[redacted]

Statistical Analysis Plan

ALZ-801-AD301
Alzheon, Inc.

A PHASE 3, MULTICENTER, RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED STUDY OF THE EFFICACY, SAFETY AND BIOMARKER EFFECTS OF ALZ-801 IN SUBJECTS WITH EARLY ALZHEIMER'S DISEASE AND APOE4/4 GENOTYPE

Phase 3

Original Protocol v1.0: 24 Dec 2020
Protocol v2.0 (Amendment 1): 9 Feb 2021
Protocol v3.0 (Amendment 2): 24 Jan 2022

Prepared by: [redacted]

Version: 2.0 (Final for IND submission)

Date: 4 APRIL 2024

[redacted]

SIGNATURE PAGE

Statistical Analysis Plan

ALZ-801-AD301

A PHASE 3, MULTICENTER, RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED STUDY OF THE EFFICACY, SAFETY AND BIOMARKER EFFECTS OF ALZ-801 IN SUBJECTS WITH EARLY ALZHEIMER'S DISEASE AND APOE4/4 GENOTYPE

Phase 3

Original Protocol v1.0: 24 Dec 2020
Final Protocol v2.0 (Amendment 1): 9 Feb 2021
Final Protocol v3.0 (Amendment 2): 24 Jan 2022

This Statistical Analysis Plan has been reviewed and approved by:

[redacted] Date

Chief Medical Officer
Alzheon, Inc.

[redacted] Date

Chief Development Officer
Alzheon, Inc.

[redacted] Date

Senior Statistical Advisor
[redacted]

[redacted]

TABLE OF CONTENTS

[redacted]

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

[redacted]

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Definition
Aβ	Amyloid-β or beta amyloid peptide
AChEI	Acetyl cholinesterase inhibitors
AD	Alzheimer's disease
ADAS-Cog	Alzheimer's Disease Assessment Scale – cognitive subscale
ADAS-Cog 11	ADAS-Cog 11 items
ADAS-Cog 13	ADAS-Cog 13 items
AE	Adverse event
A-IADL	Amsterdam Instrumental Activities of Daily Living
A-IADL-IRT	A-IADL item response theory
A-IADL-W	A-IADL weighted average scores
ALZ-801	Pro-drug of tramiprosate; tramiprosate conjugated to valine
APOE	Apolipoprotein E
APOE4	ε4 allele of apolipoprotein E gene
APOE4/4	Homozygosity for ε4 allele of apolipoprotein E gene
APOE4 carrier	Subject who is either heterozygous or homozygous for the ε4 allele of apolipoprotein E gene
ARIA-E	Amyloid related imaging abnormalities with vasogenic edema
ARIA-H	Amyloid-related imaging abnormalities due to microhemorrhages or hemosiderin deposition
BID	Twice per day
BMI	Body mass index
BP	Blood pressure
C-SSRS	Columbia-Suicide Severity Rating Scale
CBL	Change from baseline
CDR	Clinical Dementia Rating
CDR-SB	Clinical Dementia Rating– Sum of Boxes
CDR-G	Clinical Dementia Rating– Global Score
CP	Completer population
CSF	Cerebrospinal fluid
CT	Computed tomography
DAD	Disability Assessment for Dementia
DSMB	Data and Safety Monitoring Board
ECG	Electrocardiogram
EOS	End of study
EOT	End of treatment
ET	Early termination
FSH	Follicle-stimulating hormone
GFAP	Glial fibrillary acidic protein
hCG	human chorionic gonadotropin
HCV RNA	Hepatitis C virus RNA
Hep B S Ag	Hepatitis B surface antigen
Hep C Ab	Hepatitis C antibody

Abbreviation	Definition
HIV	Human immunodeficiency virus
ICH	International Council for Harmonisation
ITT	Intent-to-Treat
LP	Lumbar puncture
MCI	Mild cognitive impairment
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed-effects model repeated measures
MMSE	Mini-Mental State Examination
MRI	Magnetic resonance imaging
NfL	Neurofilament light
NIA-AA	National Institute on Aging-Alzheimer's Association
NPI	Neuropsychiatric Inventory
NPS	Neuropsychiatric symotoms
OC(s)	Observed cases
p-tau, P-tau	Phosphorylated tau
p-tau$_{181}$	Tau phosphorylated at threonine 181
p-tau$_{217}$	Tau phosphorylated at threonine 217
PET	Positron emission tomography
PK	Pharmacokinetic(s)
PPP	Per protocol population
PR interval	Time from the onset of the P wave to the start of the QRS complex
QoL-AD	Quality of Life in Alzheimer's Disease
QRS duration	Interval from the beginning of the Q wave to the termination of the S wave, representing the time for ventricular depolarization
QT interval	Interval representing the time for both ventricular depolarization and repolarization to occur
QTc	corrected QT (interval)
QTcF	corrected QT interval using Fridericia's correction
RBANS	Repeatable Battery for the Assessment of Neuropsychological Status
RDO	Retrieveddropout
RUD Lite	Resource Utilization in Dementia, Lite version
SAE	Serious adverse event
SAP	Statistical analysis plan
SD	Standard deviation
3-SPA	3-Sulfopropanioc acid
TBD	To be determined
TEAE	Treatment-emergent adverse event
t-tau, T-tau	Total tau
vMRI	volumetric magnetic resonance imaging

[redacted]

1 INTRODUCTION

This document provides a detailed description of the statistical methods and procedures to be implemented during the analysis of Study ALZ-801-AD301 (APOLLOE4). The proposed methods and approaches describe the main planned data analyses. If the data suggest that alternate methods may be more appropriate, then deviations from this plan will be considered. However, any deviations from this Statistical Analysis Plan (SAP) must be substantiated by sound statistical rationale and documented in the final clinical study report.

2 STUDY OBJECTIVES

The objective of this Phase 3 study is to prospectively evaluate clinical efficacy and the long-term safety and tolerability of ALZ-801 over 78 weeks of treatment compared to placebo. The study will also evaluate volumetric imaging endpoints, and fluid biomarker effects of oral ALZ-801 over 78 weeks of treatment as exploratory outcomes.

2.1 Primary Objectives

Primary Clinical

- To evaluate the efficacy of oral ALZ-801 on cognition in subjects with Early Alzheimer's disease (AD) who are homozygous for the ε4 variant of the Apolipoprotein E gene (APOE4 homozygous or APOE4/4) using ADAS-Cog 13
- To evaluate the safety and tolerability of ALZ-801 over 78 weeks in Early AD subjects with the APOE 4/4 genotype

2.2 Secondary Objectives

Key Secondary Clinical

- To evaluate the effects of ALZ-801 on a functional (activites of daily living) and a composite cognitive/functional outcome:
  - Amsterdam Instrumental Activities of Daily Living weighted average score (A-IADL-W)
  - Clinical Dementia Rating– Sum of Boxes (CDR-SB)

Other Secondary Clinical

- To evaluate the effects of ALZ-801 on additional measures of cognition:
  - The 11-item ADAS-cog (ADAS-Cog 11)
- To evaluate the effects of ALZ-801 on a measure of disability:
  - Disability Assessment for Dementia (DAD)

- To evaluate the effects of ALZ-801 on neuropsychiatric symptoms of AD:
  - Neuropsychiatric Inventory (NPI) (12-item form)
- To evaluate the effects of ALZ-801 on an additional measure of cognition:
  - Mini-Mental State Examination (MMSE)
- To evaluate the effects of ALZ-801 on quality of life or resource usage:
  - Quality of Life in Alzheimer's Disease (QoL-AD)
  - Resource Utilization in Dementia, Lite version (RUD Lite)

Pharmacokinetics – PK Objectives

- To analyze plasma and CSF levels of ALZ-801 and its metabolites and to build a population PK model of ALZ-801 in this AD population
- To evaluate correlations of pharmacokinetic (PK) measures to clinical efficacy and safety outcomes
- To explore correlations of PK measures to exploratory imaging and fluid biomarker outcomes

Exploratory Imaging Biomarkers

- To evaluate the effects of ALZ-801 on hippocampal volume using volumetric Magnetic Resonance Imaging (vMRI)
- To evaluate the effect of ALZ-801 on cortical thickness using vMRI
- To evaluate the effect of ALZ-801 on ventricular volume using vMRI

Exploratory Fluid Biomarkers

- Plasma Biomarkers: To evaluate the effects of ALZ-801 on plasma biomarkers of core AD pathology (Aβ and phosphorylated tau) and of astrocyte activation in all subjects
  - Aβ42, Aβ40, Aβ42/40 Ratio
  - Tau phosphorylated at threonine 181 (p-tau$_{181}$), and at threonine 217 (p-tau$_{217}$)
  - Astrocyte activation: Glial fibrillary acidic protein (GFAP)
- CSF Biomarkers: To evaluate the effects of ALZ-801 on CSF biomarkers of core AD pathology in the CSF substudy
  - CSF Aβ42, Aβ40, Aβ42/40 Ratio
  - CSF p-tau$_{181}$ and p-tau$_{217}$
- Other Biomarkers: To evaluate the effects of ALZ-801 on plasma and CSF biomarkers of synaptic toxicity, neurodegeneration, and neuroinflammation

- Synaptic toxicity: CSF neurogranin
- Neurodegeneration in AD: neurofilament light (NfL) in plasma and CSF
- Neurodegeneration in AD: CSF total tau (t-tau)
- Neuroinflammation: soluble triggering receptor expressed on myeloid cells 2 (sTREM2; microglia) and YKL-40 (astrocytes) in CSF

2.3 Additional (Exploratory) Biomarker and Imaging Objectives

- To evaluate the effects of ALZ-801 on new plasma and CSF biomarkers that may emerge as biologically relevant in AD including, but not limited to, Aβ oligomers, inflammatory markers (for example: cytokines, complement), new markers of vascular pathology or activated blood brain barrier (astrocytes, pericytes, endothelium), or levels of other misfolded proteins.
- To evaluate the effect of ALZ-801 on whole brain volume using vMRI
- To explore the effects of ALZ-801 on Diffusion Tensor Imaging (DTI) signals on MRI.

3 STUDY OVERVIEW

3.1 Study Design

This is a Phase 3, multicenter, randomized, double-blind, placebo-controlled, parallel-group, two-arm study, with a treatment duration of 78 weeks (18 months). Randomization will be stratified by use of concomitant AD medications (AChEI or none), age (50-65 years inclusive, or >65 years), gender, and by disease stage at screening (low versus high MMSE groups). Based on the screening MMSE, it is expected that approximately 60% will be AD subjects with MMSE 22-26 (inclusive), and approximately 40% will be AD subjects with MMSE 27-30 (inclusive).

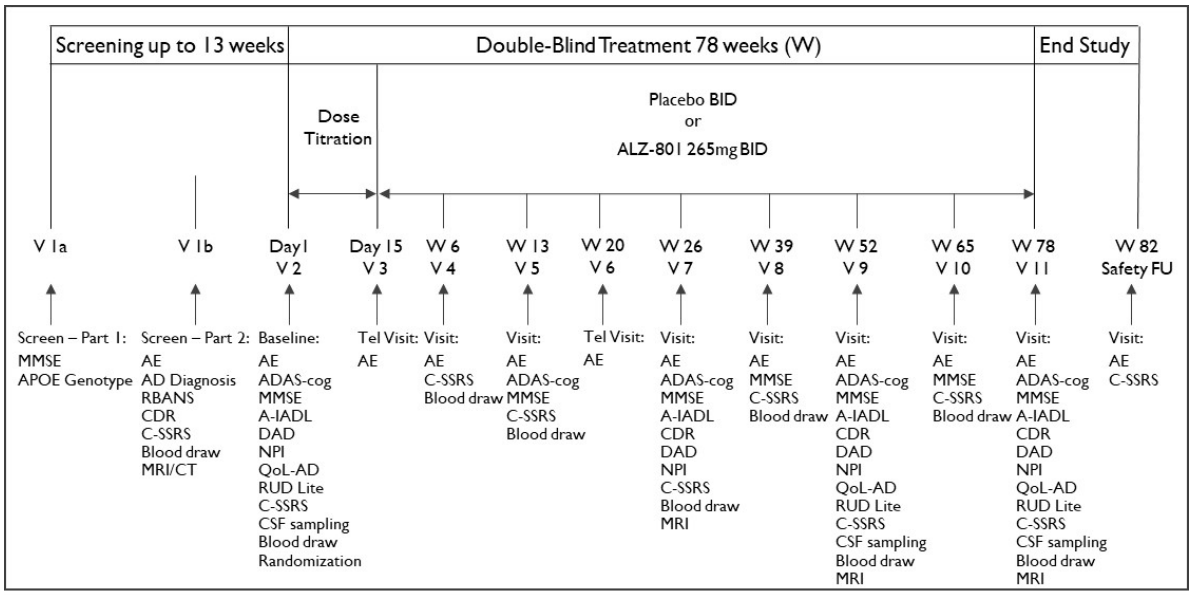
3.2 Study Procedures and Visit Structure

The study visit structure is shown in Figure 1. After the two screening visits (V1a and V1b), Day 1 of dosing is at Visit 2 (Baseline). The last visit on blinded study drug is at Week 78 or Visit 11. Subjects who complete Week 78 visit may enroll, if eligible, in a long-term extension study (Study ALZ-801-AD351 or LTE study). Subjects who enroll in the LTE study may not complete the Safety Follow up Visit and their end of study will be Week 78 or Visit 11.

The study procedures and schedule of procedures are detailed in Appendix 1.

[redacted]

Figure 1 Study Schema of Study ALZ-801-AD301: Overview of Study Structure and Visits



AD: Alzheimer's Disease; ADAS-Cog: Alzheimer's Disease Assessment Scale – cognitive subscale; AE: adverse event; A-IADL: Amsterdam Instrumental Activities of Daily Living; APOE: apolipoprotein E; BID: twice per day; CDR: Clinical Dementia Rating; CSF: cerebrospinal fluid; C-SSRS: Columbia-Suicide Severity Rating Scale; CT: computed tomography (not allowed for subjects at sites in Germany); DAD: Disability Assessment for Dementia; ECG: electrocardiogram; FU: follow-up; MMSE: Mini-Mental State Exam; MRI: magnetic resonance imaging; NPI: Neuropsychiatric Inventory; PK: pharmacokinetics; QoL-AD: Quality of Life in Alzheimer's Disease; RBANS: Repeatable Battery for the Assessment of Neuropsychological Status; RUD Lite: Resource Utilization in Dementia, Lite version; Tel: telephone; V: visit; W: week

3.3 Randomization Schedule and Blinding Procedures

All subjects will be randomized to receive double-blind treatment consisting of either ALZ-801 265mg BID or matched placebo with similar appearance. Subjects will be randomized in a 1:1 ratio to either active or placebo arms. Subjects and their study partners, study personnel at the clinical sites and all Sponsor staff will remain blinded to treatment assignement till after database lock and unblinding. Double-blind treatment will be dispensed in blister packs. Details regarding the randomization stratification can be found in Section 3.1.

During the conduct of the study, the personnel who perform the unblinded safety analysis for the Data and Safety Monitoring Board (DSMB) will be shielded from the study team by a firewall and other appropriate standard operating procedures. These steps are further described and documented in the DSMB charter.

Subjects whose treatment assignement is unblinded for safety purposes will be listed and the reasons for unblinding will be documented prior to database lock and unblinding.

In the event that prescreening/baseline data were mistakenly entered into the Interactive Response Technology for randomization and stratification purposes, a listing of all such mis-stratifications will be provided. Using the intent-to-treat (ITT) principle, efficacy analyses for such subjects will be performed "as randomized". For the statistical models used for analyses, the "as randomized" variables corresponding to the stratification factors will be used as

covariates in the model. As sensitivity analyses, the efficacy analyses will also be performed based on correct stratification factors and the correct variables included as covariates in the model.

For safety analyses, the correct stratification factors will be used to classify subjects in the actual stratification categories or subgroups.

3.4 Study Drug Dosing Regimen

The dosing regimen of ALZ-801 265mg oral tablets in the active treatment arm includes a 2-week titration from 1 tablet administered QD, followed by 1 tablet administered BID for the remainder of the study.

Blinded study drug is administered as an oral tablet beginning at the Baseline Visit (Day 1). The first dose of study drug is administered in the clinic after all pre-dose assessments (including the blood draw) are performed.

The dosing regimen of blinded study drug is one tablet orally BID except on clinic days when one tablet only is administered at the study site. In the initial 2-week titration period, the active treatment group receives placebo in the morning and ALZ-801 265mg in the evening. Thereafter, ALZ-801 265mg is administered in the morning and evening. In the placebo group, subjects always receive placebo in the morning and evening throughout the study. The last dose of study drug will be administered at the clinic on the morning of the Week 78 visit.

4 POWER AND SAMPLE SIZE

4.1 Primary Clinical Outcome

The target study sample size was 300 subjects randomized or 150 subjects per treatment group, based on ADAS-Cog 13, the primary clinical outcome. The study completed enrollment with 325 randomized subjects. The ADAS-Cog 13 is a well-validated and widely-used measure of global cognition in AD observational studies and Phase 3 drug trials. This scale evaluates multiple cognitive domains including memory, language, praxis, orientation, and executive function. To improve the sensivity of ADAS-Cog in MCI patients, two additional items were added to the original 11-item scale, including delayed word recall and digit cancellation (Mohs 1997).

Study Sample Size

Sample size determination was based on ADAS-cog results from an AD trial with tramiprosate, the active agent in ALZ-801. These efficacy data were observed in Mild AD patients (baseline MMSE =22-26) with APOE4/4 genotype who received tramiprosate 150mg BID over 78 weeks (Abushakra 2017). The tramiprosate 150mg BID dose provides plasma exposures similar to ALZ-801 at 265mg BID, the dose used in the current study.

In the 78-week tramiprosate Phase 3 study, the APOE4/4 Mild AD patients showed approximately 5.0-point difference from placebo (benefit) on ADAS-Cog. Since this Phase 3

study will include a larger number of sites across a larger number of countries, it is expected that the observed difference will be approximately 3.0 points, instead of 5.0 points.

For the primary efficacy outcome of ADAS-Cog 13, this study is powered to detect a 3.0-point difference between the active ALZ-801 treatment and the placebo in the CBL to Week 78. An increase in ADAS-Cog scores denotes cognitive worsening. It is expected that the mean baseline ADAS-Cog 13 in both groups will be 22.0 (standard deviation [SD] 6.7), which is expected to increase by 6 points in placebo (CBL: 6.0, SD 5.6), and 3 points in the active drug arm (CBL: 3.0, SD 8.1). A sample size of 125 completed subjects per arm will provide > 90% power to detect a difference of 3 points on ADAS-Cog. The drop-out rate is estimated to be approximately 17% at 78 weeks, thus a total of 300 subjects enrolled would provide approximately 250 completers (125 per arm).

Recent trials of anti-amyloid antibodies in Early AD have shown ADAS-Cog effects of ≤ 2.0 points at 78 weeks (Tolar 2020). A Phase 3 trial with aducanumab showed ADAS-cog effects versus placebo of ~ 1.8 points in APOE4 carriers (Budd-Haeberlein 2022). A Phase 2 trial with donanemab in Early AD showed ADAS-Cog effects of ~ 2.0 points (Mintun 2021). A Phase 3 trial of the approved amyloid antibody lecanemab in Early AD showed ADAS-Cog benefit of 1.44 points versus placebo (Van Dyck 2023). Therefore, we also considered a more conservative scenario where drug-placebo difference on ADAS-Cog is 2.5 points. The sample size of 125 completed subjects per arm (assuming 17% dropout rate) provides approximately 80% power to detect a difference of 2.5 points between the active treatment and the placebo at a significance level of $\alpha = 0.05$ (2-sided).

4.2 Exploratory Imaging Biomarker Outcomes

MRI Sub-study: All subjects will be included in the MRI sub-study with serial imaging at baseline, 26, 52 and 78 weeks; approximately 250 of the 300 enrolled subjects are expected to provide MRIs at the final Week 78 visit. For total hippocampal volume (HV), reduction of HV atrophy of ~20% may be expected with ALZ-801. This is based on tramiprosate Phase 3 data that suggested significant reduction of HV atrophy over 78 weeks (Gauthier 2009). In addition, the Phase 3 study of lecanemab (Van Dyck 2023) showed significant reduction of HV atrophy of ~8% over 78 weeks. No formal statistical powering was conducted for this exploratory outcome.

4.3 Exploratory Fluid Biomarker Outcomes

Exploratory Plasma Biomarker Outcomes

Plasma biomarkers are performed in all subjects. For plasma p-tau$_{181}$, a difference (reduction) of ~15% in CBL may be expected between ALZ-801 and placebo, similar to the effects seen in the lecanemab clinical trials (Leqembi FDA label 2023). No formal statistical powering was conducted for this outcome.

Exploratory CSF Biomarker Outcomes

The CSF sub-study target enrollment is a total of 120 subjects who are to provide serial CSF samples at 52 and 78 weeks, with approximately 96 subjects expected to provide CSF at Week 78 study endpoint. Early AD studies with the approved anti-amyloid antibodies (aducanumab and lecanemab) showed significant (13% to 22%) reduction in CSF p-tau$_{181}$ over 78 weeks. This CSF sub-study sample size is similar to that of the aducanumab Phase 3 studies (Budd-Haeberlein 2022). No formal statistical powering was conducted for this outcome.

5 STUDY ENDPOINTS

Rationale for Choice of ADAS-Cog Primary Outcome

In this Early AD population, the choice of ADAS-Cog 13 test as primary outcome is consistent with current FDA guidance on drug development for Early AD (FDA Draft Guidance 2018; FDA Draft Guidance 2024). ADAS-Cog is a widely used scale to assess global cognition in AD, and is the main cognitive outcome in aducanumab and lecanemab pivotal trials.

Rationale for Choice of Key Secondary Clinical Outcomes

To support the clinical relevance of drug effects on ADAS-Cog, two key secondary outcomes are selected to evalulate drug effects on daily function and a composite cognitive-functional outcome. The Amsterdam-instrumental activites of daily living (A-IADL) evaluates daily function and instrumental activites that are relevant to MCI and Mild AD subjects (Dubbelman 2022a; Dubbelman 2022b; Teng 2023). CDR-SB is a clinician's interview-based assessment of 3 cognitive and 3 functional items (Morris 1993; O'Bryant 2008). In Early AD trials of the approved anti-amyloid antibodies (lecanemab and aducanumab), both instrumental ADL scales (or ADCS-MCI-ADL) and CDR-SB have shown sensitivity to drug effects. The Disability Assessment for Dementia (DAD), a well-validated functional measure of disability has been used in Mild to Moderate AD trials (Gélinas 1999). Its sensitivity to drug effects in MCI subjects in Early AD trials is less clear. DAD was therefore moved from a key secondary outcome in the Protocol version 3.0 to a secondary outcome in this final SAP.

Rationale for Exploratory Imaging Biomarker Outcomes

In the Protocol version 3.0, hippocampal volume (HV) on MRI was designated as the primary imaging outcome. In this SAP, the HV and other imaging biomarker outcomes are being designated as exploratory outcomes. HV atrophy is an early event in AD patients, especially in APOE4/4 homozygotes who show accelerated atrophy compared to APOE3/3 patients with Early AD (Abushakra 2020). HV may be a marker of synaptic loss and neurodegeneration. Of the amyloid antibodies approved to date, only lecanemab demonstrated a significant but modest reduction (~8%) of HV atrophy in Early AD patients and correlation of HV atrophy reduction to clinical benefit of lecanemab is not reported (Van Dyck 2023). For these reasons, vMRI outcomes are now designated as exploratory imaging outcomes.

[redacted]

Rationale for Exploratory Fluid Biomarker Outcomes

In the Protocol version 3.0, CSF and plasma p-tau$_{181}$ were designated as primary fluid biomarker outcomes. In this SAP, these imaging and fluid biomarker outcomes are being designated as exploratory outcomes. In 2018, the National Institute on Aging-Alzheimer's Association (NIA-AA) proposed that research criteria for AD trials be based on positive amyloid and tau on either CSF biomarkers or PET scans (Jack 2018). In the ongoing 2023 update to these criteria, there is increased focus on the use of plasma biomarkers as research criteria for clinical trials. However, assay validation for these plasma biomarkers and development of standard cut-off values is still evolving. To date, the correlation of drug effects on the core plasma and CSF AD biomarkers to clinical benefit remains unclear. For these reasons, plasma and CSF biomarkers are designated as exploratory biomarkers.

5.1 Primary Efficacy Endpoint

The primary efficacy endpoint is the change from baseline to Week 78 in ADAS-Cog 13 scores.

5.2 Secondary Efficacy Endpoints

5.2.1 Key Secondary Endpoints

The key secondary endpoints are:

- The change from baseline (CBL) to Week 78 in A-IADL-W
- The change from baseline (CBL) to Week 78 in CDR-SB

5.2.2 Additional Secondary Efficacy Endpoints

The additional secondary efficacy endpoints are:

- The CBL to Week 78 in ADAS-Cog 11
- The CBL to Week 78 in DAD
- The CBL to Week 78 in NPI
- The CBL to Week 78 in MMSE
- The CBL to Week 78 in QoL-AD
- The CBL to Week 78 in RUD Lite
- The CBL to Weeks 13, 26, and 52 in ADAS-Cog 13 and ADAS-Cog 11 scores
- The CBL to Weeks 26 and 52 in A-IADL-W, CDR-SB, DAD, and NPI scores
- The CBL to Weeks 13, 26, 39, 52, and 65 in MMSE scores

- The CBL to Week 52 in QoL-AD (subject and informant) and RUD Lite scores

5.2.3 Pharmacokinetic Endpoints

The pharmacokinetic endpoints are:

- Plasma and CSF levels of ALZ-801, tramiprosate, and metabolite 3-sulfopropanoic acid (3-SPA) at each visit in the study, and update of the population PK model
- Correlation of PK levels to clinical efficacy and safety outcomes
- Correlation of PK levels to exploratory imaging and fluid biomarker outcomes

5.2.4 Exploratory Imaging Biomarker Endpoints

The exploratory imaging biomarker endpoints are the following volumetric MRI measures:

- CBL to Week 78 in total hippocampal volume
- CBL to Week 78 in cortical thickness
- CBL to Week 78 in ventricular volume
- CBL to Weeks 26 and 52 in total hippocampal volume
- CBL to Weeks 26 and 52 in cortical thickness and ventricular volume

Other exploratory imaging biomarkers are:

- CBL to Weeks 26, 52 and 78 in whole brain volume and DTI on MRI

5.2.4.1 Exploratory Imaging Anylysis of Right and Left Hemisphere

- CBL to Weeks 26, 52, and 78 in total hippocampal volume on right and left sides
- CBL to Weeks 26, 52, and 78 in cortical thickness on right and left sides

5.2.5 Exploratory Fluid Biomarker Endpoints

The exploratory plasma biomarker endpoints are:

- CBL to Week 78 in plasma levels of Aβ42, Aβ40, and Aβ42/40 Ratio
- CBL to Week 78 in plasma levels of p-tau$_{181}$ and p-tau$_{217}$.
- CBL to Week 78 in plasma levels of GFAP
- CBL of above plasma biomarkers to Weeks 13, 26, 52, and 65

The exploratory CSF biomarker endpoints are:

- CBL to Week 78 in CSF Aβ42, Aβ40, Aβ42/40 Ratio

- CSF to Week 78 in CSF p-tau$_{181}$ and p-tau$_{217}$
- CBL of above CSF biomarkers to Week 52

Other exploratory plasma and CSF biomarker endpoints are:

- CBL to Week 78 in CSF neurogranin
- CBL to Week 78 in plasma and CSF NfL
- CBL to Week 78 in CSF t-tau
- CBL to Week 78 in CSF sTREM2 and YKL-40
- CBL to Week 52 in CSF neurogranin, NfL, t-tau, sTREM2, and YKL-40
- CBL to Weeks 13, 26, 52, and 65 in plasma NfL

6 ANALYSIS POPULATIONS

6.1 Full Analysis Set (FAS)

The primary analysis population will be the full analysis set (FAS). The FAS includes all study subjects who received at least one dose of the study drug, had at least one baseline assessment and any post baseline efficacy assessment. All primary and secondary clinical efficacy analyses will be conducted on this population.

6.2 Safety/Intent-to-Treat (ITT) Population

The Safety/ITT population will include all study subjects who received at least one dose of study drug. All safety analyses will be conducted on this population.

6.3 Per Protocol Population (PPP)

The PPP will include all study subjects in the FAS population who are at least 80% compliant with study drug for the duration of their study participation without major protocol deviations that may impact primary or key secondary efficacy assessments. Membership in the PPP will be determined prior to database lock. The PPP will be used for supportive analyses of the primary and key secondary efficacy endpoints.

6.4 Completer Population (CP)

The CP will include all subjects in the FAS population who complete Week 78 visit and are on treatment at Week 78 ± 2, regardless of compliance with study drug. Subjects who discontinue treatment prematurely will not be included in this population even if they adhere to the scheduled visits. Membership in the CP will be determined prior to database lock. The CP will be used for supportive analyses of the primary and key secondary efficacy endpoints.

6.5 Safety MRI Population

The Safety MRI population will include all subjects with an evaluable baseline safety MRI assessment who have received at least one dose of study drug and have at least one evaluable post baseline safety MRI assessment. All Safety MRI analyses will be performed on this population.

6.6 Imaging Biomarker Population (Volumetric MRI, vMRI)

The Imaging biomarker population will include all subjects with an evaluable baseline vMRI assessment who have received at least one dose of study drug and have at least one evaluable post baseline imaging vMRI assessment. All imaging biomarker analyses will be performed on this population.

6.7 Plasma Pharmacokinetic (PK) Population

The plasma PK population will include all subjects who have received at least one dose of active treatment and have at least one evaluable post baseline PK sample. All plasma PK analyses will be performed on this population.

6.8 Plasma Biomarker Population

The Plasma biomarker population will include all subjects with an evaluable baseline sample for plasma biomarkers, received at least one dose of study drug and have at least one evaluable post baseline plasma biomarker sample. All plasma biomarker analyses will be performed on this population.

6.9 CSF PK Population

The CSF PK population will include all subjects who have received at least one dose of active treatment and have at least one evaluable post baseline CSF PK sample. All CSF PK analyses will be performed on this population.

6.10 CSF Biomarker Population

The CSF biomarker population will include all subjects with an evaluable baseline CSF biomarker sample who have received at least one dose of study drug and have at least one evaluable post baseline CSF biomarker sample. All CSF biomarker analyses will be performed on this population.

7 STUDY SUBJECTS

7.1 Subject Disposition

The number of subjects screened and randomized will be provided. For subjects who were screened but not randomized (screen failures), the counts and reasons for screen failures will be summarized. The number of subjects who withdrew from the study before Week 78 will also be provided (Early Terminations, ET) and the reasons for ET will be provided as listing

and summarized in Tables. The subset of ET subjects who discontinued study drug but attended study visits and provided further assessments (retrieved dropouts [RDOs]) will also be provided as listing and summarized in Tables. The number of subjects whose study drug was suspended by the Sponsor or investigator, and reason for study drug suspension will also be summarized.

7.2 Demographic and Baseline Characteristics

Demographics will be summarized by randomized treatment group and for the FAS, Safety, PPP, and CP populations. The following characteristics will be summarized.

- Sex (Male, Female)
- Race (White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Other)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino)
- Age
- Age group (50 – 65 years inclusive, > 65 years of age)
- Height
- Weight
- Body mass index

A summary of baseline characteristics will also include: MMSE group at screening (22-26, or >26), CDR-global (CDR-G), diagnosis of MCI or Mild AD, baseline values of all primary and secondary endpoints, use of concomitant AD medications (AChEI or none), in addition to height, weight, and body mass index. Baseline MMSE, calculated as the average of screening and V2 scores, will be summarized. Screening RBANS-delayed index score will also be summarized.

In addition, baseline values of hippocampal volume, cortical thickness, ventricular volume and whole brain volume will be summarized. For subjects in the plasma and CSF biomarker populations, the baseline values for the plasma and CSF core AD biomarkers will be provided when the analysis results become available.

The demographics and baseline characteristics will be summarized and tabulated by randomized treatment group.

7.3 Treatment Compliance

Overall treatment compliance will be calculated as a percentage using the total number of tablets that were dispensed and returned. Patients will be grouped into categories of < 70%, < 80%, 80% to 125% (inclusive), > 125%. Counts will be summed over the visits for each patient to calculate an overall compliance value.

Compliance will be summarized as described above and through descriptive statistics. The number of doses taken and number of doses should have taken are necessary to calculate compliance.

Compliance = (# of tablets dispensed - # of tablets returned) / # of tablets that should have been taken × 100%

For the calculations, the formulas are based on the assumption that patients take 2 tablets per day except for the clinical visit days.

- Number of tablets dispensed = number of kits × Number of cards per kit × 16. A Seven-card kit is to be dispensed during titration Visit 2/Baseline and 5-card kits are to be dispensed for rest of the visits.
- Number of tablets should have taken = Number of days between visits × 2 – Number of clinical visits (subject will only take one tablet during clinical visit days). See Table 1 below for details of dispensation information.

Table 1 Dispensation Schedule

Visit	Kit Type	# of Kit to Dispense	# of Blister cards per kit	# Tablets per Card
Visit 2 - Baseline	ALZ-801 265mg Titration	1	7 cards per kit	16
	Placebo Titration	1	7 cards per kit	16
Visit 4 – Week 6	ALZ-801 265mg	2	5 cards per kit	16
	Placebo	2	5 cards per kit	16
Visit 5 – Week 13 Visit 7 – Week 26 Visit 8 – Week 39 Visit 9 – Week 52 Visit 10 – Week 65* *No kits will be projected after this visit	ALZ-801 265mg	3	5 cards per kit	16
	Placebo	3	5 cards per kit	16

The number of tablets taken for kits that were not returned will be imputed to be the number of tablets that were dispensed.

7.4 Protocol Deviations

Protocol deviations by treatment will be listed and summarized for the FAS Population. The protocol deviations will be reviewed by clinical and statistical teams to select appropriate analysis subsets.

Protocol deviations include, but are not limited to:

- Inclusion/Exclusion Criteria
- Study Drug Overdose

- Other Study Drug Deviation
- Visit Window
- Restricted Concomitant Medication Change
- Study Procedure
- Other

Note that some of these deviations may not constitute major deviations from the protocol and hence, subjects with these deviations may not necessarily be excluded from the Per-Protocol Population.

7.5 Prior and Concomitant Medications

The number and percent of prior and concomitant medications will be provided by the treatment groups used for the safety analyses. A concomitant medication is defined as any medication that is taken after administration of the first dose of study drug (i.e., after receiving study drug at the Baseline Visit). Prior concomitant medications are defined as medications that are started before administration of the first dose of study drug and continue beyond that date. New concomitant medications are defined as medications that are started on or after the first dose of study drug, including those started in the Safety follow-up period. Prior medication is defined as any non-study medication that starts and stops before the first dose of study drug. Partial dates will be imputed using rules in Section 7.1.5.

Prior and concomitant medications will be coded using WHO Drug Dictionary (most current version). Summaries will be provided by anatomical therapeutic chemical classification (ATC) and preferred term.

7.6 Medical History

Medical history will be summarized by SOC and PT based on randomization and will be presented in the listings. Summary tables will be provided for the frequency and percentage of patients.

8 STATISTICAL METHODS OF ANALYSIS

8.1 General Considerations

Efficacy and Safety analyses will be performed with the FAS and Safety/ITT populations, respectively. Data will be summarized overall and by treatment arms. All Case Report Form (CRF) data collected during the study will be presented within data listings.

Should any of the statistical methods proposed prove unsuitable during the final analysis, more appropriate methods will be used, and any changes will be documented in the clinical study

report (CSR), including the rationale for use. These include the transformation of the data (for example to a logarithmic scale) in order to satisfy the model assumptions such as normally distributed residuals with constant variance; or the application of non-parametric techniques.

Following the International Council for Harmonisation (ICH) Guideline E9 (R1), an addendum adopted in November 2019, the estimand and its estimator and estimate will be defined for primary endpoint where inferential statistical comparison and hypothesis testing related to a treatment effect are of interest. When analyzing the efficacy endpoints, effort will be made to distinguish intercurrrent events from non-informative withdrawals, as well as applying various sensitivity analyses to assess the robustness of the estimate(s) for the pre-specified estimands. The definition of an intercurrent event is, "events that occur after treatment initiation and either preclude observation of the variable or affect its interpretation", e.g., use of rescue or prohibited medications, changes in permitted medications, switching treatments, discontinuing treatments, and subject deaths. Details on estimands, estimators, estimates, and intercurrent events are discussed in endpoint analysis section.

8.1.1 Descriptive Statistics

For descriptive statistical summaries, the continuous variables will be summarized by number of observations (n), mean, standard deviation (SD), standard error (SE), median, minimum (MIN), and maximum (MAX). Descriptive statistics for biomarkers will include percent coefficient of variation (%CV). For categorical variables, frequency and percentage of subjects in each category will be provided.

Minimum and maximum values will be rounded to the precision of the original value. Means and medians will be rounded to one decimal place greater than the precision of the original value. SDs, SEs, and 95% confidence intervals (CIs) will be rounded to two decimal places greater than the precision of the original value. Percentages for summarizing categorical data will be rounded to one decimal place. P-values, if needed, will be presented with 4 decimal places and values less than 0.0001 will be presented as <.0001.

The by-subject listings, including data at scheduled and unscheduled visits, will be sorted by treatment arm, subject ID, and then by date/time of the records.

8.1.2 Hypothesis Testing

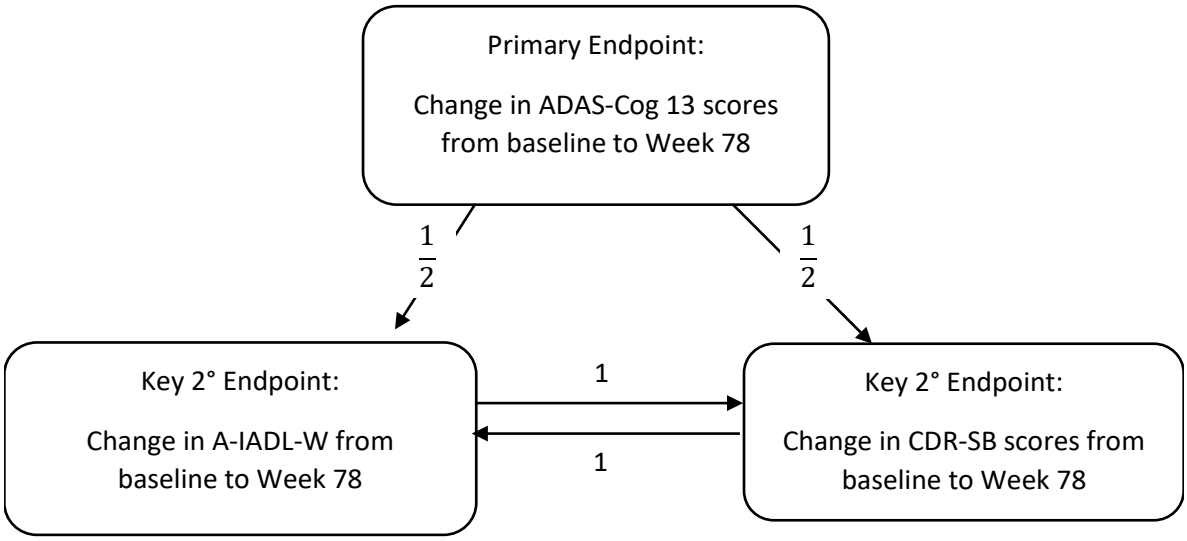
The Null Hypothesis of the study is that the CBL of ADAS-Cog 13 in ALZ-801 arm is not different from the placebo arm at 78 weeks.

The Alternate Hypothesis is that the CBL of ADAS-Cog 13 (Mohs 1997) in ALZ-801 arm is different from the placebo at 78 weeks.

The statistical testing will be two-sided and will be performed at the α=0.05 significance level. The key secondary efficacy endpoints will be compared only after the comparison of the primary endpoint has reached statistical significance. The comparisons of key secondary

endpoints will be done using a graphical approach (Bretz 2009) that strongly controls the family-wise Type 1 errors arising from multiple comparisons (Figure 2).

Figure 2 Graphical Approach Procedure for Analysis of Key Secondary Outcomes



8.1.3 Definition of Study Baseline

Unless otherwise specified, baseline is defined as the last non-missing measurement obtained prior to the first dose of the study drug for all efficacy and safety measures. Values at screening visits that are the baseline for efficacy assessments will be summarized, and all other screening values will be presented in listings. The MMSE baseline is defined as the average of the scores at the screening visit and V2 visit. If the MMSE score is available from only 1 pre-dose visit (either screening or V2), that score will be used as the baseline value.

8.1.4 Handling of Multiple Observations

For clinical efficacy data, if multiple measures were recorded for the same protocol defined visit, the later (repeated) measure will be used for data summary and analysis. Repeated data from unscheduled visits will be treated similarly.

If a subject has a scheduled visit, the assessment obtained at the scheduled visit will be used for data summary and analysis. Otherwise, if no scheduled visit assessment exists but at least one unscheduled visit assessment is available within the protocol-defined visit window, then the data at the latest unscheduled visit within the protocol-defined visit window will be used for data summary and analysis. Furthermore, in cases where both an unscheduled visit assessment and an early terminated visit assessment exist within the same analysis window, the assessment from the early terminated visit will be utilized for summarization and analysis.

When a subject undergoes multiple scheduled MMSE screening measurements, the most recent assessment will be used as the screening score. Any previous records showing MMSE scores below 22 should be regarded as instances of screen failure.

8.1.5 Multiplicity of Secondary Endpoints

The graphical approach will be used on testing multiple endpoints in a predefined order to strongly control the family-wise type 1 errors arising from multiple comparisons (Bretz 2009). The alpha recycle is described in detail in Figure 2. For the primary efficacy endpoint (CFBL to week 78 in ADAS-Cog 13, i.e., H1) hypothesis testing, the initial 2-sided significance level is set to 0.05. Only if the H1 testing is successful will the next two key secondary endpoints: CFBL to week 78 in A-IADL (H2) and CFBL to week 78 in CDR-SB (H3), be tested each at an alpha level of 0.025. If H2 shows significance, its alpha level of 0.025 is passed on to H3 for testing at an endpoint-specific alpha level of 0.025+0.025=0.05. However, if H2 is not significant, H3 is still tested at the reserved alpha level of 0.025. In this scenario, if H3 turns out to be significant, the alpha level for H2 is recycled back for re-testing at 0.05.

Since a graphical approach is used in the analysis, the overall type I error rate is strongly controlled at the 2-sided α=0.05. The hypothesis testing for the additional clinical endpoints (not presented in Figure 2) will be done without control for type I errors due to multiple comparisons and the p-values will be considered nominal.

8.2 Visit Windows and Efficacy Data Algorithms and Definitions/Handling of Dropouts or Missing Data

8.2.1 Visit Windows

Data at scheduled visits will be assigned to analysis visits as defined in the Analysis Visit Window tables below (Table 2, Table 3, Table 4, Table 5, Table 6, Table 7), to ensure that all visits have the potential to be included in the summaries. The unscheduled visits on or before day 1 will be mapped according to scheduled assessments. Visit windows herein will also be used to classify unscheduled and early termination visits.

For volumetric MRI measures (vMRI), if multiple measures were recorded for the same protocol defined visit, the later (repeated) measure will be used for data summary and analysis. If vMRI measures are missing for a protocol defined visit and an unscheduled vMRI measure is available, the unscheduled vMRI measure will be mapped to the closest missing protocol defined visit.

Table 2 Analysis Visit Windows for Assessments Done at Visit 4, 5, 7, 8, 9, 10, 11

Stage/Visit	Target Day	Study Days	
		Lower Bound	Upper Bound
Week 6 (Visit 4)	43	2	67

Stage/Visit	Target Day	Study Days	
Week 13 (Visit 5)	92	68	137
Week 26 (Visit 7)	183	138	228
Week 39 (Visit 8)	274	229	319
Week 52 (Visit 9)	365	320	410
Week 65 (Visit 10)	456	411	501
Week 78 (Visit 11)	547	502	592

Table 3 Analysis Visit Windows for Assessments Done at Visit 5, 7, 8, 9, 10, 11

Stage/Visit	Target Day	Study Days	
		Lower Bound	Upper Bound
Week 13 (Visit 5)	92	2	137
Week 26 (Visit 7)	183	138	228
Week 39 (Visit 8)	274	229	319
Week 52 (Visit 9)	365	320	410
Week 65 (Visit 10)	456	411	501
Week 78 (Visit 11)	547	502	592

Table 4 Analysis Visit Windows for Assessments Done at Visit 4, 5, 7, 9, 11

Stage/Visit	Target Day	Study Days	
		Lower Bound	Upper Bound
Week 6 (Visit 4)	43	2	67
Week 13 (Visit 5)	92	68	137
Week 26 (Visit 7)	183	138	274
Week 52 (Visit 9)	365	275	456
Week 78 (Visit 11)	547	457	592

Table 5 Analysis Visit Windows for Assessments Done at Visit 5, 7, 9, 11

Stage/Visit	Target Day	Study Days	
		Lower Bound	Upper Bound
Week 13 (Visit 5)	92	2	137
Week 26 (Visit 7)	183	138	274
Week 52 (Visit 9)	365	275	456
Week 78 (Visit 11)	547	457	592

Table 6 Analysis Visit Windows for Assessments Done at Visit 7, 9, 11

Stage/Visit	Target Day	Study Days	
		Lower Bound	Upper Bound
Week 26 (Visit 7)	183	2	274
Week 52 (Visit 9)	365	275	456
Week 78 (Visit 11)	547	457	592

Table 7 Analysis Visit Windows for Assessments Done at Visit 9, 11

Stage/Visit	Target Day	Study Days	
		Lower Bound	Upper Bound
Week 52 (Visit 9)	365	2	456
Week 78 (Visit 11)	547	457	592

8.2.2 Handling of Missing Data for Safety and AE

- Missing baseline values for vital signs, laboratory tests, ECG and safety MRIs will not be imputed..
- Missing post-baseline values for by-visit data will be summarized using the Visit Windows from Appendix 1. If a value is not available within a given window, no imputation will be done.
- Missing data for Adverse Event (AE) relationship will be imputed as "Related."
- Missing data for AE stop or start dates are imputed as follows:
  - If the imputed stop date precedes the start date (whether imputed or non-imputed), the imputed stop date will be adjusted to match the start date.
  - If the stop date is complete and the imputed start date falls after the stop date, then the start date will be imputed using the stop date.
  - Incomplete dates for AEs or non-study medications, the missing component(s) will be assumed as the most conservative database and shown in the data listings.
  - If AE is ongoing, the end date will not be imputed.
- Rules for partial dates are described in Table 8 below.

Table 8 Missing Date Imputation Rules

Parameter	Missing	Additional Conditions	Imputation
Start date for AEs	Time only		Time will not be imputed; but will be assumed to be after 1st dose date
	D only	M and Y same as M and Y of first dose date	Date of first dose
		M and/or Y not same as M and Y of first dose date	First day of non-missing month
	D and M	Y same as Y of first dose date	Date of first dose
		Y not same as Y of first dose date	Use January 1 of non-missing year
	M, D and Y	None – date completely missing	Date of first dose
Stop date for AEs	Time only		Time will not be imputed
	D only		Last day of non-missing month
	D and M		Use Dec 31 of non-missing year
	M, D and Y	Deceased	Date of death

Parameter	Missing	Additional Conditions	Imputation
		Not deceased	Date of the end of trial participation
Start date for Concomitant medications	D only	M and Y same as M and Y of first dose date	Date of first dose
		M and/or Y not same as M and Y of first dose date	First day of non-missing month
	D and M	Y same as Y of first dose date	Date of first dose
		Y not same as Y of first dose date	Use January 1 of non-missing year
	M, D and Y	None – date completely missing	Date prior to date of first dose
Stop date for Concomitant medications	D only	M and Y same as M and Y of first dose date	Last day of non-missing month
		M and/or Y not same as M and Y of first dose date	Last day of non-missing month
	D and M	Y same as Y of first dose date	Use Dec 31 of non-missing year
		Y not same as Y of first dose date	Use Dec 31 of non-missing year
	M, D and Y	None – date completely missing	Date will not be imputed

8.2.3 Handling of Missing Efficacy Data

- Missing baseline values of efficacy scales will not be imputed in any situation.
- Missing post-baseline values for by-visit data will be summarized and analyzed based on the mappings to Visit Windows (Section 8.2.1). Sensitivity analysis will be conducted to evaluate the impact of missing efficacy data (total score) as detailed in Section 8.3.

8.2.4 Efficacy Data Algorithms and Handling of Missing Sub-Items or Domains in Calculating Total Scores

If any of the individual sub-items for the primary or secondary endpoints is missing, the total score of the corresponding endpoint will be imputed by prorating the observed scores (van Ginkel 2010). The criteria for imputations are specific for each scale, as defined below.

ADAS-Cog 13/ ADAS-Cog 11: The ADAS-Cog 13 consists of 13 items for a score range of 0-85, and the ADAS-Cog 11 consists of 11 of the 13 items for a score range of 0-70, where higher scores indicate more severe cognitive deficits (Table 9).

For ADAS-Cog 13, if 3 or fewer of 13 items (<25%) are missing, the total score will be imputed by the following algorithm: ADAS-Cog 13 Total Score = Total score from the completed items × [maximum total score (=85) / maximum total score corresponding to the completed items]. The imputed number will be derived and rounded up to the nearest two decimal places. If more than 3 items are missing, the total score of ADAS-Cog 13 at that visit will be considered missing.

For ADAS-Cog 11, if 2 or fewer of 11 items (<25%) are missing, the total score will be imputed by the following algorithm: ADAS-Cog 11 Total Score = Total score from the completed items × [maximum total score (=70) / maximum total score corresponding to the completed items]. The imputed number will be derived and rounded up to the nearest two decimal places. If more than 2 items are missing, the total score of ADAS-Cog 11 at that visit will be considered missing.

Table 9 ADAS-Cog 13 and 11 Items and Algorithm for derivation of Item Scores and Total Score

Item	Algorithm	Range
Word Recall	Total number of "No" responses. Subscore is the sum the scores from trials 1, 2, and 3, divide by 3.	0-10
Commands	Total number of "No" responses from 5 tasks	0-5
Constructional Praxis	Count number of "No" responses. Subscore is 0=all 4 drawings correct; 1=1 figure drawn incorrectly; 2=2 figures drawn incorrectly; 3=3 figures drawn incorrectly; 4=4 figures drawn incorrectly; 5=no figures drawn, scribbles, parts of forms	0-5
Naming Objects / Fingers	Total number of "No" responses. Subscore is 0= 0-2 "no" responses; 1=3-5 "no" responses; 2=6-8 "no" responses; 3=9-11 "no" responses 4=12-14 "no" responses; 5=15-17 "no" responses	0-5
Ideational Praxis	Total number of "No" responses.	0-5
Orientation	Total number of "No" responses.	0-8
Word Recognition	Total # of "1" responses. If total is 12 or less, trial score=total. If total is >12, trial score=12.	0-12
Remembering Test Instruction	Subscore is 0=None, 1=Very Mild, 2=Mild, 3=Moderate, 4=Moderately Severe, 5=Severe	0-5
Comprehension	Subscore is 0=None, 1=Very Mild, 2=Mild, 3=Moderate, 4=Moderately Severe, 5=Severe	0-5
Word Finding Difficulty	Subscore is 0=None, 1=Very Mild, 2=Mild, 3=Moderate, 4=Moderately Severe, 5=Severe	0-5
Spoken Language	Subscore is 0=None, 1=Very Mild, 2=Mild, 3=Moderate, 4=Moderately Severe, 5=Severe	0-5
Number Cancellation (included in ADAS-Cog 13)	Adjusted Score = Total # correct targets crossed off minus Total # incorrect targets crossed off minus Total # times reminded of task. Then use Adjusted Score to determine subscore as follows; Adjusted Score Subscore ≥23=0; 18-22=1; 13-17=2; 9-12=3; 5-8=4; ≤4=5	0-5
Delayed Word Recall (included in ADAS-Cog 13)	Total number of "No" responses.	0-10
ADAS-Cog 13 Total	**Sum of 13-item scores**	**0-85**
ADAS-Cog 11 Total	**Sum of 11-item scores without "Number Cancellation" or "Delayed Word Recall"**	**0-70**

A-IADL: A-IADL consists of 47-70 items and is a scale administered to informant aimed at detecting deficits in instrumental and complex functions at the early stages of AD (Sikkes 2013). These activities include, but are not limited to, cooking, doing finances, and shopping. They are complex everyday tasks, determined by multiple cognitive processes and controlled processing (Dubbelman 2022a; Teng 2023). The A-IADL score can be determined using item

response theory (IRT) method (A-IADL-IRT, Sikkes 2013) or the weighted average method (A-IADL-W, Teng 2023).

A-IADL-IRT has a score range of 20-80 and will be calculated and transferred from a third party vendor for categorical responder analysis (Dubbelman 2022a). A-IADL-W has a score range of 0-100 and is calculated as follows: (sum of all scores / number of questions scored) × 25 which will be used for secondary efficacy analysis (Table 10, Teng 2023). For A-IADL-IRT, higher scores indicates better functioning, while for A-IADL-W, higher scores indicates worse functioning or more impairment.

If <25% (11 or fewer of 47; or 17 or fewer of 70 questions are answered) are answered as "don't know", A-IADL-W will be derived without imputation and considered valid. If the scores from more than 11 items are not available, A-IADL-W at that visit will be considered missing.

Table 10 A-IADL Algorithm for derivation of Total Score

Algorithm
If Yes is selected in main question with performance level: No = 0 Yes, slightly more difficult = 1 Yes, more difficult =2 Yes, much more difficult =3 Yes, he/she is no longer able to perform this task =4
If No is selected in main question with performance level: He/she was no longer able to do so due to cognitive problems = 4 He/she was no longer able to do so due to physical problems = Not scored He/she has never done that before = Not scored Other, please state = Not scored
A-IADL-W = (sum of all scores / number of questions scored) × 25

CDR-SB**:** The CDR consists of 6 boxes for a score range of 0-18 (Table 11). It is a global rating scale used to assess the stage and severity of dementia (Morris 1993). It uses a structured, clinician-rated interview that collects information on cognitive and functional capacity from both the subject and caregiver. Higher score on CDR-SB indicates more advanced AD.

The CDR-SB will be determined by adding the individual scores from each of the six domains. The same imputation technique will be applied to missing CDR-SB boxes or sub-items. If only 1 of 6 boxes (<25%) of the CDR is missing, the sum of the boxes will be imputed by prorating the sum from the other 5 boxes: CDR-SB Total Score = Total score from the completed items × [maximum total score (=18) / maximum total score corresponding to the 5 completed boxes (=15)]. The imputed number will be derived and rounded up to the nearest integer. If more than 1 box is missing, the total score of CDR-SB at that visit will be considered missing.

CDR-G: The CDR scale also provides a derived global score (CDR-G) which indicates disease stage, with CDR-G= 0 indicating no impairment, CDR-G= 0.5 indicating MCI, CDR-G=1

indicating Mild AD, and CDR-G = 2 or 3 indicating Moderate and Severe AD, respectively. CDR-G scale will be used for categorical responder analyses. In the event of a single missing box score on the CDR-SB, the aforementioned imputed CDR-SB score can be used to derive CDR-G score.

Table 11 CDR Items and Algorithm for derivation of Item Scores and Total Score

Item	Algorithm	Range
Memory	0=normal, 0.5=questionable; 1=mild, 2=moderate, 3=severe	0-3
Orientation	0=normal, 0.5=questionable; 1=mild, 2=moderate, 3=severe	0-3
Judgment and Problem Solving	0=normal, 0.5=questionable; 1=mild, 2=moderate, 3=severe	0-3
Community Affairs	0=normal, 0.5=questionable; 1=mild, 2=moderate, 3=severe	0-3
Home and Hobbies	0=normal, 0.5=questionable; 1=mild, 2=moderate, 3=severe	0-3
Personal Care	0=normal/questionable, 1=mild, 2=moderate, 3=severe	0-3
CDR-SB Total	**Sum of all items**	**0-18**

MMSE**:** The MMSE is a measure of global cognition that is widely used for clinical staging of AD (Folstein 1975). It consists of 11 domains items for a score range of 0-30 to assess general cognitive function (Table 12). Higher score on MMSE means better cognitive skills.

The same imputation algorithm will be applied to the MMSE. If 2 or fewer of 11 domains (<25%) of the MMSE are missing, the total score will be imputed by prorating the sum from the other completed domains: MMSE Total Score = Total score from the completed domains × [maximum total score (=30) / maximum total score corresponding to the completed domains]. The imputed number will be derived and rounded up to the nearest integer. If the scores from more than 2 domains are not available, the MMSE at that visit will be considered missing. For MMSE, higher total scores indicate better cognition.

Table 12 MMSE Domains and Algorithm for derivation of Domain Scores and Total Score

Domain	Algorithm	Range
Orientation to Time	5 questions: Total number of "Correct" answers	0-5
Orientation to Place	5 questions: Total number of "Correct" answers	0-5
Registration	3 words: Total number of "Correct" answers	0-3
Calculation	5 questions: Total number of "Correct" answers	0-5
Recall	3 words: Total number of "Correct" answers	0-3
Naming	2 questions: Total number of "Correct" answers	0-2
Repetition	1 question: Total number of "Correct" answer	0-1
Comprehension	3 questions: Total number of "Correct" answers	0-3
Reading	1 question: Total number of "Correct" answer	0-1
Writing	1 question: Total number of "Correct" answer	0-1
Drawing	1 question: Total number of "Correct" answer	0-1
MMSE Total	**30 questions: Total number of "Correct" answers**	**0-30**

DAD**:** The DAD consists of 40 items with a score range of 0-100 to evaluate the basic and instrumental activities in daily activities of subjects with dementia (Gélinas 1999). It is administered through an interview with the caregiver (Table 13). Higher DAD scores indicate less disability or better function.

DAD Total Score is calculated as follows: (DAD Total / DAD Number of Applicable Items) × 100. If 2 or fewer of 10 items (<25%) of the DAD are missing, the DAD Total Score will be derived without imputation and considered valid. If the scores from more than 2 items are not available, the DAD Total Score at that visit will be considered missing.

Table 13 DAD Items and Algorithm for derivation of Item Scores and Total Score

Item	Algorithm	Range
Hygiene	7 questions: Total number of "Yes" answers	0-7
Dressing	5 questions: Total number of "Yes" answers	0-5
Continence	2 questions: Total number of "Yes" answers	0-2
Eating	3 questions: Total number of "Yes" answers	0-3
Meal preparation	3 questions: Total number of "Yes" answers	0-3
Telephoning	4 questions: Total number of "Yes" answers	0-4
Going on an Outing	5 questions: Total number of "Yes" answers	0-5
Financing and Correspondence	4 questions: Total number of "Yes" answers	0-4
Medications	2 questions: Total number of "Yes" answers	0-2
Leisure and housework	5 questions: Total number of "Yes" answers	0-5
DAD Total	**40 questions: Total number of "Yes" answers**	**0-40**
DAD Number of Applicable Items	**Sum of Applicable Items**	**0-40**
DAD Total Score	**(DAD Total / DAD Number of Applicable Items)×100**	**0-100**

NPI-12**:** Early AD patients may experience various neuropsychiatric symptoms that emerge and worsen with disease progression and contribute to the morbidity and healthcare burden of AD (Forester 2019)**.** The NPI assesses the neuropsychiatric or behavioral symptoms of AD, including their frequency and severity (Cummings 1997). It is a self-administered questionnaire completed by caregivers about subjects for whom they care. The NPI-12 has 12 sub-items, each item has score range of 0-12 with a NPI-total score range of 0-144 for subject's symptoms and 0-60 for caregiver distress (Table 14). Higher scores on each NPI sub-item indicate more severe or frequent occurrence of that symptom. Higher scores on NPI-12 indicate more severe burden of neuropsychiatric or behavioral symptoms.

The same imputation algorithm will be applied to the NPI-12 and Caregiver Distress. If 2 or fewer of 12 items (<25%) of the NPI-12 are missing, the total score will be imputed by prorating the sum from the other completed items: Total score = total score from the completed items × [maximum total score (=144) / maximum total score corresponding to the completed items]. If 2 or fewer of 12 items (<25%) of the Caregiver Distress are missing, the total Caregiver Distress score will be and derived and imputed by prorating the sum from the other completed items: Total Caregiver Distress score = total score from the completed items × [maximum total score (=60) / maximum total score corresponding to the completed items].

The imputed number will be rounded up to the nearest integer. If the scores from more than 2 items are not available, the NPI-12 at that visit will be considered missing.

Table 14 NPI-12 Items and Algorithm for derivation of Item Scores and Total Score

Item	Algorithm	Range	Caregiver Distress Range
Delusions	Frequency (0-4), Severity (0-3)	0-12	0-5
Hallucinations	Frequency (0-4), Severity (0-3)	0-12	0-5
Agitation/Aggression	Frequency (0-4), Severity (0-3)	0-12	0-5
Depression/Dysphoria	Frequency (0-4), Severity (0-3)	0-12	0-5
Anxiety	Frequency (0-4), Severity (0-3)	0-12	0-5
Elation/Euphoria	Frequency (0-4), Severity (0-3)	0-12	0-5
Apathy/Indifference	Frequency (0-4), Severity (0-3)	0-12	0-5
Disinhibition	Frequency (0-4), Severity (0-3)	0-12	0-5
Irritability/Lability	Frequency (0-4), Severity (0-3)	0-12	0-5
Aberrant Motor Behavior	Frequency (0-4), Severity (0-3)	0-12	0-5
Sleep	Frequency (0-4), Severity (0-3)	0-12	0-5
Appetite Disorder	Frequency (0-4), Severity (0-3)	0-12	0-5
NPI-12 Total	**Sum of all items**	**0-144**	**0-60**

QoL-AD: The QoL-AD consists of 13 items for a score range of 13-52 for participant (subject) and informant (cargiver or study partner) (Logsdon 2002, Table 15), where higher scores indicate better quality of life.

The same imputation algorithm will be applied to the QoL-AD participant and informant. If 3 or fewer of 13 items (<25%) of the QoL-AD are missing, the total score will be imputed by prorating the sum from the other completed items: Total score = total score from the completed items × [maximum # of items (=13) / (# of completed items)]. The imputed number will be rounded up to the nearest integer. If the scores from more than 3 items are not available, the QoL-AD at that visit will be considered missing.

Table 15 QoL-AD (Participant and Informant) Items and Algorithm for derivation of Item Scores and Total Score

Item	Algorithm	Range
Physical Health	Poor (1), Fair (2), Good (3), Excellent (4)	1-4
Energy	Poor (1), Fair (2), Good (3), Excellent (4)	1-4
Mood	Poor (1), Fair (2), Good (3), Excellent (4)	1-4
Living Situation	Poor (1), Fair (2), Good (3), Excellent (4)	1-4
Memory	Poor (1), Fair (2), Good (3), Excellent (4)	1-4
Family	Poor (1), Fair (2), Good (3), Excellent (4)	1-4
Marriage	Poor (1), Fair (2), Good (3), Excellent (4)	1-4
Friends	Poor (1), Fair (2), Good (3), Excellent (4)	1-4
Self as a Whole	Poor (1), Fair (2), Good (3), Excellent (4)	1-4
Ability to Do Chores Around the House	Poor (1), Fair (2), Good (3), Excellent (4)	1-4
Ability to Do Things for Fun	Poor (1), Fair (2), Good (3), Excellent (4)	1-4
Money	Poor (1), Fair (2), Good (3), Excellent (4)	1-4
Life as a Whole	Poor (1), Fair (2), Good (3), Excellent (4)	1-4
QoL Total	**Sum of all items**	**13-52**

RUD Lite: The RUD Lite consists of two sections related tothe caregiver and to the subject (Wimo 2013). In particular, the following reported hours/minutes or reported events will be listed and summarized by treatment and by baseline MMSE using descriptive statistics. Missing data will be left blank without imputation.

- Caregiver sleep time (hours and minutes) in the last 30 days
- Caregiver time (hours and minutes) and number of days in the last 30 days in assisting subject with tasks in toilet visits, eating, dressing, grooming, walking and bathing
- Caregiver time (hours and minutes) and number of days in the last 30 days in assisting subject with tasks in shopping, food preparation, housekeeping, laundry, transporation, taking medication, and managing financial matters
- Caregiver time (hours and minutes) and number of days in the last 30 days in supervising the subject to prevent dangerous events
- Number of times the subject is admitted in a hospital (for more than 24 hours) since last visit
- Number of times the subject is admitted in a hospital emergency room (for less than 24 hours) since last visit

8.3 Efficacy Analyses

8.3.1 Analyses of the Primary Endpoint

The purpose of this study is to assess the efficacy, safety and biomarker effects in an AD population that results from patients initiating ALZ-801 versus Placebo. The estimand, estimator, and estimate of this primary efficacy endpoint are described and defined in Table 16.

Table 16 Estimand, Estimator, and Estimate of Primary Efficacy Endpoint

Estimand	Population	FAS
	Variable	Change in ADAS-Cog 13 scores from baseline to Week 78
	Intercurrent Events	Potential intercurrent events include treatment discontinuation and use of symptomatic AD medications. The "treatment-policy" strategy will be followed for the primary analysis, whereby the value for the endpoint of interest will be used regardless of occurrence of intercurrent events. Supportive analyses based on "while-on-treatment" strategy will be performed to evaluate the impact of intercurrent events on efficacy with censoring of data after intercurrent events.

	Population-level Summary	The between treatment difference in the mean change from baseline to Week 78 in ADAS-Cog 13 scores
Estimator	Analysis	Mixed-Effect Model Repeated Measure (MMRM) model with fixed class effect terms for treatment group, gender, age group, disease severity based on baseline MMSE (≤ 26 vs. > 26), use of concomitant AD medications, visit, and treatment by visit interaction and baseline ADAS-Cog 13 values as covariates.
Estimate	Primary to support the estimand	Between-treatment group comparison (effect in ALZ-801 minus effect in placebo): • LSM for the between-group difference • 95% CI of treatment difference • p-value
	Tributary	Within-treatment group comparison (effect at Week 78 minus the value at baseline): • LSM for the within-group difference • 2-sided 95% CI of the LSM • p-value

ADAS-Cog 13 scores are assessed at Day 1 (Baseline), and Weeks 13, 26, 52 and 78. Observed ADAS-Cog 13 scores including those from unscheduled visits and early discontinuation, are assigned analysis visits based on the visit window using the assessment's study days. After applying a visit window, there is only one non-missing or missing value for each analysis visit. These data are called observed data and the change from baseline in ADAS-Cog 13 scores (observed data) for efficacy analyses are derived by subtracting corresponding baseline value at each visit. Other efficacy endpoints derivation follows the same algorithm.

Unless otherwise specified, all primary efficacy analyses will be performed using the FAS population. Analyses of the primary endpoint of ADAS-Cog 13 will be performed with a likelihood based MMRM on treatment observed cases (OCs). The observed data set will consist of actual observations recorded at each visit and missing data is not explicitly imputed. The model terms will include treatment, the stratification factors of use of concomitant AD medications (AChEI or none), age group (50-65 years inclusive or > 65 years through 80 years), gender, disease severity based on baseline MMSE (≤ 26 vs. > 26), baseline ADAS-Cog 13 values, visit (VISIT), and treatment-by-visit interaction. The visit variable will be used to define the within-subject covariance for the repeated measures over time. The least squares mean (LSM) for each treatment at each visit will be generated together with its 95% confidence interval, and the LSM difference between ALZ-801 and placebo will also be calculated; together with its standard error, and the 2-sided 95% confidence interval. The MMRM has an inherent mechanism to "implicitly impute" missing values when calculating the LSM of treatment effects which assume missing at random (MAR). For the primary endpoint at 78 weeks, treatment comparisons will be done at the 0.05 significance level.

The Kenward-Roger degrees of freedom approximation and an unstructured covariance structure will be used. In case there is a convergence problem with MMRM model with the unstructured variance covariance matrix, the following structures will be used in order of 1) heterogeneous toeplitz (TOEPH), 2) heterogeneous autoregressive of order 1 (AR(1)), and 3) heterogeneous compound symmetry (CSH) and the first (co)variance structure converging to the best fit will be used as the primary analysis. If a structured covariance has to be used, the "sandwich" estimator of the standard error of the fixed effects parameters will be used in order to deal with possible model misspecification of the covariance matrix. Model estimates (LSM in treatment difference and its 95% CI) will be reported in addition to the p-value.

All data collected during the study will be used in the statistical analysis. Under the treatment policy strategy, measurements after occurrence of intercurrent events (subject withdrawal or dropouts for various reasons or use of symptomatic AD medications) will not be considered missing data and will be included in the primary analyses. For subjects who discontinue treatment prior to trial completion, every attempt will be made to have them continue with clinic visits and study assessments. Particular attention will be given to collecting Week 78 ADAS-Cog 13 assessments, regardless of when subjects discontinued treatment.

Note that only the post-baseline visits on which the endpoints are scheduled to be measured will be included in the model.

8.3.2 Supportive and Sensitivity Analyses of the Primary Efficacy Endpoint

Supportive Analyses

As supportive analysis, the primary endpoint will be analyzed using the MMRM model on the per-protocol population and on completers. Additional supportive analyses may include MMRM with Last Observation Carried Forward (LOCF).

Additional supportive analyses will include the repetition of the primary analysis using the 'while-on-treatment' strategy on the FAS population. This strategy estimates the treatment effect before the occurrence of intercurrent events. These events include treatment discontinuation and the use of symptomatic AD medications (AChE inhibitors or memantine). Any data collected after the occurrence of intercurrent events will be considered as missing.

Sensitiviy Analyses

Multiple Imputations (Missing at Random, MAR)

As a sensitivity analysis, a similar MMRM model as the one applied to the primary endpoint analysis will be repeated using multiple imputation methods under the assumption of MAR.

Before conducting the multiple imputation, the missing data pattern will be examined. A dataset is said to have monotone missing pattern if the missing values always occur at the end of the longitudinal data, e.g., missing data after withdrawal.

A dataset is said to have non-monotone missing pattern if some subjects have missing values for intermediate visits but have available data at subsequent visits. If the ADAS-Cog data are deemed having monotone missing pattern, the standard multiple imputation approach using SAS PROC MI with regression method will be applied. If ADAS-Cog data are determined to have non-monotone missing pattern, the Markov Chain Monte Carlo (MCMC) method will first be applied to impute the intermittent missing data, and then the regression method will follow to impute the monotone missing data. In the case that the regression method cannot fully impute the missing data, the MCMC method will be applied to impute both monotone and non-monotone missing patterns. The example SAS programming codes using regression method for monotone missing pattern are:

```
PROC MI DATA=DATAIN OUT=DATAOUT NIMPUTE=100 SEED=999[to be changed];
CLASS TREATMENT AChEI AGEGRP GENDER DISSTAGE;
VAR TREATMENT AChEI AGEGRP GENDER DISSTAGE BASE … ; [insert
scheduled visits for ADAS-Cog 13 up to Week 78, i.e, V5 V7 …]
MONOTONE REGRESSION;
RUN;
```

The example SAS programming codes using MCMC method for non-monotone missing pattern, follow by regression method for monotone missing pattern are:

```
PROC MI DATA=DATAIN OUT=DATAOUT_MONO NIMPUTE=100 SEED=999[to be
changed];
CLASS TREATMENT AChEI AGEGRP GENDER DISSTAGE;
VAR TREATMENT AChEI AGEGRP GENDER DISSTAGE BASE … ; [insert
scheduled visits for ADAS-Cog 13 up to Week 78, i.e, V5 V7 …]
MCMC CHAIN=MULTIPLE IMPUTE=MONOTONE;
RUN;
PROC SORT DATA=DATAOUT_MONO;
BY _IMPUTATION_ TREATMENT;
RUN;
PROC MI DATA=DATAOUT_MONO OUT=DATAOUT_REG SEED=999[to be changed]
NIMPUTE=1;
BY _IMPUTATION_;
CLASS TREATMENT AChEI AGEGRP GENDER DISSTAGE;
VAR TREATMENT AChEI AGEGRP GENDER DISSTAGE BASE … ; [insert
scheduled visits for ADAS-Cog 13 up to Week 78, i.e, V5 V7 …]
MONOTONE REGRESSION;
RUN;
```

The changes from baseline at all visits will be calculated from the imputed 100 sets of data. The MMRM model will be examined for each set of data. The results from the 100 sets of data will be combined via Rubin's rule to produce estimated LSMs and 95% CIs that incorporate the uncertainty from missing data. PROC MIANALYZE will be used. The example SAS programming codes are:

```
PROC SORT DATA=DATAOUT_REG;
BY _IMPUTATION_ TREATMENT;
RUN;
PROC MIXED DATA=DATAOUT_REG;
BY _IMPUTATION_;
CLASS TREATMENT(REF='PLACEBO')AChEI AGEGRP GENDER DISSTAGE;
```

```
MODEL CHG =TREATMENT AChEI AGEGRP GENDER DISSTAGE BASE VISIT
TREATMENT*VISIT/SOLUTION;
REPEATED VISIT/SUBJECT=USUBJID TYPE=UN;
LSMEANS TREATMENT*VISIT /ALPHA=0.05 PDIFF CL;
ODS OUTPUT LSMEANS=LS DIFFS=LSDIFF;
RUN;
PROC SORT DATA=LSDIFF;
by _IMPUTATION_;
RUN;
PROC MIANALYZE PARMS(CLASSVAR=FULL)= LSDIFF;
CLASS TREATMENT;
MODELEFFECTS TREATMENT*VISIT;
ODS OUTPUT PARAMETERESTIMATES=DIFF_MI;
RUN;
```

Multipe Imputations (Missing Not at Random, MNAR)

Additional sensitivity analysis may be provided to evaluate the robustness of inferences on the same estimand, to assess deviations from the assumptions used in the statistical model, and to address limitations in the data. The possibility of "missing not at random" (MNAR) data will also be considered. After all active subjects complete the study and database is locked, the unblinded data will be examined and the missing data pattern evaluated. Based on the missing data mechanisms in the FAS, other multiple imputations methods will be explored to select the most appropriate methods. These methods may include the "Placebo-Based Pattern-Mixture model" or the "Tipping-Point Method". These sensitivity analyses will be described and documented in the final CSR.

Other Supportive or Sensitivity Analyses

Additional supportive or sensitivity analyses using observed data only (OC analysis), may be performed. For vMRI analyses, a model of slope analysis may also be conducted (Chen 2018; McDade 2022).

If any deaths occur, the worst-rank score method (Lachin 1999) will be employed as a supportive analysis. Any missing data due to deaths would be treated as the worst outcome in the primary analysis. This method makes no assumptions about the distribution. The observed changes from baseline at Week 78 in ADAS-Cog 13 scores from both the ALZ group and the placebo group will be combined and ranked from the lowest to the highest. Since a higher score in ADAS-Cog 13 indicates more severe impairment, the scores for all deaths will be assigned the highest rank. This analysis will be performed for the Week 78 observed cases data. The Wilcoxon Rank Sum Test and Hodges-Lehmann estimate will then be used to assess the treatment differences, and their 95% confidence intervals based on the Hodges-Lehmann estimates will also be provided.

8.3.3 Analyses of Key Secondary and Other Secondary Endpoints

The same MMRM method applied to the primary efficacy endpoint will be used to evaluate the between-treatment difference for the key secondary endpoints and the other secondary endpoints using the FAS. The same treatment policy strategy as that applied to the primary efficacy endpoint will be used for the secondary efficacy endpoints analysis. Differences between the two treatment groups, the p-values, and the 2-sided 95% CIs will be calculated within the framework of MMRM. The model will include all terms as in the primary endpoint model except that the baseline values will be corresponding to the baseline value for each parameter. Note that only the post-baseline visits on which the endpoints are scheduled to be measured will be included in the model.

The main analysis population is the FAS. Similar sensitivity analyses as for primary endpoint will be repeated for the key secondary and other secondary endpoints. As supportive analysis, the key and other secondary endpoints will also be analyzed using the PPP, and CP. Additional sensitivity analyses using MMRM with Last Observation Carried Forward (LOCF) may be performed.

8.3.4 Subgroup Analyses

Efficacy analyses in the sub-groups of subjects with screening MMSE 22 to 26 (inclusive) and 27-30 (inclusive) will be performed on the FAS using the MMRM method. Efficacy analyses will also be performed in the subgroups defined by the use of AChEI (yes or no), age group (50-65 or > 65 years of age) and gender (female or male). These subgroup analyses will be performed on the FAS using the MMRM method for primary and key secondary endpoints.

8.3.5 Responder Analyses

Additional efficacy analyses may include responder analysis based on the efficacy outcomes. Since mean changes from baseline may be affected by outliers, responder analyses provide insights into drug effects that are less affected by outlier values. The following responder or categorical analyses may be conducted to complement the CBL analyses of the main efficacy outcomes including the ADAS-Cog 13, A-IADL, and CDR-SB. They may also be conducted on the MMSE and CDR-G which are part of the study inclusion criteria. Subjects with missing scores at the visits will be excluded from the responder analyses for that variable/visit. The responder categories are defined as follows:

Primary cognitive outcome, ADAS-Cog 13 (Mohs 1997):

- Number and percentage of subjects whose scores remain stable or improve from baseline (lower scores = less deficits): CBL ≤ 0 at Week 78
- Number and percentage of subjects whose scores worsen from baseline by less than:
  - ADAS-Cog 13 CBL ≤ 3.5 points at Week 78

- ADAS-Cog 13 CBL ≤ 2.5 points at Week 78
- ADAS-Cog 13 CBL ≤ 1.5 points at Week 78

Key secondary outcome, A-IADL-IRT scores for functional categories (Dubbelman 2022a):

- Number and percentage of subjects whose A-IADL-IRT scores at Week 78 remain within the same category as baseline or improve to next higher category (higher scores = less impairment), based on the following 3 functional categories:
  - Category of "No-Mild" deficits with IRT scores ≥ 50
  - Category of "Moderate" deficits with IRT scores = 40-49 inclusive
  - Category of "Severe" deficits with IRT scores < 40

Key secondary outcomes, CDR-SB for functional categories (O'Bryant 2008):

- Number and percentage of subjects whose CDR-SB values at Week 78 remain within the same category as baseline or improve to less advanced category (lower scores = lower disease severity), based on the following 3 categories or disease stages:
  - MCI or no impairment defined as CDR-SB from 0 – 4.0 inclusive
  - Mild AD defined as CDR-SB from 4.5 – 9.0 inclusive
  - Moderate AD defined as CDR-SB from 9.5 – 15.5 inclusive

Stage of AD based on CDR-G for disease stages (Morris 1993):

- Number and percentage of subjects whose CDR-G values at Week 78 remain within the same category as baseline or improve (lower scores = less severe category), based on the following 3 categories:
  - MCI or no mpairment defined as CDR-G ≤ 0.5
  - Mild AD defined as CDR-G > 0.5 and ≤ 1.0
  - Moderate AD defined as CDR-G >1 and ≤ 2.0

Stage of AD based on baseline MMSE (Average of Screening and V2 Scores):

- Number and percentage of subjects whose MMSE values at Week 78 remain within the same category as baseline or improve to the next higher category (higher scores = milder disease), based on the following 4 categories (2 distinct sets of 4 categories):
  - MMSE 28-30, inclusive
  - MMSE 24-27, inclusive
  - MMSE 22-23, inclusive

- MMSE < 22

  and

- MMSE 27-30, inclusive
- MMSE 24-26, inclusive
- MMSE 22-23, inclusive
- MMSE < 22

8.3.6 Time to Event Analyses

8.3.6.1 Time to Disease Progression

Additional efficacy analyses may include time-to-event (time to first disease progression) analyses using the CDR-SB and A-IADL-IRT scores. The CDR-SB analysis will evaluate the delay in progression from MCI to Mild AD stage or worse, and from Mild AD to Moderate or Severe AD (O'Bryant 2008). The A-IADL-IRT analysis will evaluate delay in functional decline, or progression from a category of "No-Mild" to "Moderate" or worse, or from "Moderate" to "Severe" functional deficits (Dubbelman 2022a). The time-to-event analyses will also be performed based on the CDR-G and MMSE scores. The categories for time-to-progression events are defined below for CDR-SB, CDR-G, A-IADL-IRT, and MMSE scores:

- CDR-SB: CDR-SB progresses from category of MCI (SB ≤ 4.0) to Mild AD or worse (SB ≥ 4.5); or from Mild AD (SB = 4.5-9.0 inclusive) to Moderate AD or worse (SB ≥ 9.5)
- CDR-G: CDR-G progresses from MCI (CDR-G = 0.5) to Mild AD or worse (CDR-G ≥ 1.0); or from Mild AD (CDR-G = 1.0) to Moderate AD or worse (CDR-G ≥ 2.0)
- A-IADL-IRT: The A-IADL-IRT scores change from category of No or Mild deficits (≥ 50) to Moderate deficits or worse (<50); or from Moderate deficits (40-49 inclusive) to Severe deficits (< 40)
- MMSE: MMSE progresses from MCI (MMSE >26) to Mild AD or worse (MMSE ≤ 26); or from Mild AD (MMSE 22-26, inclusive) to Moderate AD (MMSE ≤ 21)

Time-to-Event Analysis

The time-to-event is measured from day of randomization to day of the earliest progression event (to next category). Subjects who do not experience the event by the study's end or drop out for other reasons will be censored. A Cox proportional hazard regression analysis (Cox 1972) will incorporate the treatment group and four stratification factors (use of AChEI, age group, disease severity, and gender) as covariates to determine whether there is a difference in the time to event between the ALZ-801 and placebo arms. The resulting hazard ratios, 95%

confidence intervals, and p-values will be reported. In addition, we will estimate the distribution of delaying progression in each treatment group using the Kaplan-Meier method (Kaplan 1958). The Kaplan-Meier method will provide probability estimates for patients with progression. Kaplan-Meier curves will be presented for each treatment arm, along with medians for time to progression and 95% confidence intervals.

8.3.6.2 Time to Emergence or Worsening of Neuropsychiatric/Behavioral Symptoms

The NPI scale will be used to evaluate the emergence or worsening of specific neuropsychiatric symptoms (NPS) of interest. Emergence of a NPS is defined as an NPI sub-item score = 0 at baseline that becomes ≥ 3 at subsequent study visit. The NPS of interest are the following:

- Cluster 1: Irritability, anxiety, depression/dysphoria
- Cluster 2: Agitation/Aggression, aberrant motor behaviours
- Cluster 3: Delusions, hallucinations

The catergories for NPS emergence or worsening are defined by NPI sub-item score (frequency × severity, Table 14):

- Baseline NPI sub-item score < 3 increases to NPI sub-item score ≥ 3 on one or more of Cluster 1 NPS
- Baseline NPI sub-item score < 3 increases to NPI sub-item score ≥ 3 on either agitation/aggression or aberrant motor behaviours of Cluster 2 NPS
- Baseline NPI sub-item score < 3 increases to NPI sub-item score ≥ 3 on either delusions or hallucinations of Cluster 3 NPS

Time-to-Event Analysis

The time-to-event is measured from day of randomization to day of earliest emergence or worsening of a NPS (NPI sub-item score) based on the above definitions. Emergence is defined as a NPS with NPI sub-item score = 0 at baseline and becomes > 0 at a subsequent study visit. Worsening is defined as a NPS with NPI sub-item score of 1 or 2 at baseline and becomes ≥3 at a subsequent study visit. Subjects who do not experience the event by the study's end or drop out for other reasons will be censored. The same statistical analysis methods will be employed as in Section 8.3.6.1.

8.3.7 Additional Cognitive / Functional / Quality of Life Analysis

Additional composite scores that utilize sub items from cognitive, functional, and/or quality of life scales may be derived and analyzed, guided by emerging validation studies of composite scales in AD studies.

8.3.8 PK Analysis and Population PK Model

The plasma and CSF concentrations of ALZ-801, tramiprosate, and the metabolite (3-SPA) will be summarized and tabulated. Population PK model will be developed by a dedicated group of PK and modelling scientists according to a dedicated SAP for PK/PD modeling.

8.3.9 Correlation of PK levels to Efficacy, Safety Parameters, and Exploratory Biomarker Outcomes

- The correlation of PK parameters to clinical efficacy and safety parameters will be analyzed by estimating coefficients for meta regressions using all collected data points with subject as a random effect.
- The correlation of PK parameters to exploratory imaging and fluid biomarker outcomes will be analyzed using the same methods. Mediation analyses may be conducted to explore relationships between clinical outcomes and imaging or fluid biomarkers.

8.4 Exploratory Imaging and Biomarker Analyses

8.4.1 Analyses of Exploratory Imaging Biomarker Endpoints

Analysis of the volumetric MRI outcomes will be performed using the same MMRM approach described for analysis of the primary clinical endpoint. The MMRM model will include treatment, the use of concomitant AD medications (AChEI or none), age group (50 through 65 years or > 65 years), gender, disease severity based on baseline MMSE, baseline volumetric MRI value, visit, and treatment by visit interaction. The MRI magnet strength (1.5 or 3 Tesla) will also be included as a covariate in the model.

The imaging biomarker analyses will be performed on the imaging biomarker population.

8.4.2 Analyses of Exploratory Plasma Biomarker Endpoints

Serial plasma samples from screening and each subsequent visit will be stored and batched, under GLP conditions, for analyses after the end of study. Samples from multiple subjects and study sites will be batched and analyzed at a single central laboratory using validated assays, or well performing, widely accepted plasma assays.

The plasma biomarker analyses will be performed on the plasma biomarker population. The change from baseline and percent change from baseline in plasma biomarkers will be summarized by visit and treatment group. The change from baseline and percent change from baseline in each plasma biomarker will be analyzed using an MMRM model. The model will include treatment, the use of concomitant AD medications (AChEI or none), age group (50 through 65 years or > 65 years), gender, disease severity based on baseline MMSE, baseline plasma biomarker level, visit, and treatment by visit interaction in the model.

[redacted]

8.4.3 Analyses of Exploratory CSF Biomarker Endpoints

Serial CSF samples from the baseline, Week 52 and Week 78 visits will be stored and batched for analyses. Samples from multiple subjects and study sites will be analyzed in batches, at a single central laboratory, using standardized procedures and validated assays, or well performing, widely accepted CSF assays.

The CSF biomarker analyses will be performed on the CSF biomarker population. The change from baseline and percent change from baseline in CSF biomarkers will be summarized by visit and treatment group. The between groups comparisons of change from baseline and percent change from baseline in each CSF biomarker will be analyzed using an MMRM model with treatment, the use of concomitant AD medications (AChEI or none), age group (50 through 65 years or > 65 years), gender, disease severity based on baseline MMSE, baseline CSF biomarker level, visit, and treatment by visit interaction in the model.

8.5 Safety Analyses

8.5.1 Extent of Exposure

The extent of study medication exposure will be listed for each subject as the number of days on study medication, and summary statistics will be provided for each treatment group. Duration of exposure (in days) will be calculated as (last dose date – first dose date + 1).

The extent of exposure for each treatment group will also be summarized categorically. The categories of exposure are defined as follows: exposures of ≥ 13 weeks, ≥ 26 weeks, ≥ 39 weeks, ≥ 52 weeks, ≥ 65 weeks and ≥ 78 weeks.

8.5.2 Adverse Events

Adverse event tables will include summaries of TEAEs by treatment arms. A TEAE is defined as any new AE that begins, or any preexisting condition that worsens in severity, after at least one dose of study treatment has been administered.

An overview table containing the number and percent of the following will be included and presented by treatment arms:

- Number of total AEs, TEAEs, and deaths
- Incidence of patients with at least one TEAE, drug-related TEAE, serious AE (SAE), drug-related SAE, and non-serious TEAEs
- Incidence of patients who discontinued due to TEAE, drug-related TEAE, SAE, and drug-related SAE
- Incidence of deaths and deaths due to drug-related TEAE

The following tabular summaries of TEAEs will be provided by treatment arms:

- TEAEs summarized by SOC and PT, descending frequency of PT, by maximum severity, by age group (50 – 65 inclusive, > 65 years), and by disease stage (MCI, Mild AD).
- Adverse events leading to discontinuation summarized by SOC and PT, and by descending frequency of PT.
- Drug-related AEs summarized by SOC and PT, and by descending frequency of PT.
- Serious adverse events summarized by SOC and PT, descending frequency of PT, and will include a summary of drug related events.
- SAE will also be summarized by age groups and by disease stage (as defined above).

MRI Safety Analsyes

The safety assessments will also include analyses of MRI findings (ARIA-E and ARIA-H) based on the central MRI readings by treatment arms.

Below are the rules to follow for AE summaries:

- A drug-related AE is defined as an AE with an assigned relationship of "possibly related," "probably related," "definitely related," or missing.
- When assessing severity, the TEAE with the worst severity will be chosen for a subject.
- AEs with missing severity will be summarized and presented separately.

AEs will be coded using most current Medical Dictionary for Regulatory Activities (MedDRA) version.

8.5.3 Evaluation of ARIA on Brain MRI

Evaluation of amyloid-related imaging abnormalities or ARIA on Safety MRIs (Sperling 2011) will include evaluation of ARIA-E (brain edema or effusion), ARIA-H (microhemorrhages), ARIA-H (superficial siderosis) and ARIA-H (macrohemorrhages). Safety MRIs will also be analyzed for severity of white matter disease (WMD) in the periventricular and deep white matter, using the Fazekas severity scores. Other ischemic abnormalities such as lacunes or infarcts will also be evaluated. Shift tables for the 4 types of ARIA, WMD parameters and the other ischemic lesions will be provided by visit. Contiuous variables for the ARIA-H will also be summarized by visit.

Drug Suspension due to MRI Findings

Some subjects who develop MRI findings of interest may be instructed to suspend study drug till resolution of the MRI finding. The number of subjects whose study drug was suspended by the Sponsor or investigator for MRI finding, and the reason for study drug suspension [ARIA-

E or ARIA-H with macrohemorrhage or ARIA-H with siderosis, or a combination of these lesions)] will also be summarized.

Clinical Laboratory Evaluations

Clinical lab results will be reported for hematology, clinical chemistry, coagulation tests, and urinalysis and will be summarized by visit.

Parameters will be summarized descriptively through change from baseline (CFBL) and percent CFBL for numeric values. If a parameter is categorical, it will be listed only.

Out-of-range values will be assessed through shift tables. Each lab value will be assessed as low, normal or high based on the normal ranges provided by the central lab. Frequencies of each combination of shifts will be provided.

Potentially clinically significant (PCS) tables will also be used to summarize out-of-range values. Summaries will be given for any time post-baseline. The denominator for the percentages will be the number of patients with a result at the specific visits for each parameter.

All laboratory results will be listed, including scheduled and unscheduled/repeat measurements (if any). Laboratory assessments that are outside of normal ranges and/or with potential clinical significance will be flagged in the listings.

Clinically notable changes in LFTs will be summarized and listed for individual study subjects. The criteria for clinically significant LFT results are as follows:

- ALT > 5 × ULN,
- AST > 5 × ULN, or
- TBL > 3 × ULN

8.5.4 Evaluation of Weight Loss

Maximum percent change from baseline in weight, with weight loss > 8%, >9%, and >12% will be summarized, including a breakdown by age, gender, baseline BMI, by use of AChEI, and by disease severity at baseline (MMSE ≤ 26, MMSE > 26).

8.5.5 Evaluation of Electrocardiogram (ECG) Findings

Electrocardiogram (ECG) Findings

Electrocardiograms will be assessed by a central reader and will be recorded at the following visits:

- Screening/Visit 1
- Visit 4/Day 43
- Visit 5/Day 92

- Visit 7/Week 26
- Visit 9/Week 52
- Visit 11/Week 78

The following quantitative parameters will be reported by the central reader: heart rate, PR interval, QRS duration, QT interval (uncorrected), and QT interval with Fridericia's correction (QTcF). Change from baseline and percent change from baseline by visits will be calculated for each parameter. The number and percent of patients within each category (normal, abnormal not clinically significant, and abnormal clinically significant) will also be summarized by visit. A standard triplicate 12-lead ECG will be performed at the Screening, Visit 4 and Visit 5. At subsequent visits, a single 12-lead ECG will be performed unless there is evidence of QT prolongation, in which case a triplicate ECG must be performed. Assessment of clinically notable abnormalities will be based on the average of triplicate ECG readings. To be conservative, the worst interpretation of the three readings will be used to generate the summary of ECG clinically notable abnormalities table. The denominator for the percentages will be the number of patients with a result at the specific visits for each parameter.

Clinically notable changes in ECGs will be listed for individual study subjects and tabular summaries provided per visit. The criteria for clinically significant ECGs results are as follows:

- PR interval ≥ 220 msec
- PR interval ≤ 120 msec
- QRS duration ≥ 120 msec and increase of ≥ 20 msec from baseline
- QT interval >500 msec
- QTcF interval > 500 msec and increase of ≥ 60 msec from baseline

Categorical Analyses of QTcF Interval

At each visit with ECG, observed QTcF data will be summarized into the following categories and shift tables provided:

- Categories of QTcF ≤ 450msec, 451-480 msec, 481-500 msec, and > 500 msec
- Categories of QTcF > 450 msec, > 480 msec, and > 500 msec
- Categories of CBL in QTcF: ≤ 10 msec, 11-30 msec, 31-60 msec
- Categories of CBL in QTcF: > 30 msec, > 60 msec

Vital Signs

Vital signs will be assessed at all visits except Visit 3 and Visit 6. The following parameters will be summarized: systolic blood pressure (SBP), diastolic blood pressure (DBP), body

temperature and respiratory rate. Weight and Height will also be summarized for available visits. These parameters will be summarized through change from baseline and percent change from baseline in similar fashion as the ECG parameters.

Vital signs will be assessed through potentially clinically significant (PCS) criteria. Patients will be counted if they meet the criteria at any time post-baseline. The denominator for the percentages will be the number of patients with a result at the specific visits for each parameter.

Clinically notable systolic and diastolic BP will be listed for individual subjects and summarized using descriptive statistics. Clinically notable criteria are described in Table 17.

Table 17 Criteria for Clinically Notable Vital Sign Abnormalities

Parameter	High Threshold	Low Threshold
Systolic BP	> 30 mm Hg increase from baseline	> 30 mm Hg decrease from baseline
Diastolic BP	> 20 mm Hg increase from baseline	> 20 mm Hg decrease from baseline
Pulse rate	> 120 beats/min and an increase in pulse rate of ≥ 15 beats/min from baseline	< 40 beats/min and a decrease in pulse rate of ≥ 15 beats/min from baseline
Body temperature	> 39.0°C (≥ 102.2°F)	< 35.0°C (≤ 95.0°F)

BP: blood pressure

Physical Examinations

Physical examinations will be performed at all visits except Visit 3 and Visit 6. All data will be listed. The number and percentage of subjects with abnormal neurological examination findings will be summarized.

Neurological Examinations

Neurological exam will be performed at the following visits:

- Screening/Visit 1
- Visit 7/Week 26
- Visit 9/Week 52
- Visit 11/Week 78
- Safety Follow up Visit/Week 82

All data will be listed. The number and percentage of subjects with abnormal neurological examination findings will be summarized.

[redacted]

8.5.6 Other Safety Analyses

C-SSRS suicidal behavior, suicidal ideation, and intensity of most severe ideation will be listed and summarized by visit and by questionnaire types (baseline and since last visit). The number and percentage of subjects with abnormal C-SSRS scores will be summarized.

9 INTERIM ANALYSIS

No interim analyses are planned or were conducted.

10 STATISTICAL SOFTWARE

All analyses will be performed using SAS Version 9.4.

[redacted]

11 REFERENCE LIST

Abushakra S, Porsteinsson AP, Scheltens P, et al. (2017) Clinical effects of oral tramiprosate in APOE4/4 homozygous patients with mild Alzheimer's disease suggest disease modification. *J Prev Alzheimers Dis* **4**:149-56.

Abushakra S, Porsteinsson AP, Sabbagh M. et al. (2020) APOE4/4 homozygotes with early Alzheimer's disease show accelerated hippocampal atrophy and cortical thinning that correlates with cognitive decline. *Alzheimer's Dement.* 2020;6:e12117.

Aduhelm (aducanumab-avwa) injection for intravenous use [prescribing information]. Cambridge, MA: Biogen Inc; 2021. Available from: https://www.aduhelm.com/.

Bretz F, Maurer W, Brannath W, et al. (2009) A graphical approach to sequentially rejective multiple test procedures. Stat Med; 28(4):586-604. ICH E9(R1) Guideline. Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials.

Budd-Haeberlein S, Aisen PS, Barkhof F, et al. (2022) Two randomized phase 3 studies of aducanumab in early Alzheimer's disease. *J Prev Alzheimers Dis* **9**:197-210.

Chen Y-F, Nia X, Fleisher AS, et al. (2018) A simulation study comparing slope model with mixed-model repeated measure to assess cognitive data in clinical trials of Alzheimer's disease. *Alzheimers Dement (N Y)* **4**:46-53.

Cox DR (1972) Regression Models and Life-Tables. *J R Stat Soc: Series B (Methodol)* **34**:187–220.

Cummings JL (1997). The Neuropsychiatric Inventory: assessing psychopathology in dementia patients. *Neurology*. **48**:S10-6.

Dubbelman MA, Terwee CB, Verrijp M, et al. (2022a) Giving meaning to the scores of the Amsterdam instrumental activities of daily living questionnaire: a qualitative study. *Health Qual Life Outcomes* **20**:47.

Dubbelman MA, Verrijp M, Terwee CB, et al. (2022b) Determining the Minimal Important Change of Everyday Functioning in Dementia, Pursuing Clinical Meaningfulness. *Neurology* 99: 9 (e955-e964).

Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**:189-98.

Forester BP, Vahia I (2019) Behavioral and Psychological Symptoms - An Emerging Crisis of the Alzheimer Dementia Epidemic. *JAMA Netw Open* **2**:e190790.

Gauthier S, Aisen PS, Ferris SH, et al. (2009) Effect of tramiprosate in patients with mild-to-moderate Alzheimer's disease: exploratory analyses of the MRI sub-group of the Alphase study. *J Nutr Health Aging* **13**:550–7.

Gélinas I, Gauthier L, McIntyre, et al. (1999) Development of a functional measure for persons with Alzheimer's disease: the disability assessment for dementia. *Am J Occup Ther* **53**:471-81.

Jack CR, Bennett DA, Blennow K, et al. (2018) NIA-AA Research Framework: toward a biological definition of Alzheimer's disease. *Alzheimers Dement* **14**:535-62.

Kaplan EL, Meier P (1958) Nonparametric Estimation from Incomplete Observations. *J Am Stat Assoc* **53**:457-81.

Lachin, JM (1999) Worst-Rank Score Analysis with Informatively Missing Observations in Clinical Trials. *Control Clin Trials*, **20**: 408–422.

Logsdon RG, Gibbons LE, McCurry SM, et al. (2002) Assessing quality of life in older adults with cognitive impairment. *Psychosom Med*. **64**:510-9

Leqembi (lecanemab-irmb) injection for intravenous use [prescribing information]. Nutley, NJ: Eisai Inc; 2023. Available from: https://www.leqembi.com/.

McDade E, Cummings JL, Dhadda S. et al. (2022) Lecanemab in patients with early Alzheimer's disease: detailed results on biomarker, cognitive, and clinical efects from the randomized and open-label extension of the phase 2 proof-of-concept study. *Alzheimers Res Ther* **14**:191.

Mintun MA, Lo AC, Evans CD, et al. (2021) Donanemab in early Alzheimer's disease. *N Engl J Med* **384**:1691-704.

Mohs, RC, Knopman, D, Petersen, RC, et al. (1997) Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the Alzheimer's Disease Assessment Scale that broaden its scope. The Alzheimer's Disease Cooperative Study. *Alzheimer Dis Assoc Disord,* **11 Suppl 2**:S13–21.

Morris JC (1993) The Clinical Dementia Rating (CDR): current vision and scoring rules. *Neurology.* **43**:2412-4

O'Bryant SE, Waring SC, Cullum CM, et al. (2008) Staging dementia using Clinical Dementia Rating Scale Sum of Boxes scores: a Texas Alzheimer's research consortium study. *Arch Neurol* **65**:1091-5.

Sikkes SA, Knol DL, Pijnenburg YA et al. (2013) Validation of the Amsterdam IADL Questionnaire, a new tool to measure instrumental activities of daily living in dementia. *Neuroepidemiology* **41**:35-41.

Sperling RA, Jack CR, Black SE, et al. (2011) Amyloid-related imaging abnormalities in amyloid-modifying therapeutic trials: recommendations from the Alzheimer's Association Research Roundtable Workgroup. *Alzheimers Dement* **7:**367-85.

Tolar M, Abushakra S, Hey JA, et al. (2020) Aducanumab, Gantenerumab, BAN2401, and ALZ-801—the first wave of amyloid-targeting drugs for Alzheimer's disease with potential

for near term approval. *Alzheimers Res Ther* **12**:95.

Teng E, Li Y, Manser PT. et al. (2023) Cross-sectional and longitudinal assessments of function in prodromal-to-mild Alzheimer's disease: A comparison of the ADCS-ADL and A-IADL-Q scales. *Alzheimers Dement (Amst)* **15**:e12452.

US Department of Health and Human Services, FDA (February 2018). Early Alzheimer's Disease: Developing Drugs for Treatment Guidance for Industry, Draft Guidance.

US Department of Health and Human Services, FDA (March 2024). Early Alzheimer's Disease: Developing Drugs for Treatment Guidance for Industry, Draft Guidance.

Van Dyck CH, Swanson CJ, Aisen P, et al. (2023) Lecanemab in early Alzheimer's disease. *N Engl J Med* **388**:9-21.

Van Ginkel JR, Sijtsma K, et al. (2010) Incidence of missing item scores in personality measurement, and simple item-score imputation. Methodology: *Eur J Res Methods Behav Soc Sci* **6**:17-30.

Wimo A, Gustavsson A, Jonsson L, et al. (2013) Application of resource utilization in dementia (RUD) instrument in a global setting. *Alzheimers Dement* **9**:429-35.

[redacted]

12 APPENDICES

Appendix 1 Schedule of Evaluation (Protocol v3.0)

	Screening*			Double-Blind Treatment										
Visit (V)	**V 1a**	**V 1b**	**V 2**	**V 3 Phone**	**V 4**	**V 5**	**V 6 Phone**	**V 7**	**V 8**	**V 9**	**V 10**	**V 11 EOT**	**ET**	**Safety FU EOS**
Visit Time	**Scr – Part 1**	**Scr – Part 2**	**#Bsl Day1**	**W 2 Day15**	**W 6 Day43**	**W 13 Day92**	**W 20 Day141**	**W 26 Day183**	**W 39 Day274**	**W 52 Day365**	**W65 Day456**	**W 78 Day547**	**TBD**	**W 82 Day574**
Window	**Up to -13 weeks**	–	–	**± 2 d**	**± 7 d**	**± 14 d**	**± 2 d**	**± 14 d**	**± 14 d**	**± 14 d**	**± 14 d**	**± 14 d**	–	**± 7 d**
Informed consent/ caregiver/study partner consent[1]	X	X												
MMSE Score	X		X			X		X	X	X	X	X	X	
Prohibited medications	X													
Demographics, medical history	X													
APOE testing	X													
Vital signs[2]		X	X		X	X		X	X	X	X	X	X	X
Physical examination		X	X		X	X		X	X	X	X	X	X	X
Neurological examination		X						X		X		X	X	X
Height		X												
Weight		X	X		X	X		X	X	X	X	X	X	X
Prior/concomitant medications		X	X	X	X	X	X	X	X	X	X	X	X	X
AD diagnostic criteria		X												
RBANS		X												
Evidence of progressive memory loss over the last 12 months		X												
12-Lead ECG[3]		X			X	X		X		X		X	X	
Future potential genomic testing		X												
Safety lab tests[4]		X			X	X		X	X	X	X	X	X	
Serum FSH/hCG test[5]		X												
Immunology (HIV, Hep B S Ag, Hep C Ab, HCV RNA if applicable)		X												
Urinalysis		X			X	X		X		X		X	X	
Urine Pregnancy test[6]			X		X	X		X	X	X	X	X	X	
Urine drug screen		X												
Adverse events	X	X	X	X	X	X	X	X	X	X	X	X	X	X
MRI[7]		X						X		X		X	X	
Inclusion/exclusion criteria			X											
Randomization			X											
ADAS-Cog			X			X		X		X		X	X	
A-IADL			X					X		X		X	X	
CDR[8]		X						X		X		X	X	
DAD			X					X		X		X	X	

	Screening*			Double-Blind Treatment										
Visit (V)	**V 1a**	**V 1b**	**V 2**	**V 3 Phone**	**V 4**	**V 5**	**V 6 Phone**	**V 7**	**V 8**	**V 9**	**V 10**	**V 11 EOT**	**ET**	**Safety FU EOS**
Visit Time	**Scr – Part 1**	**Scr – Part 2**	**#Bsl Day1**	**W 2 Day15**	**W 6 Day43**	**W 13 Day92**	**W 20 Day141**	**W 26 Day183**	**W 39 Day274**	**W 52 Day365**	**W65 Day456**	**W 78 Day547**	**TBD**	**W 82 Day574**
Window	**Up to -13 weeks**	–	–	**± 2 d**	**± 7 d**	**± 14 d**	**± 2 d**	**± 14 d**	**± 14 d**	**± 14 d**	**± 14 d**	**± 14 d**	–	**± 7 d**
NPI			X					X		X		X	X	
QoL-AD			X							X		X	X	
RUD Lite			X							X		X	X	
C-SSRS[9]		X	X		X	X		X	X	X	X	X	X	X
CSF sampling[10]			X							X		X	X	
Plasma biomarkers			X		X	X		X	X	X	X	X	X	
PK sample collection[11]			X		X	X		X	X	X	X	X	X	
Time of last dose prior to clinic visit					X	X		X	X	X	X	X	X	
Study drug return, accountability, compliance					X	X		X	X	X	X	X	X	
Study drug dosing on site[12]			X		X	X		X	X	X	X	X		
Dispense study drug[13]			X		X	X		X	X	X	X			

ADAS-Cog: Alzheimer's Disease Assessment Scale – cognitive subscale; A-IADL: Amsterdam Instrumental Activities of Daily Living; APOE: apolipoprotein E; BP: blood pressure; Bsl: baseline; CDR: Clinical Dementia Rating; CSF: cerebrospinal fluid; C-SSRS: Columbia-Suicide Severity Rating Scale; CT: computed tomography; d: Day; DAD: Disability Assessment for Dementia; ECG: electrocardiogram; EOS: end of study; EOT: end of treatment; ET: early termination; FSH: follicle-stimulating hormone; FU: Follow-up; hCG: human chorionic gonadotropin; HCV RNA: hepatitis C virus RNA; Hep B S Ag: hepatitis B surface antigen; Hep C Ab: hepatitis C antibody; HIV: human immunodeficiency virus; LP: lumbar puncture; MMSE: Mini-Mental State Examination; MRI: magnetic resonance imaging; NPI: Neuropsychiatric Inventory; PK: pharmacokinetics; QoL-AD: Quality of Life in Alzheimer's Disease; QTc: corrected QT (interval); Scr: Screen; RBANS: Repeatable Battery for the Assessment of Neuropsychological Status; RUD Lite: Resource Utilization in Dementia Lite version; TBD: to be determined; V: visit; W: week.

* Previous evidence of APOE4/4 status may be used as a reference and allow subjects to combine Screening – Par 1 Visit and Screening – Part 2 Visit together at the same day if the site and subject prefer. Screening procedures may be conducted over several visits during the Screening period provided that the results are available to evaluate inclusion and exclusion criteria before randomization.

1. Informed consent at Screening – Part 1 Visit is for APOE genotyping, MMSE, demographics, medical history and prohibited medications. Complete consent from subject and caregiver/study partner at the Screening – Part 2 Visit or at the Screening – Part 1 Visit if APOE4/4 status is known or confirmed using a rapid genotype test.
2. Vital sign measurements (BP, pulse, respiratory rate, and body temperature) will occur prior to dosing, LP, and blood draw, where applicable. BP and pulse will be measured after the subject has been sitting for 5 minutes. All BP measurements should be performed on the same arm, preferably by the same person, using an inflatable cuff and calibrated manometer. After last blood collections, subjects will be observed for approximately 15 minutes. BP and pulse while sitting will be measured and documented at the end of this observation period.
3. ECGs should be obtained after the subject has rested quietly in the supine position for at least 5 minutes. Triplicate ECG will be performed at screening, Week 6, and Week 13. Triplicate ECG at Week 26, Week 52, Week 78, and ET will be performed only if QTc prolongation observed in current and prior visits.
4. Lab tests include clinical chemistry, hematology, and coagulation tests, and will be done after ADAS-Cog and CDR-SB assessments.
5. Serum FSH levels will be measured only for female subjects of non-childbearing potential. Serum hCG levels will be measured only for female subjects of childbearing potential.
6. Urine pregnancy tests will be conducted only for female subjects of childbearing potential.
7. For those who cannot undergo MRI, CT is acceptable at Screening but requires prior approval by Medical Monitor. These subjects will not be scheduled for further CT imaging and will not contribute to the assessment of the volumetric imaging biomarkers. Note: CT is not allowed for subjects of sites at Germany.
8. At screening, CDR global score and memory box score are calculated for screening purposes; the sum of boxes score is recorded at subsequent visits.
9. C-SSRS: B/S = baseline/screening version, SLV = since last visit version.
10. This CSF sub-study intends to enroll 60 subjects per treatment group. For each subject, LP will be performed within the same 2-hour window on each of the 3 visits regardless of dosing time and other assessments. In other words, for any two of the three CSF collections, the maximal time difference, in terms of the time of day, is within 2 hours from each other. Time of CSF collection (not time of preparation) will be recorded.
11. Plasma PK samples will be collected at pre-dose at each clinic visit. Additional, a PK sample at ⩾ 1 hour post-dose during clinic visit will be collected at Week 65.
12. The times of study drug administration when taken at the study site will be recorded.
13. Last study drug administration will occur on the morning of the Week 78 visit. One dose is administered at the site during clinic visits; the evening dose is skipped on clinic visit days.

#Baseline MMSE used for change from baseline analyses will be the average of V1a and V2 scores.