Official Title: A Phase III, Multicenter, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group, Efficacy, and Safety Study of Gantenerumab in Patients With Early (Prodromal to Mild) Alzheimer's Disease

NCT Number: NCT03444870 (WN29922)

Document Date: SAP Version 3: 05-October-2022

STATISTICAL ANALYSIS PLAN

STUDY TITLE: **TWO PHASE III, MULTICENTER, RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED, PARALLEL-GROUP EFFICACY AND SAFETY STUDIES OF GANTENERUMAB IN PATIENTS WITH EARLY (PRODROMAL TO MILD) ALZHEIMER'S DISEASE**

STUDY NUMBER: WN29922, WN39658

STUDY NAME: GRADUATE I, GRADUATE II

VERSION NUMBER: *3*

ROCHE COMPOUND(S): Gantenerumab (RO4909832)

EUDRACT NUMBER: 2017-001364-38 (WN29922)
2017-001365-24 (WN39658)

IND NUMBER: 102,266

NCT NUMBER: NCT03444870 (WN29922)

PLAN PREPARED BY: [redacted]

STATISTICAL ANALYSIS PLAN AMENDMENT APPROVAL

SPONSOR: F. Hoffmann-La Roche Ltd
LEGAL REGISTERED ADDRESS: Grenzacherstrasse 124
4070 Basel, Switzerland

DATE FINAL: See electronic date stamp on the last page of this document

STATISTICAL ANALYSIS PLAN VERSION HISTORY

This SAP was developed based on Roche SAP model document v2.0, revised 26 October 2020.

SAP Version	Approval Date	Based on Protocol (Version, Approval Date)
3	see electronic date stamp on last page of the document	Version 5.0, 2 August 2021
2	29 July 2022	Version 5.0, 2 August 2021
1	12 October 2021	Version 5.0, 2 August 2021

STATISTICAL ANALYSIS PLAN AMENDMENT RATIONALE

Key changes to the SAP version 2.0, along with the rationale(s) for each change, are summarized below.

Section	Description of Change	Rationale for Change
Section 4	In Table 2, Neurogranin has been added to the definition of the CSF-longitudinal analysis set	It was a typo not including Neurogranin it in the previous version of the SAP.
Section 5.3.3	In the 'analysis step of the 'description if the primary estimator' the percent relative difference was updated to be defined as 100 times the estimated treatment effect divided by the absolute value of the placebo arm estimate	It was a typo not to considering the 'absolute value' for the denominator in the previous version of the SAP.
Section 5.3.5.2	A new supplementary analysis has been added, 'Treatment effect in relationship to a post-baseline timepoint', to estimate the percent relative difference in CDR-SB between week 52 and week 116, as well between week 24 and week 116 and week 76 and week 116	In order to assess the treatment effect using different starting points relative to the scheduled start of the target dose (with Weeks 24, 52 and 76 being respectively 12 weeks before, 16 weeks after and 40 weeks after the scheduled start of the target dose). This analysis may help to better understand the time dynamics of a treatment effect.
Section 5.3.6	A new section was added to describe what to do in case of the rare circumstances that the imputation model (see step 1, imputation model, in Section 5.3.3) would fail to converge.	To add clarity
Section 5.4	More details have been added to describe the estimator used for the secondary efficacy endpoints.	To add clarity
Section 5.4.1	Given that the 'confirmatory secondary endpoints' are a subset of the secondary endpoints, now the SAP is referencing to the secondary endpoints for the details of the estimand's strategy,	To add clarity
Section 5.5	More details have been added to describe the estimator used for the exploratory efficacy endpoints.	To add clarity
Section 5.7.7	The CSF biomarkers analysis plan has been updated to now use two different	Although those two models will provide the same exact p-

	ANCOVA models: one based on the change from baseline and the other one based on the value at visit	value in characterizing the treatment effect, they also allow to estimate complementary summary statistics.
Section 5.7.8	The plasma biomarkers analysis plan has been updated to now use two different MMRM models: one based on the change from baseline and the other one based on the value at visit	Although those two models will provide the same exact p-value in characterizing the treatment effect, they also allow to estimate complementary summary statistics.

Additional minor changes have been made throughout the document to improve clarity and consistency.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation or Term	Description
Aβ	amyloid-beta
AD	Alzheimer's disease
ADA	anti-drug antibody
ADAS-Cog11 / Cog13	Alzheimer disease assessment scale – cognition, subscale 11 / 13
ADCS-ADL	Alzheimer disease cooperative study - activities of daily living
ADL	activities of daily living
AE	adverse event
AESI	adverse event of special interest
ANCOVA	analysis of covariance
APOE	apolipoprotein ε
ARIA-E	amyloid-related imaging abnormalities – edema/effusion
ARIA-H	amyloid-related imaging abnormalities – hemosiderin deposition
CDR-GS	clinical dementia rating – global score
CDR-SB	clinical dementia rating – sum of boxes
CIR	copy increments from reference
CNS	central nervous system
COVID-19	coronavirus disease 2019
CSF	cerebrospinal fluid
CSR	Clinical Study Report
C-SSRS	Columbia-suicide severity rating scale
DTI	diffusion tensor imaging
ECG	Electrocardiogram
eCRF	electronic case report form
EMA	European Medicines Agency
EQ-5D	EuroQol-Five Dimensions
FAQ	functional activities questionnaire
FDA	(U.S.) Food and Drug Administration
ICE	intercurrent event
ICH	International Council on Harmonization
iDMC	independent Data Monitoring Committee
IRC	independent review charter

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS (Cont'd)

Abbreviation or Term	Description
ISR	injection-site reaction
ITT	intent to treat
IxRS	interactive voice/web-based response system
MAR	missing at random
MedDRA	Medical Dictionary for Regulatory Activities
mITT	modified intent to treat
MMRM	mixed effects model repeated measures
MMSE	mini-mental state examination
MNAR	Missing not at random
MRI	magnetic resonance imaging
NMPA	National Medical Products Administration
NPI-Q	neuropsychiatric inventory-questionnaire
NSDCR	not study drug or condition related
OLE	open-label extension
PET	Positron emission tomography
PK	Pharmacokinetic
pTau	phosphorylated tau
Q2W	every 2 weeks
Q4W	every 4 weeks
QoL	quality of life
QoL-AD	quality of life – Alzheimer's disease
RUD-Lite	resource utilization in dementia - lite
SAE	serious adverse events
SAP	Statistical Analysis Plan
SC	Subcutaneous
SDCR	study drug or condition related
SUV	standard uptake value
SUVR	standard uptake value ratio
tTau	total tau
ZCI-AD	Zarit caregiver interview – Alzheimer's disease

1. INTRODUCTION

This document describes the statistical analyses that will be reported in the primary Clinical Study Reports (CSR) of Studies WN29922 (hereafter referred to as "GRADUATE I") and WN39658 (hereafter referred to as "GRADUATE II"). The descriptions, methodology, and analyses presented in this document applies to both studies unless otherwise specified. The efficacy estimands and safety endpoints that will be the basis for comparing the two treatment arms will be defined in full in this document along with the populations of participants that are to be used in the analyses.

This Statistical Analysis Plan (SAP) covers analyses planned for the double-blind treatment period and the safety follow-up across both studies. Analyses for the OLE phase will focus on safety and will be listed directly in the corresponding List of Planned Outputs (LoPO). Pharmacokinetic (PK) data will be reported in a separate population PK report and thus is not covered in this document. Similarly, health economic data (such as utility values derived from the EQ-5D-5L and the RUD-lite) will be analyzed and reported separately from the Clinical Study Report and are therefore not covered in this document.

The description of layouts for the CSR outputs, the details about the underlying analysis datasets and programs, and the linking of production outputs to sections in the CSR are not within the scope of this document and will be covered in separate documents, i.e., Data Analysis Plan Module 2 and 3.

The language used in this SAP supersedes that in the protocol and protocol synopsis.

An early draft of this SAP was presented to U.S. Food and Drug Administration (FDA) in the context of a Type C meeting (Written Response Only procedure, December 21, 2020, Ref ID: 4720726) and to the European Medicines Agency (EMA) in the context of Scientific Advice procedure (EMA Written Advice received on 29 January 2020; EMA//SA/0000046418). These procedures focused on the proposed primary estimand, the estimator and other aspects of the analysis plan. Both agencies in principle agreed on the primary question of interest in the context of the estimand framework (ICH E9[R1]). There was also agreement on the proposed hierarchy of secondary endpoints.

A more advanced version of the SAP was submitted to the FDA for their review in December 2021. The feedback received during these health authority interactions was duly considered and informed the current version of this SAP.

1.1 OBJECTIVES

This study will evaluate the efficacy and safety of gantenerumab compared with placebo in participants with early (prodromal to mild) AD.

Table 1 Objectives and Corresponding Endpoints

Primary Objective(s)	Corresponding Endpoint(s)
• To evaluate the efficacy of gantenerumab administered by SC injection compared with placebo	• The change in global outcome from baseline (Day 1) to Week 116, as measured by the CDR-SB
Secondary Objective(s)	**Corresponding Endpoints**
• To evaluate the efficacy of gantenerumab versus placebo on cognition and/or function	The change from baseline to Week 116 in cognition and/or function, as measured by: • MMSE total score • ADAS-Cog11 and ADAS-Cog13 • Verbal Fluency Task • Coding • FAQ • ADCS-ADL total score and instrumental score
Exploratory Objective(s)	**Corresponding Endpoints**
• To evaluate the efficacy of gantenerumab versus placebo	The change from baseline to Week 116 in the following: • Clinically evident decline as measured using the CDR • Severity, as assessed by the CDR Global Score • Dependence level, as derived from the ADCS-ADL score • Health-related quality of life, as assessed by the QoL-AD scale • Behavioral and neuropsychiatric symptoms of AD, as measured by the NPI-Q • Study partner burden, as assessed by the ZCI-AD scale • Elements of resource utilization, as assessed by the RUD-Lite • Health outcomes in participant and study partner, as measured by EQ-5D questionnaire

Table 1 Objectives and Corresponding Endpoints (cont.)

Safety Objective(s)	Corresponding Endpoints
• To evaluate the safety of gantenerumab compared with placebo	• Nature, frequency, severity, and timing of adverse events and serious adverse events • Physical examinations (including neurological systems), vital signs, blood tests, ECGs, and C-SSRS • Nature, frequency, severity, and timing of MRI findings: ARIA-E and ARIA-H • Nature, frequency, severity, and timing of injection-site reactions • Presence of ADAs during the study relative to the presence of ADAs at baseline (in active drug group only)
Pharmacodynamic Biomarker Objective	**Corresponding Endpoints**
• To evaluate the effect of gantenerumab compared with placebo	• Change from baseline to Week 116 in brain amyloid load, as measured by amyloid PET scan in a subset of participants • Change from baseline to Week 116 in brain tau load, as measured by tau PET scan in a subset of participants • Change from baseline to Week 116 in cerebrospinal fluid markers of disease in a subset of participants, including, but not limited to total tau, and phosphorylated tau
Exploratory Biomarker Objective	**Corresponding Endpoints**
• To evaluate the effect of gantenerumab compared with placebo	• Change over time in plasma and other CSF biomarkers • Change from baseline to Week 116 in functional brain connectivity, as measured by resting-state functional MRI (where available) • Change from baseline to Week 116 in integrity of white matter, as measured by DTI-MRI (where available) • Change in MRI-derived measurements over time, such as volumetric changes in whole brain, ventricles, hippocampus, or other structures, in all participants

AD = Alzheimer's disease; ADA = anti-drug antibody; ADAS-Cog11 = Alzheimer Disease Assessment Scale-Cognition, Subscale 11; ADAS-Cog13 = Alzheimer Disease Assessment Scale-Cognition, Subscale 13; ADCS-ADL = Alzheimer Disease Cooperative Study Group-Activities of Daily Living; ARIA-E = amyloid-related imaging abnormalities-edema/effusion; ARIA-H = amyloid-related imaging abnormalities-hemosiderin deposition; CDR = Clinical Dementia Rating; CDR-GS = Clinical Dementia Rating global score; CDR-SB = Clinical Dementia Rating-Sum of Boxes; CSF = cerebrospinal fluid; C-SSRS = Columbia-Suicide Severity Rating Scale; DTI = diffusion tensor imaging; EQ-5D = EuroQol-Five Dimensions; FAQ = Functional Activities Questionnaire; MMSE = Mini-Mental State Examination; MRI = magnetic resonance imaging; NPI-Q = Neuropsychiatric Inventory-Questionnaire; PET = positron emission tomography; QoL-AD = Quality of Life-Alzheimer's Disease; RUD-Lite = Resource Utilization in Dementia-Lite; SC = subcutaneous; ZCI-AD = Zarit Caregiver Interview-Alzheimer's Disease.

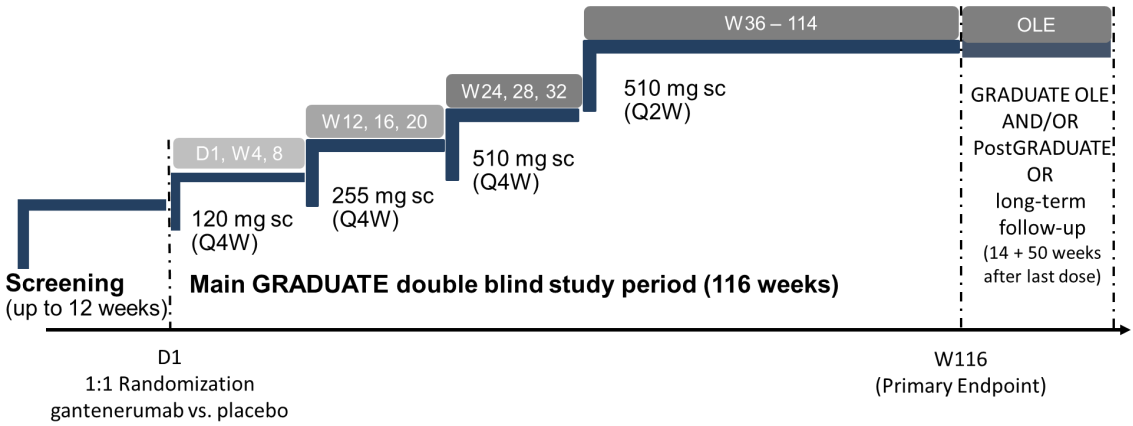
1.2 STUDY DESIGN

GRADUATE I and GRADUATE II are two identical Phase III, multicenter, randomized, double-blind, placebo-controlled, parallel-group studies designed to evaluate the efficacy and safety of gantenerumab in participants with early (prodromal to mild) AD.

The planned number of participants for the global enrollment phase for each study is approximately 1016 participants randomized in a 1:1 ratio to receive active drug or placebo (508 participants randomized to active drug and 508 randomized to placebo; see Section 3). To maintain a balanced number of participants enrolled in each treatment arm and to ensure that the arms are comparable, randomization will be stratified by stage of the disease (prodromal AD vs. mild AD), apolipoprotein ε4 (APOE) allele status (presence vs. absence of the ε4 allele), use of AD medication (presence vs. absence), geographic region (Western Europe and Australia vs. Rest of the World vs. North America), and participation in the longitudinal amyloid and tau PET substudies. Approximately 175 centers in approximately 15 countries worldwide will participate in these studies.

Due to the global impact of the coronavirus disease (COVID-19) pandemic and the resulting disruption in study drug administration, the duration of the double-blind treatment period was extended by 12 weeks, with Protocol Amendment 4. The optional scenario of a further extension of 12 weeks – resulting in a final efficacy and safety visit at Week 128 – was not implemented. An overview of the study design is provided in Figure 1.

Figure 1 Overall Study Design



Each study consists of three distinct periods:

Screening (including an optional prescreening): The screening period may last up to 12 weeks for each eligible participant.

Double-blind treatment period: After screening, participants who meet all eligibility criteria will be randomly assigned to one of two arms (active drug or placebo) in a 1:1 ratio. Following baseline assessments, each participant will receive a minimum of 9 subcutaneous (SC) administrations every 4 weeks (Q4W) of study drug (uptitration period), followed by up to 40 administrations every 2 weeks (Q2W) of study drug at target dose in the double-blind treatment period. The last dose of study drug will be administered at Week 114. At the end of the double-blind treatment period, 2 weeks after the last dose, all participants will undergo the final efficacy and safety study visit. Participants who have already completed the double-blind treatment period prior to implementation of Protocol Version 4 will have received up to 34 SC Q2W administrations of study drug during the double-blind treatment period. The last dose will have been administered at Week 102, and their final efficacy and safety visit will be at Week 104.

The Sponsor will emphasize to Investigators the importance of collecting data for the primary endpoint through Week 116, even if participants withdraw from treatment but do not withdraw from the study.

Post-double-blind treatment period: After the final efficacy and safety study visit for the double-blind treatment period, all participants will be asked to come back for the long-term follow-up visits or to continue in the open-label extension (OLE). Participants will either directly enter the separate WN42171 (hereafter referred to as "POSTGRADUATE") OLE study or enter a parent study OLE period. If entering the parent study OLE period, they will be required to complete the uptitration period (a minimum 36 weeks) following which they will then be able to roll over to the

POSTGRADUATE OLE study. This second option is for participants at sites where POSTGRADUATE is not yet approved when they have reached the end of the double-blind treatment period.

China Enrollment Plan

Based on historical data, participant recruitment is expected to take longer in China and therefore, a specific China enrollment plan has been established. Thus, if at least 1 participant is enrolled at sites in mainland China, Hong Kong, and Taiwan that are recognized by the National Medical Products Administration (NMPA) during the global enrollment phase, additional participants may be enrolled in a China extension to ensure a total enrollment that is sufficient to support registration in China. Only participants enrolled at NMPA-recognized sites, during the global enrollment phase, will be included in the primary analysis. All participants enrolled at NMPA-recognized sites, either during the global enrollment phase or the China extension phases, will be included in a China-specific analysis.

The China-specific analysis will be described in a separate SAP and therefore is not covered in this document.

1.2.1 Treatment Assignment and Blinding

Randomization will be performed centrally using an interactive voice or Web-based response system (IxRS). After screening, participants who meet all eligibility criteria will be randomly assigned to one of two treatment groups (active drug or placebo). The ratio will be 1:1; one active to one placebo. The randomization method will be stratified block-randomization. Randomization to treatment allocation will be stratified by geographic region (Western Europe and Australia vs. Rest of the World vs. North America), participant APOE ε4 status (carrier vs. non-carrier), participant stage of disease (prodromal vs. mild AD), use of AD medication (presence vs. absence), and participation in the longitudinal amyloid and tau PET substudies. Except in circumstances in which a Health Authority, Ethics Committee, or Institutional Review Board requires it, a participant will not be told of his or her APOE ε4 status. Individual participant APOE ε4 genotype results will be blinded to participants, Investigators, and the Sponsor. APOE ε4 status information will be supplied directly to the IxRS vendor by the central testing laboratory so that the information can be incorporated at the time of randomization. For participants for whom APOE ε4 status is already known, the results will be blinded to the Sponsor and, as much as possible, to the site and central MRI reader.

The study is to be conducted in a double-blind manner to minimize potential bias from Investigators and participants. The Sponsor will be blinded to study treatment. In the OLE phase, the Sponsor, participants, and site staff will remain blinded to previous treatment allocation.

The randomization method implemented in the China extension cohort will be the same as that implemented in the global population.

1.2.2 Independent Review Facility

MRI scans will be reviewed by a central MRI core laboratory, which will perform the diagnostic reads and assessment of MRI outcome measures.

Central facilities will be used for PET assessments (see Independent Review Charter [IRC]).

1.2.3 Data Monitoring Committee

The iDMC will evaluate participant safety and efficacy on a regular basis. In addition to the data defined in the iDMC Charter (e.g., the incidence and nature of adverse events [AE], serious adverse events [SAE], adverse events of special interest [AESI], amyloid-related imaging abnormalities-edema/effusion [ARIA-E], amyloid-related imaging abnormalities-hemosiderin deposition [ARIA-H], and injection-site reactions [ISR]), the iDMC will review all necessary cumulative data at regular intervals during the study, including efficacy when needed. It is anticipated that these assessments will occur approximately every 3 months. At the time of each review, the iDMC will make appropriate recommendations (e.g., the study should continue as planned, enrollment in a specific arm should be discontinued, the treatment regimen should be modified, the protocol should be amended, enrollment should be held pending further safety evaluations). The iDMC will also evaluate planned or unplanned interim analyses for efficacy or futility.

2. STATISTICAL HYPOTHESES

The primary efficacy analysis will compare the active drug arm to the placebo arm at Week 116 with a two-sided test corresponding to the following null hypothesis, H_0, and alternative hypothesis, H_1:

$$H_0: \mu_{active} = \mu_{placebo}$$

$$H_1: \mu_{active} \neq \mu_{placebo}$$

Where μ_{active} and $\mu_{placebo}$ are the mean change from baseline to Week 116 in the CDR-SB score for each arm.

3. SAMPLE SIZE DETERMINATION

Determination of sample size is based on participants enrolled in the global enrollment phase. In each study, approximately 1016 participants will be enrolled and randomized in 1:1 ratio to each treatment arm (active drug or placebo) during the global enrollment phase. The original planned sample size was of 760 participants, but it was increased to 1016 participants based on considerations from external studies (in protocol version 3).

Additional participants may be randomized during the China extension if at least 1 participant is enrolled at sites in mainland China, Hong Kong, or Taiwan that are recognized by the NMPA during the global enrollment phase.

The estimated sample size required to demonstrate efficacy with regard to Clinical Dementia Rating-Sum of Boxes (CDR-SB) is based on the following assumptions:

- the mean change in the CDR-SB from baseline to Week 104 is 2.5 points in the placebo arm
- a common standard deviation across treatment arms for change of approximately 2.97 from baseline to Week 104 in the CDR-SB
- active drug has a true effect of a 30% relative reduction in deterioration of CDR-SB

Based on these assumptions, and using a student's T-test with equal variance, a sample size was calculated for 90% power to detect a true treatment effect of 30% relative reduction in deterioration of CDR-SB at a two-sided α of 0.05.

Uncertainties on parameters affecting the power (e.g., dropout rate, study drug discontinuation rate, standard deviation, and placebo decline) were assumed to be equivalent to a drop in sample size of up to around 35% and were corrected for accordingly. This calculation leads to a total of 1016 participants in the study.

At the date of writing Protocol Version 4, it was expected that participants would miss an average of 8 weeks of study drug administration over the course of the original two-year study due to the COVID-19 pandemic. This had the potential to decrease the power of the study from ~90% to ~80%. To mitigate the impact of missed administrations, the double-blind treatment period has been extended by 12 weeks. Under current assumptions, this brings the study power back to the originally planned ~90%.

4. ANALYSIS SETS

The following analysis sets are defined:

Table 2 Analysis Sets

Analysis set	Definition	Scope
All enrolled participants	All participants randomized during the global enrollment phase whether or not the participant received the assigned treatment. Analysis using this analysis set will be performed by randomized treatment.	Participant disposition report will be based on this analysis set
Intent-to-treat (ITT)	All participants randomized during the global enrollment phase, who received at least one dose of study drug. Analysis using this analysis set will be performed by randomized treatment.	All efficacy analyses, including the primary estimand, will be based on this analysis set.
CSF modified intent-to-treat (CSF-mITT)	All participants in the ITT analysis set who had at least one valid quantitative cerebrospinal fluid (CSF) measurement Analysis using this analysis set will be performed by randomized treatment.	All analyses of CSF biomarkers will be based on this analysis set
CSF-longitudinal	All participants in the CSF-mITT who had at least one valid quantitative Week 116 CSF measurement of phosphorylated tau (pTau-181), tTau*, Neurogranin* or NFL.	Summary of treatment group comparability will also be repeated on this analysis set.
Plasma-longitudinal	All participants in the ITT analysis set who had at least one valid quantitative post-baseline plasma measurement of Amyloid beta (1-42) (Abeta-42) or phosphorylated tau (pTau-181)	Summary of treatment group comparability will also be repeated on this analysis set.
MRI modified intent-to-treat (MRI-mITT)	All participants in the ITT analysis set who had at least one valid volumetric MRI quantitative measurement. Analysis using this analysis set will be performed by randomized treatment.	All analyses of volumetric MRI parameters will be based on this analysis set.
MRI-longitudinal	All participants in the MRI-mITT with at least one valid post-baseline volumetric MRI quantitative measurement.	Summary of treatment group comparability will also be repeated on this analysis set.
Tau PET modified intent-to-treat (Tau-PET-mITT)	All participants in the ITT analysis set who participated in the Tau PET sub-study and who had at least one Tau PET scan with a valid quantitative measurement and who did not withdraw from the Tau PET substudy before randomization Analysis using this analysis set will be performed by randomized treatment.	All analyses of Tau PET parameters will be based on this analysis set.
Tau-PET-longitudinal	All participants in the Tau-PET-mITT with a valid post-baseline quantitative Tau PET measurement.	Summary of treatment group comparability will also be repeated on this analysis set.

Table 2 Analysis Sets (cont.)

Analysis set	Definition	Scope
Amyloid PET modified intent-to-treat (Amyloid-PET-mITT)	All participants in the ITT analysis set who participated in the Amyloid PET sub-study and who had at least one Amyloid PET scan with a valid quantitative measurement performed with either florbetaben or flutemetamol and who did not withdraw from the Amyloid PET substudy before randomization. Analysis using this analysis set will be performed by randomized treatment.	All analyses of Amyloid PET parameters will be based on this analysis set
Amyloid-PET-longitudinal	All participants in the Amyloid-PET-mITT with a valid post-baseline quantitative amyloid PET measurement performed with either florbetaben or flutemetamol.	Summary of treatment group comparability will also be repeated on this analysis set.
Safety-evaluable	All participants randomized during the global enrollment phase who received at least one dose of study drug. Any participant randomized to placebo who received at least one dose (any dose) of active drug will be summarized as having received the active drug. Analysis using this analysis set will be performed by treatment actually received.	All safety analyses (with the exception of Safety MRI) will be based on this analysis set.
MRI Safety-evaluable	All participants in the Safety-evaluable analysis set who had at least one post-baseline safety MRI scan. Analysis using this analysis set will be performed by treatment actually received	All analyses of safety MRI will be based on this analysis set.
Amyloid-PET Safety Evaluable	All participants in the Safety-evaluable analysis set who received at least one dose of radioligand Analysis using this analysis set will be performed by treatment actually received	All safety analysis for the amyloid-PET sub-study will be based on this analysis set.
Tau-PET Safety Evaluable	All participants in the Safety-evaluable analysis set who received at least one dose of radioligand. Analysis using this analysis set will be performed by treatment actually received	All safety analysis for the tau-PET sub-study will be based on this analysis set.

5. STATISTICAL ANALYSES

5.1 GENERAL CONSIDERATIONS

The clinical cutoff date for the primary analysis is defined by the date of the last randomized participant plus 116 weeks.

In the following sections, for all continuous variables for which descriptive statistics are indicated, the following statistics will be reported: the number of observations, the mean, median, standard deviation, and minimum and maximum. The 25th and 75th percentiles

(Q1 and Q3) will also be reported for selected tables. Descriptive summaries of discrete data will include frequencies expressed in terms of number and percentage of participants.

For clinical endpoints and biomarkers assessments, the baseline is defined as the assessment taken at the baseline visit (Day 1) up to and including the day of first study drug intake. If no assessment is reported at the baseline visit up to the day of first study drug intake, the earliest assessment reported after Day 1 and up to the day of second dose or Day 35, whichever is earlier, will be used as baseline. If no assessment is reported either at baseline visit or up to Day 35, an assessment reported at screening will be used as baseline. Day 35 is the latest timepoint allowed for Week 4 visit as per protocol.

For all other assessments, the baseline is defined as the last available assessment prior to first study drug intake, unless specified otherwise.

For biomarkers (CFS and plasma), values below the lower limit of quantification (LLOQ) will be set to 0.5*LLOQ, while values above the upper limit of quantification (ULOQ) will be set to ULOQ. A summary table will be provided to summarize the values below LLOQ and above ULOQ for each treatment arm.

The efficacy analyses will be based on the ITT analysis set (see Table 2) and will compare the active drug arm against the placebo arm with regards to mean change from baseline to Week 116. Two-sided test hypotheses will be defined in the following sections and the type I error level will be 5%. There are two identical Phase III studies, for each respective study the type I error level will be 5%. In order to protect the overall type I error rate (i.e., at each study level) when incorporating the hypothesis testing of multiple endpoints into the analysis, a fixed sequence testing procedure (Westfall and Krishen 2001) will be used to adjust for multiple comparisons. Testing of each hypothesis will follow a pre-specified order such that an endpoint would only be tested if the preceding one in the hierarchy was significant at 5% alpha level. The endpoint hierarchy, starting with the primary endpoint and including only confirmatory secondary endpoints, is the following:

1. CDR-SB (Clinical Dementia Rating, Sum of Boxes)
2. ADAS-Cog 13 (Alzheimer's Disease Assessment Scale, Cognitive subscale, 13-item)
3. ADCS-ADL (Alzheimer's Disease Cooperative Study, Activities of Daily Living scale) total score
4. FAQ (Functional Activities Questionnaire)

Missing outcome data will be handled using data imputation aligned with the estimand, see Section 5.3.3 for a detailed description of the analysis strategy.

When using a statistical model with baseline covariate adjustment, missing baseline covariate data other than the baseline outcome measure will be imputed to the overall median value for continuous covariates, or will be imputed to the most frequent category for categorical variables. In addition, baseline covariates will be derived from information collected into the eCRF, unless otherwise specified.

In statistical models using change from baseline of a given outcome measure as the dependent variable, there will be no imputation of the baseline outcome value, with the consequence that participants missing the baseline outcome measure will not contribute to the analysis.

The impact of the COVID-19 pandemic on the study and its conduct will be monitored and the overall impact will be assessed and described in the Clinical Study Reports.

5.2 PARTICIPANT DISPOSITION

The analysis of participant disposition will be based on all enrolled participants (see Analysis sets in Table 2). The number of participants enrolled will be tabulated by country, site, and treatment arm. Participant disposition (the number of participants randomized and completing the different study periods) will be tabulated by treatment arm. Premature study drug discontinuation and study discontinuation, as well as reasons for discontinuations, will be summarized.

Major protocol deviations, including major deviations with regard to the inclusion and exclusion criteria, will be summarized by treatment arm. Separate tables will be provided for COVID-19-related major protocol deviations and reasons for COVID-19-related major protocol deviations.

5.3 ENDPOINT ANALYSIS

5.3.1 Definition of Primary Endpoint

As detailed in the study protocols, the primary endpoint is the change from baseline to Week 116 in the CDR-SB which is a global scale covering both functional and cognitive domains.

The CDR-SB is a detailed quantitative general index that is scored from 0 to 18 with higher scores indicating greater impairment. The CDR characterizes six domains: memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care. The necessary information to make each rating is obtained through a semi-structured interview with the patient and a reliable informant or collateral source (e.g., a study partner).

5.3.2 Definition of Primary Estimand

The clinical question of interest is to assess the effect of the active drug on disease progression at Week 116, irrespective of use or initiation of symptomatic treatments for AD, in the absence of a substantial impact of the COVID-19 pandemic.

In alignment with the Addendum to ICH E9, the primary efficacy estimand is described by the following attributes:

- **Target Population**:
  Early (prodromal to mild) AD population as identified by the inclusion and exclusion criteria, see Sections 4.1.1 and 4.1.2 of the study protocols
- **Variable**:
  Change from baseline to Week 116 in Clinical Dementia Rating – Sum of Boxes (CDR-SB)
- **Treatment**:
  Prescribed study drug including uptitration to the target dose and safety-related dose modifications, irrespective of use or initiation of symptomatic treatment for AD
- **Population-level summary**:
  Difference in variable means between treatment arms

Intercurrent events (ICE):

The ICEs are classified into two categories: those that are Study Drug or Condition Related (SDCR) and those that are not (NSDCR). The list of main anticipated ICEs, and their classification as SDCR or NSDCSR, is presented below (Table 3). The final list of ICEs may need to be adapted in case unanticipated ICEs emerge during the study conduct. The classification of each ICE into SDCR or NSDCR will be completed and documented prior to the unblinding.

The ICE of substantial reduction in drug exposure due to the COVID-19 pandemic is defined as 4 or more dose-months (i.e., 16 weeks, not necessarily consecutive) of treatment missed due to COVID-19-related reasons. This definition will be equally applied to placebo and active arms. One dose-month is defined as 4 weeks of dosing, i.e., one dose with a Q4W dosing frequency (mostly during uptitration) or two doses with a Q2W dosing frequency (at target dose). The threshold of four missed dose-months was determined based on the following reasons:

– A 12-week extension to the study was implemented (in protocol amendment version 4) to mitigate the impact of the COVID-19 pandemic. Therefore, treatment interruptions up to 3 dose-months (i.e., 12 weeks) are already accounted for in the study design.

– Based on the half-life of approximately 24 days of gantenerumab, plasma concentration after a 4 months' interruption at the target dose is expected to be below the observed concentration with the dose of 225 mg Q4W which was

shown to be ineffective in Studies WN25203 (SCarlet RoAD) and WN28745 (Marguerite RoAD).

– A 20% difference is the usual acceptability threshold to establish PK bioequivalence. The protocol includes 20 dose-months (80 weeks) at target dose after the up-titration period. Missing less than 4 dose-months results in an overall reduction in drug exposure of less than 20% of the total planned dose.

As a conservative approach, all withdrawals from study treatment due to an AE will be classified as SDCR for the purpose of the primary analysis. This includes withdrawal from study treatment due to suspected or confirmed COVID-19 infection AE, because the relationship of these events to the participant condition may be ambiguous.

All SDCR ICEs will be handled with a treatment policy strategy, while NSDCR ICEs will be handled with a hypothetical strategy. The frequencies of ICE will be summarized by treatment arm.

In this study, given the disease stage at baseline of the target population, death is expected to be a rare event and mostly not considered related to treatment or disease progression, and as such the corresponding ICE of death will be handled with a hypothetical strategy for the primary estimand. Of note, a supplementary estimand which defines all ICEs, including death, as SDCR (and thus using the same imputation strategy as for the other SDCR ICEs; see Section 5.3.3), is described in Section 5.3.5.

Table 3 Intercurrent Events Impacting Primary Analysis

Intercurrent Event	SDCR/NSDCR	Estimand Approach
Withdrawal from study treatment due to lack of efficacy	SDCR	Treatment Policy
Withdrawal from study treatment due to safety or tolerability reason (NOTE: This will include discontinuation due to AE, incl. suspected or confirmed COVID-19 AEs)	SDCR	Treatment Policy
Withdrawal from study treatment with no informative reason given	SDCR	Treatment Policy
Withdrawal from study treatment due to the COVID-19 pandemic	NSDCR	Hypothetical Strategy
Substantial reduction in drug exposure due to the COVID-19 pandemic (defined as $\geq$ 4 missed dose-months)	NSDCR	Hypothetical Strategy
Withdrawal from study treatment due to purely administrative reason	NSDCR	Hypothetical Strategy
Death	NSDCR	Hypothetical Strategy
Withdrawal from study treatment due to use or initiation of protocol prohibited medication	SDCR	Treatment Policy
Withdrawal from study treatment due to other SDCR ICEs	SDCR	Treatment Policy

AE = adverse event; COVID-19 = coronavirus disease 2019; NSDCR = Non Study Drug or Condition Related; SDCR = Study Drug or Condition Related

ICE Derivation

Identification and categorization of the ICEs will be done based on fully blinded data only and will be finalized before database lock.

The number of missed dose-months due to the COVID-19 pandemic will be derived from the standardized eCRF data fields. More specifically, any Q4W missed dose due to COVID will count as 1 missed dose-months, while any Q2W missed dose due to COVID will count as 0.5 missed dose-months. The information whether a dose was missed due to COVID will be derived from the corresponding “reason for missed visit” and “reason for missed dose” fields in the eCRF.

All occurrences of “withdrawal from study treatment” ICE, as per Table 3, will be identified and categorized as SDCR/NSDCR based on the standardized reason reported in the eCRF “Study Drug Completion/Early Discontinuation” form. In case of ambiguity (i.e., if the reason for study drug discontinuation as reported in the eCRF is either “Protocol deviation”, “Withdrawal by subject”, “Physician decision” or “Other”), an adjudication committee will review the dedicated eCRF free text field and assign the withdrawal to a pre-specified ICE and a corresponding SDCR or NSDCR category. The

adjudication committee may also introduce additional ICEs to the list, in case where no appropriate fit is found in Table 3 or Appendix 2 . The adjudication committee members must not have been involved in the conduct of Studies GRADUATE I and GRADUATE II and must not have been exposed to unblinding data from these studies. The members of the adjudication committee must have signed the charter of the adjudication committee, in Appendix 3.

5.3.3 Main Analytical Approach for Primary Estimand and Primary Endpoint

Primary study hypothesis

The primary efficacy analysis for this study will test the superiority of the active drug over the placebo at Week 116 with a two-sided test corresponding to the following null hypothesis, H_0, and alternative hypothesis, H_1:

$$H_0: \mu_{active} = \mu_{placebo}$$

$$H_1: \mu_{active} \neq \mu_{placebo}$$

Where μ_{active} and $\mu_{placebo}$ are the mean changes from baseline to Week 116 in the CDR-SB score for each arm.

Time Windowing

For the primary endpoint (and in general for clinical efficacy endpoints) the following time windows will be used for analyses (see Table 4), based on study days. Study days are defined based on days on study since the date of the first dose, with the day of the first dose being study day 1.

Table 4 Time windows for clinical endpoint assessments

Visit	Target day	Time window (in study days*)
Baseline	1	Up to Day 35
Week 24	169	72, 266
Week 52	365	267, 448
Week 76	533	449, 630
Week 104	729	631, 770
Week 116	813	771, earliest between 897 or first dose of OLE

In cases where more than one assessment falls within a time window, the assessment with the date closest to the target day is selected whether it's a scheduled, unscheduled or an early termination visit.

With regards to an ICE of withdrawal from study treatment, an assessment will be regarded as having happened after the ICE of withdrawal from study treatment if it is collected >28 days since the last dose.

General strategy to address the ICEs

For the primary estimand, a treatment policy approach will be used for all SDCR ICEs. In line with the clinical question of interest, the aim is to estimate a treatment effect irrespective of the occurrence of these ICEs. This approach is largely in line with the EMA guideline on the clinical investigation of medicines for the treatment of AD (EMA 2018).

All NSDCR ICEs will be handled using a hypothetical approach. The aim is to estimate a treatment effect "as if" the ICE had not happened. Post-ICE outcome values compatible with a hypothetical strategy are not directly observable. Consequently, any observed outcome values after NSDCR ICEs will be removed and treated as missing data for analysis purposes.

For participants with multiple ICEs, the type of the first ICE will determine the strategy to be considered.

Missing data assumptions for the primary estimator

For intermittent missing data (i.e., for participants with non-missing Week 116 data but missing data at other visits), missing data not associated with an ICE (e.g., for participants who had completed Week 104 visit before protocol v4 was implemented), and for missing data after NSDCR ICEs (handled with a hypothetical strategy), the missing data are assumed to be similar to those from the other participants in the same treatment group with no such missing data. This is compatible with a missing-at-random (MAR) assumption.

All observed data after SDCR ICEs will be included in the analysis. If data after SDCR ICEs are missing, they will be assumed to be similar to those in the placebo group for both study arms. Specifically, data will be imputed based on the placebo arm using reference based imputation with a Copy Increments from Reference (CIR) assumption (Carpenter et al, 2013 and Cro et al, 2020). This approach appears conservative yet plausible for the study drug. CIR assumes that changes in the primary endpoint after the ICE in a participant randomized to active drug can be represented by, i.e., imputed from, that of participants randomized to placebo. It therefore assumes no treatment effect after the ICE. In the placebo arm, this is compatible with a MAR assumption whereas in the active drug arm, the imputation is under a Missing Not At Random (MNAR) assumption.

Further details about the implementation of the missing data imputation are provided below. Sensitivity analyses for the missing data assumptions are discussed in Section 5.3.4.

Description of the primary estimator

The primary estimator will be applied to the ITT analysis set (see Table 2) and it will be implemented using four steps. First, an imputation model will be fitted to the data. Second, imputation of missing data will be performed based on the parameter estimates from the imputation model. Third, the completed data will be analyzed using an analysis of covariance (ANCOVA) model. Finally, inference will be performed based on resampling techniques. All four steps are described and justified in a published manuscript (Wolbers et al. 2022) which provides a more detailed justification of the statistical methodology and supportive simulations. Considerations about the control of type-I error for the primary estimator and supportive simulations, mimicking the setting of the GRADUATE I and GRADUATE II studies and exploring an extended range of plausible scenarios for missing data, are provided in the Appendix 1.

1. Imputation model

The imputation model is a mixed effects model for repeated measures (MMRM) with the longitudinally assessed change from baseline in CDR-SB as the dependent variable. Its purpose is to estimate mean trajectories and covariance matrices of longitudinal outcomes in the placebo and active drug arms, respectively, while subjects remain on their randomized treatment. Therefore, all data after withdrawal from study treatment will be removed and considered as missing for the purpose of estimating the imputation model, and for this purpose only. If these data were not excluded, then the imputation model would estimate mean trajectories based on a mixture of observed pre- and post-discontinuation data. These would not be compatible with the CIR assumption employed in the subsequent imputation step, which requires combining mean trajectories while on active drug up to the ICE with increments while on placebo thereafter, respectively.

The imputation model includes the following covariates: treatment group, visit, and treatment-by-visit interaction, baseline CDR-SB score and baseline CDR-SB score-by-visit interaction, baseline ADAS-Cog 13 total score, baseline ADCS-ADL total score and the randomization stratification factors, namely: disease stage (from the eCRF), geographic region (from the IxRS), the use of AD medication at baseline (derived from the eCRF based on a search of medications; see below for more details) and the APOE ε4 status (from the Vendor). For geographic region, IxRS is preferred over the eCRF source, to keep the “starting” region for analysis, in case a participant moved to a different region during the study (the eCRF would otherwise reflect the “final” region). AD medication at baseline will be defined as any use of donepezil, galantamine, memantine or rivastigmine prior to randomization and with an

end date not before the randomization date. An unstructured variance-covariance structure will be applied to model the within-subject errors across visits. If the model fails to converge, then a heterogeneous Toeplitz covariance structure will be used instead and if this still fails, then a compound symmetry covariance structure will be used.

Imputations will be based on restricted maximum likelihood (REML) estimation of the regression and covariance parameters from the imputation model (von Hippel and Bartlett 2021; Wang and Robins 1998).

2. Imputation step

The imputation model implies a marginal multivariate normal distribution of the longitudinal outcome values across all visits based on a participant's assigned treatment arm and covariate values. This marginal imputation distribution will be used for all participants in the placebo arm and all participants in the active drug arm without an SDCR ICE. For participants in the active drug arm with an SDCR ICE, the mean of the marginal imputation distribution for outcomes after the SDCR ICE will be modified as per the CIR assumption (Carpenter et al. 2013).

For each participant, the conditional imputation distribution of their missing outcome values is defined as the marginal imputation distribution conditional on the participant's observed outcomes (including observed post-SDCR ICE outcome assessments). A single deterministic imputation using the conditional mean from the conditional imputation distribution for each participant with missing outcomes will be used.

3. Analysis step

The completed data (using conditional mean imputation as described above) will be analyzed using an ANCOVA model with the change from baseline in CDR-SB at the Week 116 visit as the dependent variable. This analysis model includes treatment group as the primary covariate with adjustment for the same set of covariates as for the imputation model described above, namely baseline CDR-SB score, baseline ADAS-Cog 13 total score, baseline ADCS-ADL total score and the randomization stratification factors namely: disease stage (from eCRF), geographic region (from the IxRS), the use of AD medication at baseline (derived from eCRF based on a search of medications; see below for more details) and the APOE ε4 status (from the Vendor). For geographic region, IxRS is preferred over the eCRF source, to keep the "starting" region for analysis, in case a participant moved to a different region during the study (the eCRF would otherwise reflect the "final" region). AD medication at baseline will be defined as any use of donepezil, galantamine, memantine or rivastigmine prior to randomization and with an end date not before the randomization date. The primary treatment effect estimator is defined as the regression coefficient associated with the treatment group. The treatment effect will be reported as a difference in adjusted means. The treatment

effect at the other visits will be estimated in the same way and will be reported as supplementary analyses of the primary endpoint (see Section 5.3.5).

Percent relative difference, defined as 100 times the estimated treatment effect divided by *the absolute value of the* placebo arm estimate, will also be reported, for all visits, as a point estimate for descriptive purposes. Importantly, ANCOVA is applied to a complete dataset after appropriate missing data imputation. For complete data, ANCOVA applied to outcomes from a single visit is equivalent to a more complex MMRM model. It can be demonstrated that it leads to identical parameter estimates as a corresponding MMRM model with an arbitrary covariance structure if separate regression coefficients are estimated at different visits for all covariates (Amemiya 1985, p. 197).

4. Inference step

Inference will be based on resampling techniques as recommended by von Hippel and Bartlett 2021. Specifically, the jackknife (Efron and Tibshirani, 1994) will be used to estimate the standard error of the primary treatment effect estimator and the test of the primary statistical hypothesis will be based on the corresponding *Z*-score. Compared to other resampling techniques, the jackknife has the advantage of providing a deterministic standard error estimate and, hence, removing any simulation randomness from the procedure.

5.3.3.1 Software Implementation and Validation

The reference based imputation methodology will be implemented by an internally developed R package "rbmi" ("reference-based multiple imputation"). The package will comply with the ICH guidance document "ICH E 9: Statistical principles for clinical trials" which states that: "The computer software used for data management and statistical analysis should be reliable, and documentation of appropriate software testing procedures should be available."

The testing strategy for the package consists of defining and documenting the expected input and output of each function and implementing unit tests to prove that the package performs as expected. All implemented methods will be recorded and referenced against the literature with unit tests and simulations put in place to show that known values can be recovered as well as showing consistency with similar software from other languages (most notably the "5-Macros" implemented by the Drug Information Association Scientific Working Group on Estimands and Missing Data in SAS).

The package, including all documentation and testing materials, is available on GitHub.com (https://github.com/insightsengineering/rbmi) and on CRAN repository (https://cran.r-project.org/package=rbmi) to allow for unrestricted access and enable public scrutiny of the code and methods. Description of the rbmi package is also available on a published manuscript (Gower-Page et al. 2022).

The rbmi R package was validated by the in-house tool, Autovalidate R, which performs general R package quality checks and performs testing for expected behaviors.

5.3.4 Sensitivity Analyses for Primary Endpoint

Impact of COVID-19 pandemic

The following sensitivity analyses will be performed to evaluate the impact of the COVID-19 pandemic:

- Vary threshold on number of missed dose-months due to the pandemic in the definition of "substantial reduction in drug exposure" ICE. The "substantial reduction in drug exposure" ICE is defined with a threshold $\geq$ 4 dose-months (not necessarily consecutive). A sensitivity analysis will be performed using different thresholds: 1, 2, 3, 5 and 6 dose-months. All other aspects of the primary estimator will remain the same.
- Exclude participants based on site closure information and apply a treatment policy strategy to all other COVID-19 related ICEs. In this analysis, all participants enrolled before a site closure due to the pandemic will be removed from the analysis set. In alignment with the primary estimand, only a site closure of 16 weeks or more during the double-blind treatment period and without any access to study drug (no home nursing) will be considered. In accordance with the FDA's guidance for Industry "Statistical Consideration for Clinical Trial During the COVID-19 Public Health Emergency (June 2020)", participants excluded from this analysis can be identified using baseline data only: randomization date and site number. The site closure information is an administrative site level information, independent of the conduct of the trial, collected using the eCRF for the purpose of this study and analysis. The substantial reduction in drug exposure due to the COVID-19 pandemic (defined as $\geq$ 4 missed dose-months) will not be considered as an ICE. All other aspects of the primary estimator will remain the same.
- Remote scale administration. Remote scale administration was authorized in exceptional cases at Weeks 104 and 116 visits due to COVID-19 related restrictions. In this analysis, all CDR-SB assessments performed remotely will be excluded from the analysis and treated as missing data. All other aspects of the primary estimator will remain the same.
- A subgroup analysis based on the date of randomization. This will allow estimating the treatment effect for subgroups of participants randomized at least 12 months before the onset of the global COVID pandemic (11 March 2020 according to the World Health Organization (WHO) opening remarks), in $\geq$5 to 12 months prior to the onset of the global COVID pandemic and within less than 5 months from the onset of the global COVID pandemic (see Section 5.3.5.1 for an exact definition of each subgroup).

Impact of missing data handling methods

The following sensitivity analyses will be performed to evaluate the impact of missing data handling method:

– Tipping point analysis

  This analysis stress tests the CIR assumption by imputing worse outcomes after SDRC ICEs in the active drug arm than predicted by the CIR assumption. This will be implemented via a marginal δ-approach as described in the 'rbmi: advanced functionality' vignette of the rbmi's package and in Cro et al, 2020. Specifically, the imputation step will be performed as for the primary estimator and, after imputation is completed, a constant δ will be added to the imputed week 116 outcomes occurring after SDCR ICEs in the active drug arm. The subsequent analysis step of these δ-adjusted imputed datasets is as for the primary estimator.

  To determine the tipping point, the constant δ will be increased in small steps starting from a value of 0 (corresponding to the primary estimator). The tipping point will then be defined as the value of δ at which the p-value for the treatment effect estimator first becomes greater than 5%.

Impact of potential outliers or extreme observations

In order to assess the impact of a potentially small number of "extreme observations" or "outlier points" (e.g., rapid progressors) on the treatment effect, the ANCOVA analysis model will be replaced by a robust linear regression. Robust regression will produce treatment effect estimates less contaminated by highly influential observations. Specifically, robust regression using M estimation will be used. Standard errors and confidence intervals will be based on jackknife as described previously.

5.3.5 Supplementary Analyses for Primary Endpoint

5.3.5.1 Subgroup Analyses for Primary Endpoint

The generalizability of the CDR-SB results when comparing the active drug arm to the placebo arm will be investigated by estimating the treatment effect in the following subgroups:

- Demographics:
  - Age, two age categories cut by the median
  - Sex
  - Geographic Region, as per IxRS
- Baseline disease severity:

  CDR-GS = 0.5 vs CDR-GS > 0.5
  - Prodromal vs Mild (as per eCRF)

- APOE ε4 genotype
  - Carrier/Non Carrier
- Use of symptomatic AD medication at baseline Yes/No (derived from eCRF based on the search below)
  - Symptomatic AD medication is defined as any one of: donepezil, galantamine, memantine or rivastigmine
- Randomization date, three subgroups defined by the following dates:
  - *before* 11 March 2019
  - *in between* (and including) 11 March 2019 and 1 October 2019
  - *after* 1 October 2019

Summaries of the treatment effect for the change in CDR-SB from baseline to Week 116 by these subgroups will be provided in forest plots.

5.3.5.2 Other Supplementary Analyses for Primary Endpoint(s)

Treatment effect estimates before Week 116

In this supplementary analysis, the same analysis strategy as described for the primary estimator will be used to estimate the effect of the active drug on disease progression defined as a change in CDR-SB at other time points, other than Week 116.

The treatment effect on the adjusted mean change in CDR-SB from baseline to Week 24, 52, 76 and 104 will be provided.

Treatment effect in relationship to a post-baseline timepoint

In this supplementary analysis, the same analysis strategy as described for the primary estimator will be used to estimate the percent relative difference between Week 52 and Week 116 as well as between Week 24 and Week 116, and between Week 76 and Week 116.

The relative difference between e.g., Week 52 and Week 116 (and similarly for the other considered post-baseline timepoints) will be defined as:

*[(Delta_wk_116 - Delta_wk_52) / (Placebo_wk_116 - Placebo_wk_52)] *100*

where Delta_wk_x is the estimated treatment effect at Week x, and Placebo_wk_x is the estimated change from baseline for the placebo arm at Week x. All those quantities will be obtained from the same model used for the primary estimand.

Treatment policy estimand

All ICEs will be handled with a treatment policy strategy regardless of whether being SDCR or NSDCR. All observed data will be included regardless of occurrence of any ICE. Missing values following all ICEs will be imputed with the method used in the primary estimator for missing data following an SDCR ICE, see Section 5.3.3. Missing values, not following an ICE, will be imputed under MAR (similarly to the primary estimands analysis, see "Missing data assumptions for the primary estimator" in Section 5.3.3). Note that the attributes of population, variables, treatment, and population level summary will remain the same as for the primary estimand.

Concomitant AD treatment estimand

In this supplementary analysis, the treatment effect will be evaluated in the hypothetical scenario that no post-baseline initiation or modification of the use of other approved AD medication has happened.

All attributes of this estimand except Treatment will be identical as for the primary estimator. The treatment attribute will be: "Prescribed study drug including uptitration to the target dose, irrespective of concomitant use of symptomatic treatment for AD at baseline, but assuming no initiation or change in symptomatic treatment after baseline". In this supplementary analysis, all data following ICEs "Starting another treatment for AD" and "Changing the dose of a symptomatic treatment for AD" will be set to missing and imputed under a MAR assumption, in line with a hypothetical strategy. The analysis methods will be the same as described for the primary estimand.

MMRM

This analysis aims to provide a reference point analysis method described in early versions of the protocol (up to version 4) and to other external analyses where MMRM was considered the default and primary analysis. All available data will be used in the analysis. There will be no missing data imputation or consideration for any intercurrent events. The model will include the following covariates: treatment group, visit, and treatment-by-visit interaction, baseline CDR-SB score and baseline CDR-SB score-by-visit interaction, baseline ADAS-Cog 13 total score, baseline ADCS-ADL total score and the randomization stratification factors, namely: disease stage (from eCRF), geographic region (from the IxRS), the use of AD medication at baseline (from eCRF) and the APOE ε4 status (from the Vendor). An unstructured variance-covariance structure will be applied to model the within-subject errors across visits. In the case of non-convergence, compound symmetry will be used, together with a robust estimator of standard error ("sandwich" estimator).

Clinically evident decline: a progressor analysis

In this supplementary analysis, CDR-SB progression will be defined as a change from baseline in CDR-SB greater than or equal to a threshold *x*. The primary threshold of interest is $x = 2.5$. The work justifying the threshold will be presented, as an oral presentation, at the Alzheimer's Association International Conference (AAIC), 31 July–4 Aug 2022; San Diego, CA and online (title of the oral presentation: *Selecting appropriate meaningful change thresholds for trials of early (prodromal-to-mild) AD: A caregiver-rated, anchor-based analysis based on the Tauriel Study*).

The CDR-SB progression endpoint will be analyzed using the same conditional mean imputed dataset as for the primary estimator of the primary estimand (Section 5.3.3).

The probability of CDR-SB progression by Week 116, in each treatment arm, will be estimated by the proportion of participants with a progression at any time during the double blinded period. The probability of progression by Week 116 will be obtained for different values of the progression threshold *x*. A plot of the probability of CDR-SB progression by Week 116 versus threshold values *x* will be generated.

In addition, a Cox proportional hazard model to estimate the treatment effect on time to first CDR-SB progression (based on the primary threshold of interest), may also be considered. In this case, the same baseline covariates as for the primary estimator of the primary estimand will be used (Section 5.3.3). An unadjusted model or a stratified analysis may be considered as well. Corresponding standard errors and confidence intervals will be calculated using the jackknife as previously described.

Impact of ARIA-E MRI finding on the primary outcome

In this supplementary analysis, the objective is to estimate the treatment effect, in the hypothetical scenario where ARIA-E would not have occurred. This analysis aims at removing the potential impact of ARIA-E on the primary endpoint.

All the attributes from the primary estimand will remain the same, but an ICE of 'ARIA-E occurrence' will be added. The 'ARIA-E occurrence' will be handled with hypothetical strategy. All data collected after the occurrence of an ARIA-E will be removed and treated as missing for this data analysis purpose. Post- ARIA-E data will then be imputed under the Missing at Random assumption unless ARIA-E is not the first ICE. In this case, the imputation will be determined by the strategy for the first ICE as described for the main estimator.

5.3.6 ***Convergence Considerations for RBMI-Based Analyses***

In the rare situation that the imputation model (see step 1, imputation model, in Section 5.3.3) would fail to converge despite the attempt to simplify the variance-covariance

matrix, it will be investigated whether removing some covariates and/or pooling factor levels of categorical covariates would solve the issue.

In case the root cause of the problem could not be attributed to the choice of covariates, the following approaches will be tried in sequence:

- *Replace jackknife with bootstrap, and accept up to a maximum of 5% failed runs (out of the total number of bootstrap runs). Runs that did fail would be discarded.*
- *Derive only a point estimate based on the conditional mean imputation, but no standard error (as jackknife and bootstrap did not work)*

5.4 SECONDARY ENDPOINTS ANALYSES

The same primary estimand's analysis strategy *(see Section 5.3.2)* will be applied to secondary endpoints listed in Table 5*, with the following exceptions for the corresponding estimator:*

- *. the imputation model will not include, as covariates, the baseline ADAS-Cog 13 total score and the baseline ADCS-ADL total score*

  The imputation model will thus include treatment group, visit, and treatment-by-visit interaction, the baseline score and the baseline score-by-visit interaction, and the randomization stratification factors, namely: disease stage, geographic region, the use of AD medication at baseline and the APOE ε4 status

- *the analysis model will not include, as covariates, the baseline ADAS-Cog 13 total score and the baseline ADCS-ADL total score*

  The analysis model will thus include treatment, the baseline score and the randomization stratification factors, namely: disease stage, geographic region, the use of AD medication at baseline and the APOE ε4 status

In the following, confirmatory secondary endpoints refer to endpoints included in the type I error control procedure. Other important secondary endpoints not subject to the type I error control procedure are considered as supportive secondary endpoints. For all secondary endpoints, the treatment effect over time will also be considered and the change from baseline to Weeks 24, 52, 76, and 104 will be provided as supplementary analyses.

Table 5 Secondary Endpoints

Secondary Efficacy Endpoint	Confirmatory	Type
Alzheimer's Disease Assessment Scale, Cognitive subscale, 13-item (ADAS-Cog 13)	yes	Continuous
Alzheimer's Disease Cooperative Study, Activities of Daily Living scale (ADCS-ADL) total score	yes	Continuous
Functional Activities Questionnaire (FAQ)	yes	Continuous
Mini Mental State Examination (MMSE)	no	Continuous
Alzheimer's Disease Assessment Scale, Cognitive subscale, 11-item (ADAS-Cog 11)	no	Continuous
Coding (Digit Symbol Substitution Test [DSST])	no	Continuous
Verbal Fluency Task	no	Continuous
Alzheimer's Disease Cooperative Study, Activities of Daily Living scale (ADCS-ADL) instrumental score	no	Continuous

ADAS-Cog 13= Alzheimer's Disease Assessment Scale, Cognitive subscale, 13-item; ADCS-ADL= Alzheimer's Disease Cooperative Study, Activities of Daily Living scale; ADAS-Cog 11 = Alzheimer's Disease Assessment Scale, Cognitive subscale; DSST= Digit Symbol Substitution Test; FAQ= Functional Activities Questionnaire; MMSE= Mini Mental State Examination.

5.4.1 Confirmatory Secondary Endpoints

The confirmatory secondary endpoints are provided to increase the confidence in the treatment effect observed on the primary endpoint. For these confirmatory secondary endpoints, *as already outlined in the beginning of Section 5.4, the same primary estimand's strategy will be applied, with some changes to the corresponding estimator (see Section 5.4 for the details).*

For these confirmatory secondary endpoints, MMRM analyses will also be considered as supplementary analyses. In this case, the change from baseline in the confirmatory secondary endpoints will be adjusted on the following covariates: treatment group, visit, treatment-by-visit interaction, the respective baseline score for each secondary endpoint, the baseline score-by-visit interaction and the randomization stratification factors as specified for the primary endpoint, see Section 5.3.4, paragraph on MMRM.

The method for controlling the overall Type I error is described in section 5.1.

5.4.1.1 Alzheimer's Disease Assessment Scale, Cognitive Subscale, 11- item and 13-Item (ADAS-Cog 11/ ADAS-Cog 13)

The ADAS-Cog is the most frequently used scale to assess cognition in clinical trials for mild to moderate AD dementia (Rozzini et al. 2007; Connor and Sabbagh 2008; Ihl et al.

2012). More specifically, the ADAS-Cog is a participant-based assessment that measures learning and memory, language production, language comprehension, constructional praxis, ideational praxis, and orientation. The modified version will be used; it has 13 items and includes the addition of delayed word recall and a number of cancellation subtests, as well as use of only one trial for word recognition (Mohs et al. 1997). Equivalent alternate forms of the word recall, word recognition, and number cancellation subtests will be used in successive test administrations. The ADAS-Cog 11 and 13 will be used in this study, with ADAS-Cog 13 considered as a confirmatory secondary endpoint. Individual item scores are based on errors and generally range from 1 to 5, although some items have smaller or larger score ranges. The ADAS-Cog 13 total score ranges from 0-85, with higher scores reflecting greater impairment. It takes approximately 45 min to administer the ADAS-Cog 13.

5.4.1.2 Alzheimer's Disease Cooperative Study, Activities of Daily Living Scale (ADCS-ADL), Total Score

The ADCS-ADL (Galasko et al. 1997) is the scale most widely used to assess functional outcome in participants with AD (Vellas et al. 2008). The ADCS-ADL is a 23-item informant-based questionnaire that covers both basic activities of daily living (ADL) (e.g., eating and toileting) and more complex ADL or instrumental ADL (e.g., using the telephone, managing finances, preparing a meal). It has a 4-week recall period. Total scores range from 0-78, with higher scores indicating better functioning.

5.4.1.3 Functional Activities Questionnaire (FAQ)

The FAQ (Pfeffer et al. 1982) is an informant-based measure of functional abilities. Informants provide performance ratings of the target person on ten complex higher-order activities. The FAQ is a 30-point scale, the higher the score the worse the performance.

5.4.2 Supportive Secondary Endpoints

For context, additional clinical endpoints collected longitudinally in the studies will be provided (see Table 5). The Sponsor proposes not to rank these hierarchically as for confirmatory secondary endpoints.

5.4.2.1 Mini Mental State Examination (MMSE)

The MMSE is a set of standardized questions used to evaluate possible cognitive impairment and help stage the severity level of this impairment (Folstein et al. 1975). The questions target six areas: orientation, registration, attention, short-term recall, language, and constructional praxis/visuospatial abilities. The MMSE is a participant-based assessment. The score ranges from 0-30, with lower values indicating a greater impairment.

5.4.2.2 Digit Symbol Substitution Test (DSST)

Coding, also called Digit Symbol Substitution Test (or DSST), is a subtest from the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV; Wechsler 2008). Coding is a participant-based assessment that measures speed of processing and associative

memory. The test is known to be sensitive to disease progression (Lezak et al. 2004). The 120-second version of the test will be used in this study.

5.4.2.3 Verbal Fluency Task

The category verbal fluency (e.g., animals) is a participant-based assessment that measures speed and flexibility of verbal thought. Verbal fluency tests are sensitive tools for detecting dementia (Pasquier et al. 1995; Lezak et al. 2004) and for monitoring decline over time (Clark et al. 2009).

5.4.2.4 Alzheimer's Disease Cooperative Study, Activities of Daily Living Scale (ADCS-ADL), Instrumental Score

The ADCS-ADL instrumental score is a sub score of the ADCS-ADL scale, see Section 5.4.1.2. The ADCS-ADL instrumental score used in this study is the sum of items 6a and 7 to 23.

5.4.3 COVID-19 Related Sensitivity Analyses for Secondary Endpoints

During the COVID-19 pandemic, many countries took movements' restriction policy. Among the clinical outcomes collected in the study, some questionnaires may be particularly impacted. These questionnaires include items describing actions that are strongly constrained or prohibited by restriction policies.

We define modified versions of the total score for these scales. These modified scales scores may be used to perform sensitivity analysis aiming at understanding and mitigating the impact of the COVID-19 pandemic on study results and interpretation, in line with the primary clinical question of interest.

5.4.3.1 ADCS-ADL COVID-19 Modified Total Score

The original ADCS-ADL is a 23 item scale with a total score range of 0-78. The following items were identified as being particularly impacted by the pandemic related restrictions:

- Item 2: Optimal walking performance, maximum 3 points
- Item 15: Optimal performance getting around/travelling outside the home, maximum 4 points
- Item 16a/b: Shopping trips - selecting items and paying without supervision, maximum 4 points
- Item 18: Left away from home, maximum 3 point

A sensitivity analysis may be conducted, using a modified version of the ADCS-ADL total score after removing of these four items, resulting in a 19-item scale with a score range of 0-64. This alternative version will be referred to as "ADCS-ADL COVID-19 modified total score".

5.4.3.2 FAQ COVID-19 Modified Total Score

The original FAQ is a 10-item scale with a score range of 0-30. The following items were identified as being particularly impacted by the pandemic related restrictions:

- Item 3: Shopping alone, maximum 3 points
- Item 10: Travelling outside of the neighborhood, maximum 3 points

A sensitivity analysis may be conducted, using a modified version of the FAQ total score after removing of these 2 items, resulting in an 8-item scale with a score range of 0-24. This alternative version will be referred to as “FAQ COVID-19 modified total score”.

5.5 EXPLORATORY ENDPOINT(S) ANALYSIS

The same primary estimand’s analysis strategy *(see Section 5.3.2)* will be applied to continuous exploratory endpoints listed in Table 6, *with the following exceptions for the corresponding estimator:*

- *. the imputation model will not include, as covariates, the baseline ADAS-Cog 13 total score and the baseline ADCS-ADL total score*

  The imputation model will thus include treatment group, visit, and treatment-by-visit interaction, the baseline score and the baseline score-by-visit interaction, and the randomization stratification factors, namely: disease stage, geographic region, the use of AD medication at baseline and the APOE ε4 status

- *the analysis model will not include, as covariates, the baseline ADAS-Cog 13 total score and the baseline ADCS-ADL total score*

  The analysis model will thus include treatment, the baseline score and the randomization stratification factors, namely: disease stage, geographic region, the use of AD medication at baseline and the APOE ε4 status

. For all continuous exploratory endpoints, the ANCOVA analysis model will include the baseline score of the exploratory endpoint, the disease stage, geographic region, the use of AD medication at baseline and the APOE ε4 status as covariates.

For ordinal endpoints, only descriptive analyses will be considered, summarizing the frequencies of the different categories, as well the proportion of participants with a certain shift in categories.

The Resource Utilization in Dementia Scale–Lite (RUD-Lite) and the EuroQoL–5 Dimensions (EQ-5D-5L) scales will be used in this study for informing pharmacoeconomic evaluations and will be reported separately.

Table 6 Exploratory Endpoints

Exploratory Efficacy Endpoint	Type
Clinical Dementia Rating-Global Score (CDR-GS)	Ordinal
CDR-Individual Components	Continuous
Dependency Level, as assessed by the Alzheimer disease cooperative study - activities of daily living (ADCS-ADL) score	Ordinal
Integrated AD Rating Scale (iADRS)	Continuous
Quality of Life–Alzheimer's Disease (QoL-AD)	Continuous
Neuropsychiatric Inventory Questionnaire (NPI-Q)	Continuous
Zarit Caregiver Interview–Alzheimer's Disease (ZCI-AD)	Continuous

ADCS-ADL = Alzheimer disease cooperative study-activities of daily living; ADCOMS = AD Composite Score; CDR-GS =Clinical Dementia Rating-Global Score; EQ-5D= EuroQoL–5 Dimensions; iADRS = Integrated AD Rating Scale; NPI-Q= Neuropsychiatric Inventory Questionnaire ; QoL-AD =Quality of Life–Alzheimer's Disease; ZCI-AD=Zarit Caregiver Interview–Alzheimer's Disease.

5.5.1 Clinical Dementia Rating–Global Score (CDR-GS) and Individual Components of the CDR scale

The Washington University CDR is a global assessment instrument that yields global scores (GS) and sum of boxes (SOB) scores. The CDR is derived from a semi-structured interview with the participant and an appropriate informant, and it rates impairment in six categories (Memory, Orientation, Judgment and Problem Solving, Community Affairs, Home and Hobbies, and Personal Care) on a 5-point scale for which 0 = no impairment, 0.5 = questionable impairment, and 1, 2, and 3 = mild, moderate, and severe impairment, respectively. From the six individual category ratings, or box scores, the CDR-GS is established by clinical scoring rules, for which CDR 0 = no dementia and CDR 0.5, 1, 2, or 3 = questionable, mild, moderate, or severe dementia, respectively (Morris 1993).

5.5.2 Dependence Level Assessed by the ADCS-ADL Score

To calculate dependence levels, scores on the ADCS-ADL can be transformed into discrete levels of disability via an algorithm developed initially by Kahle-Wrobleski (2015). Items from the ADSC-ADL were mapped to 6 levels of dependence derived from the Dependence Scale (Zhu et al., 2009), ranging from Level 0: no impairment in instrumental or basic ADLs to Level 5: complete incontinence or inability to transfer. Four subscales were used to aid the construction of dependence levels, including bADLs, domestic/household activities, communication/engagement and outside

activities. The mapping of items to dependence levels was validated using additional clinical and economic measures.

A revised algorithm, developed to a) remove ambiguity regarding the contribution of some items and b) add clarity on the handling of missing data, will be used to calculate the dependence levels. Progression to greater levels of dependence is indicative of disease progression and is informative for a variety of care providers and stakeholders. This algorithm is provided in appendix, see Appendix 2.

5.5.3 Integrated AD Rating Scale (iADRS)

The iADRS is a composite of cognition and function that combines scores from the ADAS-Cog-13 (cognition) and the instrumental component of the ADCS-ADL (function) (Wessels et al., 2018). A sum score of the total scores of both components is calculated (ADAS-Cog is reversed) using the following formula:

iADRS = [-(ADAS-Cog13) + 85] + ADCS-iADL.

Total score range from 0 to 141. The iADRS total score will be generated and results may be reported in the CSR.

5.5.4 AD Composite Score (ADCOMS)

The ADCOMS was developed to assess cognition and function in early stages of AD. It is a composite score that combines 12 items from existing AD measures, specifically the ADAS-Cog (Delayed word recall, Orientation, Word recognition, Word finding difficulty), MMSE (Orientation time and Constructional praxis) and all CDR-SB items (Wang et al., 2016). The ADCOMS score was built using a linear longitudinal model to characterize the relationship between disease progression and the individual items from existing AD clinical scales. A PLS regression procedure was used to identify individual clinical scale items that represent AD-related clinical decline over time to calculate their respective weighting factors. The resulting composite ADCOMS score is a weighted linear combination of the individual scale items selected in the fitted PLS model. Items with small contribution to the PLS model were removed according to Wold's criterion (a Variable Importance of Projection below 0.8). Total score range from 0 to 1.97, with lower scores indicating greater impairment. The ADCOMS total score will be generated and results may be reported in the CSR.

5.5.5 Quality of Life–Alzheimer's Disease (QoL-AD)

The Quality of Life-Alzheimer's Disease (QoL-AD) was developed to assess quality of life (QoL) in participants who have dementia (Logsdon et al. 1999, Logsdon et al 2002). The QoL-AD consists of 13 items covering aspects of participants' relationships with friends and family, physical condition, mood, concerns about finances, and overall assessment of QoL. The total score is the sum of the 13 items and ranges from 13 to 52, with higher scores indicating better health-related QoL. In this study, the QoL-AD

will be administered in a standardized, structured interview format to participants by investigative staff in order to gather participant responses about QoL. The study partner will also complete the study partner version of the questionnaire to enable proxy responses from the study partner.

5.5.6 Neuropsychiatric Inventory Questionnaire

The Neuropsychiatric Inventory-Questionnaire (NPI-Q) (Kaufer et al. 2000) was developed to assess a wide range of behaviors encountered in dementia participants, to provide a means of distinguishing severity of behavioral changes, and to facilitate rapid behavioral assessment through the use of screening questions. It is an informant-based instrument in which 12 behavioral domains are evaluated. The recall period is the past month, and severity scores range from 0 to 36, with higher scores indicating greater severity. The study partner's distress portion of the scale was not used in this study.

5.5.7 Zarit Caregiver Interview–Alzheimer's Disease (ZCI-AD)

The Zarit Caregiver Interview-Alzheimer's Disease (ZCI-AD) is a modified version of the Zarit Burden Interview 22-item version, which was originally designed to reflect the stresses experienced by caregivers for people with dementia (Zarit and Zarit 1990). The modified version includes modifications in item and title wording (e.g., removal of "your relative "to refer directly to the participant, removal of "burden" from title), inclusion of additional items, the use of 11-point numerical rating scales for each item and a 4 week recall period. The ZCI-AD measure consists of 27 items covering 13 domains (i.e., humanistic impact domain (14 items) including the domains physical (3 items), emotional (4 items), social (3 items), and daily life (4 items), and the additional domains exhaustion (2 items), dependence (2 items), worry (2 items), role perception (3 items), financial impact (1 items), difficulty with medication (1 item), overall difficulty of caregiving (1 item), and sadness (1 item)). The ZCI-AD is completed by the study partner without involvement from the site staff. The ZCI-AD is scored on a domain level with each domain score ranging from 0 to 100 with higher scores indicating higher level of impact. All item responses are re-coded on a 0 to 4 scale (response category 0 = 0; response category 1,2, and 3 = 1; response category 4,5,and 6 = 2; response category 7 and 8 = 3; response category 9 and 10 = 4) and items of a domain are summed up and transformed to 0 to 100 score range. Domain scores are only calculated if responses of at least 80% of items of the respective domain are available. The ZCI-AD has been validated in prodromal, mild and moderate stages of AD (Bernaards, C et al.)

5.6 SAFETY ANALYSES

Descriptive statistics will be used to analyze all safety data collected in the double-blind treatment period in the safety-evaluable analysis set, unless otherwise specified.

Safety data collected from the day of the first dose of blinded study drug up to 14 weeks after the last dose of blinded study drug (but no later than the day before the first OLE dose for the participants who entered the OLE period) will be included in the

beyond the 14-week post last dose period and until the end of the study will be included in the follow-up period analyses.

For participants who entered the OLE period, safety data collected from the day of the first OLE dose up to 14 weeks after the last OLE dose will be included in the OLE period analyses

Safety analyses will be assessed through summaries of exposure to study treatment, adverse events, changes in laboratory test results (including shift tables), MRI findings, changes in vital signs and ECGs, and changes in C-SSRS scores.

5.6.1 Extent of Exposure

Exposure to study drug information will be descriptively summarized by treatment as follows:

- Treatment duration (in weeks)
- Total number of administrations
- Total cumulative dose (mg)
- Number of dose-administrations at each dose level

5.6.2 Adverse Events

All verbatim AEs terms will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) version that is current at the time of the analysis (Version 25.0 or higher), and AE severity will be graded according to the scale defined in Table 5 in Section 5.3.3 of the study protocol (mild/moderate/severe). For each treatment group, the frequency of each AE preferred term will be defined as the number of participants experiencing at least one occurrence of the event. Each table will present the overall number and percentage of participants experiencing at least one AE and the total number of AEs reported. Percentages will be based on the number of participants in the safety-evaluable analysis set. In summary tables, AEs will be sorted by body system (in decreasing order of overall incidence), then by preferred term (in decreasing order of overall incidence). The summary tables will be restricted to treatment-emergent AEs, i.e., AEs that occur on or after the day of first dose of study drug. Non-treatment-emergent AEs (with onset before the first dose) will be listed.

The following safety information will be summarized by treatment group for the double-blind treatment period:

- AEs, AEs by intensity, AEs related to study drug
- Deaths
- SAEs, SAEs related to study drug
- AEs leading to discontinuation of study treatment

- AEs leading to dose modifications (dose interruption, dose reduction or delayed up-titration). Delayed up-titration at any given visit is defined as the simultaneous occurrence of the following two tick-boxes in the eCRF Adverse Events form:
  - Action take with Gantenerumab due to AE/SAE: Dose Not Changed
  - Was dose regimen modified from protocol schedule? Yes"
- Injection site reaction (ISR) signs and symptoms
- Systemic injection reactions (AEs with eCRF tickbox “systemic injection reaction” selected)
- ‘Hypersensitivity reactions’

Protocol-specified adverse events of special interest (AESI) will be listed.

The impact of the COVID-19 pandemic on the safety data will be assessed by reviewing the following:

- Confirmed or suspected COVID-19 AEs
- AEs associated with COVID-19
- Potential long COVID-19 symptoms

The following data handling rules will be applied for all AE summary tables:

- Events that are missing both at onset and at end dates will be considered to have started after the first dose of study drug and the duration will be set to missing.
- If the onset date is missing, and the end date is on or after the first dosing date or unresolved or missing, then the event will be considered to have started after the first dose of study drug.

The following data handling rules will also be applied for specific tables:

- An AE will be included in the summary table of AEs leading to study drug discontinuation if the “action taken with blinded gantenerumab” drop-down menu on the AE eCRF is checked “drug withdrawn”.
- In the summary table of AEs by intensity, if a participant has more than one occurrence of an event, the event with the most severe intensity will be counted. If the intensity of an AE is missing, then the AE will be included only in the total number of events column, and not in the count of participants with the event by intensity.
- In the summary table of AEs related to study drug, if a participant has more than one occurrence of an event, the related event will be counted if applicable. If the relationship of an AE is missing, then the AE will not be included.

5.6.3 Magnetic Resonance Imaging Safety Findings

ARIA-E and ARIA-H are identified risks associated with gantenerumab. Sites were asked to capture all ARIAs as AEs in the eCRF that met any of the following criteria:

- Symptomatic ARIA-E (i.e., accompanied by CNS symptoms), and/or

- ARIA findings that result in a change in study treatment (e.g., dosage modification, treatment interruption, or treatment discontinuation), and/or
- Findings that are otherwise clinically significant in the investigator's judgment

Not all ARIA MRI findings qualify as AE. ARIA analyses will be mainly based on ARIA MRI findings. ARIA AEs will also be reported. Based on MRI data, the incidence, and severity (based on the Barkhof Grand Total Score [BGTS]) ARIA-E and the incidence of ARIA-H will be summarized by treatment group and within this also by APOE ε4 genotype (by number of alleles) and by dose level. Additionally, the timing of ARIA-E and the timing to meet the protocol-defined criteria for permanent discontinuation due to ARIA-H will be also summarized by descriptive statistics and eventually by Kaplan-Meier methods. Recurrence of ARIA-E will be summarized by treatment group and within this also by APOE ε4 genotype (by number of alleles). ARIA-E associated with CNS symptoms (see Section 5.6.3.1) and with serious CNS symptoms will be summarized by treatment group and within this also by APOE ε4 genotype. Temporal co-occurrence of ARIA-E and ARIA-H will be summarized by treatment group and within this also by APOE ε4 genotype (by number of alleles). Temporal co-occurrence is defined as an MRI scan showing new ARIA-H that occurs between ARIA-E onset and resolution (inclusive), irrespective of the brain region.

MRI findings other than ARIA will also be summarized.

5.6.3.1 CNS Symptoms Temporally Associated with ARIA-E MRI Findings

CNS symptoms temporally associated with ARIA-E are defined as onset or worsening of CNS symptom(s) that is/are temporally associated with ARIA-E MRI findings. CNS symptoms experienced by the participant that are new or worsened since the last MRI without ARIA-E are collected in a CNS Symptoms Request Form before the MRI takes place at a visit. To identify CNS symptoms temporally associated with ARIA-E MRI findings, the following definitions will be used:

NEW CNS symptoms: If there is any AE reported in the eCRF with "Reported on the MRI CNS symptoms request form" = Y that is [new since date of most recent MRI scan showing no ARIA-E findings] AND is [ongoing or ends between the date of most recent site visit prior to date of MRI scan showing new ARIA-E (MRI) and date of MRI scan showing ARIA-E resolution (MRI)] then ARIA-E should be classified as associated with CNS symptoms

OR

WORSENED CNS symptoms: If there is any AE reported in the eCRF with "Reported on the MRI CNS symptom request form" = Y that is [started before the date of most recent MRI scan showing no ARIA-E findings] AND is [ongoing or ends between the date of

most recent site visit prior to date of MRI scan showing new ARIA-E (MRI) and date of MRI scan showing ARIA-E resolution from MRI data] AND [there is an increase in severity grading] then ARIA-E should be classified as associated with CNS symptoms.

The CNS symptoms temporally associated with ARIA-E MRI findings will be listed and summarized by treatment group and within this also by APOE ε4 genotype (number of alleles).

5.6.4 Laboratory Data

Laboratory data will be summarized by treatment group for each assessment visit using descriptive statistics of absolute values and change from baseline values. In addition, the frequency of participants with abnormal laboratory values will be summarized by treatment group.

5.6.5 Vital Signs

Vital signs assessments include systolic blood pressure, diastolic blood pressure, and pulse rate measured throughout the study. Vital sign measurements will be summarized by treatment group for each assessment visit using descriptive statistics of absolute values and change from baseline values. In addition, the frequency of participants with abnormal results will be summarized by treatment group.

5.6.6 ECGs

ECG data will be summarized by treatment group for each assessment visit using descriptive statistics of absolute values and change from baseline values for the following parameters:

- Heart rate
- QRS duration
- RR interval
- PR interval
- QT intervals (including QTcF)

For QTcF, the summary will also include 2-sided 90% confidence interval at each time-point.

In addition, ECG overall interpretations will be summarized by treatment group and visit.

5.6.7 Columbia-Suicide Severity Rating Scale (C-SSRS)

The C-SSRS is an assessment tool used to assess the lifetime suicidality of a participant (C-SSRS at baseline) as well as any new instances of suicidality (C-SSRS since last visit). The structured interview prompts recollection of suicidal ideation, including the intensity of the ideation, behavior, and attempts with actual or potential lethality.

Suicidal ideation, suicidal behavior, and self-injurious behavior without suicidal intent will be summarized by treatment group. In addition, change from baseline to worst post-baseline assessment in suicidality categories will be summarized by treatment group.

5.7 OTHER ANALYSES

5.7.1 Summaries of Conduct of Study

The summary of study conduct will include a description of the following items by treatment arm:

- Number of participants enrolled and randomized
- Number of participants included in each analysis set
- Number and percentage of participants who prematurely withdrew from the study or from study treatment (including the reasons for discontinuation and the distribution of these discontinuations by time-windowed visit)
- Incidence of protocol deviations – overall and by four main categories (inclusion criteria, exclusion criteria, medication and procedural)
- Stratification factor reported in IxRS and used for randomization
- Number of participants with home nursing
- Number of participants initiating or changing symptomatic treatment of AD during the study

Major protocol deviations and premature withdrawals will be listed.

5.7.2 Summaries of Treatment Group Comparability

Demographics and baseline characteristics (including age, sex, medical history, randomization stratification factors) will be summarized descriptively by treatment arm as assigned for the ITT analysis set using:

- means, SDs, medians, and ranges for continuous variables
- counts and proportions for categorical variables

Exposure to AD concomitant medication (including post-baseline initiation or change in dose) will be summarized by treatment arm for the ITT analysis set.

5.7.3 Immunogenicity Analyses

Immunogenicity analyses include the evaluation for antibodies against gantenerumab, including the determination of antibody titers. The results of the confirmatory assay will be presented as a frequency table summarizing baseline and post-baseline results.

A listing of participants with positive ADA status per confirmatory assay and titer result will be provided.

5.7.4 Summaries of COVID-19 Impact on the Trials

Studies GRADUATE I and GRADUATE II were ongoing during the COVID-19 pandemic. Consequently, to monitor the potential impact of the pandemic on the trials, we will provide a specific set of descriptive analyses related to COVID-19 by treatment arm for the ITT analysis set (see Table 2), including:

- Demographic and Baseline Characteristics in Participants with Confirmed/Suspected COVID-19
- COVID-19 AEs (see Section 5.6.2)
- COVID-19 related Protocol Deviations
- Missed doses due to COVID-19
- Study discontinuations due to COVID-19 as determined by the adjudication committee for ICEs
- Study drug administrations of 1020 mg Q4W
- Remote scale administrations
- Duration of study site closure

5.7.5 Amyloid PET Substudy

The objective of the GRADUATE I and GRADUATE II longitudinal amyloid PET substudies is to assess changes in brain amyloid load over time using the change in florbetaben or flutemetamol from baseline to the last amyloid PET visit in the Standard Uptake Value Ratio (SUVR).

Two amyloid PET ligands are allowed in the GRADUATE I and GRADUATE II longitudinal amyloid PET substudies to provide assessment of β-amyloid protein deposition according to country and site availability: radiopharmaceuticals florbetaben-F18 and flutemetamol-F18. However, the same ligand has to be used for the same participant throughout the study (e.g., if a participant has been enrolled in the main study with a positive florbetaben PET scan, only florbetaben will be allowed and used for the longitudinal follow-up scans for the participant).

Centiloid mapping will be completed for SUVR data from the two amyloid PET ligands. The primary amyloid PET outcome measure is the change in amyloid PET Centiloid from baseline to Week 116.

5.7.5.1 General considerations on Amyloid PET statistical analyses

With the Centiloid endpoint, data from both tracers will be pooled and analyzed together. Separate analysis by tracer with the Centiloid endpoint may also be conducted as appropriate. When SUVR metrics is the endpoint, the analysis will be done separately and reported separately for each tracer, when possible.

Missing values will not be imputed.

The amyloid PET analyses will be performed on the Amyloid-PET-mITT analysis set (see Table 2), unless otherwise specified, and participants will be analyzed according to the treatment assigned at randomization by IxRS.

5.7.5.2 Summaries of Treatment Group Comparability

Demographics and baseline characteristics (including age, sex, medical history, randomization stratification factors) will be summarized descriptively by treatment arm as assigned for both the Amyloid-PET-mITT analysis set and the Amyloid-PET-longitudinal analysis set (see Table 2) using:

- means, SDs, medians, and ranges for continuous variables
- counts and proportions for categorical variables

5.7.5.3 Visit Windowing

For amyloid PET assessments, due to the long time between scheduled assessments, time windows based on study days, as defined in Table 7, will be used. Study days are defined based on days on study since the date of the first dose, with the day of the first dose being study day 1.

Table 7 Time Windows for Amyloid PET Assessments

Visit	Target day	Time window
Baseline	1	Up to day 35
Week 52	365	281, 449
Week 116	813	645, 897

For Week 116, the time window will cover the week 104 timepoint as well. In case of more than one assessment within a time window, the assessment with the date closest to the target day is selected. For the combined window of Week 104 and 116, the target day is the Week 116 assessment day.

Because of visit windowing, data collected at an early termination visit will be summarized at the corresponding visit as defined by the time window. For participants who have discontinued treatment early, if a PET scan is performed more than 56 days (early termination visit expected 14 days after last dose, followed by time window per protocol for early termination is ±42 days) after the date of last dose, the PET scan will not be used for the analysis.

The end of a substudy is defined as the date when all participants enrolled in the substudy have:

1. completed the last required amyloid PET scan (Week 116), or
2. completed an early termination scan, or
3. discontinued from the main study.

5.7.5.4 Definition of the Estimand for Amyloid PET

The scientific question of interest for the amyloid PET substudies is to assess the effect of the intended study treatment on the PD endpoint, change from baseline in amyloid load burden at Week 116, in the absence of a substantial impact of the COVID-19 pandemic and as if treatment discontinuation would not have occurred.

The primary Amyloid PET estimand is thus described by the following attributes:

- **Target Population**:
  Early (prodromal to mild) AD population as identified by the inclusion and exclusion criteria, see Sections 4.1.1 and 4.1.2 of the substudy protocols and Sections 4.1.1 and 4.1.2 of the main study protocols
- **Variable**:
  Change from baseline to Week 116 in amyloid PET Centiloid
- **Treatment**:
  Prescribed study drug including up-titration to the target dose
- **Population-level summary**:
  Difference in variable means between treatment arms
- Intercurrent events
  - Substantial reduction in drug exposure due to the COVID-19 pandemic (defined as $\geq$4 missed dose-months)
  - Treatment discontinuation due to any reason

Amyloid PET is listed in the protocol under “Pharmacodynamic Biomarker Objective”. Considering that Amyloid PET primarily aims at estimating the Pharmacodynamic effect of the drug, rather than a direct measure of clinical efficacy, reporting treatment effect estimated "as if there were no treatment discontinuations and no substantial reduction in exposure due to COVID-19 pandemic" (thus using a hypothetical strategy) addresses the scientific question of interest.

Details for Definition of Variable

The Centiloid variable will be used rather than the original SUVR, because it allows to combine data from different tracers, by mapping SUVR values to a standardized scale. The Centiloid variable is the current common standard in the scientific community.

The primary SUVR measure of interest is computed using a weighted composite target region and whole cerebellum as reference region. The weighted composite target region is composed of (both left and right side):

- frontal lobe,
- parietal lobe,
- temporal lobe lateral,

- cingulum posterior and
- anterior cingulate gyrus

Each region is weighted by its own volume. The Centiloid conversion is a linear transformation of SUVR with tracer-specific parameters that are given below:

Centiloid Equation:

$$CL = SlopeCL \times SUVR + InterceptCL$$

CL = Centiloids; SlopeCL = slope; SUVR = standard uptake value ratio of the target region; InterceptCL = intercept.

The pertinent values for the two tracers are:

Table 8 Primary Centiloid Equation Parameters

Tracer	Reference	Slope	Intercept
Florbetaben-F18	whole cerebellum	175.6	-174.2
Flutemetamol-F18	whole cerebellum	143.5	-141.1

5.7.5.5 Main Analytical Approach

An MMRM will be used to estimate the mean change in Centiloid from baseline to Week 116 within each of the substudies. The model will include the change from baseline in Centiloids as the dependent variable, while adjusting for treatment arm (as categorical), visit (as categorical), APOE ε4 status (as categorical; carrier vs non-carrier), type of tracer (as categorical; Florbetaben vs Flutemetamol), baseline Centiloid, baseline Centiloid-by-visit and treatment-by-visit interaction. Visit will be treated as the repeated variable within a participant. An unstructured variance-covariance matrix will be applied to model the within-participant errors; in the case of non-convergence, compound symmetry will be used, together with a robust estimator of standard error ("sandwich" estimator).

In line with the estimand definition, data acquired after the ICE "Substantial reduction in drug exposure due to the COVID-19 pandemic" (as defined in Table 3) or more than 56 days from treatment discontinuation will be excluded from the analysis and treated as missing for the primary analysis purposes.

There will be no data imputation for missing data. However, the MMRM provides unbiased estimates and valid inference in the presence of missing data, under the MAR (missing at random) assumption.

The difference in least squares means between active drug and placebo arms at Week 116 will be estimated and presented alongside p-value and the 95% confidence interval for treatment difference. A p-value <0.0001 will be regarded as a statistically significant evidence, irrespective of the results of the other primary/secondary endpoints

(along the lines of the Haybittle–Peto rule, and thus maintaining the overall type-I error within the study).

5.7.5.6 Supplementary analyses

Changes in weighted Composite Summary region SUVR (for each tracer) and Centiloids from baseline will be summarized using descriptive statistics.

Based on the same model as specified in the Section 5.7.5.4, the change from baseline to Week 52 in Centiloids will also be reported as supplementary analysis.

Additionally, the number and proportion of participants with values below or equal to the positivity threshold will be summarized for each assessment time point. Centiloid zero is the mean amyloid burden for a typical population of young healthy controls, and 100 Centiloid is the typical mean of a population with AD. The Centiloid value of 24 is consistent with the diagnostic amyloid positivity threshold (Klunk et al. 2015; Navitsky et al. 2018) (see Section 5.7.5.3 for definition of Centiloid**Error! Reference source not found.**).

A chi-square test (or Fisher's exact test, whichever is appropriate) will be used to compare the proportions of participants with values below or up to the positivity threshold between the active drug and placebo arm.

In order to account for the potential impact of missing values, a completers analysis will be performed, i.e., restricting the analysis to participants who completed the visit with non-missing data.

5.7.6 Tau PET Substudy

There is a single tau PET substudy enrolling subjects from both studies (WN29922/WN39658 Longitudinal Tau PET Substudy). This substudy utilizes [^{18}F]GTP1 (RO6880276) as tau PET radioligand.

Statistical analyses will be conducted on tau PET Median Standardized Uptake Value Ratios (SUVR) in the following four target regions of interest. In composite target regions, each region is weighted by its own volume.

- A temporal composite target region. This region is composed of (both left and right):
    - anterior and posterior superior temporal gyrus,
    - posterior temporal lobe,
    - fusiform gyrus,
    - middle and inferior temporal gyrus;

- A medial temporal composite region not including the hippocampus, composed of (both left and right):
    - Amygdala,
    - Parahippocampus,
    - Anterior medial and lateral temporal lobe;
- Frontal lobe (both left and right);
- Parietal lobe (both left and right);

The inferior cerebellar grey matter will be used as reference region for the calculation of median SUVRs for all four target regions considered.

All the statistical analyses will be based on the Tau-PET-mITT analysis set (see Table 2) unless otherwise specified.

5.7.6.1 Summaries of Treatment Group Comparability

Demographics and baseline characteristics (including age, sex, medical history, randomization stratification factors) will be summarized descriptively by treatment arm as assigned for both the Tau-PET-mITT analysis set and the Tau-PET-longitudinal analysis set (see Table 2) using:

- means, SDs, medians, and ranges for continuous variables
- counts and proportions for categorical variables

5.7.6.2 Definition of the Estimand

The scientific question of interest for the Tau PET substudy is to assess the effect of the intended study treatment on the PD endpoint, change from baseline in tau PET median SUVR at Week 116, in the absence of a substantial impact of the COVID-19 pandemic and as if treatment discontinuation would not have occurred. The same estimand will be considered for all four median SUVRs regions.

The primary Tau PET estimand is thus described by the following attributes:

- **Target Population**:
  Early (prodromal to mild) AD population as identified by the inclusion and exclusion criteria, see Sections 4.1.1 and 4.1.2 of the substudy protocols and Sections 4.1.1 and 4.1.2 of the main study protocols
- **Variable**:
  Change from baseline to Week 116 in tau PET median SUVR
- **Treatment**:
  Prescribed study drug including up-titration to the target dose
- **Population-level summary**:
  Difference in variable means between treatment arms
- Intercurrent events

- Substantial reduction in drug exposure due to the COVID-19 pandemic (defined as ≥4 missed dose-months)
- Treatment discontinuation due to any reason

Tau PET is listed in the protocol under “Pharmacodynamic Biomarker Objective”. Considering that tau PET aims at estimating the Pharmacodynamic effect of treatment, reporting treatment effect estimated "as if there were no treatment discontinuations and no substantial reduction in exposure due to COVID-19 pandemic" (thus using an hypothetical strategy) addresses the scientific question of interest.

Since there are not individual Tau PET substudies for GRADUATE I and GRADUATE II, but rather a single Tau PET substudy across all patients from GRADUATE I and GRADUATE II, Tau PET will be analyzed at the pooled level of GRADUATE I and GRADUATE II.

5.7.6.3 Main Analytical Approach

An MMRM analysis will be used to estimate the mean change from baseline to Week 116 for each of the median SUVRs defined. The model will include the change from baseline in median SUVR as the dependent variable, while adjusting for treatment arm (as categorical), visit (as categorical), APOE ε4 status (as categorical; carrier vs non-carrier), baseline median SUVR, baseline median SUVR-by-visit, study, study-by-visit and treatment-by-visit interaction. Visit will be treated as the repeated variable within a participant. An unstructured variance-covariance matrix will be applied to model the within-participant errors; in the case of non-convergence, compound symmetry will be used, together with a robust estimator of standard error (“sandwich” estimator).

In line with the estimand definition, data acquired after the ICE “Substantial reduction in drug exposure due to the COVID-19 pandemic” (as defined in Table 3) or more than 56 days from last dose will be excluded from the analysis and treated as missing for the primary analysis purposes.

There will be no data imputation for missing data. However, the MMRM provides unbiased estimates and valid inference in the presence of missing data, under the MAR (missing at random) assumption. A subgroup analysis by parent study may be also considered in order to derive a study-specific treatment effect.

Based the same MMRM model already described in this section, the change from baseline to week 52 in tau PET will also be reported as supplementary analysis.

5.7.6.4 Visit Windowing

For tau PET assessments, due to the long time between scheduled assessments, time windows based on study days, as defined in Table 9, will be used. Study days are

defined based on days on study since the date of the first dose, with the day of the first dose being Study Day 1.

Table 9 Time Windows for tau PET Assessments

Visit	Target study day	Time window
Baseline	1	Up to day 35
Week 52	365	281, 449
Week 116	813	645, 897

5.7.7 Cerebrospinal Fluid (CSF) Analyses

The main analysis for the CSF biomarkers will be based on the pooled dataset of GRADUATE I and GRADUATE II studies (due to the small number of CSF samples within each study). Analyses at individual study level will be conducted as well.

Analysis of CSF biomarkers will be based on the CSF-mITT analysis set unless otherwise specified.

CSF biomarker data will be summarized by treatment group and visit (see schedule of assessment in the study protocol).

For each CSF biomarker, two statistically equivalent ANCOVA models will be considered:

1. *using the change from baseline to Week 116 as dependent variable, with covariates: treatment arm (as categorical), APOE ε4 status (as categorical; carrier vs non-carrier) and baseline biomarker.*

2. *using the value at Week 116 as the dependent variable, with covariates: treatment arm (as categorical), APOE ε4 status (as categorical; carrier vs non-carrier) and baseline biomarker.*

Both models will use log (base 10) transformed biomarker data. The following summary statistics will be estimated after the appropriate back-transformation of the original model parameters:

- The geometric mean at Week 116 for each treatment arm, obtained as 10^LSM, where LSM is the least square mean for the corresponding treatment arm based on model 2.
- The geometric mean ratio (gantenerumab relative to placebo) at week 116, obtained as 10^LSM_GvsP, where LSM_GvsP is the least square means of the contrast of Gantenerumab against placebo based on model 2.
- The %-difference in geometric mean (relative to placebo) at week 116, obtained as [(10^LSM_GvsP) - 1]*100 based on model 2.

- *The %-change from Baseline to Week 116 in the Geometric Mean for each treatment arm, obtained as [10^LSM -1]*100, where LSM is the least square mean for the corresponding treatment arm at the corresponding visit based on model 1*

There will be no data imputation for missing data.

The following CSF biomarkers will be analyzed:

- Total tau (tTau)
- Phosphorylated tau (pTau-181)
- Neurogranin
- Neurofilament light (NFL)

Other exploratory CSF biomarker may be reported separately.

Summaries of Treatment Group Comparability

Demographics and baseline characteristics (including age, sex, medical history, randomization stratification factors) will be summarized descriptively by treatment arm as assigned for both the CSF-mITT analysis set and the CSF-longitudinal analysis set (see Table 2) using:

- means, SDs, medians, and ranges for continuous variables
- counts and proportions for categorical variables

Visit Windowing

For CSF assessments time windows based on study days, as defined in Table 10, will be used. Study days are defined based on days on study since the date of the first dose, with the day of the first dose being Study Day 1.

Table 10 Time Windows for CSF Assessments

Visit	Target study day	Time window
Baseline	1	Up to day 35
Week 116	813	645, 897

5.7.8 Plasma Biomarker Analyses

Plasma biomarkers will be analyzed separately for each study (GRADUATE I and GRADUATE II).

Analysis of plasma biomarkers will be based on the ITT analysis set unless otherwise specified.

The plasma biomarkers will be summarized by treatment group and visit (see schedule of assessment in the study protocol).

For each plasma biomarker, two statistically equivalent MMRM models will be considered:

1. *using the change from baseline as dependent variable, with covariates:* treatment arm (as categorical), visit (as categorical), treatment-by-visit interaction, APOE ε4 status (as categorical; carrier vs non-carrier), baseline biomarker and baseline biomarker-by-visit.

2. *using the value at the post-baseline visit as the dependent variable, with covariates: treatment arm (as categorical), visit (as categorical), treatment-by-visit interaction, APOE ε4 status (as categorical; carrier vs non-carrier), baseline biomarker and baseline biomarker-by-visit.*

For both models, visit will be treated as the repeated variable within a participant. An unstructured variance-covariance matrix will be applied to model the within-participant errors; in case of non-convergence, compound symmetry will be used, together with a robust estimator of standard error ("sandwich" estimator).

Both models will *use* log *(base 10)* transformed biomarker data. *The following summary statistics will be estimated after the appropriate back-transformation of the original model parameters:*

- *The geometric mean at each visit for each treatment arm, obtained as 10^LSM, where LSM is the least square mean for the corresponding treatment arm at the corresponding visit based on model 2.*
- *The geometric mean ratio (gantenerumab relative to placebo) at each visit, obtained as 10^LSM_GvsP, where LSM_GvsP is the least square means of the contrast of Gantenerumab against placebo at the corresponding visit based on model 2.*
- *The %-difference in geometric mean (relative to placebo) at each visit, obtained as [(10^LSM_GvsP) - 1]*100, at the corresponding visit, based on model 2.*
- *The %-change from Baseline to each visit in the Geometric Mean, obtained as [10^LSM -1]*100, where LSM is the least square mean for the corresponding treatment arm at the corresponding visit based on model 1.*

There will be no data imputation for missing data. However, the MMRM provides unbiased estimates and valid inference in the presence of missing data, under the MAR (missing at random) assumption.

The following plasma biomarkers will be analyzed:

- Amyloid beta (1-42) (Abeta-42)
- Phosphorylated tau (pTau-181)

Other exploratory plasma biomarkers may be reported separately.

Summaries of Treatment Group Comparability

Demographics and baseline characteristics (including age, sex, medical history, randomization stratification factors) will be summarized descriptively by treatment arm as assigned for both the ITT analysis set and the Plasma-longitudinal analysis set (see Table 2) using:

- means, SDs, medians, and ranges for continuous variables
- counts and proportions for categorical variables

Visit Windowing

For Plasma assessments time windows based on study days, as defined in Table 11, will be used. Study days are defined based on days on study since the date of the first dose, with the day of the first dose being Study Day 1.

Table 11 Time windows for Plasma assessments

Visit	Target study day	Time window
Baseline	1	Up to Day 35
Week 24	169	85, 253
Week 52	365	281, 449
Week 104	729	645, 770
Week 116	813	771, 897

5.7.9 Volumetric MRI Analyses

Structural MRI will be analyzed separately for each study (GRADUATE I and GRADUATE II).

Analysis of structural MRI (volumetric MRI) will be based on the MRI-mITT analysis set unless otherwise specified.

Volumetric MRI, for each brain region, will be summarized by treatment group and visit using descriptive statistics for the absolute volume at baseline and percent change from baseline at post-baseline visits.

An MMRM analysis will be used to estimate the mean percent change from baseline to Week 116 (as well Week 48 and 104) in Volumetric MRI for each brain region. The model will include the percent change from baseline in volumetric MRI as the dependent variable, while adjusting for treatment arm (as categorical), visit (as categorical), treatment-by-visit interaction, baseline age, gender (as categorical), APOE ε4 status (as categorical; carrier vs. non-carrier) and disease stage (as categorical). Visit will be treated as the repeated variable within a participant. An unstructured variance-covariance matrix will be applied to model the within-participant errors; in the case of non-convergence, compound symmetry will be used, together with a robust estimator of standard error (“sandwich” estimator).

There will be no data imputation for missing data. However, the MMRM provides unbiased estimates and valid inference in the presence of missing data, under the MAR (missing at random) assumption.

The following brain regions will be considered:

- Whole brain
- Ventricles

- Hippocampus right and left
- Cortical gray matter

Other exploratory analyses based on other brain regions may be reported separately.

Summaries of Treatment Group Comparability

Demographics and baseline characteristics (including age, sex, medical history, randomization stratification factors) will be summarized descriptively by treatment arm as assigned for both the MRI-mITT analysis set and the MRI-longitudinal analysis set (see Table 2) using:

- means, SDs, medians, and ranges for continuous variables
- counts and proportions for categorical variables

Visit Windowing

For Volumetric MRI assessments time windows based on study days, as defined in Table 12, will be used. Study days are defined based on days on study since the date of the first dose, with the day of the first dose being Study Day 1.

Table 12 Time windows for Volumetric MRI assessments

Visit	Target study day	Time window
Baseline	1	Up to Day 35
Week 48	337	253, 421
Week 104	729	645, 770
Week 116	813	771, 897

5.8 INTERIM ANALYSES

Details of interim analysis plans are described in a separate interim analysis SAP (iSAP), providing information about a futility interim analysis based on the primary efficacy endpoint CDR-SB.

Other than the futility analysis, there is no plan for an efficacy interim analysis based on the primary endpoint. The primary analysis of the clinical efficacy endpoints will be performed only once, after completion of the efficacy data collection at the end of the double blind part of the study (as described in this SAP) and it will be the only opportunity to formally reject the primary null hypothesis of the trial.

6. SUPPORTING DOCUMENTATION

This document is part of a broader Data Analysis Plan that has several documents, including:

- Graduates studies interim analysis SAP
- Graduates studies Data Analysis Plan Module 2
- Graduates studies Data Analysis Plan Module 3

7. REFERENCES

[EMA] European Medicines Agency. Committee for Medicinal Products for Human Use. Guideline on the clinical investigation of medicines for the treatment of Alzheimer's disease [resource on the Internet]. 22 February 2018 [cited: 20 May 2020]. Available at: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-investigation-medicines-treatment-alzheimers-disease-revision-2_en.pdf.

Amemiya TA, Takeshi A. Advanced econometrics. Harvard university press; 1985.

Bartlett, Jonathan W, and Rachael A Hughes. 2020. "Bootstrap Inference for Multiple Imputation Under Uncongeniality and Misspecification." Statistical Methods in Medical Research 29 (12): 3533–46.

Bernaards, C, Fischer, K, et al. Development and Psychometric Validation of the 27 Item Zarit Caregiver Interview for Alzheimer's Disease (ZCI-AD-27)) (submitted)

Carpenter JR, Roger JH, Kenward MG. Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. J Biopharm Stat 2013;23(6):1352-71.

Clark LJ, Gatz M, Zheng L, et al. Longitudinal verbal fluency in normal aging, preclinical, and prevalent Alzheimer's disease. Am J Alzheimers Dis Other Demen 2009;24:461-8.

Connor DJ, Sabbagh MN. Administration and scoring variance on the ADAS-Cog. J Alzheimers Dis 2008;15:461–4.

Columbia-Suicide Severity Rating Scale (C-SSRS; Columbia Protocol). The Columbia Lighthouse Project [resource on the Internet]. 2016 [Accessed: 28 July 2022]. Available from: https://cssrs.columbia.edu/

Cro S, Morris TP, Kenward MG, Carpenter JR. Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: a practical guide. Statistics in medicine. 2020;39(21):2815-42.

Efron B, Tibshirani RJ. An introduction to the bootstrap. CRC press; 1994 May 15.

EMA (EMA//SA/0000046418). Scientific Advice procedure (EMA Written Advice received on 29 January 2020; EMA//SA/0000046418).

FDA (4720726). Type C meeting, Written Response Only procedure, 21 December, 2020, Ref ID: 4720726.

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

Galasko D, Bennett D, Sano M, et al. An inventory to assess activities of daily living for clinical trials in Alzheimer's disease. The Alzheimer's Disease Cooperative Study. Alzheimer Dis Assoc Disord 1997;11(2):S33-9.

Gower-Page C, Noci A, Wolbers M. rbmi: AR package for standard and reference-based multiple imputation methods. Journal of Open Source Software. 2022 Jun 15;7(74):4251.

Haybittle J. Repeated assessment of results in clinical trials of cancer treatment. Brit J Radiol 1971; 44: 793–797

ICH E 9: Statistical principles for clinical trials: Available at: https://database.ich.org/sites/default/files/E9_Guideline.pdf

Ihl R, Ferris S, Robert P, Winblad B, Gauthier S, Tennigkeit F. Detecting treatment effects with combinations of the ADAS‐cog items in patients with mild and moderate Alzheimer's disease. Int. J. Geriatr. Psychiatry. 2012;27(1):15-21.

International Council of Harmonization. ICH E9 (R1) addendum on estimands and Sensitivity Analysis in Clinical Trials to the guideline on statistical principles for clinical trials. EMA/CHMP/ICH/436221/2017.

Kahle-Wrobleski K, Fillit H, Kurlander J, et al. Methodological challenges in assessing the impact of comorbidities on costs in Alzheimer's disease clinical trials. The European Journal of Health Economics. 2015;16(9):995-1004.

Kaufer DI, Cummings JL, Ketchel P, et al. Validation of the NPI-Q, a brief clinical form of the Neuropsychiatric Inventory. J Neuropsychiatry Clin Neurosci 2000;12:233-9.

Klunk WE, Koeppe RA, Price JC, et al. The Centiloid Project: standardizing quantitative amyloid plaque estimation by PET. Alzheimers Dement. 2015;11(1):1-15.e154. doi:10.1016/j.jalz.2014.07.003.

Lezak MD, Howieson DB, Loring DW. Neuropsychological assessment (4th revised edition). New York: Oxford University Press, 2004.

Liu, G Frank, and Lei Pang. 2016. "On Analysis of Longitudinal Clinical Trials with Missing Data Using Reference-Based Imputation." Journal of Biopharmaceutical Statistics 26 (5): 924–36.

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

Logsdon RG, Gibbons LE, McCurry SM, et al. Quality of life in Alzheimer's disease: patient and caregiver reports. J Mental Health Aging 1999;5:21-32.

Lu, Kaifeng, and Devan V Mehrotra. 2010. "Specification of Covariance Structure in Longitudinal Data Analysis for Randomized Clinical Trials." Statistics in Medicine 29 (4): 474–88.

Mallinckrodt, Craig H, Christopher J Kaiser, John G Watkin, Michael J Detke, Geert Molenberghs, and Raymond J Carroll. 2004. “Type i Error Rates from Likelihood-Based Repeated Measures Analyses of Incomplete Longitudinal Data.” Pharmaceutical Statistics 3 (3): 171–86.

Mallinckrodt, Craig H, Peter W Lane, Dan Schnell, Yahong Peng, and James P Mancuso. 2008. “Recommendations for the Primary Analysis of Continuous Endpoints in Longitudinal Clinical Trials.” Drug Information Journal 42 (4): 303–19.

Mohs RC, Knopman D, Petersen RC, et al. 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 1997;11(2):S13–21.

Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 1993;43:2412–4.

Navitsky, M, Joshi, A.D, Kennedy I, et al. (2018), Standardization of amyloid quantitation with florbetapir standardized uptake value ratios to the Centiloid scale. Alzheimer's & Dementia, 14: 1565-1571.

Pasquier F, Lebert F, Grymonprez L, et al. Verbal fluency in dementia of frontal lobe type and dementia of Alzheimer type. J Neurol Neurosurg Psychiatry 1995;58:81-4.

Peto, R; Pike, MC; Armitage, P; et al. (1976). "Design and analysis of randomized clinical trials requiring prolonged observation of each patient.. Introduction and design". Br. J. Cancer. 34 (6): 585–612

Pfeffer RI, Kurosaki TT, Harrah CH Jr, et al. Measurement of functional activities in older adults in the community. J Gerontol 1982;37:323–9.

rbmi: advanced functionality, available at https://cran.r-project.org/web/packages/rbmi/vignettes/advanced.html (accessed 28 July 2021)

Robins, James M, and Naisyin Wang. 2000. “Inference for Imputation Estimators.” Biometrika 87 (1): 113–24.

Roger, James. Reference-based MI via Multivariate Normal RM (the “five macros” and MIWithD). Available at: https://www.lshtm.ac.uk/research/centres-projects-groups/missing-data#dia-missing-data (accessed 04 August 2021)

Rozzini L, Costardi D, Chilovi BV, Franzoni S, Trabucchi M, Padovani A. Efficacy of cognitive rehabilitation in patients with mild cognitive impairment treated with cholinesterase inhibitors. Int. J. Geriatr. Psychiatry. 2007;22(4):356-60.

Siddiqui, Ohidul. 2011. “MMRM Versus MI in Dealing with Missing Data—a Comparison Based on 25 NDA Data Sets.” Journal of Biopharmaceutical Statistics 21 (3): 423-36.

Vellas B, Andrieu S, Sampaio C, et al., for the European Task Force Group. Endpoints for trials in Alzheimer's disease: a European Task Force Consensus. Lancet Neurol 2008;7:436-50.

von Hippel, Paul T, and Jonathan W Bartlett. 2021. "Maximum Likelihood Multiple Imputation: Faster Imputations and Consistent Standard Errors Without Posterior Draws." Statistical Science 36 (3): 400–420.

Wang J, Logovinsky V, Hendrix SB, Stanworth SH, Perdomo C, Xu L, Dhadda S, Do I, Rabe M, Luthman J, Cummings J. ADCOMS: a composite clinical outcome for prodromal Alzheimer's disease trials. Journal of Neurology, Neurosurgery & Psychiatry 2016;87(9):993-9.

Wang N, Robins JM. Large-sample theory for parametric multiple imputation procedures. Biometrika 1998;85(4):935-48.

Wechsler D. Wechsler adult intelligence scale-Fourth Edition (WAIS-IV). San Antonio, TX: NCS Pearson, 2008.

Wessels AM, Andersen SW, Dowsett SA, Siemers ER. The Integrated Alzheimer's Disease Rating Scale (iADRS) Findings from the EXPEDITION3 Trial. J Alzheimers Dis. 2018;5(2):134-6.

Westfall, PH, Krishen, A. Optimally weighted, fixed sequence, and gatekeeping multiple testing procedures. J Stat Plan Inference 2001;99:25-40.

WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.

Wolbers M, Noci A, Delmar P, Gower-Page C, Yiu S, Bartlett JW. Standard and reference-based conditional mean imputation. Pharmaceutical Statistics, 2022. DOI: 10.1002/pst.2234

Zarit SH, Zarit JM. The memory and behavior problems checklist and the burden interview. Gerontology Center, The Pennsylvania State University, 1990.

Zhu CW, Leibman C, Townsend R, et al. Bridging from clinical endpoints to estimates of treatment value for external decision makers. J Nutr Health Aging 2009;13:256-9.

Appendix 1 Type I Error Control Considerations for Reference-Based Conditional Mean Imputation Combined with the Jackknife for Inference

Introduction

In the GRADUATE trials, hypothetical and treatment policy strategies are applied to Non Study Drug or Condition Related (NSDCR) intercurrent events (ICEs) and Study Drug or Condition Related (SDCR) ICEs, respectively That is, the clinical interest is in a scenario where SDCR ICEs but not NSDCR ICEs do occur. For estimating the treatment effect, data after NSDCR are assumed to missing at random (MAR) whereas missing data after SDCR are imputed based on the placebo group in both treatment groups using the copy-increments-in-reference (CIR) imputation method, a specific type of reference-based imputation. We refer to Section 5.3.3 for further details.

The primary analysis of the GRADUATE trials is implemented using conditional mean imputation of missing data combined with the jackknife for inference. This methodology has been described and justified in Wolbers et al. (2022). In this appendix, we provide further details regarding type I error control of the method including additional simulations targeted to the setting of the GRADUATE trials.

Theoretical and published simulation evidence for type I error control

As demonstrated in Section 2.5 of Wolbers et al. (2022), the proposed conditional mean imputation methodology corresponds to a computationally efficient implementation of maximum likelihood multiple imputation. Estimators based on maximum likelihood imputation are asymptotically normal and unbiased if the imputation model and the associated missing data assumptions are correctly specified (Wang and Robins (1998), Robins and Wang [2000]). Moreover, standard error estimates based on resampling methods such as the jackknife or the bootstrap are consistent (Bartlett and Hughes (2020), von Hippel and Bartlett (2021), Wolbers et al. (2022).

Therefore, large-sample (asymptotic) type I error control is guaranteed if the imputation model and the associated missing data assumptions are correctly specified. This is the case if the following conditions hold:

- All ICEs are correctly identified and classified as NSDCR or SDCR.
- Observed outcome data prior to an ICEs follow a multivariate normal mixed model for repeated measures (MMRM) with an unstructured covariance matrix.
- Missing outcome data prior to an ICE are missing at random (MAR).
- Missing outcome data after NSDCR ICEs compatible with a hypothetical strategy is also MAR.
- Missing outcome data after SDCR ICEs are compatible with the copy-increments-in-reference (CIR) assumption.

In contrast, the commonly used MMRM model requires that all missing data are MAR (Mallinckrodt et al. [2008]). In many settings, a reference-based missing data assumption for SDCR ICEs such as CIR is arguably more plausible and conservative than the MAR assumption of the classical MMRM model.

In order to study type I error control for finite sample sizes, simulation studies are required. A simulation study reported in Wolbers et al. (2022) assessed type I error via simulation (based on 100,000 simulations) for a setting with a relatively low sample size (100 subjects per group), a large proportion of ICEs (34% and 24% in the active and placebo groups, respectively), and a large probability of study drop-out after the occurrence of the ICE of 75%. For standard MAR and all reference-based imputation methods, inference based on conditional mean imputation and jackknifing strictly protected type I error. Simulation results reported in Liu and Pang (2016) for a similar method also found no evidence of type I error inflation for reference-based methods or the MMRM model if the missing date mechanism was correctly specified. These simulation results are also consistent with simulation studies which report that the MMRM model (or an asymptotically equivalent multiple imputation model) provides adequate type I error control if all missing data are MAR (Siddiqui [2011], Lu and Mehrotra [2010]).

In contrast, statistical methods cannot be expected to strictly control type I error if the missing data assumptions are not correctly specified. For example, simulations reported in Liu and Pang (2016) demonstrate type I error inflation for both reference-based methods and the MMRM model if missing data are simulated under certain missing not at random (MNAR) scenarios. Similarly, Mallinckrodt et al. (2004) demonstrated type I inflation of the MMRM model under MNAR scenarios but type I error control of the MMRM model was much closer to nominal values compared to naive approaches to missing data such as the last observation carried forward (LOCF) approach. Whether or not the missing data assumptions is correctly specified is by nature something that cannot be verified

In conclusion, reference-based imputation methods based on conditional mean imputation and jackknifing control type I error if the imputation model and the associated missing data assumptions are correctly specified. Neither reference-based imputation methods nor the MMRM model can guarantee strict type I error control if the missing data assumptions are not correctly specified. Therefore, sensitivity analyses such as tipping point analyses should always be performed to assess the robustness of the results to deviations from the missing data assumptions.

Additional Simulations Targeted to the Setting of the GRADUATE Trials

Methods

We conducted a simulation study to further investigate the type I error control of the reference-based conditional mean imputation combined with the jackknife for inference. The set-up of the simulation study is similar to the simulations reported in Wolbers et al. (2022), but the simulation parameters and the scenarios considered are adapted to the GRADUATE trials. All simulations were for 1:1 randomized placebo-controlled trials with 508 subjects per group and visits at baseline and at 24, 52, 76, 104, and 116 weeks of follow-up. All simulations were performed under the null hypothesis of no difference in mean outcome trajectories between the groups. Scenarios including different probabilities of ICEs in the two groups and/or misclassification of SDCR ICEs as NSDCR ICEs were also included as described below.

Specifically, the other simulation parameters were chosen as follows:

- The mean outcome trajectory in both groups increased linearly from 3.65 to 4.65 during the first 52 weeks from baseline (i.e., an increase of 1 point in the first year), and increased linearly by 1.5 points per year afterwards.
- The variances of the outcomes at baseline and follow-up visits in both groups were: 1.53, 2.08, 2.65, 3.19, 4.26, 4.26.
- The correlation matrix of the baseline and follow-up values in both groups is shown on Table 1.

Table 1 Correlation Matrix

	Baseline	Week 24	Week 52	Week 76	Week 104	Week 116
Baseline	1	0.72	0.60	0.54	0.52	0.52
Week 24	0.72	1	0.72	0.60	0.54	0.54
Week 52	0.60	0.72	1	0.72	0.60	0.60
Week 76	0.54	0.60	0.72	1	0.72	0.72
Week 104	0.52	0.54	0.60	0.72	1	0.85
Week 116	0.52	0.54	0.60	0.72	0.85	1

- Two types of intercurrent events were simulated: Study Drug or Condition Related ICEs (SDCR) and Non-Study Drug or Condition Related ICEs (NSDCR).
- Simulation of SDCR ICEs:
  - The probability of an SDCR ICE after each visit was calculated according to a logistic model, which also depended on the observed outcome at that visit.
  - The visit-wise probability of an SDCR ICE for a subject with an observed outcome at that visit of 3.65 was varied across 4 different scenarios as reported in Table 2.
  - The odds of an SDCR ICE further increased by 45% for each 1 point increase in the observed outcome.
  - In the placebo group, an SDCR ICE had no effect on the mean trajectory. In the active group, subjects who experienced an SDCR ICE followed the slope of the mean trajectory from the placebo group from that time point onward (CIR).
  - Study drop-out after the SDCR ICE visit occurred with a probability of 80% leading to missing outcome data from that time point onward.
- Simulation of NSDCR ICEs:
  - The probability of an NSDCR ICEs after each visit was assumed to be independent of the visit and the observed outcome. The specification of these probabilities for each of the 4 scenarios is presented in Table 2.
  - NSDCR ICEs always led to missing outcome data from that time point onward.
- If both SDCR and NSDCR ICEs were simulated to occur for a subject, then it was assumed that only the earlier of them counted. In case both ICEs were simulated to occur at the same time, the event was considered a SDCR ICE. This means that a single subject could experience either a SDCR or a NSDCR ICE, but not both of them.
- Additional missing data unassociated with an ICE was simulated by assuming that a subject missed any visit with a probability of 5%.

Table 2 Specifications of ICEs Probabilities in Each Group for the Four Simulation Scenarios. Only the Parameters that Varied Across Different Scenarios are Presented in this Table.

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Probability of a SDCR ICE after each visit for a subject with outcome equal to 3.65 (PLACEBO)	2%	1.5%	3%	2.5%
Probability of a SDCR ICE after each visit for a subject with outcome equal to 3.65 (ACTIVE)	2%	2.5%	3%	3.5%
Probability of a NSDCR ICE after each visit (PLACEBO)	2.5%	2%	0%	0%
Probability of a NSDCR ICE after each visit (ACTIVE)	2.5%	3%	0%	0%

Scenario 1 specifies equal ICE probabilities for the two groups, while Scenario 2 specifies a higher probability of both SDCR and NSDCR events in the active group compared to the placebo group. Scenario 3 simulates only SDCR events, with equal probabilities across the two groups. Scenario 4 also simulates only SDCR events, but with a higher probability in the active group.

Results

The overall probability for a subject to experience either a SDCR or a NSDCR event during the trial under the four scenarios is reported in Table 3. These probabilities were estimated via simulation of a large trial with a sample size of 100'000 subjects per group.

Table 3 Overall Probability for a Subject to Experience Either a SDCR or a NSDCR Event during the Trial under all the Scenarios Considered. Probabilities were Estimated via Simulation of a Large Trial with a Sample Size of 100'000 Subjects Per Group.

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Probability of SDCR ICE (PLACEBO)	21%	17%	30%	27%
Probability of SDCR ICE (ACTIVE)	21%	24%	30%	33%
Probability of NSDCR ICE (PLACEBO)	11%	9%	0%	0%
Probability of NSDCR ICE (ACTIVE)	11%	12%	0%	0%
Overall ICE probability (PLACEBO)	32%	26%	30%	27%
Overall ICE probability (ACTIVE)	32%	36%	30%	33%

For each scenario considered, we assumed that a given fraction of SDCR ICEs had been misclassified as NSDCR ICEs. That is, we will apply a hypothetical strategy not only to those ICEs that are truly NSDCR, but also to a given fraction of SDCR events. The misclassification rate varied as follows: 0%, 15%, 30%, 45%, where 0% means that there is no misclassification of any SDCR event, while 45% means that 45% of SDCR events will be classified (and handled) as NSDCR events.

Type I error was simulated based on 10'000 simulated trials for each scenario. The respective misclassification rates were then applied to these data before the analysis. The analysis was consistent with the primary analysis of the GRADUATE trial, i.e., conditional mean imputation with jackknife-based inference (as described in Wolbers et al. [2022]) with copy-increments-in-reference (CIR) imputation for missing data after SDCR ICEs.

The simulation results are summarized in Table 4. Simulated type I errors ranged from 4.69% to 5.18% across scenarios and misclassification rates. Simulations based on 10,000 simulated data sets provide a Monte Carlo standard error for type I error estimates of approximately $\pm 0.22\%$. Therefore, these results are fully consistent with strict type I error control at the 5% significance level.

Table 4 Type I Error Rate Estimates for Each of the Scenarios and Misclassification Rates. Simulations are based on 10,000 Simulated Data Sets which Provide a Monte Carlo Standard Error for Type I Error Estimates of Approximately 0.22%.

Scenario	ICE misclassification rate (from SDCR to NSDCR)	Type I error
Scenario 1	0%	4.95%
	15%	4.84%
	30%	4.97%
	45%	4.83%
Scenario 2	0%	5.11%
	15%	5.00%
	30%	5.18%
	45%	5.13%
Scenario 3	0%	4.98%
	15%	4.81%
	30%	4.82%
	45%	4.69%
Scenario 4	0%	4.73%
	15%	4.78%
	30%	5.00%
	45%	5.01%

Conclusion

The primary analysis described in this SAP is based on the method described in Wolbers et al. (2022). This publication describes and justifies the method in detail and provides theoretical and simulation support that it controls type I error.

Clarification and details on the theoretical argument were provided and additional simulations were conducted. These simulations explored plausible scenarios calibrated to the GRADUATE studies. They also assessed the impact of ICE misclassification.

Based on this body of evidence, the proposed method for primary analysis (using conditional mean imputation of missing data combined with the jackknife for inference) adequately controls type I error.

Appendix 2 ADCS-ADL Dependence Scale Algorithm Version 2.1

Item Domains Used in the Algorithm

- bADL - basic Activities of Daily Living: Questions 1-5, 6B
- iADL – instrumental Activities of Daily Living
  - Household Activities: Questions 6A, 7, 10-14, 23
  - Communication and Engagement: Questions 17, 21, 22
  - Outside Activities: Questions 15, 16A, 16B, 18

Note that 4 items included in the original communication and engagement algorithm are not included in the updated algorithm. Specifically concentrating on a television programme (Q8), participating in small talk (Q9), talking about current events (Q19), recalling information recently read in book/magazine (Q20) were excluded. Whilst these concepts are clearly important to quality of life, they do not require supervision and are therefore not considered fundamental to independence.

The algorithm starts by checking the requirements for Level 5 which is the highest level of impairment and continues to check each lower level until the patient meets the requirements of a level and is assigned to that level. If a patient does not meet the requirements for any of Levels 1-5, then the patient is assigned Level 0.

- **Level 0:** is the dependence level assigned when a patient has no recorded impairment
- **Level 1**: Item score $=2$ on one or more items from only one of the following clusters: Household Activities, Communication and Engagement, Outside Activities. There should be no bADL impairment.
- **Level 2:** Item Score $=2$ on one or more items from two or more of the following clusters: Household Activities, Communication and Engagement, Outside Activities, OR Item Score ≤ 1 on one or more items from any of the following clusters: Household Activities, Communication and Engagement. (There should be no bADL impairment).
- **Level 3:** Item Score ≤ 2 for all items from the following clusters: Household Activities, Communication and Engagement, Outside Activities, OR Item Score ≤ 1 on one or more items from Outside Activities, OR Item Score $=2$ for Bathing (Q4), Score $=2$ for toileting (Q3), Score $=3$ for dressing, score $=1$ or 2 for eating OR Item score $=2$ for walking
- **Level 4**: Any one of the following: Item Score ≤ 2 for Grooming (Q5), Item score ≤ 1 for Bathing (Q4) item score $=1$ for toileting (Q3) or Item score ≤ 2 for dressing, OR item score $=0$ for Eating (Q1) or Item score $=1$ for walking
- **Level 5**: Item Score $=0$ for either Walking (Q2) or Toileting (Q3)

Item recoding in preparation for application of the algorithm

Item 16B with a binary response (0/1) was recoded to 0/3 so that the “no impairment” level was consistent across items.

Item 21 was recoded such that a score of 2 (ability to write short notes or messages that others understood) would not be considered “impaired” for the purposes of the dependence scale and was collapsed with response option 3 (letters or long notes that others understood). This was achieved by recoding 2 to 3.

Missing data was not imputed. “Don’t know” responses were also treated as missing and not imputed. The dependence scale score was not scored if there were more than 2 missing items and/or any of the bADL items were missing.

Appendix 3 Charter for Adjudication Committee for Intercurrent Events

CHARTER FOR ADJUDICATION COMMITTEE FOR INTERCURRENT EVENTS

TITLE:	**TWO PHASE III, MULTICENTER, RANDOMIZED, DOUBLE BLIND, PLACEBO-CONTROLLED, PARALLEL-GROUP, EFFICACY, AND SAFETY STUDIES OF GANTENERUMAB IN PATIENTS WITH EARLY (PRODROMAL TO MILD) ALZHEIMER'S DISEASE**
PROTOCOLS:	WN29922, WN39658
AUTHOR:	[redacted]
IND NUMBERS:	102,266
EUDRACT NUMBERS:	2017-001364-38 (Study WN29922) 2017-001365-24 (Study WN39658)
SPONSOR:	F. Hoffmann-La Roche Ltd
DATE FINAL:	01 September 2021

TABLE OF CONTENTS

1. VERSION HISTORY

Version	Date	Details
Version 1	01September2021	Creation

ADJUDICATION COMMITTEE FOR INTERCURRENT EVENTS MEMBER SIGNATURES

TITLE: **TWO PHASE III, MULTICENTER, RANDOMIZED, DOUBLE BLIND, PLACEBO-CONTROLLED, PARALLEL-GROUP, EFFICACY, AND SAFETY STUDIES OF GANTENERUMAB IN PATIENTS WITH EARLY (PRODROMAL TO MILD) ALZHEIMER'S DISEASE**

PROTOCOLS: WN29922
WN39658

I have read this Charter and confirm that, to the best of my knowledge, it accurately describes the conduct of the Adjudication Committee for Intercurrent Events.

DocuSigned by:
[redacted]
B97097216CF14A1...

[redacted], MD
[redacted], GDL
F. Hoffmann-La Roche Ltd
(ACI [redacted])

Date 9/1/2021

ADJUDICATION COMMITTEE FOR INTERCURRENT EVENTS MEMBER SIGNATURES

TITLE: **TWO PHASE III, MULTICENTER, RANDOMIZED, DOUBLE BLIND, PLACEBO-CONTROLLED, PARALLEL-GROUP, EFFICACY, AND SAFETY STUDIES OF GANTENERUMAB IN PATIENTS WITH EARLY (PRODROMAL TO MILD) ALZHEIMER'S DISEASE**

PROTOCOLS: WN29922
WN39658

I have read this Charter and confirm that, to the best of my knowledge, it accurately describes the conduct of the Adjudication Committee for Intercurrent Events.

DocuSigned by:
[redacted]
C73603349D0D4C4...

[redacted]
F. Hoffmann-La Roche Ltd

Date 9/2/2021

ADJUDICATION COMMITTEE FOR INTERCURRENT EVENTS MEMBER SIGNATURES

TITLE: **TWO PHASE III, MULTICENTER, RANDOMIZED, DOUBLE BLIND, PLACEBO-CONTROLLED, PARALLEL-GROUP, EFFICACY, AND SAFETY STUDIES OF GANTENERUMAB IN PATIENTS WITH EARLY (PRODROMAL TO MILD) ALZHEIMER'S DISEASE**

PROTOCOLS: WN29922
WN39658

I have read this Charter and confirm that, to the best of my knowledge, it accurately describes the conduct of the Adjudication Committee for Intercurrent Events.

DocuSigned by:
[redacted]
538E17B187EB44B...

[redacted]
F. Hoffmann-La Roche Ltd

Date 9/2/2021

2. INTRODUCTION

RO4909832 (gantenerumab) is a fully human monoclonal antibody targeting aggregated forms of amyloid-β including oligomers, fibrils, and plaques. Studies WN29922 and WN39658, defined as GRADUATE studies, will evaluate the efficacy and safety of gantenerumab compared with placebo for the treatment of patients with early (prodromal to mild) Alzheimer's disease.

Intercurrent events (ICEs) are defined as events occurring after treatment initiation that affect either the interpretation or the existence of the measurements associated with the clinical question of interest. In order to estimate precisely the treatment effect as described as the primary estimand, it is crucial to correctly identify and address these ICEs. ICEs will either be considered Non Study Drug or Condition Related (NSDCR) or Study Drug or Condition Related (SDCR). For the primary estimand of the GRADUATE studies, the Sponsor is proposing a treatment policy approach for all SDCR ICEs. All remaining NSDCR ICEs will be handled using a hypothetical approach.

This Charter contains a description of the adjudication committee for intercurrent events (ACI) membership and operations for Studies WN29922 and WN39658. The ACI will review the ICEs related to study treatment discontinuation where ambiguity exists in order to support the study team classifying them as NSDCR or SDCR as per the SAP.

Terms and abbreviations used in this Charter are defined in Table 1.

Table 1 Terms and Abbreviations

Term and Abbreviation	Definition
ACI	Adjudication committee for intercurrent events
eCRF	electronic Case Report Form
ICEs	Intercurrent events
NSDCR	Non Study Drug or Condition Related
SDCR	Study Drug or Condition Related
Sponsor	F. Hoffmann-La Roche Ltd
Study Team	Team composed of Sponsor employees directly involved with the study leadership team (SLT)
unblinding data	data for which treatment assignment is identified

3. ROLE OF THE COMMITTEE

The Study Team delegates to the ACI the responsibility to review and sort ICEs as SDCR or NSDCR according to prespecified ICE categories defined in the study SAP for those cases of study treatment discontinuation where reasons are not precisely captured by the eCRF. These cases will be identified by the GRADUATE Study Team after completion of data cleaning efforts

including medical data review. The predefined ICE categories are the following (specific ICEs may be added to the list if deemed necessary by the ACI):

Table 2 ICE Categories

Intercurrent Event (ICE)	SDCR/NSDCR
Withdrawal from study treatment due to lack of efficacy	SDCR
Withdrawal from study treatment due to safety or tolerability reason (NOTE: This will include discontinuations due to AE, incl. suspected or confirmed COVID-19 AEs)	SDCR
Withdrawal from study treatment with no informative reason given	SDCR
Withdrawal from study treatment due to the COVID-19 pandemic	NSDCR
Significant reduction in drug exposure due to the COVID-19 pandemic (defined as ≥ 4 missed dose-months)	NSDCR
Withdrawal from study treatment due to purely administrative reason	NSDCR
Death	NSDCR
Withdrawal from study treatment due to use or initiation of protocol prohibited medication	SDCR
Withdrawal from study treatment due to other SDCR ICEs	SDCR

The ACI will review discontinuation cases where the reason for study treatment discontinuation and substudy discontinuation is non informative and captured in the eCRF as 'Protocol deviation', 'Withdrawal by subject', 'Physician decision' or 'Other'. For these cases, free text is captured in the eCRF.

The ACI will review data provided by the GRADUATE Study Team (described in section 4.2). By carefully reviewing data relative to the ICEs, the ACI will help the GRADUATE Study Team achieve an objective classification of ICEs as SDCR or NSDCR.

4. COMMITTEE MEMBERSHIP

4.1 MEMBERS

The ACI is composed of a [redacted] and two additional members. The [redacted] has the responsibility to digitally sign the ICEs Categorization Report and the form documenting that a meeting of the ACI took place.

Members: [redacted], MD
[redacted], GDL
F. Hoffmann-La Roche Ltd
(ACI [redacted])

[redacted]
F. Hoffmann-La Roche Ltd

[redacted]
F. Hoffmann-La Roche Ltd

4.2 ACI MEMBERS SELECTION CRITERIA

The ACI members may be employees of the Sponsor or any contract research organization that works with the Sponsor. The committee should include three members representing at least one of the following line functions: Clinical Science, Data Science, and Safety Science. ACI members should not have been involved in the conduct of Studies WN29922, WN39658, and related substudies, should not have been exposed to unblinding data of the studies in scope, and should have a minimum of two-years experience in clinical trials conduct. Based on the aforementioned criteria, ACI members will be selected by the GRADUATE Study Team.

Members of the ACI who do not fulfill all the selection criteria and whose ACI membership may materially affect objectivity will be asked to resign from the committee and will be replaced.

4.3 DURATION OF THE ACI MEMBERSHIP

The membership will extend for the duration of the Studies in scope (see Section 1), at least up to the time the database for primary analysis is locked and the study is unblinded to the Sponsor. If a member leaves the ACI, the GRADUATE Study Team will select a replacement based on the criteria described in section 3.2.

5. COMMITTEE MEETINGS

5.1 ORGANIZATIONAL MEETING

A first introduction meeting will formally establish the ACI and acquaint the ACI with the process that will be followed. In advance of the organizational meeting the committee will have received the study protocols, the IBs, the blank eCRF and the SAP (the latest draft if not yet final).

5.2 SCHEDULED MEETINGS

The number of meetings will depend on the amount of data to be reviewed by the ACI. The Study Team and the ACI will agree on the number of meetings to be held during the organizational meeting.

The Study Team will prepare reports including data to be reviewed. These reports will be provided to the ACI at least three business days prior to each meeting (Appendix 1). The content of these reports is limited to eCRF data extracted from:

- "Study Drug Completion/Early Discontinuation" form including :
  - "Completion/discontinuation reason" item
  - Free text field " If primary reason is protocol deviation, withdrawal by subject, physician decision or other, specify"

The data will be extracted from the eCRF and provided in a tabular format to the committee. ACI meetings will not be attended by the GRADUATE Study Team.
For ACI meetings to take place all three members should be attending. The decisions should be made in a unanimous way. However if this is not possible in some cases, the Chair has the casting vote.

6. COMMUNICATION AND DATA FLOW

6.1 COMMUNICATION

The GRADUATE study team will communicate to the ACI the meeting dates and the SPA responsible will extract data to be reviewed. The ACI will communicate the adjudicated ICEs to the GRADUATE Study Team and the SPA responsible (see section 5.2).

ACI members are to treat all communications regarding these clinical studies, including reports, data, review meeting discussions, teleconferences, and meeting minutes, as confidential material.

All communications relative to these meetings will be archived in the eTMF.

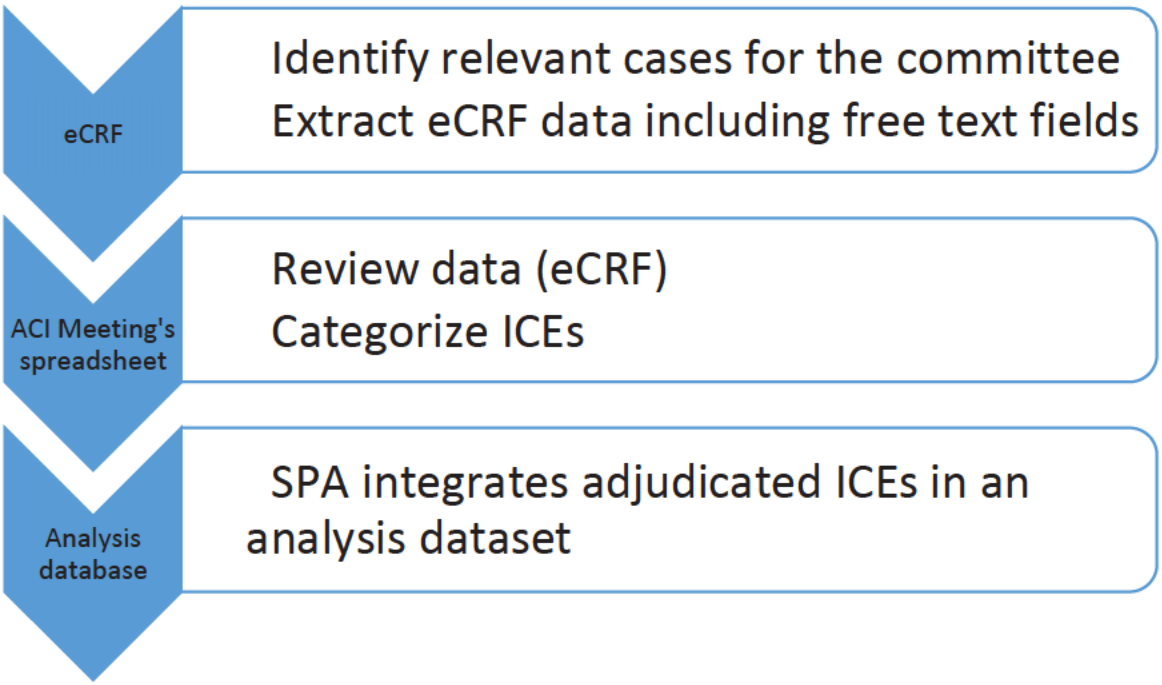
6.2 ICEs CATEGORIZATION REPORT

After each scheduled meeting, ACI will provide input on ICEs categorization during the meetings to the GRADUATE Study Team within seven business days. The format of ICEs Categorization Report will be in a tabular format. An example is presented in Appendix 2.

The GRADUATE Study Team will collect the outcome of ACI meetings, integrate them in an analysis dataset, and archive the documents in the eTMF.

7. APPENDIX 1

Organization flowchart



8. APPENDIX 2

The format of ICEs Categorization Report

Treatment discontinuation reason	Reason specification	ICE Categorization
Protocol deviation / Withdrawal by subject / Physician decision / Other	<eCRF free text>	Withdrawal from study treatment due to lack of efficacy / Withdrawal from study treatment due to safety or tolerability reason / etc.
etc.		

Signature Page for Final SAP WN29922/WN39658 (Graduate) v3
System identifier: RIM-CLIN-453772

Approval Task	[redacted] Company Signatory 05-Oct-2022 07:46:31 GMT+0000