items, an imputed total score will be calculated as a prorated total based on the observed items:

$$30 * \left(\frac{\text{Total number of correct answers}}{\text{Total number of non-missing items}}\right)$$

Imputed total scores will be rounded to 1 decimal place. If there are 10 or more missing items, the total score will be missing.

Change from baseline at each time point will be calculated for the MMSE total score. Values and changes of the MMSE total score at each time point will also be standardized by dividing by the baseline SD. For off-treatment visits, change from the off-treatment baseline (see Section 2.4) will also be calculated.

5.4.1.5. Cognitive Function Index (CFI)

See Section 5.3.1 for a description of the CFI. Changes from baseline in informant-reported CFI total score at each time point are secondary efficacy endpoints.

5.4.1.6. Clinical Dementia Rating (CDR)

The CDR is administered during screening, annually after randomization, and at Month 54 (or early termination). Based on semi-structured interviews of both the subject and the subject's informant carried out by a trained rater, 3 domains of cognition (memory, orientation, judgment/problem solving) and 3 domains of function (community affairs, home/hobbies,

personal care) are rated on a 5-point scale (no impairment = 0, questionable impairment = 0.5, mild impairment = 1, moderate impairment = 2, severe impairment = 3. Note: 'questionable impairment' is not an available choice for the personal care domain).

The CDR Sum of Boxes (CDR-SB) is the sum of the 6 domain (or "box") scores. A Global CDR is derived from the individual domain scores according to an algorithm[10] that yields a rating on a similar 5-point scale as each domain (no dementia = 0, questionable dementia = 0.5, mild dementia = 1, moderate dementia = 2, severe dementia = 3). The memory domain is the primary category driving the Global CDR. Each domain score will be recorded on an electronic device and the CDR Sum of Boxes and the Global CDR will be provided by the electronic device.

Change from baseline at each time point will be calculated for each domain score, CDR-SB, and the Global CDR. For off-treatment visits, change from the off-treatment baseline (see Section 2.4) will also be calculated.

5.4.1.7. ADCS Activities of Daily Living – Prevention Instrument (ADCS-ADL-PI)

The ADCS-ADL-PI is administered 5 to 10 days before randomization, then every 3 months through Month 54 (or early termination if not performed at the previous visit). The ADCS-ADL-PI includes 15 questions related to activities of daily living and 3 questions related to high-level function[2]. High-level function items ask about the use of electronic devices such as cellphones/smartphones, computers/tablets, and e-readers (such as a Kindle or Nook). Study subjects and their informants independently rate the subject's ability to perform each activity over the past 3 months (with no difficulty = 3, with some difficulty = 2, with a lot of difficulty = 1, did not do/don't know = 0). Informants are additionally asked to evaluate whether activities were completed less often, required more time to complete, or if any errors were made performing the task. High-level function items are rated as "yes" or "no".

The ADL total score is the sum of the scores of the 15 activities of daily living questions (range: 0-45) with higher scores indicating less impairment. If there are fewer than 5 (<30%) missing items, an imputed total score will be calculated as a prorated total based on the observed items:

$$15 * \left(\frac{\text{Total score of non-missing items}}{\text{Total number of non-missing items}}\right)$$

Imputed total scores will be rounded to 1 decimal place. If there are 5 or more missing items, the total score will be missing. Both a subject- and informant-reported ADL total score will be calculated. Change from baseline at each time point will be calculated for each activity of daily living (both subject- and informant-reported) and for the subject- and informant-reported ADL total scores. For off-treatment visits, change from the off-treatment baseline (see Section 2.4) will also be calculated.

5.4.1.8. Repeatable Battery for the Assessment of Neuropsychological Status (RBANS)

The RBANS is administered during screening prior to amyloid testing, 5 to 10 days before randomization, Month 3 after randomization, then every 6 months through Month 51 (or early termination, if the PACC components were administered at the subject's previous visit). The RBANS is a 20- to 25-minute battery developed for cognitive assessment, detection, and characterization of dementia in the elderly, as well as for neuropsychological screening for younger patients[12]. Three versions of the RBANS will be administered in an alternating manner during the study. The RBANS includes 12 subtests that measure the following 5 indices:

Immediate Memory

The Immediate Memory index is composed of the List Learning and Story Memory subtests.

In the List Learning subtest, a list of 10 words is read and the subject is asked to immediately recall as many of the words as the subject can. Four trials of the same list of words are conducted. The List Learning total score is the sum of correctly recalled words over the 4 trials (range: 0 to 40).

In the Story Memory subtest, a short story is read and the subject is immediately asked to recall as much of the story as the subject can. Two trials of the same story are conducted. The Story Memory total score is the total number of items of information (out of 12) the subject correctly recalled over the 2 trials (range: 0-24).

Based on the List Learning and Story Memory total scores, an Immediate Memory Index Score is determined from a mapping table appropriate for the subject's age. For example, for a subject between 60-69 years with a List Learning total score of 30 and a Story Memory total score of 15, the Immediate Memory Index Score is 100. The Immediate Memory Index Score ranges from 40 to 152.

Visuospatial/Constructional

The Visuospatial/Constructional index is composed of the Figure Copy and Line Orientation subtests.

In the Figure Copy subtest, subjects are shown a figure composed of 10 items and are asked to immediately copy the figure from memory. For each figure item, it is determined whether the item was correctly drawn and correctly placed. The Figure Copy total score is the sum of the number of correctly drawn and number of correctly placed items (range: 0-20).

In the Line Orientation subtest, subjects are shown a pair of lines and asked to identify the matching lines from a set of 13 lines. Ten pairs of lines are shown. The Line Orientation total score is the total number of correct responses (range: 0-20).

Based on the Figure Copy and Line Orientation total scores, a Visuospatial/Constructional Index Score is determined from a mapping table appropriate for the subject's age. For example, for a

subject between 60-69 years with a Figure Copy total score of 18 and a Line Orientation total score of 18, the Visuospatial/Constructional Index Score is 100. The Visuospatial/Constructional Index Score ranges from 50 to 131.

Language

The Language index is composed of the Picture Naming and Semantic Fluency subtests.

In the Picture Naming subtest, subjects are shown 10 pictures and asked to name the item pictured. The Picture Naming total score is the total number of correct responses (range: 0-10).

In the Semantic Fluency subtest, subjects are asked to name all of the fruits and vegetables they can think of. The Semantic Fluency total score is the number of correct responses in 1 minute, to a maximum of 40.

Based on the Picture Naming and Semantic Fluency total scores, a Language Index Score is determined from a mapping table appropriate for the subject's age. For example, for a subject between 60-69 years with a Picture Naming total score of 6 and a Semantic Fluency total score of 31, the Language Index Score is 100. The Language Index Score ranges from 40 to 134.

Attention

The Attention index is composed of the Digit Span and Coding subtests.

In the Digit Span subtest, subjects are read a string of single-digit numbers and asked to repeat the string. The test begins with 2 strings of length 2 then continues with pairs of strings of increasing length from 3 to 9. The Digit Span total score is the total number of correctly recalled strings (range: 0-16).

In the Coding subtest subjects are shown a 'key' that contains boxes with the numbers 1 to 9 paired with symbols. The remainder of the page contains 89 boxes with symbols and blank boxes below the symbols. The subjects are directed to fill in the blank boxes with the number corresponding to the symbol as quickly as possible working from left to right without skipping numbers or rows. The Coding total score is the total number of correctly entered numbers in 90 seconds.

Based on the Digit Span and Coding total scores, an Attention Index Score is determined from a mapping table appropriate for the subject's age. For example, for a subject between 60-69 years with a Digit Span total score of 13 and a Coding total score of 32, the Attention Index Score is 100. The Attention Index Score ranges from 40 to 150.

Delayed Memory

The Delayed Memory index is composed of the Delayed List Recall, Delayed List Recognition, Delayed Story Recall, and Delayed Figure Recall subtests. These tests are administered after all of the above subtests have been administered.

In the Delayed List Recall subtest, subjects are asked to recall as many of the words that were read to them in the Immediate List Recall subtest. The Delayed List Recall total score is the number of correctly recalled words (range: 0-10).

In the Delayed List Recognition subtest, subjects are read a list of 20 words and asked if the word was or was not in the list that was read to them in the Immediate List Recall subtest. The Delayed List Recognition total score is the number of correct responses (range: 0-20).

In the Delayed Story Recall subtest, subjects are asked to recall as many details of the short story that was read to them in the Immediate Story Memory subtest. The Delayed Story Recall total score is the number of the 12 items of information that are correctly recalled (range: 0-12).

In the Delayed Figure Recall subtest, subjects are asked to draw the same figure shown in the Figure Copy subtest. The Delayed Figure Recall total score is the sum of the number of correctly drawn and number of correctly placed items (range: 0-20).

Based on the sum of the Delayed List Recall, Delayed Story Recall, and Delayed Figure Recall total scores and the Delayed List Recognition total score a Delayed Memory Index Score is determined from a mapping table appropriate for the subject's age. For example, for a subject between 60-69 years with a Delayed List Recall total score of 9, a Delayed Story Recall total score of 10, a Delayed Figure Recall total score of 18 (sum = 37) and a Delayed List Recognition total score of 17, the Delayed Memory Index Score is 100. The Delayed Memory Index Score ranges from 40 to 133.

Higher scores indicate less impairment. The Sum of Index Scores is the sum of the five index scores and the Sum of Index Scores is converted to a Total Scale value via a mapping table. In the examples above, the Sum of Index Scores is 500, which is mapped to a Total Scale value of 100. The Total Scale is a norm-based t-score based on a distribution with a mean of 100 and standard deviation of 15.

The RBANS item scores will be entered on an electronic device and all subtest total scores, all index scores, the Sum of Index Scores, and the Total Scale will be provided by the electronic device.

Change from baseline at each time point will be calculated for each subtest total score, each of the 5 index scores, and the Total Scale. For off-treatment visits, change from the off-treatment baseline (see Section 2.4) will also be calculated.

5.4.1.9. Neuropsychological Assessment Battery - Daily Living Tests (NAB-DLTs)

The Daily Living tests of the Memory and Attention NAB[19] modules are administered during screening prior to amyloid testing, 5 to 10 days before randomization, and after randomization at Months 9, 21, 33, 45, and 54 (or early termination, if not performed at previous visit). Two versions of the NAB-DLTs will be administered in an alternating manner during the study.

Daily Living Memory

- **Medication Instructions – Immediate Recall (MED-irc)**: In this test, subjects are shown and read a 2-sentence medication instruction including 9 items of information and immediately asked to recall as much of the instruction as they can. Three trials are conducted. The MED-irc score is the total number of correctly recalled items over the 3 trials (range: 0-27).
- **Name/Address/Phone Number – Immediate Recall (NAP-irc)**: In this test, subjects are shown and read a name, address, and phone number, including 8 items of information, and immediately asked to recall as much of the information as they can. Three trials are conducted. The NAP-irc score is the total number of correctly recalled items over the 3 trials (range: 0-24).
- **Medication Instructions – Delayed Recall (MED-drc)**: After both immediate recall tests are administered, subjects are asked to recall as much of the medication instructions as they can. The MED-drc score is the number of correctly recalled items (range: 0-9).
- **Medication Instructions – Delayed Recognition (MED-drg)**: In this test, subjects are shown several versions of each sentence of the medication instructions and asked to identify the correct version. The MED-drg total score is the number of sentences correctly identified (0, 1, or 2).
- **Name/Address/Phone Number – Delayed Recall (NAP-drc)**: Subjects are asked to recall as much of the name, address, and phone number as they can. The NAP-drc total score is the number of correctly recalled items (range: 0-8).
- **Name/Address/Phone Number – Delayed Recognition (NAP-drg)**: In this test, subjects are shown several versions of each item of information in the name, address, and phone number and asked to identify the correct version. The NAP-drg total score is the number of items correctly identified (range: 0-8).

Daily Living Attention – Driving Scenes (DRV)

In this test, subjects are shown a picture of a driving scene as viewed from behind the steering wheel of a car. Then another version of the driving scene is shown and subjects are asked to identify any new, different, or missing items compared to the previous scene. This is repeated for 5 scenes and the DRV total score is the total number of correctly identified new, different, and missing items over all of the scenes (range: 0-70).

For all tests, higher scores indicate less impairment. DLT item scores will be recorded on an electronic device and the total scores for each test will be provided by the electronic device.

Each total score will be converted to a standardized score based on demographically corrected norms. Change from baseline at each time point will be calculated for each raw total score and each standardized score. For off-treatment visits, change from the off-treatment baseline (see Section 2.4) may also be calculated.

5.4.2. Analysis Methods

Analyses of the changes from baseline of each of the 4 PACC components, the changes from baseline of each of the 5 RBANS index scores, the change from baseline of the RBANS Total Scale, changes from baseline in informant-reported CFI total score, and change from baseline in ADCE-ADL-PI total scores (both the subject- and informant-reported) will be performed using a similar MMRM model as described above for the primary efficacy endpoint (Section 5.2.3). For each endpoint, baseline score in the model will be the baseline score for the given endpoint. For endpoints measured at unequally spaced time points, time (month) will be included as a continuous variable. For the individual PACC components, summaries based on both the original and standardized scale will be provided. For off-treatment period, changes from the off-treatment period baseline will be summarized.

For other secondary efficacy endpoints that are continuous parameters, change from baseline at each on-treatment time point will be summarized with descriptive statistics. Inferential analyses may be performed. The distribution of responses to the ADCS-ADL-PI high-level function items will be summarized at each time point by treatment group.

Plots of mean (±SE) values and changes over time by treatment group will be provided for selected secondary efficacy endpoints. Change from the off-treatment baseline may also be summarized for each off-treatment time point and plotted.

For Global CDR ≥0.5, the number and percentage of subjects meeting the criterion at any post-baseline on-treatment time point will be summarized. A similar summary based on off-treatment time points may also be provided. Descriptive statistics of the PACC change score may be summarized by treatment group for subjects who did and did not have the Global CDR ≥0.5 event at Month 12.

Association between the change from baseline in the CFI subject total score and the PACC change score, the CFI (actual) subject total score and the PACC change score, and the change from baseline in ADCS-ADL-PI subject total score and the PACC change score will be explored. Association between the functional measure and other cognitive measures will be similarly analyzed. Association between the total scores of subject-reported functional measures and the informant-reported functional measures will be explored as well.

5.5. Exploratory Efficacy Variables

5.5.1. Definition

5.5.1.1. Amyloid Disclosure Assessments

Concerns about Alzheimer's Disease Dementia

The Concerns about Alzheimer's Disease Dementia scale is administered during screening prior to and after amyloid disclosure and after randomization at Months 1, 3, 9, 21, 42, and 54 (or early termination). The scale is a short self-report instrument that assesses an individual's concern about developing Alzheimer's disease dementia. The questionnaire asks subjects to rate

their agreement with 6 statements related to concerns about AD dementia with ratings that range from 1 (strongly disagree) to 5 (strongly agree).

Change from baseline at each time point will be calculated for each statement score. For off-treatment visits, change from the off-treatment baseline (see Section 2.4) will also be calculated.

Future Time Perspective (FTP) Scale

The FTP scale[5] is administered during screening prior to and after amyloid disclosure and after randomization at Months 1, 3, 9, 21, 42, and 54 (or early termination). The FTP scale asks subjects to rate their agreement with 10 statements related to perception of time with ratings that range from 1 (very untrue) to 7 (very true). For the first 7 statements, higher scores indicate a positive future time perspective (e.g., "My future is filled with possibilities"). For the last 3 statements, higher scores indicate a negative time perspective (e.g., "There are only limited possibilities in my future"). An overall FTP score is the average of the 10 individual statement scores, after reverse-scoring the last 3 statement scores. A higher overall FTP score indicates a more positive future time perspective.

Change from baseline at each time point will be calculated for each statement score and for the overall FTP score. For off-treatment visits, change from the off-treatment baseline (see Section 2.4) will also be calculated.

5.5.1.2. Computerized Cognitive Battery

The CogState Brief Battery[6] will be administered during screening prior to amyloid testing and 5 to 10 days before randomization, and every 6 months after randomization through Month 54 (or early termination). The CogState Brief Battery includes tasks that use playing cards as stimuli: detection, identification, one card learning, and one back. For each, the test software measures the speed and accuracy of the subject's responses.

- The **detection** task is a measure of information processing speed and uses a well-validated simple reaction time paradigm. In this task, the playing cards are all jokers. The subject is asked to press the "yes" key as soon as the card in the center of the screen turns face up. The reaction time is measured in milliseconds and $\log_{10}$-transformed. The overall score is the average of all log-transformed reaction times. A lower score is better.
- The **identification** task is a measure of visual attention and uses a well-validated choice reaction time paradigm. In this task, the playing cards are all either red or black jokers. The subject is asked whether the card currently being presented in the center of the screen is red. The subject responds by pressing the "yes" key when the joker card is red and "no" when it is black. The reaction time is measured in milliseconds and $\log_{10}$-transformed. The overall score is the average of all log-transformed reaction times of correct responses. A lower score is better.
- The **one card learning** task is a measure of visual recognition memory and uses a well-validated pattern separation paradigm. In this task, the subject is asked whether the playing card presented in the center of the screen was seen previously in this task. The subject responds by pressing the "yes" or "no" key. Because no card has been presented yet, the first

response is always "no". The overall score is the arcsine transformation of the square root of the proportion of correct responses. A higher score is better.

- The **one back** task is a measure of working memory and uses a well-validated n-back paradigm. In this task, the subject is asked whether the playing card being presented is the same as the one presented immediately previously. The subject responds by pressing the "yes" or "no" key. Because no card has been presented yet, the first response is always "no". The reaction time is measured in milliseconds and $\log_{10}$-transformed. The overall score is the average of all log-transformed reaction times of correct responses. A lower score is better.

In addition to the CogState Brief Battery above, computerized testing also includes the Face Name Associative Memory Exam[14]:

- The **Face Name Associative Memory Exam** begins with an exposure in which the subject is shown 16 faces, one at a time, for 2 seconds in a random order. For each face the subject is asked to decide whether "the name goes with that face". Once this is completed, each of the 16 faces is presented again. For each face the subject must recall the name that was associated with that face on the initial trial and indicate this by typing the name into the computer. The correct number of face-name pairs is recorded as an initial learning score. After a delay the subject is shown each face in turn with 3 names given beneath that face. The subject must select the name that was initially paired with the face. The number of correct responses is recorded (range: 0 to 16; higher score is better).

Change from baseline at each time point will be calculated for each score. For off-treatment visits, change from the off-treatment baseline (see Section 2.4) may also be calculated.

5.5.1.3. Cognitive Function Index-acute (CFI-a)

The CFI-a is a modified version of the CFI, a subject and informant-reported outcome measure developed by Janssen. The subject-reported and informant-reported CFI-a's are assessed on the same schedule: 5 to 10 days before randomization, then every 3 months through Month 54 (or early termination if not performed at the previous visit). This assessment includes 14 questions similar to those in the CFI (see Section 5.4.1.5) that evaluate the subject's perceived ability to perform high-level functional tasks in daily-life and sense of overall cognitive functional ability. Study subjects and their informants independently rate the subject's abilities based on their current or most recent experience. The scale uses frequency (never = 0, rarely = 1, sometimes = 2, often = 3, or always = 4) to measure impairment, with "always" indicating greatest impairment. Responses of "I do not drive", "Does not apply", or "I do not work or volunteer" will be scored as 0.

The CFI-a total score is the sum of the 14 questions, with higher scores indicating greater impairment (range = 0 to 56). If there are fewer than 5 (<30%) missing items, an imputed total score will be calculated as a prorated total based on the observed items:

$$14 * \left(\frac{\text{Total score of non-missing items}}{\text{Total number of non-missing items}}\right)$$

Imputed total scores will be rounded to 1 decimal place. If there are 5 or more missing items, the total score will be missing. A subject-reported and an informant-reported total score will be calculated. Change from baseline in the CFI-a total score will be calculated at each time point. For off-treatment visits, change from the off-treatment baseline (see Section 2.4) will also be calculated.

5.5.1.4. Financial Capacity Instrument – Short Form (FCI-SF)

The FCI-SF is administered at selected centers during screening 5 to 10 days before randomization, annually after randomization through Month 48, and at Month 54 (or early termination, if not collected at the subject's previous visit). The FCI-SF is a 37-item psychometric instrument that measures financial capacity. Items are summed to form 5 component performance scores:

- Mental calculation (2 items, range: 0-4),
- Financial conceptual knowledge (4 items, range: 0-8),
- Single checkbook/register (10 items, range: 0-20),
- Complex checkbook/register (14 items, range: 0-28),
- Bank statement management (7 items, range: 0-14)

Correct responses are scored as 2 and incorrect responses are scored as 0 (a partial credit score of 1 is possible for some items). Each item in the mental calculation, financial conceptual knowledge, and bank statement management tasks has a time limit for providing a response. For the checkbook tasks, there is a time limit to complete all items within the task. If a response was given after the time limit, the item was scored as 0 (even if the response was correct).

An FCI-SF total score is calculated as the sum of all 37 items. A component score and the Total Score are missing if any item in the component score is missing.

The time, in seconds, to complete the items listed below is recorded. If the completion time exceeded the time limit for the item(s), the maximum allowed time was recorded.

- Medical deduction problem (maximum time: 90 seconds)
- Income tax problem (maximum time: 90 seconds)
- Single checkbook/register (maximum time to complete all items: 240 seconds)
- Complex checkbook/register (maximum time to complete all items: 300 seconds)

A checkbook/register composite time will be calculated as the sum of the single and complex checkbook/register times and a total composite time will be calculated as the sum of all 4 individual component times.

Change from baseline at each time point will be calculated for each performance and timing parameter. For off-treatment visits, change from the off-treatment baseline (see Section 2.4) may also be calculated.

5.5.1.5. Amsterdam IADL Questionnaire (A-IADL-Q)

The A-IADL-Q will be administered at selected centers only during screening, annually after randomization through Month 48, and at Month 54 (or early termination, if not collected at the subject's previous visit). The A-IADL-Q is a computerized questionnaire aimed at measuring difficulties with complex daily activities. It is completed by the informant of the subject. The A-IADL-Q consists of 70 items and for each item difficulty is rated on a 5-point scale. To optimize individual differences in premorbid IADL activities, items are tailored to individual responses. If the patient had not performed the main activity, more detailed items on that activity are skipped.

The total score is calculated using an item response theory method of scoring. Item response theory assumes that ordered-categorical item responses represent an underlying construct. In this case, the construct is IADL functioning, ranging from disability to ability. The total score is normally distributed (mean = 50, standard deviation = 10), with higher scores indicating better functioning[4].

Change from baseline at each time point will be calculated for the IADL total score. For off-treatment visits, change from the off-treatment baseline (see Section 2.4) may also be calculated.

5.5.2. Analysis Methods

All exploratory efficacy endpoints will be summarized with descriptive statistics. Inferential analyses may be performed to evaluate the treatment effect. The endpoints may also be summarized for the off-treatment period. Association between the endpoints and the cognitive endpoints will be explored.

6. SAFETY

All safety summaries will be presented for the on-treatment period and the off-treatment period,

6.1 Adverse Events

The verbatim terms used in the CRF by investigators to identify adverse events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). Any AE occurring on or after the date of the initial administration of study medication through the day of last dose plus 7 days is considered to be treatment emergent. If the event date is recorded as partial or completely missing, then the event will be considered to be treatment emergent unless it is known to be prior to the first administration of study medication based on a partial onset date or resolution date. All reported treatment-emergent adverse events will be included in the analysis. For each adverse event, the number and percentage of subjects who experience at least 1 occurrence of the given event will be summarized by treatment group. AEs in the off-treatment period are the AEs with onset more than 7 days after the last dosing.

Summary tables will be provided for:

- AEs
- AEs with a fatal outcome

- Serious AEs (SAEs)
- AEs leading to discontinuation of study medication (events with action taken of 'drug withdrawn')
- AEs leading to termination of study participation (events associated with a trial termination reason of 'adverse event')
- AEs by severity

AEs will also be summarized by age group, by gender, and by race. In addition to the summary tables, listings will be provided for subjects who:

- Died
- Had SAEs
- Had AEs leading to discontinuation of study medication or termination of study participation

The above listings will include events with onset during the screening phase.

Incidence of other treatment-emergent adverse events of special interest will be summarized. Hypopigmentation related and retinal related adverse events are identified by the investigator on the AE eCRF page and include:

- Hypopigmentation Related Adverse Events
  - Hair and skin hypopigmentation (events related to lighting hair, skin coloration, or iris coloration)

- Retinal Related Adverse Events
  - Retinal changes

Hepatic related treatment-emergent adverse events will also be summarized. Adverse events related to abnormal LFTs are defined as follows:

Events belonging to the MedDRA Drug Related Hepatic disorders SMQ – comprehensive search

In addition, other adverse event clinical MedDRA PT clusters will be summarized by treatment during the on-treatment phase and off-treatment phase of the study (including but not limited to rash related, sleep disorder related, depression related, anxiety related adverse events).

6.1. Clinical Laboratory Tests

Clinical laboratory blood samples will be collected at screening; at randomization; after randomization every 2 weeks through Month 3, then monthly through Month 6, then every 2 months through Month 12, then every 3 months through Month 54 (or early termination). Urinalysis samples will be collected at screening and after randomization at Months 12, 27, 42, and 54 (or at early termination).

Creatinine clearance will be calculated using the Cockcroft-Gault equation:

$$\frac{(140 - age) * wt}{72 * (CRT/88.4)} [* 0.85, \text{if female}]$$

Where age is age, wt is body weight in kg on the date of the laboratory sample (or on the most recent prior visit), and CRT is creatinine in µmol/L.

Descriptive statistics will be presented for all chemistry, hematology, and urinalysis (pH and specific gravity) laboratory parameters at scheduled time points for on-treatment period.

Change from baseline to each time point and the endpoint in the on-treatment period will be summarized for chemistry, hematology, and urinalysis (pH and specific gravity) parameters and displayed by treatment group.

Shift tables, by treatment group, will be provided summarizing the shift in laboratory values from baseline to each time point and to the endpoint in the on-treatment period with respect to the central laboratory reference ranges (low, normal, high).

Number and percentage of subjects with post-baseline potentially clinically important laboratory values will be presented by treatment group. Definitions of potentially clinically important laboratory values are provided in Attachment 6.

All above summaries will also be presented for the off-treatment period.

6.1.1 Liver function tests

Change from the baseline in ALT, AST, alkaline phosphatase, total bilirubin, lymphocytes, and eosinophils will be summarized. Descriptive statistics, shift tables, and percentage of subjects with potentially clinically important values based on change from the baseline will be summarized for these tests as well. Box-Whisker plots will be provided. These summaries and plots will also be presented for the off-treatment period.

The number and percentage of subjects meeting the below liver enzymes criteria during the on-treatment period will be presented by treatment group:

- ALT >8x the upper limit of normal (ULN) or AST >8x ULN.
- ALT >5x ULN or AST >5x ULN.
- Total bilirubin >2x ULN and either ALT >3x ULN or AST >3x ULN.
- Prothrombin International Normalized Ratio (INR) >1.5 and either ALT >3x ULN or AST >3x ULN.

For the off-treatment period, number and percent of subjects with values within normal range at the off-treatment baseline and met the above criteria post baseline will be summarized.
The distribution of the time to first occurrence of an ALT or AST value >ULN while on study medication will be displayed with Kaplan-Meier curves. Subjects who complete or prematurely

discontinue the study medication without an ALT or AST value >ULN will be censored and the end of treatment date will serve as the time of censoring. The distribution of the time to first occurrence of an ALT or AST value >3x ULN will be summarized in the same way. The peak TB (total bilirubin) times the upper limit of the reference range (ULRR) will be plotted against the peak ALT times the upper limit of the reference range (ULRR), on a log10 scale (eDISH plot). These time-to-event plots and eDISH plot will be presented for the off-treatment period as well.

A listing of subjects with clinically important laboratory values will be provided that displays all values over time for subjects with at least one post-baseline clinically important laboratory value.

6.2. Vital Signs

Blood pressure, pulse, and body temperature will be measured at screening; at randomization; every 3 months after randomization through Month 12, then every 6 months through Month 54 (or early termination). At each visit, blood pressure and pulse will be measured in both supine and standing positions. Weight will be measured at screening; at randomization; then every 6 months after randomization through Month 54 (or early termination). Height will be measured at screening and Month 54 (or early termination).

Continuous vital sign variables including pulse, blood pressure (systolic and diastolic), temperature, weight, and Body Mass Index (BMI) will be summarized with descriptive statistics at each on-treatment time point and for the on-treatment endpoint. BMI will be calculated as $[\text{weight (kg)} / (\text{height (m)})^2]$ at each time point that body weight is measured. The height measurement collected at screening will be used in the calculation. Changes from baseline will be summarized for each on-treatment time point and for the on-treatment endpoint. Descriptive statistics (mean, standard deviation, median, minimum and maximum) will be presented. Weight will be plotted over time for the on-treatment period and for the off-treatment period.

Incidence of treatment-emergent potentially clinically important abnormalities in vital signs while on treatment, as defined in Table 4, will be summarized for subjects who had a baseline assessment and at least one post-baseline, on-treatment assessment for that vital sign. The incidence will also be summarized for the off-treatment visits.

For off-treatment period, the vital sign variables will be similarly summarized.

Table 4: Potentially Clinically Important (PCI) Criteria for Vital Signs

Vital sign	Criteria for Low PCI Values	Criteria for High PCI Values
Systolic BP (mmHg)	< 90	>180
Diastolic BP (mmHg)	<50	>105
Pulse (bpm)	<50	>120
Temperature (C)	<36	>38
Weight (kg)	Decrease of 10% relative to baseline	Increase of 10% relative to baseline
Orthostatic hypotension	A decrease in systolic (>20 mm Hg) or diastolic (>10 mm Hg) blood pressure after standing for at least 2 minutes relative to supine position with an increase in pulse rate of >15 beats per minute.	

6.3. Electrocardiogram (ECG)

ECGs will be performed at screening; at randomization; after randomization at Month 1, Month 3, and Month 6; then, every 6 months through Month 54 (or early termination).

The ECG variables that will be analyzed are heart rate, PR interval, RR interval, QRS interval, QT interval, and QTc using the following correction methods:

- Bazett's formula: QTcB (msec) = QT (msec) / $(60/HR(bpm))^{1/2}$;
- Fridericia's formula: QTcF (msec) = QT (msec) / $(60/HR(bpm))^{1/3}$;
- Linear Derived formula: QTcLD(msec) = QT(msec) + β * (1-60/HR(bpm)), where β is the estimated slope derived from the linear regression model QT(msec) = α + β * 60/HR(bpm) using all observations prior to treatment initiation of all subjects excluding screen failures;
- Sagie formula: QTcl (msec) = QT(msec) + 154 * (1 - 60/HR(bpm)))

QTcB and QTcF will be provided by the central ECG vendor and not recalculated. Triplicate ECGs, completed within a 4-minute time period, are collected at each visit. The value for each parameter on a given date will be the average of the values obtained from the triplicate ECGs.

Descriptive statistics will be presented by treatment group for the above ECG variables at each scheduled time point and for the on-treatment endpoint. Change from baseline to each scheduled on-treatment time point and to the on-treatment endpoint will be descriptively summarized by treatment group. For off-treatment visits, change from the off-treatment baseline (see Section 2.4) will also be calculated.

The number and percentage of subjects with treatment-emergent potentially clinically important ECG values as defined in Table 5 below will be presented for each on-treatment time point by treatment group.

Table 5: Potentially Clinically Important (PCI) Criteria for Selected ECG Parameters

Parameter	Criteria for Low PCI Values	Criteria for High PCI Values
HR	< 50	> 100
PR	< 120	> 200
QRS	< 70	> 120
QT		> 500

The number and percentage of subjects with corrected QT values within each of the categories defined below will be presented for the maximum post-baseline, on-treatment value and for each on-treatment time point by treatment group.

- QTc Interval:
  - Males
    - ≤450 msec
    - >450 to ≤480 msec
    - >480 to ≤ 500 msec

  - >500 msec
- Females
  - ≤470 msec
  - >470 to ≤ 500 msec
  - >500 msec

- QTc change from baseline
  - ≤30 msec
  - >30 – 60 msec
  - >60 msec

The distribution of the investigator interpretations (not evaluable; normal; abnormal, not clinically significant; abnormal, clinically significant) will be summarized at each on-treatment time point and for the on-treatment endpoint.

In addition, the number and percentage of subjects with ECG abnormalities according to the central ECG vendor interpretation will be summarized by type of abnormality (conduction, rhythm, etc.) and specific diagnosis for the double-blind phase. Listings for subjects with a corrected QT value >450 for males or >470 for females or change from baseline >30 will be provided.

Similar analyses will be performed for the off-treatment period.

6.4. Other Safety Parameters

6.4.1. Assessment of Psychological Well Being

The Assessment of Psychological Well Being, which includes the 15-item short version of the Geriatric Depression Scale (GDS-15)[23] and the 6-item short version of the state scale of the State-Trait Anxiety Inventory (STAI)[9], will be performed during screening; at randomization; and after randomization at Month 3, Month 9, then every 3 months through Month 39, then at Months 45, 48, and 54 (or early termination).

The GDS-15 is a subset of items from the GDS-30 (see Section 4.1.6). The items included in the GDS-15 and the scoring for each item are shown in Attachment 3. The GDS-15 total score is the sum of the 15 individual item scores. Higher scores indicate greater levels of depression.

For the STAI, subjects are read 6 statements related to anxiety (“I feel calm”, “I am tense”, “I feel upset”, “I am relaxed”, “I feel content”, and “I am worried”) and asked to rate how they feel “right now, at this moment” on a 4-point scale (“not at all” = 1, “somewhat” = 2, “moderately” = 3, “very much” = 4). The STAI total score is the sum of the individual scores after reverse-scoring the “calm”, “relaxed”, and “content” items. Higher scores indicate more anxiety.

Item scores for both scales are recorded in an electronic device which will provide the total scores for both scales.

Change from baseline will be calculated for both total scores at each time point. Values and changes from baseline at each time point will be summarized for both total scores with descriptive statistics by treatment group. For off-treatment visits, change from the off-treatment baseline (see Section 2.4) will also be calculated and summarized with descriptive statistics.

6.4.2. Columbia Suicide Severity Rating Scale (C-SSRS)

The C-SSRS is conducted during screening; at randomization; and after randomization every 2 weeks through Month 3, then every month through Month 6; at Months 8, 9, 10, and 12; then every 3 months through Month 54 (or at early termination). The C-SSRS is a measure of the spectrum of suicidal ideation and behavior that was developed in the National Institute of Mental Health Treatment of Adolescent Suicide Attempters Study to assess severity and track suicidal events through any treatment[11]. It is a semi-structured clinician-administered questionnaire designed to solicit the occurrence, severity, and frequency of suicide-related ideation and behaviors during the assessment period.

The following are C-SSRS categories and have binary responses (yes/no). Each category is based on a direct question in the C-SSRS. The categories are ordered by increasing seriousness.

Suicidal Ideation (1-5)

- Category 1: Wish to be dead
- Category 2: Non-specific active suicidal thoughts
- Category 3: Active suicidal ideation with any methods (not plan) without intent to act
- Category 4: Active suicidal ideation with some intent to act, without specific plan
- Category 5: Active suicidal ideation with specific plan and intent

Suicidal Behavior (6-10)

- Category 6: Preparatory acts or behavior
- Category 7: Aborted attempt
- Category 8: Interrupted attempt
- Category 9: Non-fatal suicide attempt
- Category 10: Completed suicide

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

The Suicidal Ideation Score at a time point is the maximum suicidal ideation category with a "yes" response at that time point. A score of 0 is assigned if no suicidal ideation is present. The Suicidal Ideation Score ranges from 0 to 5.

The number and percentage of subjects with "yes" responses will be summarized for each category at each time point by treatment group. In addition, the number and percentage of subjects with any suicidal ideation, with any suicidal behavior, with either suicidal ideation or

behavior, and with any non-suicidal, self-injurious behavior will also be summarized at each time point and over the on-treatment period by treatment group.

The number and percentage of subjects meeting the following criteria will be summarized by treatment group:

- An increase in Suicidal Ideation Score from baseline at any post-baseline, time point in the // on-treatment period (treatment-emergent suicidal ideation compared to baseline).
- An increase in Suicidal Ideation Score from <4 at baseline to ≥4 at any post-baseline, time point in the on-treatment period (treatment-emergent serious suicidal ideation compared to baseline).
- An increase in Suicidal Ideation Score from 0 at baseline to ≥4 at any post-baseline, time point in the on-treatment period (emergence of serious suicidal ideation compared to baseline).

In these summaries, the denominator is the number of subjects meeting the baseline criterion.

Listings including additional C-SSRS details will be created for subjects with any suicidal ideation or behavior.

The above summaries will be presented for off-treatment time points, using the off-treatment baseline as the reference time point.

6.4.3. MRI (Magnetic Resonance Imaging)

All subjects will have an MRI during screening and after randomization at Months 6, 12, 24, 36 and 54 (or at early termination, if not performed at the previous visit). MRI image evaluation will be performed by a core imaging laboratory.

The number and percentage of subjects with evidence of a finding at any time during the double-blind period and at each time point will be summarized by treatment group for each of the following:

- ARIA-E (Amyloid-related imaging abnormalities which represent vasogenic edema and related extravasated fluid phenomena), including parenchymal hyperintensity, sulcal hyperintensity, and gyral swelling.
- ARIA-H (Amyloid-related imaging abnormalities which represent microhemorrhages and hemosiderin deposits), including microhemorrhages (mH), macrohemorrhages, and superficial siderosis.
- Lacunar infarct
- Territorial infarct outside of the cerebellum
- Full territory cerebellar infarct

Similar summaries for ARIA-E and ARIA-H finding will be provided. A listing of the subjects with any of the findings listed above will be provided.

Similar analyses will be performed for the off-treatment period.

6.4.4. Neurological Examination

A neurological examination will be performed during screening; at randomization; after randomization at Month 6 and Month 12, then annually through Month 48, and at Month 54 (or early termination). The neurological examination includes an evaluation of mental status, cranial nerves, motor ability (including strength, tone, and involuntary movements), coordination (including finger-to-nose, gait, and postural reflexes), and sensation (including proprioception, cold, light touch, and deep tendon reflexes). The number and percentage of subjects with abnormal findings will be summarized by category and by treatment group for both on and off-treatment periods.

6.4.5. Physical Examination

Physical examination will be performed during screening; at randomization; after randomization at Month 6 and Month 12, then annually through Month 48, and at Month 54 (or early termination). The number and percentage of subjects with abnormal findings will be summarized by body system and by treatment group for both on and off-treatment periods.

6.4.6. Dermatological Examination

A dermatological examination will be performed during screening and after randomization at Month 12 (or early termination if no Month 12 examination).

The distribution of the dermatological examination result (normal; abnormal, not clinically significant; abnormal, clinically significant) will be summarized at screening, Month 12, and double-blind phase endpoint. A listing of subjects with changes from normal at screening to abnormal during the double-blind phase will be provided and will include the description of the abnormality.

7. PHARMACOKINETICS/PHARMACODYNAMICS

7.1. Pharmacokinetics

Blood samples for pharmacokinetics (PK) will be collected at randomization and after randomization at Month 3, Month 6, and then every 6 months until Month 54. Collection at randomization is pre-dose, collection at post-randomization time points is both pre-dose and within 1 to 4 hours post-dose. For subjects participating in CSF collection, samples are collected during screening and post-randomization at Month 12 and Month 36 (and early termination if not collected at the previous visit). Concentrations below the limit of quantification (LOQ), will be treated as zero. No analysis is planned for PK parameters.

7.2. Pharmacodynamics

Biomarkers measured in this study include the following:

- Fluid biomarkers

- CSF: A subset of subjects will have longitudinal CSF samples taken during the study at randomization (if a CSF screening sample was taken, it was not necessary to have a repeat sample at randomization), Month 12, Month 36 (or early termination if not collected at the previous visit). After the biomarkers interim analyses (Section 3.1.1), the Month 12 sample could be taken any time between baseline and Month 36. The key parameter of interest is $A\beta_{1-40}$. $A\beta_{1-37}$, $A\beta_{1-38}$, $A\beta_{1-42}$, p-tau, and t-tau will also be measured.

- Imaging biomarkers
  - Amyloid PET: A subset of subjects will have longitudinal amyloid PET scans during the study at randomization (if an amyloid PET scan was performed during screening, it was not necessary to have a repeat scan at randomization), Month 24 and Month 48 (or early termination). The parameters of key interest are the composite region of interest (ROI) SUVr and regional SUVr's.

  - MRI: All subjects will have an MRI during screening and after randomization at Months 6, 12, 24, 36 and 54 (or early termination, if not performed at the previous visit). The parameters of key interest are brain volumes of the cortical and sub-cortical regions.

Biomarker analyses will be based only on those subjects with at least 85% treatment compliance (see Section 4.3) at the analysis time point.

Changes from baseline and percent changes from baseline will be calculated for all parameters.

Values of each parameter will be summarized with descriptive statistics at each time point by treatment group. Change from baseline will also be summarized for each parameter, at each time point by treatment group. Percent change from baseline in CSF and plasma $A\beta_{1-40}$ will be summarized at each time point by treatment group. Plots of mean (±SE) values, changes, and/or percent changes over time by treatment group will be provided for selected biomarkers.

For off-treatment visits, change and percent change from the off-treatment baseline (see Section 2.4) may also be calculated and summarized with descriptive statistics.

Exploratory evaluations of an association between biomarker data and changes in cognitive endpoints may be performed graphically with: 1) plots of biomarker results over time with mean (±SE) change score of the cognitive endpoints over time overlaid; 2) scatterplots of the change score of the cognitive endpoints versus baseline values of selected biomarkers by treatment group; and 3) scatterplots of the change scores of the cognitive endpoints versus changes in biomarkers by treatment group at corresponding time points.

7.3. Pharmacokinetic/Pharmacodynamic Relationships

No analysis is planned for PK/PD relationships,

8. HEALTH ECONOMICS AND HEALTH OUTCOMES

Only on-treatment analyses will be performed for health economics and health outcomes.

8.1. SF-36 Health Survey

The SF-36 Health Survey is a self-administered, generic, 36-item questionnaire designed to measure 8 domains of functional health status and well-being. It is collected during screening and after randomization at Months 12, 27, 42, and 54 (or early termination).

Eight subscales are calculated from the 36 items: physical functioning (PF), role functioning (RF), bodily pain (BP), general health (GH), vitality (VT), social functioning (SF), role emotional (RE), and mental health (MH). Higher scores indicate better health. Following the algorithm by Ware, et al[21], the subscales are sums of selected items in the survey. The sums are then transformed to a 0-100 scale and the transformed values are converted to z-scores using the US population mean and standard deviation for the subscale. Finally, the z-scores are converted to norm-based scores by multiplying by 10 and adding 50 (this gives the corresponding quantile for the z-score from a normal distribution with mean 50 and standard deviation 10).

Two additional subscales, the Physical Component Summary (PCS) and Mental Component Summary (MCS), will be calculated as weighted linear combinations of the 8 subscale z-scores with weights as defined by Ware, et al[21]. The resulting value is also a z-score which is converted to a norm-based score as described above.

Derivation of the 8 subscales and 2 summary scores will be performed using Quality Metrics' proprietary software.

Change from baseline will be calculated for each of the 8 subscales and the 2 summary scores at each time point. Values and changes from baseline at each time point will be summarized for each SF-36 subscale and the 2 summary scores with descriptive statistics by treatment group.

8.2. EQ-5D-5L

The EQ-5D-5L (EuroQol – 5 dimensions – 5 levels) is a self-administered, standardized measure of health status. The EQ-5D-5L is collected during screening and after randomization at Months 12, 27, 42, and 54 (or early termination).

The EQ-5D-5L includes 5 items that each rate a single dimension of health status (Mobility, Self-Care, Usual Activities, Pain/Discomfort, and Anxiety/Depression) on a 5-point scale with 1 = no problems, 2 = slight problems, 3 = moderate problems, 4 = severe problems, and 5 = unable or extreme problems. Subjects select an answer for each of the 5 dimensions considering the response that best matches their health "today". The frequency distribution of scores for each item will be summarized at each time point by treatment group.

EQ-5D-5L also includes a visual analog scale (VAS) on which subjects rate their health from 0 = "Worst imaginable health state" to 100 = "Best imaginable health state". Change from baseline will be calculated for the VAS. Values and changes from baseline at each time point for the VAS will be summarized with descriptive statistics by treatment group.

8.3. Healthcare Resource Utilization Questionnaire (HRUQ)

A Healthcare Resource Utilization Questionnaire (HRUQ) will be administered during screening, after randomization at Months 1, 2, and 3, and then every 3 months through Month 54 (or early termination). At each visit whether or not a subject was hospitalized, visited an emergency room (ER), required the service of a day hospital, required the service of an adult day center, or had an outpatient consultation related to the subject's disease or treatment since the subject's last visit (or in the previous 6 months for the screening visit) will be recorded. Additional information about each healthcare resource utilized will be collected:

- Hospitalizations: Type of hospital, type of ward, reason for hospitalization (psychiatric, neurologic, social reason, or other medical reason), and start and end dates.
- ER visits: Reason for visit (psychiatric, neurologic, social reason, or other medical reason) and date of visit.
- Day hospital: Type of hospital, frequency/week, reason for visit (psychiatric, neurologic, social reason, or other medical reason), and start and end dates.
- Adult day center: Frequency/week, reason for visit (psychiatric, neurologic, social reason, or other medical reason), and start and end dates.
- Outpatient consultation related to the subject's disease or treatment: Kind of healthcare professional consulted (psychiatrist, neurologist, home care nurse, primary care physician, social worker, personal care assistant, other) and number of consultations.

The number and percentage of subjects with any healthcare resource use and with each type of healthcare resource use in the recall period will be summarized at each time point by treatment group. The number and percentage of subjects with new hospitalizations (with a start date after the previous visit) will be summarized at each time point by treatment group. Number of ER visits, frequency of use of a day hospital, frequency of use of an adult day center, and number of outpatient consultations will be summarized at each time point with descriptive statistics by treatment group.

8.4. Accommodation and Occupational Status

Subjects' usual accommodation status and current occupational status will be collected during screening. Changes in usual accommodation status or current occupational status, and the reason for the change (psychiatric, neurologic, social reason, other), will be collected after randomization at Months 1, 2, and 3, and then every 3 months through Month 54 (or early termination). Choices of accommodation status are: at home, alone; at home, with family or friends; homeless; psychiatric institution; sheltered living; prison; or other. Choices of occupational status are: full-time employment or gainfully self-employed; part-time employment; casual employment; sheltered work; unemployed, but seeking work; unemployed, but not seeking work; retired; housewife or dependent husband; or other.

For subjects who are employed (full-time employment or gainfully self-employed; part-time employment; casual employment; or sheltered work), whether or not the subject lost any working days because of the subject's disease since the last visit (or in the previous 6 months at

screening) and, if so, how many, will be recorded. For unemployed subjects, the number of weeks unemployed in the recall period will be recorded.

The number and percentage of subjects who changed accommodation status from living at home (alone or with family or friends) at the previous visit to not living at home will be summarized at each time point by treatment group. The number and percentage of subjects who changed occupational status from employed (as defined above) to unemployed (unemployed and either seeking or not seeking work) and to retired will be summarized at each time point by treatment group. The number of working days lost and the number of weeks unemployed will be summarized at each time point with descriptive statistics by treatment group.

REFERENCES

1. Data Monitoring Committee Charter, 54861911ALZ2003. Janssen Research & Development (29 August 2016).
2. Galasko D, Bennet DA, Sano M., et. al. ADCS Prevention Instrument Project: Assessment of instrumental activities of daily living for community-dwelling elderly individuals in dementia prevention clinical trials. Alzheimer Disease & Associated Disorders; 2006; 20:S152-S169.
3. Horowitz M, Wilner N, Alvarez W. Impact of Event Scale: A measure of subjective stress. Psychosomatic Medicine; 1979; 41:209-218.
4. Koster N, Knol DL, Uitdehaag BM, Scheltens P, Sikkes SA. The sensitivity to change over time of the Amsterdam IADL Questionnaire. Alzheimer's & Dementia; 2015; 11:1231-1240.
5. Lang FR, Carstensen LL. Time counts: future time perspective, goals and social relationships. Psychology and Aging; 2002; 17:125-139.
6. Lim YY, Ellis KA, Harrington K, Ames D, Martins RN, Masters CL, Rowe C, Savage G, Szoeke C, Darby D, Maruff P & the AIBL Research Group. Use of the CogState Brief Battery in the assessment of Alzheimer's disease related cognitive impairment in the Australian Imaging, Biomarkers and Lifestyle (AIBL) study. Journal of Clinical and Experimental Neuropsychology; 2012; 34:345-358
7. Lu K, Luo X, Chen PY. Sample size estimation for repeated measures analysis in randomized clinical trials with missing data. International Journal of Biostatistics; 2008; 4:Article 9.
8. Marson DC, Sawrie SM, Snyder S, et. al. Assessing financial capacity in patients with Alzheimer disease: A conceptual model and prototype instrument. Archives of Neurology; 2000; 57:877-84.
9. Marteau T, Bekker H. The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). British Journal of Clinical Psychology; 1992; 31:301-306.
10. Morris J. The Clinical Dementia Rating (CDR): Current version and scoring rules. Neurology; 1993; 43:2412-2414.
11. Posner K, Brown GK, Stanley B, et. al. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. American Journal of Psychiatry; 2011; 168:1266-1277.
12. Protocol, REGISTRYALZ0001. Cognitive Health in Ageing Register: Investigational, Observational, and Trial studies in dementia research (CHARIOT): Prospective Readiness Cohort Study (PRO). Janssen Research & Development (30 June 2016).
13. Randolph, C. Repeatable Battery for the Assessment of Neuropsychological Status (RBANS); 1998; San Antonio: Harcourt, TX: The Psychological Corporation.
14. Rentz, DM, Amariglio RE, Becker JA, et. al. Face-name associative memory performance is related to amyloid burden in normal elderly. Neuropsychologia; 2011; 49:2776-2783.
15. Rosen, WG, Terry RD, Fuld PA, Katzman R, & Peck A. Pathological verification of ischemic score in differentiation of dementias. Annals of Neurology; 1980; 7:486-488.
16. Rowe CC, Ellis KA, Rimajova M, et al. Amyloid imaging results from the Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging. Neurobiology of Aging; 2010; 31:1275-1283.
17. SAS Institute Inc. SAS/STAT ® 9.2 User's Guide, Second Edition. Cary, NC.; 2009;.3839-3890.
18. Sperling RA, Aisen PS, Beckett LA, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & Dementia; 2011; 7:208-292.
19. Stern RA, & White T. Neuropsychological Assessment Battery. In L. FL, & F. L. Lutz, Psychological Assessment Resources; 2003.
20. Walsh SP, Raman R, Jones KB, & Aisen PS. ADCS Prevention Instrument Project: the Mail-In Cognitive Function Screening Instrument (MCFSI). Alzheimer Disease & Associated Disorders, 2006; 20:S170-S178.

21. Ware J, et al. User's Manual for the SF-36v2 Health Survey (Second Edition). Lincoln, RI: QualityMetric Incorporated, 2007.

22. Yesavage JA, Brink TL, Rose TL, et. al. Development and validation of a geriatric depression screening scale: a preliminary report. Journal of Psychiatric Research; 1982-3; 7:37-49

23. Yesavage J. Geriatric depression scale. Psychopharmacology Bulletin; 1988; 24:709-711.

ATTACHMENTS

Attachment 1: Statistical Operational Plan (access to study data and flow of information)

The operational procedures between the independent external Data Monitoring Committee (DMC)/Statistical Support Group (SSG) and Janssen Research and Development (JRD) to perform interim data monitoring and interim analyses are described below including the roles and responsibilities for each party involved in the data flows.

Of the interim analyses planned in the protocol, only the futility interim analysis of $A\beta_{1-40}$ in CSF was performed because of the cessation of screening, randomization, and dosing.

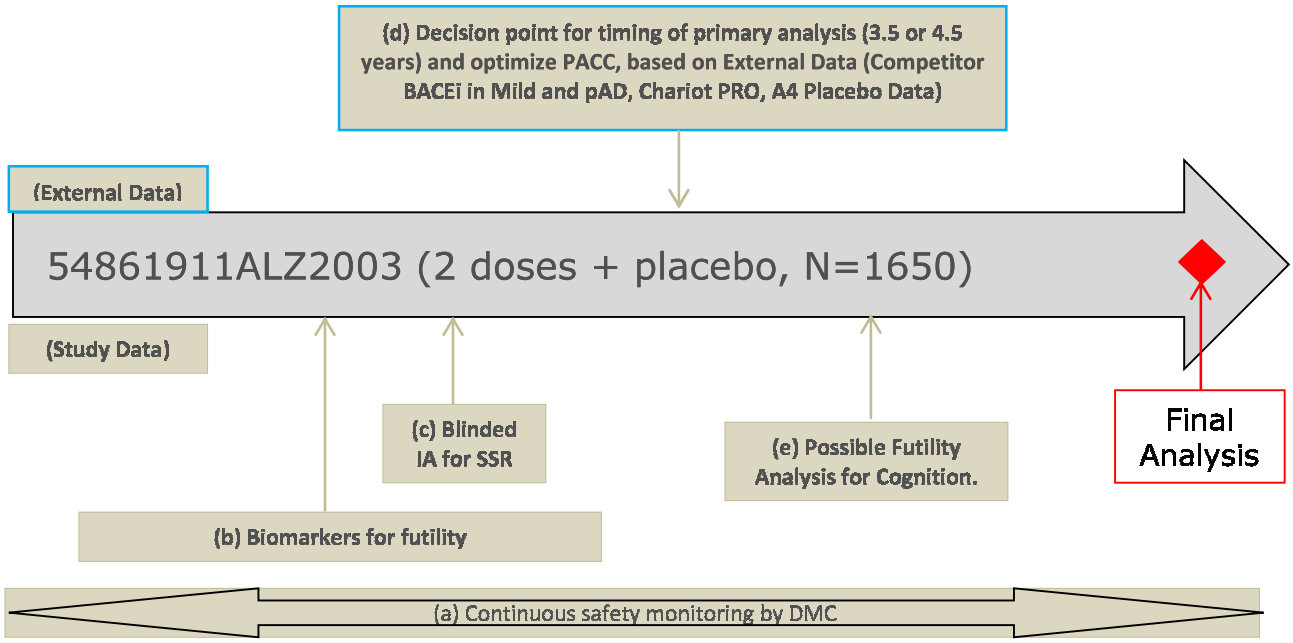
The operational plan describes how it will be ensured that no post-baseline data from Study 54861911ALZ2003 (ALZ2003) will be used in the evaluation of external data for potential optimization of the primary efficacy endpoint, changing the time point of the primary efficacy analysis, or changing or removing the key secondary efficacy endpoint.

The operational plan also protects against operational bias by not allowing access to blinded, post-baseline cognitive data to those individuals with access to potentially unblinding safety data who interact frequently with investigators and investigative site staff to discuss patient safety. At the same time, certain R&D operations functions will have access to all blinded data to allow for quality study monitoring, data management, and analysis planning.

Certain internal and/or external functions are permitted to be unblinded to randomization group assignment under conditions defined in Janssen R&D Standard Operating Procedures (SOPs) (e.g., Global Medical Safety, Clinical Supply, Bioanalytical Laboratories). Unblinded access for these functions is governed and documented according to those SOPs and is not covered in this operational plan.

Figure A1 below depicts the general timing of interim blinded and unblinded data monitoring and interim analyses.

Figure A1: Chronologic Summary of Potential Decisions



Key: pAD = prodromal Alzheimer's disease; SSR = sample size re-estimation

The level of access to ALZ2003 study data depends on the activity an individual performs. Groups of individuals defined by their activity and level of access are described in the table below. An 'X' indicates that the group has access to the data in that column, otherwise the group does not have access to that data.

Group		Activity	Access to ALZ2003 Data[a]					
			Unblinded	Blinded				
			Any subject level data	Subject level baseline data (cognitive, safety, and biomarker)	Subject level, post-baseline biomarker data	Subject level post-baseline cognitive data	Subject level post-baseline safety data	QC summaries of aggregated cognitive data
A	External data statisticians/clinicians	Optimization of the primary efficacy endpoint; change/remove key secondary endpoint; estimate placebo decline; decide on mitigations due to non-compatible PACC translations. All based on sources external to ALZ2003		X[c]			X	X
B	Independent external Data Monitoring Committee(DMC)/Statistical Support Group (SSG)	Safety monitoring; conduct interim analyses of biomarkers and cognitive endpoint (if performed) and make recommendations based on prospectively planned criteria	X[b]					
C	Independent external Secure Data Office	Receive unblinded randomization and unblinded biomarker results from vendors; transfer unblinded data to Group D and SSG. No involvement in the conduct of the study	X					
D	Biomarker scientists	Outlier detection/quality review of biomarker assay results		X	X			
E	Sample size re-estimation statistician (one individual)	JRD statistician independent of the study team: Calculate sample size increase using variance estimate provided by Group F and revised estimate of placebo decline provided by Group A.		X			X	
F	Study team statisticians/data managers/programmers	Analysis planning; data quality review; estimate variance of PACC from aggregated study data (efficacy and safety)		X		X	X	X
G	Cognitive data central monitoring	Quality review of cognitive scale data		X		X		X
H	Study team clinicians/Trial management	Data quality review; trial conduct monitoring; medical monitoring; site interaction		X			X	X
I	Clinical research associates/Medical science liaisons	On-site source document verification		X		X[d]	X	

Table footnotes

[a] **Statistical programming support may be shared among Groups A, E, F, and G.**
[b] **Per the DMC Charter, the JRD Sponsor Committee may request unblinded results from the DMC Chair, but this would only be expected if the DMC recommends a major change in the study conduct or design.**
[c] **Group A will not have access to baseline PACC, RBANS, CFI, or ADCS-ADL-PI data.**
[d] **Group I will not have access to post-baseline PACC, RBANS, or CDR scores, but may have access to secondary cognitive and functional measures collected on the eCRF at their assigned sites.**

Definitions

- 'Unblinded/Blinded' refers to access to randomization group assignments. Only Groups B and C will have access to unblinded data.
- Cognitive data includes: PACC components (FCSRT, WAIS, WMS, MMSE), CDR, RBANS, CFI, ADCS-ADL-PI, CFI-a, FCI, NAB-DLT, A-IADL-Q, CogState Battery.
- QC summaries of blinded, aggregated cognitive data (which could incorporate post-baseline data) would include descriptive statistics (n, mean, standard deviation) and/or proportions only. Summaries of post-baseline will not describe a longitudinal trend. Summaries of post-baseline data shared with Group A will not include any information that could identify a site or subject.

The appropriate level of access is maintained by the following measures:

- Management of randomization and biomarker data, including documentation of what data is blinded, is governed by trial-specific Data Transfer Agreements (tsDTAs) between the vendors and the independent external Secure Data Office (Group C) and between the vendors and IQVIA data management.
- The flow of randomization codes and biomarker results between the independent external Secure Data Office (Group C) and the independent external DMC/SSG (Group B) is documented in the DMC Charter. Transfer of data between the Secure Data Office and the DMC/SSG is via a shared location that only Groups B and C have access to.
- The subject level, post-baseline biomarker data provided to Biomarker scientists (Group D) by the vendor will be anonymized by replacing ALZ2003 subject identifiers with dummy identifiers and Group D will not have access to subject-level, post-baseline ALZ2003 data. Group D will not share subject level, post-baseline biomarker data with any other group.
- Read access to ALZ2003 study data on JRD-managed platforms is restricted to Study team statisticians/data managers/programmers (Group F). Group F will provide screening/baseline summaries to other groups, provide post-baseline cognitive data to Cognitive data central monitoring (Group G) via secure communication (e.g., password-protected files), and provide the estimated standard deviation of the PACC to the SSR statistician (Group E). Group F will not share data with any other group that that group is not allowed access to.
- Groups A, E, and H (External data statisticians/clinicians, SSR statistician, Study team clinicians/trial management, respectively) are prevented from viewing subject-level, post-baseline cognitive data in the following ways:
  - Group G (cognitive data central monitoring) will provide only QC summaries of aggregated cognitive data to Group H (study team clinicians/trial management).
  - They do not have read privileges to ALZ2003 subject-level data on JRD-managed platforms.
  - They do not have access to cognitive data that may be available in any vendor's portal. Only members of Group G will have user accounts on such portals. Cognitive data will not be available in any data visualization tool members of Groups A, E, and H have

access to. Reports generated by vendors that may list subject-level, post-baseline cognitive data will not be provided to Groups A, E, or H.

  - They do not have access to eCRF pages on which cognitive data is recorded.

- Clinical research associates and Medical science liaisons (Group I) are responsible for only a select few sites each and are not involved in any decision related to the modification of the primary outcome.

Individual members of each group (from JRD, IQVIA, and ATRI) will be identified in separate documentation. They will be trained on this operational plan and records of that training will be maintained.

Documentation indicating that the above measures were implemented will also be maintained.

Further details about the interim analyses and study-external analyses are outlined below.

Continuous Safety Monitoring:

Timing	Quarterly throughout the double-blind portion of the study
Data	Limited to safety data (blinded or potentially unblinded): AE, clinical lab tests, electrocardiograms (ECG), vital signs, physical examination, Magnetic Resonance Imaging (MRI), and other safety data
Conduct the monitoring	Group B: Independent external statistical support group (SSG) will prepare the safety monitoring reports. Independent external Data Monitoring Committee (DMC) will review the data and make recommendations.
Potential recommendations	1. Continue the study unmodified 2. Continue the study with modifications 3. Discontinue or halt the study
Approval of the recommendations and documentation	The sponsor committee (SC) will decide whether to adopt the recommendations. The SC is not involved in the day-to-day conduct of the study.
Dissemination of the decision	Decisions (2) and (3) will be communicated to the study team, the study sites, and all necessary parties.

Interim analysis of biomarkers:

CSF Aβ

Timing	Expected to occur around 2 years after First Subject Dosed (FSD), when 60 subjects (20/group) have 12 month CSF Aβ data.
Data	unblinded CSF and plasma Aβ data and baseline characteristics
Conduct the IA	Group B: Independent external SSG and DMC
Potential recommendations	Stop the study due to futility
Approval of the recommendations and documentation	The sponsor committee (SC) will decide whether to adopt the recommendations. The SC is not involved in the day-to-day conduct of the study.
Dissemination of the decision	If no action required, communication will be limited to key members of the study team. No further communication unless the decision is to stop the study.

Other biomarkers

Timing	Expected to take place between 2-3 years after first subject dosing and possibly again approximately 3 months before the tentative cognitive endpoint IA. • Amyloid PET: approximately 168 subjects (56 /group) with 2-year data. • CSF p-tau: approximately 56 subjects/group with 1 year CSF p-tau data. • Tau PET: approximately 56 subjects/group with 18 months data. The number of subjects may change depending on the emerging external data. A second IA could include biomarker results from subsequent time points (e.g., Month 48 amyloid PET).
Data	Unblinded biomarker data and baseline characteristics
Conduct the IA	Group B: Independent external SSG and DMC
Potential recommendations	<u>First IA</u>: • Conduct a subsequent additional futility analysis on biomarkers. • Do not conduct a subsequent additional futility analysis on biomarkers or a futility analysis on cognitive endpoints. <u>Second IA:</u> • Conduct a subsequent futility analysis on cognitive endpoints. • Do not conduct a subsequent futility analysis on cognitive endpoints. • No recommendation if no biomarker is technically feasible.
Approval of the recommendations and documentation	The sponsor committee (SC) will decide whether to adopt the recommendations. The SC is not involved in the day-to-day conduct of the study.
Dissemination of the decision	Key members of the study team will be informed.

Sample size re-estimation:

Timing	During the enrollment period
Data	Blinded, aggregated study data: PACC, the revised PACC, RBANS, and dropout rate
Conduct the IA	Group E: A JRD statistician independent of the study team, will perform the SSR assessment using the provided information: • Group F (Study team statisticians/data managers/programmers) will obtain standard deviation estimates from blinded, aggregated study data and provide to Group E. • Group A (External data statisticians/clinicians) will provide a revised estimate of placebo decline, based on external data, to Group E.
Potential recommendations	No sample size increase or increase by x number of subjects.
Approval of the recommendations and documentation	The JRD study team and JRD management will approve the sample size adjustment recommendation.
Dissemination of the decision	If JRD study team and JRD management decide to increase the sample size (with enrollment capped at 2,400 subjects) as specified in the protocol, the SC, DMC, IWRS, drug supply, and key GCO personnel will be informed. There will be no communication to the IRBs or the sites since this is not a change to the protocol.

Modify time point of primary efficacy analysis; modify components and/or component weights of PACC score based on external data:

Optimize PACC

Timing	During the study period, before the unblinded futility IA of the cognitive endpoints if sufficient external data is available.
Data	External data only
Conduct the assessment	Group A: Janssen R&D statisticians and clinicians and external experts with limited access to ALZ2003 study data (access to baseline data, safety data, and outcome measure QC findings).
Potential recommendations	Switch to the optimized PACC
Approval of the recommendations and documentation	Group A
Dissemination of the decision	The JRD study team, SC, and DMC will be informed. There will be no further communication.

Modify time point of primary efficacy analysis

Timing	During the study period, before the unblinded futility IA of the cognitive endpoints if sufficient external data is available.
Data	External data only
Conduct the assessment	Group A: Janssen R&D statisticians and clinicians and external experts with limited access to ALZ2003 study data (access to baseline data, safety data, and outcome measure QC findings).
Potential recommendations	Shorten the timing of primary endpoint to 3.5 years treatment, and plan for an early filing
Approval of the recommendations and documentation	Group A
Dissemination of the decision	Regardless of the decision, there will be no communications to the study sites. The SC and DMC will be notified.
The logistics in the event of an early filing	• A separate detailed operational plan will be in place for the logistics of the early filing. • Two teams will be formed. All filing related activities (i.e. prepare the database, the analyses, and the dossier) will be handled by the filing team. A separate team will assume the execution of the remaining on-going study. This team will remain blinded. The filing team will be unblinded. However, they will not be involved in the conduct of the remaining on-going study. • There will be no changes to the conduct of the study due to the filing. The remaining on-going study will still be conducted blindly according to the protocol. All subjects will still be treated as planned for 4.5 years before they can enter the long-term follow-up trial.

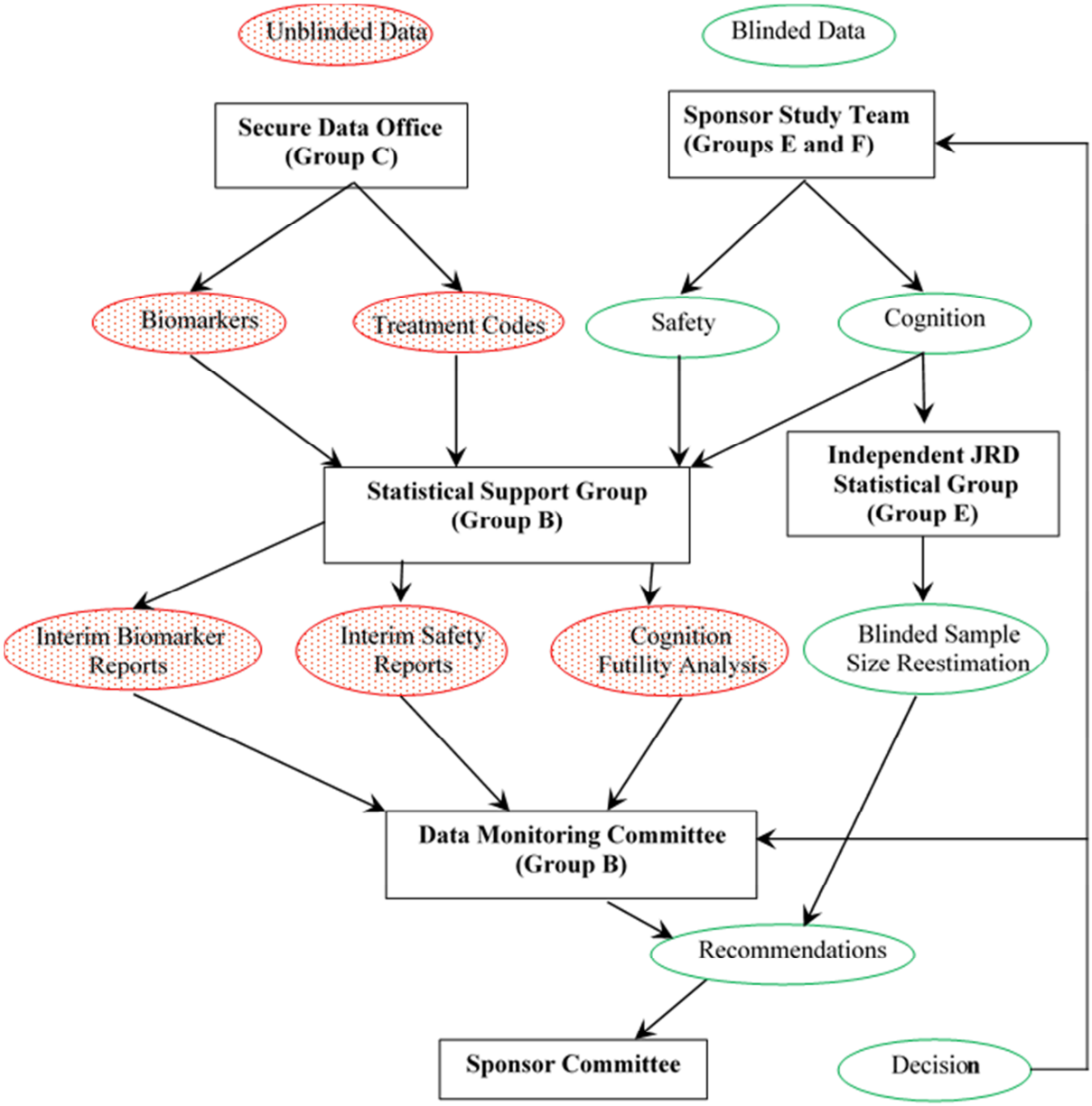
Futility interim analysis of cognitive endpoint:

Timing	If triggered by biomarker IA, may occur at 40 months after first subject dosing, or a later time. Will require at least 240 subjects/group with 4-year treatment.
Data	Unblinded study data: PACC, RBANS, other cognitive endpoints, and baseline characteristics.
Conduct the IA	Group B: Independent external SSG and DMC
Potential recommendations	• Stop the study due to lack of efficacy • Continue the study with no changes
Approval of the recommendations and documentation	The sponsor committee.
Dissemination of the decision	Key members of the study team will be informed. No further communication unless the decision is to stop the study.

The separation of responsibilities, data flows, and which groups have access to blinded and unblinded data are shown in Figure A2 below for the interim data monitoring and interim analyses performed using ALZ2003 study data.

Procedures for maintaining the integrity of blinded data and reports, and for providing interim data monitoring/interim analysis documentation after study completion, are outlined in the DMC Charter. The SSG is responsible for adhering to these procedures.

Figure A2: Data Flow Between DMC and Sponsor

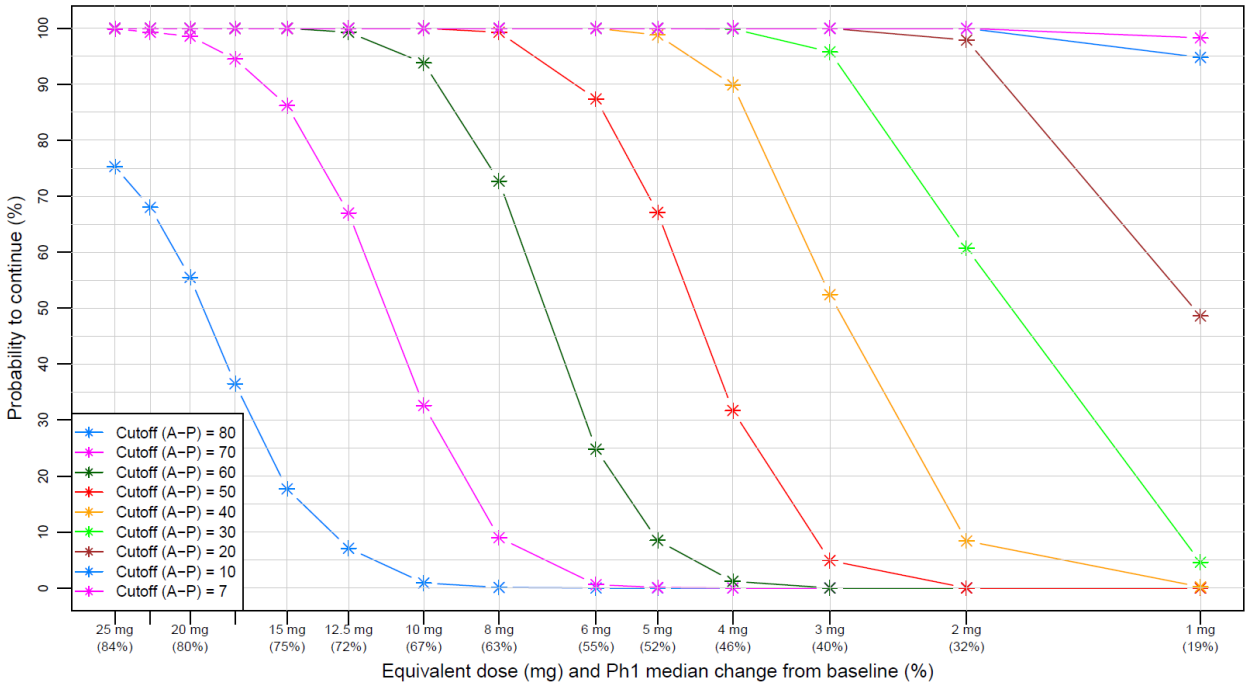


Attachment 2: Sample Size Justification for Biomarker Interim Analyses

CSF $A\beta_{1-40}$ interim analysis

Twenty subjects per group was chosen to provide adequate control of false positive and false negative error for the interim futility analysis, assuming the SD in each treatment is 20% or less, as shown in the Figure A1 below.

Figure A1: Operating characteristics of the CSF $A\beta_{1-40}$ futility criterion

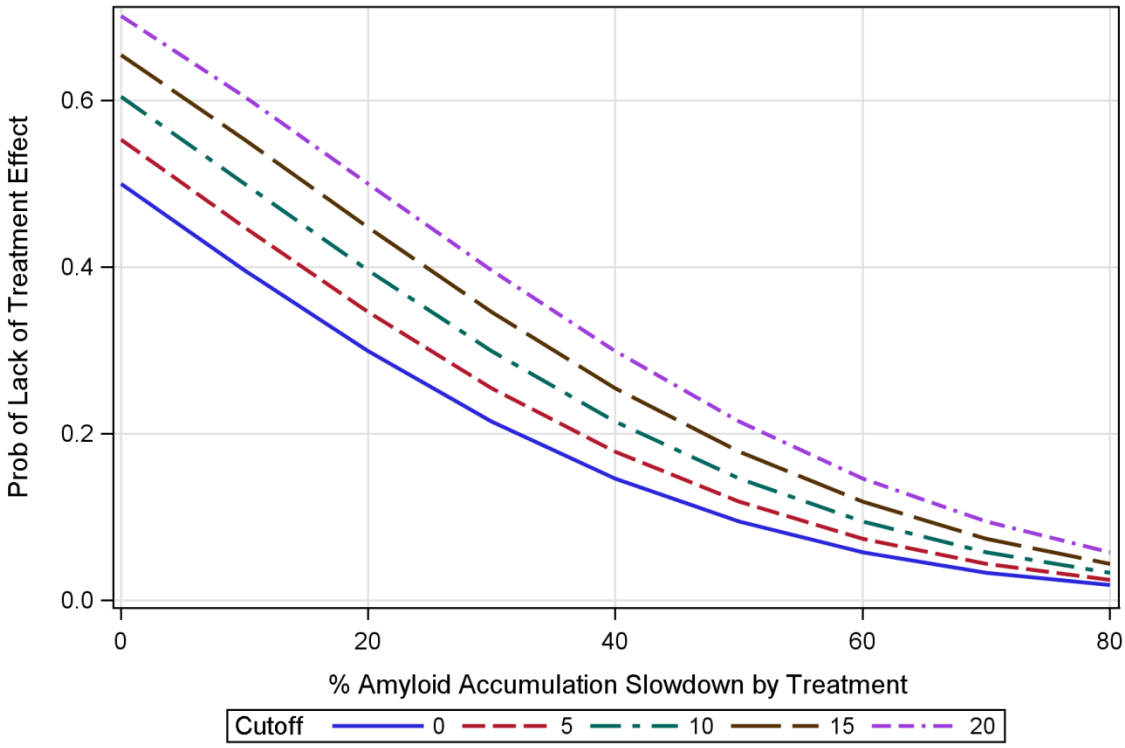


Assumptions:
N=20 per group, SD=20%.

Interim analysis of other biomarkers

In the latest Alzheimer's Disease Neuroimaging Initiative (ADNI) database, the (Mean/SD) ratio for change from baseline in amyloid accumulation at Month 24 was around 0.5 in ADNI subjects with positive AV45 amyloid PET baseline status (see **Error! Reference source not found.**). Assuming amyloid PET accumulation in 24 months, tau PET accumulation in 18 months and CSF p-tau increase in 12 months in the study population are similar to the amyloid accumulation data in ADNI, a total of 168 subjects in the IA dataset for each of the biomarkers (56 subjects per group) are expected to provide adequate control of false positive and false negative error for the interim analysis as shown in the Figure A2 below.

Figure A2: Operating characteristics of the futility criterion for amyloid PET



Assumptions:

(Mean/ SD) ratio of change in placebo is 0.5.

2-sample t-test with N=56 per group

% Amyloid Accumulation Slowdown = (LS mean change in pbo – LS mean change in active)/ LS mean change in pbo x100%

Prob (lack of effect) = Prob (% amyloid accumulation slowdown < cutoff | Assumed % amyloid accumulation slowdown)

Attachment 3: Geriatric Depression Scale (GDS)

Item (**bolded items** belong to the 15-item GDS short version)	Scoring	
	Yes	No
Are you basically satisfied with your life?	**0**	**1**
Have you dropped many of your activities and interests?	**1**	**0**
Do you fell that your life is empty?	**1**	**0**
Do you often get bored?	**1**	**0**
Are you hopeful about the future?	0	1
Are you bothered by thoughts you can't get out of your head?	1	0
Are you in good spirits most of the time?	**0**	**1**
Are you afraid that something bad is going to happen to you?	**1**	**0**
Do you feel happy most of the time?	**0**	**1**
Do you often feel helpless?	**1**	**0**
Do you often get restless and fidgety?	1	0
Do you prefer to stay at home, rather than going out and doing new things?	**1**	**0**
Do you frequently worry about the future?	1	0
Do you feel you have more problems with memory than most?	**1**	**0**
Do you think it is wonderful to be alive now?	**0**	**1**
Do you often feel downhearted and blue?	1	0
Do you feel pretty worthless the way you are now?	**1**	**0**
Do you worry a lot about the past?	1	0
Do you find life very exciting?	0	1
Is it hard for you to get started on new projects?	1	0
Do you feel full of energy?	**0**	**1**
Do you feel that your situation is hopeless?	**1**	**0**
Do you think that most people are better off than you are?	**1**	**0**
Do you frequently get upset over little things?	1	0
Do you frequently feel like crying?	1	0
Do you have trouble concentrating?	1	0
Do you enjoy getting up in the morning?	0	1
Do you prefer to avoid social gatherings?	1	0
Is it easy for you to make decisions?	0	1
Is your mind as clear as it used to be?	0	1

Attachment 4: Impact of Events Scale

Item	Subscale	
	Intrusion	Avoidance
I thought about it when I didn't mean to.	X	
I avoided letting myself get upset when I thought about it or was reminded of it.		X
I tried to remove it from my memory.		X
I had trouble falling asleep or staying asleep because of pictures or thoughts about it that came to my mind.	X	
I had waves of strong feelings about it.	X	
I had dreams about it.	X	
I stayed away from reminders of it.		X
I felt as if it hadn't happened or wasn't real.		X
I tried not to talk about it.		X
Pictures about it popped into my mind.	X	
Other things kept making me think about it.	X	
I was aware that I still had a lot of feelings about it, but I didn't deal with them.		X
I tried not to think about it.		X
Any reminder brought back feelings about it.	X	
My feelings about it were kind of numb.		X

Attachment 5: Rules for Partial Concomitant Medication Dates

Impute incomplete concomitant medication start date:

- If both the year and month are known but the day is missing, impute the day by the first day of the month.
- If only the year is known, impute the month and day by January 1st.
- If the year is missing, then the date will be considered missing.

Impute incomplete concomitant medication end date:

- If both the year and month are known but the day is missing, impute the day by the last day of the month.
- If only the year is known, impute the month and day by December 31st.
- If the year is missing, then the date will be considered missing.

Attachment 6: Potentially Clinically Important Criteria for Laboratory Tests

Test	Criteria for Low PCI Values	Criteria for High PCI Values
Hematology		
WBC Leukocytes (x10E9/L)	<3	>16
Eosinophils (x10E9/L)	NA	> 0.6
Lymphocytes (x10E9/L)	<0.8	NA
Monocytes (x10E9/L)	NA	> 1.0
Neutrophils, Segmented (x10E9/L)	<1.5	> 8.0
Platelets (x10E9/L)	< 75	> 700
Hematocrit (fraction)	<0.3	> 0.5 females, > 0.55 males
Hemoglobin (g/L)	value <100 or change from baseline <= - 20	>165 females, >185 males
MCV (fL)	< 75	>105
Chemistry		
Serum Albumin (g/L)	< 30	NA
Alkaline Phosphatase (U/L)	NA	> 3 ULN
ALT (U/L)	NA	> 3 ULN
AST (U/L)	NA	> 3 ULN
Bicarbonate (mmol/L)	< 17	> 32
Bilirubin (umol/L)	NA	> 3 ULN
BUN (mmol/L)	NA	> 1.5 ULN
Calcium (mmol/L)	< 2.0	> 2.7
Chloride (mmol/L)	< 94	> 112
Creatine Kinase (U/L)	NA	> 1.5 ULN
Creatinine (umol/L)	NA	> 1.5 ULN
GGT (U/L)	NA	> 3 ULN
Glucose (mmol/L)	< 3.9	> 5.6
Phosphate (mmol/L)	< 0.7	> 1.7
Potassium (mmol/L)	< 3	> 5.5
Protein (g/L)	< 60	> 80
Sodium (mmol/L)	< 135	> 145
Urate (umol/L)	< 150	> 500
Urinalysis		
pH	< 4	> 8
Specific Gravity	NA	> 1.035