Statistical Analysis Plan: ATH-1017-AD-0201

Study Title:	A Randomized, Placebo-Controlled, Double-Blind Study of ATH-1017 Treatment in Subjects with Mild to Moderate Alzheimer's Disease
Study Number:	ATH-1017-AD-0201
Study Phase:	Phase 2/3
Sponsor:	Athira Pharma, Inc. 18706 North Creek Parkway Suite 104 Bothell, WA 98011 USA
NCT #:	NCT04488419
Version:	Final 1.1
Date:	13 August 2024

Confidentiality Statement

1 TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

2 SIGNATURE PAGE

Study Title: A Randomized, Placebo-Controlled, Double-Blind Study of ATH-1017 Treatment in Subjects with Mild to Moderate Alzheimer's Disease

Study Number: ATH-1017-AD-0201

Prepared by: [redacted] Date: Aug 13, 2024
[redacted]
VP, Biometrics & Data Management
Athira Pharma, Inc.

Reviewed by: [redacted] Date: Aug 13, 2024
[redacted]
Associate Manager, Biostatistics
MMS Holdings, Inc.

3 LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Aβ	amyloid-β
AChEI	acetylcholinesterase inhibitor
AD	Alzheimer's Disease
AE	adverse event
AEC	absolute eosinophil counts
ADAS-Cog$_{11}$	Alzheimer's Disease Assessment Scale-Cognitive Subscale
ADCS-ADL23	Alzheimer's Disease Cooperative Study – Activities of Daily Living, 23-item version
ADCS-CGIC	Alzheimer's Disease Cooperative Study – Clinical Global Impression of Change
AIC	Akaike's Information Criteria
ANOVA	analysis of variance
ApoE	apolipoprotein E
ATC	anatomical therapeutic class
BIC	Bayesian Information Criteria
BMI	body mass index
BP	Blood pressure
BPSD	Behavioral and Psychological Signs of Dementia
C-SSRS	Columbia-Suicide Severity Rating Scale
CDR	Clinical Dementia Rating Scale
CFB	change from baseline
CI	confidence interval
COWAT	Controlled Oral Word Association Test
CP	conditional power
CRF	case report form
CT	Computed tomography
DSMB	Data Safety Monitoring Board
ECG	12-lead electrocardiogram
EQ-5D-5L	EuroQol Group 5-Dimension 5 Level Questionnaire
FSH	Follicle-Stimulating Hormone
fT3	Free Triiodothyronine

fT4	Free Thyroxine
GDS	Geriatric Depression Scale
GFAP	glial fibrillary acidic protein
GST	Global Statistical Test
HR	Heart rate
ICE	intercurrent event
ICF	informed consent form
IDMC	Independent Data Monitoring Committee
ISR	injection site reaction
ITT	intent-to-treat
LFT	liver function test
LLN	lower limit of normal
MCMC	Markov Chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
mITT	modified intent-to-treat
MMRM	Mixed Models for Repeated Measures
MMSE	Mini-Mental State Examination
N/A	Not Applicable
NfL	neurofilament light chain
NIA-AA	National Institute on Ageing – Alzheimer's Association
NPI	Neuropsychiatric Inventory
p-tau	phosphorylated tau
OD	once daily
PK	pharmacokinetic(s)
PR	interval from the P wave (atrial contraction or depolarization) to the onset of the Q wave in the measurement of electrical activity of the myocardium
PT	preferred term
QD	once a day
QRS	Q wave, R wave, and S wave
QTcB	QT interval using Bazett's formula
QTcF	QT interval using Fridericia's formula

RR interval	the time elapsed between two successive R-waves of the QRS signal on the electrocardiogram
RUD-Lite®	Resource Utilization in Dementia lite version
SAE	serious adverse event
SAP	statistical analysis plan
SAS	statistical analysis system
SC	subcutaneous
SD	standard deviation
SOC	system organ class
SSR	sample size re-estimation
TEAE	treatment-emergent adverse event
TSH	Thyroid-Stimulating Hormone
ULN	upper limit of normal
US	United States
VAS	visual analog scale
WHO	World Health Organization
ZBI	Zarit Burden Interview

4 INTRODUCTION

The purpose of this Statistical Analysis Plan (SAP) is to describe the framework for the reporting, summarization and statistical analysis methodology of the safety and efficacy parameters measured throughout the study. It is based on Protocol v8.0 dated 21 Mar 2024.

5 STUDY OBJECTIVES AND ENDPOINTS

Table 1 Study Objectives and Associated Endpoints

Primary Efficacy Objectives	Primary Efficacy Endpoint*
To evaluate the clinical efficacy of ATH-1017 in participants not on background acetylcholinesterase inhibitors (AChEIs)	The Global Statistical Test (GST) score (O'Brien, 1984) is a composite of cognition and function, calculated as the average of two change from baseline z-scores; the z-scores are calculated for the change from baseline scores for cognition (Alzheimer's Disease Assessment Scale-Cognitive Subscale [ADAS-Cog_{11}]) and function (Alzheimer's Disease Cooperative Study – Activities of Daily Living, 23-item version [ADCS-ADL23] score) at Week 26
Key Secondary Efficacy Objectives	**Key Secondary Efficacy Endpoints***
To evaluate the clinical efficacy of ATH-1017 separately on: cognition and activities of daily living	ADAS-Cog_{11} score: change from baseline at Week 26 ADCS-ADL23 score: change from baseline at Week 26
To determine the effect of ATH-1017 on plasma neurofilament light chain (NfL) levels (a biomarker of neurodegeneration with literature support for being a biomarker of Alzheimer's disease [AD] progression)	Change from baseline in the NfL concentration at Week 26
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

Safety Objectives	Safety Endpoints
To determine the safety and tolerability of ATH-1017	Analysis of adverse events (AEs), including injection site AEs; changes from baseline for the following variables: vital signs, 12-lead electrocardiogram (ECG), and laboratory tests (chemistry, hematology, urinalysis); concomitant medication assessments, physical and neurological exams, Columbia-Suicide Severity Rating Scale (C-SSRS), and Geriatric Depression Scale (GDS)

* The comparison between ATH-1017 40 mg and placebo in participants not on background AChEIs will serve as the primary efficacy comparison for the primary and key secondary endpoints. [redacted]

6 STUDY DESIGN CONSIDERATIONS

The study (ATH-1017-AD-0201) underwent several changes during conduct. The original high level study design is provided in Section 6.1 with key changes summarized in Section 6.1.1. Additional protocol changes are described in Section 8.

6.1 Original Study Design

This study (ATH-1017-AD-0201) is designed to demonstrate efficacy in mild to moderate Alzheimer's Disease (AD) participants and establish long-term safety information. Eligible participants will receive once daily (OD) subcutaneous (SC) injections of ATH-1017 (40 mg or 70 mg) or placebo, over a 26-week double-blind period, followed by a 4-week safety follow-up. Participants who complete the double-blind study will have the option to roll over into a long-term (26-week) open-label extension study.

This is a Phase 2/3 multicenter, randomized, double-blind, placebo-controlled, parallel-group, dose-ranging study comparing ATH-1017 40 mg/day and ATH-1017 70 mg/day with placebo in participants with a clinical diagnosis of mild to moderate AD, diagnosed on a 'probable' level according to National Institute on Ageing – Alzheimer's Association (NIA-AA) criteria (McKhann, 2011). The study will be conducted at a total of approximately 70 centers in the United States (US). Participants and their caregivers will be required to sign an informed consent form (ICF) and will be evaluated against the inclusion/exclusion criteria during a screening period. Those who meet all inclusion/exclusion criteria will be randomized in a ratio of 1:1:1 to 3 parallel arms, either the active treatment (ATH-1017 40 mg/day or ATH-1017 70 mg/day) or to the placebo group. During randomization, participants will be stratified by screening Mini-Mental State Examination (MMSE) severity: mild (MMSE: 20-24) versus moderate (MMSE: 14-19). All eligible participants will be tested for apolipoprotein E (ApoE) genotype.

6.1.1 Changes to Original Study Design Important for this SAP

Patients receiving background acetylcholinesterase inhibitors (AChEIs) were no longer enrolled following implementation of the protocol v5 amendment (13 September 2022). This amendment also enabled an interim analysis, as described in Section 9.4. The protocol was further amended (03 May 2023) to discontinue enrollment at 70 mg/day, and patients enrolling after implementation of this amendment will be randomized in a 1:1 ratio to the two remaining parallel arms, i.e., ATH-1017 40 mg/day or placebo. The open label extension study was lengthened to 30 months. Further information on protocol amendments relevant to this SAP are provided in Section 8. All of these changes, except for the sample size increase defined in the Interim Analysis SAP, were based on data external to the LIFT trial or blinded data. [redacted]

6.1.2 Safety Review Committee/Data Monitoring Committee

An independent Data Safety Monitoring Board (DSMB) is planned to conduct periodic review and assessments of unblinded safety data (adverse events [AEs], labs, 12-lead electrocardiogram

[ECG], etc.) throughout the study to ensure the safety of study participants (see Protocol Section 7.3). DSMB review plans, timings and details will be documented in DSMB charter separately.

6.2 Justification of Sample Size

The sample size for the trial was chosen to ensure adequate power for the primary endpoint (Global Statistical Test [GST] score). The following treatment effect assumptions were made to support sample size calculations:

- Change from baseline (CFB) to Week 26 in Alzheimer's Disease Assessment Scale-Cognitive Subscale ($ADAS\text{-}Cog_{11}$): Mean treatment difference (standard deviation [SD]): 1.8 (6).

- CFB to Week 26 in Alzheimer's Disease Cooperative Study – Activities of Daily Living, 23-item version (ADCS-ADL23): Mean treatment difference (SD): 2.7 (9).

The sample size of 149 evaluable patients (patients in the primary analysis population) per trial arm would yield at least 85% power with a two-sided 0.05 level. A positive outcome will be achieved if a significant treatment effect based on the primary endpoint is established.

The power calculations were performed using a simulation-based approach with 10,000 simulation runs based on the parameters listed above (in addition, it was assumed that the endpoint-specific statistics follow a multivariate normal distribution with a common correlation coefficient of 0.5). Power was estimated as the fraction of the simulation runs where the condition was met for the primary endpoint.

6.3 Clinical Measures

The protocol specifies that a qualified, trained, and certified rater will administer questionnaires to the study participant and/or dedicated support person/caregiver.

6.3.1 Cognitive Variables

6.3.1.1 Global Statistical Test score

A composite approach will be used to facilitate the assessment of an overall change in disease status/trajectory in the trial. The GST score will be defined as a single outcome variable based on standardizing and then combining individual patient-level z-score of CFB cognition ($ADAS\text{-}Cog_{11}$) and functional (ADCS-ADL23) scores. The GST score will be determined for each patient at each time point and the resulting scores will define the efficacy outcome variable to be used in the primary efficacy analysis. A formal definition of the GST score is given in Appendix 1. A lower GST score indicates a better result.

6.3.1.2 Alzheimer's Disease Assessment Scale – Cognitive Subscale

The $ADAS\text{-}Cog_{11}$ is designed to measure cognitive symptom change in participants with AD (Rosen, 1984). The standard 11 items are word recall, commands, constructional praxis, naming

objects and fingers, ideational praxis, orientation, word recognition, spoken language ability, comprehension of spoken language, word-finding difficulty, and remembering test instructions.

The test includes 7 performance items and 4 clinician-rated items, with a total score ranging from 0 (no impairment) to 70 (severe impairment). Therefore, higher scores indicate more severe cognitive impairment.

[redacted]

6.3.1.4 Mini–Mental State Examination (MMSE)

The MMSE (Folstein, 1975) is a widely used test of overall cognitive function, assessing memory, orientation, and praxis in a short series of tests. The score is from 0 to 30 with 30 being the best possible and 0 being the worst possible score. A higher MMSE score indicates a better cognitive function.

The MMSE is administered at Screening with a score of 14 to 24 inclusive for participant eligibility and at Baseline.

6.3.1.5 Clinical Dementia Rating Scale (CDR)

The CDR (Hughes, 1982) is a global rating of the function of AD participants assessed in 6 categories: memory, orientation, judgment, and problem solving, community affairs, home and hobbies, and personal care. It is based on a semi-structured interview conducted with the participant and caregiver. Each category has scores from 0 (no symptoms) to 3 (severe) from which the overall CDR global score is derived. The CDR is administered at the Screening visit with a score of 1 or 2 required for participant eligibility. A lower CDR global score indicates a better function.

6.3.2 Functional Outcome Measurements

[redacted]

[redacted]'s illness has improved or worsened relative to a baseline state at the beginning of the [redacted]

[redacted]

6.3.2.2 Alzheimer's Disease Cooperative Study – Activities of Daily Living, 23-item Version

The ADCS-ADL23 (Galasko, 1997) is a 23-item assessment of functional impairment in terms of activities of daily living administered to the support person/caregiver. It comprises 23 questions about the participant's involvement and level of performance across items representing daily living. The questions range from basic to instrumental activities of daily living. Each item is rated from the highest level of independent performance to complete loss. The total score range is from 0-78, with lower scores indicating greater functional impairment.

[redacted]

[redacted]'s disabilities on their life. The score for each [redacted]

[redacted]

[redacted]

6.3.6 Biomarkers

Several plasma biomarkers relevant to AD are being measured in this trial. [redacted]

7 STUDY POPULATIONS

7.1 Analysis Populations

7.1.1 Intent-to-Treat (ITT) Population

The ITT population will consist of all randomized participants regardless of whether or not the participant received test article. A participant is considered randomized when the test article assignment (ie, completes a randomization transaction) is provided. No analyses are planned for the ITT population.

7.1.2 Modified Intent-to-Treat (mITT) Population

The mITT population is a subset of the ITT population, which includes all randomized participants who took at least one dose of the study medication and who, during at least one post-baseline visit, completed both an ADAS-Cog_{11} and ADCS-ADL23 assessment, as well as at baseline. Analyses will be based on the randomized treatment assignment.

7.1.3 Primary Analysis Population

The primary analysis population is comprised of all members of the mITT population who were not taking AChEIs < 28 days before the first dose of study drug and were randomized to either ATH-1017 40 mg or placebo. Primary analyses will be based on the randomized treatment assignment.

7.1.4 Per Protocol Population

The per protocol population will include all randomized participants who took the assigned medication, completed both an ADAS-Cog_{11} and ADCS-ADL23 assessment during at least one post-baseline visit as well as at baseline, and did not have any major protocol deviations that are deemed to potentially interfere with efficacy as evaluated at a classification meeting prior to unblinding. Analyses will be based on the actual treatment received.

7.1.5 Safety Population

The Safety population will include all randomized participants who received at least one dose of the study medication. Participants will be analyzed based on actual treatment received.

[redacted]

8 CHANGES IN CONDUCT OR PLANNED ANALYSES FROM THE PROTOCOL

The following changes have been made in the trial's design and analytical strategies:

- ADCS-CGIC, as a component of the GST score, was replaced by ADCS-ADL23
- Patients receiving background AChEIs were no longer enrolled following implementation of the protocol v5 amendment (13 September 2022).
- Patient randomization in the ATH-1017 70 mg/day arm discontinued (amendment v7, 03 May 2023). Moving forward, patients are randomized in a 1:1 ratio to the two remaining parallel arms, i.e., ATH-1017 40 mg/day or placebo, maintaining the same ratio between those two arms. The comparison between ATH-1017 40 mg and placebo in participants not on background AChEIs will serve as the primary efficacy comparison. The comparison between ATH-1017 70 mg and ATH-1017 40 mg, between ATH-1017 70 mg and placebo, and between pooled ATH-1017 and placebo will be treated as an exploratory comparison.
- The ADCS-ADL23 evaluation at Week 20 was added with protocol amendment v5 (13 September 2022); this results in some designed missingness for the exploratory endpoints GST and ADCS-ADL23 at Week 20
- Samples for biomarker analysis at Week 12 were added with protocol v6 (07 December 2022); this results in some designed missingness for the exploratory biomarker endpoints at Week 12.
- Samples for plasma biomarker analysis became mandatory with protocol amendment v7 (May 2023).
- The efficacy evaluations will be performed at a two-sided 0.05 level at the final analysis, e.g., the overall Type I error rate for the analysis of the primary and key secondary endpoints will be controlled at a two-sided 0.05 level.
- NfL CFB to Week 26 has been included as a key secondary endpoint.
- The testing strategy for the primary and key secondary endpoints has been revised to include all primary and key secondary endpoints. The updated strategy is defined in Section 10.6.
- The final analysis methodology has been modified to apply the combination function approach at the final analysis regardless of the value of conditional power (CP) at the interim analysis. The updated methodology is defined in Section 10.6.
- Provided additional details and clarify on the following:

- Visit windowing for GST score: Components will be visit windowed first and GST will be calculated only where ADAS-Cog_{11} and ADCS-ADL are available at the same visit window.

- ICE definition: Clarification with regards to how ICE-1 for lack of efficacy will be defined.

- GST score and ADCS-ADL analysis: Clarification that the analysis will not include Week 20 for those included in the interim analysis, since Week 20 was not collected prior to the interim analysis.

- Subgroup analysis: Cleaned up language to maintain consistency within the document about how to handle subgroup analysis (using the BY approach without subgroup by treatment interaction terms).

- Multiple imputations: Clarification of how the multiple imputations analysis for sensitivity analysis will be performed.

- Additional baseline characteristic for Years since diagnosis and calculation details

- Laboratory measurements (safety labs): The data handling for values above or below the limit of quantification was clarified.

9 OVERALL STATISTICAL CONSIDERATIONS

9.1 General Conventions

Efficacy and safety data may be summarized and presented by treatment group and time point (Visit) in summary tables. Continuous variables will be presented by descriptive statistics: n, mean, SD, median, minimum, and maximum. Categorical variables will be tabulated by frequency count and percentage. Unless otherwise stated, all statistical tests will be two-sided hypothesis tests performed at the 5% level of significance and all confidence intervals (CIs) will be two-sided 95%CIs. The efficacy and baseline summary statistics tables, as described in Section 10.2, will include statistics for the difference vs placebo.

Means, medians and percentiles will usually be displayed to one more decimal place than the reported data, dispersion statistics (e.g., SD) will have two more decimal places, and the minimum and maximum will be displayed to the same number of decimal places as reported in the raw data. Percentages will be displayed with one decimal place. P-values will be quoted to 4 decimal places consistent with statistical analysis system (SAS) PVALUEw.d format set to PVALUE6.4. P-values < 0.0001 will be presented as p<0.0001.

All data will be listed and patients who were not randomized will be displayed after the randomized treatment.

All data will be analyzed using the SAS® Version 9.4 or later.

9.2 Baseline Definition

For safety summaries, the last measurement prior to first dose is defined as the baseline value. GST score is a change score, and Appendix 1 describes calculation of a baseline GST value for the purpose of use as a covariate variable. [redacted] For all other efficacy measures, baseline is defined as either the measurement on Day 1, or else the last measurement before Day 1 if there is no Day 1 measurement. If no test article is received, baseline is defined as the value closest to but prior to randomization. In all cases, to capture necessary data for split visits, the baseline window will include the first 3 days after randomization. Meaning, if there is missing data at the date of randomization but data captured one to 3 days after, those are to be considered the baseline values. Otherwise, missing baseline values will not be imputed.

9.3 Handling of Partial Dates

If partial dates are recorded for safety outcomes, then partially missing start/beginning date (e.g. AE/concomitant medication start date) will fill in the missing month with January and missing day with 1. For example, if month and day were both missing, then the date would be filled in with January 1st. Partially missing end/finishing date (e.g., AE/concomitant medication end date) will be filled in with December and missing day with the last day of the month. For example, if month and day were both missing, then the date would be filled in with December 31st. For other outcomes (e.g. date of vital signs collection), if only the day is missing it will be imputed as the

15th. When both month and day are missing, the missing month and day will be July 1st. For safety outcomes, (e.g. AE/concomitant medication start date) if the start date is entirely missing, it will be filled in with study start date as a conservative approach. Details on imputation of partial dates for AEs and prior/concomitant medications are available in Appendix 4.

9.4 Interim Analysis

An adaptive design with a single interim analysis was employed in this trial. The interim analysis was performed after approximately 100 patients (about 45% of the enrollment which was planned at the time of the interim) completed the trial. For this purpose, "completed" meant no further participation in the trial, either due to early termination or completion of the Week 26 visit. The unblinded interim data were examined by an Independent Data Monitoring Committee (IDMC) to determine if the sample size in the trial needed to be increased if the original treatment effect assumptions were not aligned with the interim data. The IDMC recommended sample size was later superseded by a small additional sample size increase based on simulations, in view of both a revised statistical testing hierarchy (moving CFB NfL to last in the hierarchy), and no longer randomizing participants to 70 mg ATH-1017, as described in Section 6.2.

9.5 Pooling Strategy for Study Sites

This schema is based on pooling geographically similar sites. The pooled site term (Region) shall be included as a fixed effect variable in the Mixed Models for Repeated Measures (MMRM) model. All sites who use Spanish data collection forms for assessments for any participants will comprise their own Region pool and not be included in the other Regions. [redacted] Analyses based on this SAP will comprise of all sites per pooled site. Post-hoc analyses may be produced excluding Spanish assessment sites.

Sites in States (USA)	Pooled Site (Region)
FL, GA, NC, TN, AL, LA, MS	SE
OH, NM, CO, IL, IN, TX, MO	MW
NJ, NY, MA, ME, PA, RI, MD	NE
WA, CA, NV, OR, AZ, HI, UT	W
Spanish Assessment Site	SP

MW, Midwest; NE, Northeast; SE, Southeast; SP, Spanish Assessment Site; W, West

9.6 Visit Windows/Unscheduled Visits

For listings, all timepoint (including unscheduled and early termination visits) will be included. For analyses that use visit categories (summary tables, figures, and statistical analysis), visit windowing will be applied. All visits (including early termination assessments and unscheduled visits) will be put in a visit window based on the number of days since Day 1 and will be used to identify visit categories. If there are multiple assessments that fall within the same visit window,

the scheduled visit will supersede other visits (that is, only the scheduled visit will be considered for analysis if assessed within the window). Otherwise, the visit closest to the scheduled post-baseline visit will be used.

The analysis visit windows for ADAS-Cog11 [redacted] will be followed as:

Analysis Visit	Protocol-Specified Day	Analysis-Specified Interval
Week 2	Day 14	Day 2-28
Week 6	Day 42	Day 29-63
Week 12	Day 84	Day 64-112
Week 20	Day 140	Day 113-161
Week 26	Day 182	Day 162-196
Week 30	Day 210	Day 197-217

The analysis visit windows for C-SSRS, Physical and Neurological exams, ECG, Vital Signs, and safety labs will be followed as:

Analysis Visit	Protocol-Specified Day	Analysis-Specified Interval
Week 2	Day 14	Day 2-28
Week 6	Day 42	Day 29-63
Week 12	Day 84	Day 64-98
Week 16	Day 112	Day 99-126
Week 20	Day 140	Day 127-161
Week 26	Day 182	Day 162-196
Week 30	Day 210	Day 197-217

The analysis windows for GDS, Height, Weight will be followed as:

Analysis Visit	Protocol-Specified Day	Analysis-Specified Interval
Week 12	Day 84	Day 2-133
Week 26	Day 182	Day 134-196
Week 30	Day 210	Day 197-217

The analysis visit windows for ADCS-ADL23 will be followed as:

Analysis Visit	Protocol-Specified Day	Analysis-Specified Interval
Week 12	Day 84	Day 2-112
Week 20	Day 140	Day 113-161
Week 26	Day 182	Day 162-217

Note that a separate visit windowing for GST is not defined. For GST score, the visit windowing of the components will be implemented first. Then, GST score will be derived when both the ADAS-Cog_{11} and ADCS-ADL23 are available at the same analysis visit.

The analysis windows for [redacted] biomarkers [redacted] will be followed as:

Analysis Visit	Protocol-Specified Day	Analysis-Specified Interval
Week 12	Day 84	Day 2-133
Week 26	Day 182	Day 134-217

[redacted]

9.7 Determination of Participants on AChEIs

If a participant received AChEIs < 28 days before the first dose of study drug, the participant is classified as "on AChEIs." Otherwise, the participant will be classified as "not on AChEIs." (If a participant was on AChEIs after the first dose of study drug, this will be reviewed at the protocol deviations (PD) review meeting before unblinding and determined if the participant is excluded from per protocol population or not.)

10 STATISTICAL ANALYSIS METHODS

10.1 Subject Disposition

A summary table will be provided for all patients to display the number of participants randomized, Safety Population, mITT Population, Primary Analysis Population, Per Protocol population, and patients who completed or discontinued the study. Listings will be provided for the different reasons for screen failure, discontinuation, and exclusion from the analysis populations. Reasons for discontinuation will be presented as reported. A summary table will be provided for all randomized patients to include those with one or more major protocol deviations. Supportive listings will present all information on the Study Completion / Early Termination and the Protocol Deviations.

10.2 Demographics and Baseline Characteristics

A summary table and listing will be provided for the Safety Population to summarize demographics and baseline characteristics. Listings will include all measurements, including entry criteria measurements and efficacy measurements, at screening and Day 1. Summary tables will be provided for baseline, for efficacy defined as either the measurement on Day 1, or else the last measurement before Day 1 if there is no Day 1 measurement.

The demographics and baseline characteristics will include Years since Diagnosis which will be calculated as follows:

(Treatment Start Date - Date of Diagnosis) / 365.25

Partial dates for the date of diagnosis will be imputed as follows (using the first month/date principle):

	Example				Example		
Partial Date	Year	Month	Date	Imputation Rule	Year	Month	Date
Only year is available	2024	**??**	**??**	Impute with first month and first day of the month	2024	**01**	**01**
Only year and month are available	2024	08	**??**	Impute with first day of the month	2024	08	**01**

10.3 Medical History

A summary table and listing for medical history will be generated. The medical history findings will be coded with the Medical Dictionary for Regulatory Activities (MedDRA).

10.4 Prior and Concomitant Medications and Procedures

A summary table for concomitant medications will be created using World Health Organization (WHO) Drug Dictionary and summarized by anatomical therapeutic class (ATC) level 2

(therapeutic main group) and preferred term (PT). Listings of Prior and Concomitant Medications and Procedures will also be created.

10.5 Treatment Compliance and Exposure

A summary table will be provided for the Safety Population grouped by treatment group to summarize the duration of treatment, total injections received, average number of injections per week and compliance as reported on case report form (CRF). The count and percentage of participants with compliance below 80% and above 110% will be provided as an assessment of under/overdosing.

10.6 Efficacy

All statistical tests will be carried out using a two-sided 5% significance level and all comparisons between ATH-1017 and placebo will be reported with 95% CIs, unless otherwise stated. The comparison between ATH-1017 40 mg and placebo among participants not receiving AChEIs will serve as the primary efficacy comparison. The comparisons between ATH-1017 70 mg and placebo, between ATH-1017 70 mg and ATH-1017 40 mg, between ATH-1017 and placebo, and between pooled ATH-1017 and placebo will be treated as exploratory comparisons. [redacted]. Due to the designed missingness of ADCS-ADL23 at Week 20, that visit may not be included in the MMRM primary analysis of GST score and ADCS-ADL23. [redacted]

10.6.1 Primary Analysis

The primary efficacy endpoint will be analyzed using a MMRM with the GST score (a composite of the CFB z-scores for the variables ADAS-Cog$_{11}$ and ADCS-ADL23) as the outcome variable and the following fixed terms: baseline GST score, treatment arm, visit, treatment by visit interaction, ApoE4 carrier status, baseline age (continuous) and the baseline MMSE stratification factor. Study site pool (with sites grouped as described in Section 9.5) will also be included as a fixed effect factor. This will be performed for the primary analysis population.

The correlations among repeated observations within a patient will be accounted for by specifying an unstructured covariance matrix. If this model fails to converge, the following covariance matrices may be evaluated in the following order to achieve model convergence: Heterogeneous Toeplitz, Toeplitz, first-order autoregressive, and compound symmetry. The objective is to achieve a covariance model with reasonable parsimony and without noteworthy model failure (this includes note messages such as “final Hessian is not positive definite”).

The Kenward-Roger adjustment for the degrees of freedom will be applied. The least squares means and treatment differences will be estimated from the MMRM model. Appendix 2 describes how the p-value for evaluation of primary endpoint success will be derived. This involves the weighting of Wald statistics from two analyses performed on subsets of the primary analysis population: participants who were included in the interim analysis and those who were not. The

MMRM model used for those two subsets will use the same structure using the order as defined above for which both subsets converge. Further information on this may also be found in Section 10.6.8.

As noted in Section 10.6, the analysis for GST score for those included in the interim analysis will not include Week 20 (but will only include Weeks 12 and 26), since at the time of the interim analysis ADCS-ADL23 was not collected at Week 20. The analysis for GST for those not in the interim analysis will include Weeks 12, 20 and 26.

As such, the p-value for GST score based on the weighted test statistics will also not be reported for Week 20, since the Week 20 results will not be available for those included in the interim analysis.

10.6.1.1 Primary Estimand

The primary estimand will be defined as follows:

- Treatment: Randomization to ATH-1017 or placebo.
- Population: Patients in the primary analysis population with mild to moderate AD defined by the protocol's inclusion/exclusion criteria.
- Patient-level outcome variable: The GST score, which is a composite of CFB components.
- Intercurrent events (ICEs): Two types of ICEs are defined below.
- Population-level summary measure: Between treatment group difference expressed in terms of the mean CFB at Week 26.

The following ICEs will be defined for the primary estimand and analysis:

- ICE-1: Discontinuation of treatment due to (1) an AE or (2) lack of efficacy. This event (1) is defined as selection of "Any Adverse Event" as a reason for study discontinuation/withdrawal, even if other reasons are selected. The (2) lack of efficacy will be determined via the selection of "Lack of efficacy" as a reason for study discontinuation/withdrawal, even if other reasons are selected.
- ICE-2: Discontinuations due to other reasons other than (1) an AE or (2) lack of efficacy.

A treatment policy strategy will be applied to both types of ICEs, i.e., all data observed up until the last day of Week 26 analysis window (Day 196) will be included in the analysis irrespective

of whether or not the ICE occurred without imputing missing outcomes, and the treatment effect will be evaluated using the model defined in Section 10.6.1.

[redacted]

[redacted]

- [redacted] "control based" hypothetical strategy [redacted]
- [redacted]

[redacted]

- [redacted]
- [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

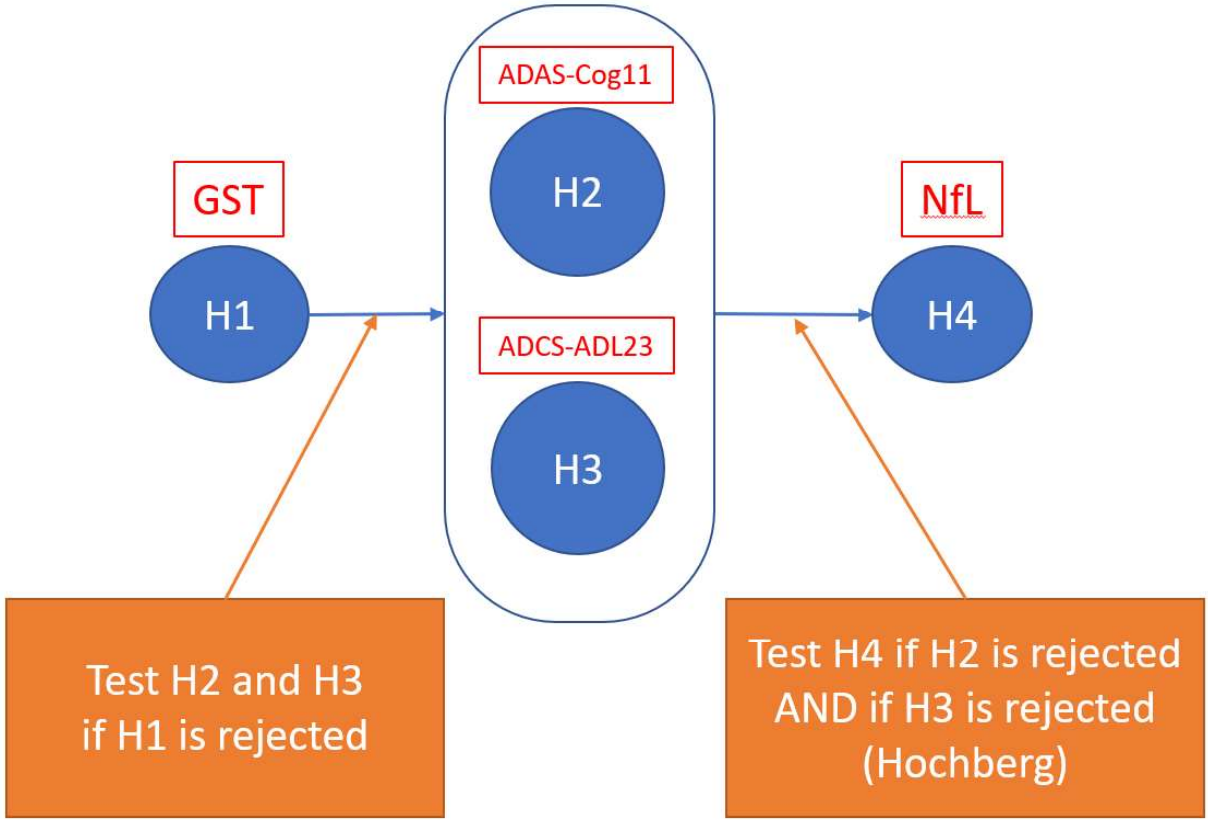
[redacted]

[redacted]

- [redacted]
- [redacted]

[redacted]

Figure 1 Statistical Gatekeeping Strategy



ADAS-Cog11, Alzheimer's Disease Assessment Scale-Cognitive Subscale; ADCS-ADL23, Alzheimer's Disease Cooperative Study – Activities of Daily Living, 23-item version; GST, Global Statistical Test; NfL, neurofilament light chain

The null hypothesis of no effect is presented for each comparison of ATH-1017 40 mg/qd versus placebo as follows: H1 = GST score; H2 = ADAS-Cog11 score; H3 = ADCS-ADL23 score; H4 = NfL

Four null hypotheses of interest will be defined for the comparison of ATH-1017 40 mg versus placebo based on the key secondary endpoints and selected exploratory endpoints:

H_1: Null hypothesis of no difference between ATH-1017 40 mg vs placebo with respect to the GST score.

H_2: Null hypothesis of no difference between ATH-1017 40 mg vs placebo with respect to ADAS-Cog_{11}.

H_3: Null hypothesis of no difference between ATH-1017 40 mg vs placebo with respect to ADCS-ADL23.

H_4: Null hypothesis of no difference between ATH-1017 40 mg vs placebo with respect to NfL.

The hypotheses will be grouped into three families:

Family 1: Hypothesis H_1.

Family 2: Hypothesis H_2 and H_3.

Family 3: Hypotheses H_4.

The families will be tested using a sequentially rejective approach starting with Family 1. The null hypotheses in Family 2 will be tested using the standard Hochberg test and the null hypothesis in Family 3 will be tested only if both null hypotheses are rejected in Family 2.

As an illustration, consider the gatekeeping procedure that would have been applied to test the four hypotheses of interest if only the first source of multiplicity had been present in the trial. Let p_1 through p_4 denote the two-sided p-values for the four null hypotheses, respectively, let the two-sided Type I error rate be denoted by $\alpha = 0.05$.

The following testing algorithm would have been used in this case:

- Step 1. Let $\alpha_1 = \alpha$ and reject the hypothesis H_1 if $p_1 \leq \alpha_1$. Proceed to Step 2 if the hypothesis is rejected.
- Step 2. Let $\alpha_2 = \alpha_1$ and apply the Hochberg test to the hypotheses H_2 and H_3, i.e., reject both hypotheses if the larger p-value is $\leq \alpha_2$; otherwise reject the associated hypothesis if the smaller p-value is $\leq \alpha_2/2$. Proceed to Step 3 if H_2 and H_3 are both rejected.

Step 3. Let $\alpha_3 = \alpha_2$ and reject the hypothesis H_4 if $p_4 \leq \alpha_3$.

Since there are two sources of multiplicity in the trial, the gatekeeping procedure will be applied in conjunction with the combination function approach, see Section 10.6.8.

[redacted]

[redacted]

[redacted]

10.6.6 Pharmacokinetic Analyses

A separate document will describe the planned PK analyses, which are not in scope for this present document.

10.6.7 Interim Analysis

An adaptive design with a single interim analysis was employed in this trial. The interim analysis corresponded to approximately a 45% information fraction and supported an option to apply the futility stopping rule and SSR rule.

10.6.7.1 Interim Analysis Rules

The futility stopping and SSR rules were based on the CP calculated for the primary endpoint (GST). CP was defined using a pooled analysis of ATH-1017 40 mg/day and ATH-1017 70 mg/day at a one-sided 0.05 level under the assumption, at that time, that the total final sample in the trial would be 175 patients. The resulting futility and SSR rules are defined in Table 2.

Table 2 Futility and SSR rules at the interim analysis

Condition	Interim decision
CP < 19.8%	Stop the trial for futility
19.8% ≤ CP < 66.4%	Increase the total sample size to 275
66.4% ≤ CP < 70.7%	Increase the total sample size to 250
70.7% ≤ CP < 75.3%	Increase the total sample size to 225
75.3% ≤ CP < 80%	Increase the total sample size to 200
80% ≤ CP	Continue to the planned sample size of 175

CP, conditional power; SSR, sample size re-estimation

Based on these rules, the IDMC recommended that the total sample size should be increased to 275 patients.

10.6.7.2 Updated Primary Analysis Population

Protocol amendment v7 (03 May 2023) discontinued further randomization of participants to 70 mg/ once a day (QD), leading to a 1:1 40 mg/once a day (QD) vs placebo randomization, and a target cohort size of N=149 for a total of 298 evaluable participants for the final analysis. This also led to removal of 70 mg/once a day (QD) participants from the primary analysis population and an updated statistical testing hierarchy.

10.6.8 Final Analysis Rules

The final assessment of the primary and key secondary endpoints will be performed using the Cui, Hung and Wang (CHW) approach (Cui, 1999). It is important to note that this approach guarantees overall error rate control for any interim decision rule as long as the combination weights are prospectively defined.

The combination function approach will be applied in conjunction with the gatekeeping procedure defined in Section 10.6.3 to ensure that the overall Type I error rate is preserved at a two-sided 0.05 level with respect to the two sources of multiplicity in the trial. The two sources of multiplicity will be accounted for analogous to the method developed in Kordzakhia, 2018. The details of this method are presented in Appendix 2. This approach requires that the primary and key secondary endpoint analyses will be conducted for two sets of participants: participants in the primary analysis set who 1) were included in the interim analysis, and 2) were not included in the interim analysis. The Wald statistic from the first will receive weight square root of 0.45, and the Wald statistic from the second will receive weight square root of 0.55. The resulting test statistic will be evaluated as per asymptotic normality.

10.7 Safety and Tolerability

10.7.1 Adverse Events

AEs reported on CRFs will be coded into system organ classes and PTs using MedDRA v23.1 or later. A treatment-emergent adverse event (TEAE) is defined as an AE with an onset date on or after the start of dosing. Any AEs that are not considered treatment-emergent will be provided in data listings only.

All safety summaries will be based on the safety set. In addition, for selected safety summaries, the primary analysis set will be used, as specified in the Table of Contents for the Tables, Listings, and Figures. For each treatment group, AE incidence rates will be summarized with frequency and percentage by MedDRA system organ class (SOC) and PT, with all participants treated in that treatment group as the denominator, unless otherwise specified. In addition, TEAE incidence rates will also be summarized by severity grade and relationship to study drug. Treatment-related AEs are those judged by the Investigator to be at least related to the blinded study treatment. Participants with multiple occurrences of events will only be counted once at the maximum severity to study drug for each PT, SOC, and overall. The number and percentage of participants with the following TEAEs will be summarized:

Overall summary of TEAEs: participants experiencing at least one AE, any serious adverse event (SAE), any treatment-related AE, any treatment-related SAE, any discontinuations due to an AE, and any deaths

TEAEs by SOC and PT

TEAEs by SOC, PT, and maximum severity

Drug-related TEAEs by SOC and PT

Drug-related TEAEs by SOC, PT, and maximum severity

TEAEs that led to study drug interruption by SOC and PT

TEAEs that led to study drug discontinuation by SOC and PT

TEAEs that led to study drug discontinuation by SOC, PT, and maximum severity.

Treatment-emergent SAEs by SOC and PT

Treatment-emergent SAEs by SOC, PT, and maximum severity

Drug-related Treatment-emergent SAEs by SOC and PT

TEAEs that led to study discontinuation

TEAEs leading to death

Deaths and cause of death will be summarized.

Most commonly reported TEAEs or treatment-emergent SAEs will be defined as an event of interest reported with a frequency of > 5% and >2% respectively per treatment group.

Also, data listings by participant, treatment group, verbatim term, and PT for all these AEs will be created as appropriate.

10.7.2 Clinical Laboratory

Continuous blood clinical laboratory analytes absolute values and CFB values will be summarized by analyte and visit using descriptive statistics (mean, median, SD, minimum, maximum, and number of participants). Categorical laboratory analytes, classified as normal or abnormal, will be summarized by analyte and visit using the number and percentage of participants in each category and in each treatment group. The denominators for calculating the percentages will be based on the number of participants with non-missing assessments in each treatment group at a particular visit for the safety population. The latest non-missing clinical laboratory tests collected prior to dosing will be used as the baseline values.

Shifts to values outside of the normal range will be presented by analyte and will be summarized by the number and percentage of participants with shifts. Shifts will be determined for analytes in which both the baseline value and the termination value are recorded. The denominators for calculating the percentages will be based on the number of participants with non-missing assessments for a particular analyte.

Clinical laboratory results will be provided in data listings by participant, visit and analyte. Abnormal lab results (either below lower limit of normal [LLN] or above upper limit of normal [ULN]) will be provided in a separate listing by participant, center, analyte, treatment group and visit.

Listings will be provided for clinical laboratory results for hematology, blood chemistry and urinalysis, and abnormal laboratory results.

Laboratory values that are below or above the limit of quantification will be handled as follows:

Limit of quantification	Values to use
Below the limit of quantification	Use the lower limit of quantification
Above the limit of quantification	Use the upper limit of quantification

10.7.3 Electrocardiograms

ECG values and CFB values will be summarized by visit using descriptive statistics for (interval from the P wave [atrial contraction or depolarization] to the onset of the Q wave in the measurement of electrical activity of the myocardium [PR]) interval, QT interval, QT interval using Bazett's formula (QTcB) interval, QT interval using Fridericia's formula (QTcF) interval, the time elapsed between two successive R-waves of the QRS signal on the electrocardiogram (RR interval), mean heart rate, and Q wave, R wave, and S wave (QRS) duration. ECG abnormalities will be summarized as the count and percentage of participants in each treatment group. CFB will be summarized in a shift table crossing baseline and each visit result. The denominators for calculating the percentages will be the number of participants in each treatment group who have an evaluation for both the screening and each post-baseline visit in the safety population. These results will be analyzed descriptively.

10.7.4 Vital Signs

Each vital sign will be summarized by treatment and by visit, using descriptive statistics (mean, median, SD, minimum, maximum, and number of participants) for the safety population. Additionally, descriptive summaries will be provided for CFB values for each treatment by visit for vital sign measurements collected during the study.

The latest non-missing vital sign value collected prior to dosing will be used as the baseline values. The baseline values will usually be the vital signs recorded at the baseline visit. In the case of repeated vital signs, the last collected values within that visit will be used for the summary tables.

Vital signs will be provided in a data listing by participant, treatment group, visit, and parameter.

10.7.5 Physical and Neurological Examinations

Physical and neurological examination findings will be listed, and also summarized as the count and percentage of participants in each treatment group.

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

Results from the C-SSRS questionnaire will be listed and summarized using descriptive statistics by treatment group and visit. Additionally, descriptive summaries will be provided for CFB values for each treatment by visit.

10.7.7 Geriatric Depression Scale (GDS)

Results from the GDS questionnaire total score will be listed and summarized using descriptive statistics by treatment group and visit. Additionally, descriptive summaries will be provided for CFB values for each treatment by visit.

10.8 Pharmacokinetics/Pharmacodynamics

These will be addressed in a SAP specific for PK analyses.

10.9 Other Relevant Data Analyses/Summaries

N/A

11 REFERENCES

[redacted]

Cui, L., Hung, H.M., Wang, S.J. (1999).
Modification of sample size in group sequential clinical trials. Biometrics. 55, 853-857.

[redacted]

Folstein M.F., Folstein S.E., McHugh P.R. (1975).
"Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatric Research 12(3), 189-198.

Galasko D., Bennett D., Sano M., Ernesto C., Thomas R., Grundman M., et al. (1997).
An inventory to assess activities of daily living for clinical trials in Alzheimer's disease. The Alzheimer's disease Cooperative Study. Alzheimer's Disease and Associated Disorders, 11(S2), S33-S39.

Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. (1982).
A new clinical scale for the staging of dementia. Br J Psychiatry 140, 566-572.

Kordzakhia, G., Dmitrienko, A., Ishida, E. (2018).
Mixture-based gatekeeping procedures in adaptive clinical trials. Journal of Biopharmaceutical Statistics. 28, 129-145.

Lipkovich, I. and A. Dmitrienko (2014).
"Strategies for identifying predictive biomarkers and subgroups with enhanced treatment effect in clinical trials uses SIDES." J Biopharm Stat 24(1): 130-153.

McKhann, G.M., et al. (2011).
The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging – Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & Dementia: The Journal of the Alzheimer's Association 7, 263-269.

O'Brien, P.C. (1984).
Procedures for comparing samples with multiple endpoints. Biometrics. 40, 1079-1087.

Rosen WG, Mohs RC, Davis KL (1984).
A new rating Scale for Alzheimer's disease. Am J Psychiatry 141:1356-1364.

[redacted]

S.P. Dickson, B. Haaland, C.H. Mallinckrodt, et al.
"Time Saved" Calculations to Improve Decision-Making in Progressive Disease Studies, J Prev Alz Dis 2024;3(11):529-536; http://dx.doi.org/10.14283/jpad.2024.64

[redacted]

Wassmer, G., Brannath, W. (2016).
Group Sequential and Confirmatory Adaptive Designs in Clinical Trials. New York: Springer.

[redacted]

12 APPENDICES

Appendix 1 Derivation of GST scores.

A composite approach will be utilized to derive the GST score for each patient at each visit in order to support an overall assessment of disease based on the combination of the cognition and functional scores (z-score of CFB ADAS-Cog$_{11}$ and z-score of CFB ADCS-ADL23 scores).

The GST score follows the principles articulated in O'Brien (1984), using z-scores to achieve comparable scales for effect sizes (relative to the variation among participants for that measure) in the components of this composite score. For this trial, we will use the principle that for pooling used to estimate SD, one does not include participants in a pool if they have different values of a variable being tested; in this case we will not put participants in the same pool if they differ for 1) treatment arm, 2) AChEI usage, or 3) before protocol amendment v5 vs after protocol amendment v5. This results in 9 pools (each treatment arm +/-AChEI before/after protocol amendment v5, noting that there were no +AChEI participants after protocol amendment v5). This approach leads to the calculation of ADAS-Cog$_{11}$ and ADCS-ADL23 z-scores and the resulting GST scores once, for all analyses of GST score, for each participant-visit combination; these "one time" calculations will use the overall visit mean (without regard to treatment arm, +/-AChEI usage, or before/after protocol amendment v5), and the nine pools just described for estimation of a pooled SD.

In the development below, the notation includes a subscript for VISIT, and the algorithm is applied at each VISIT without consideration of the other VISITs. This means that each VISIT, an overall mean and pooled SD are calculated for both CFB ADAS-Cog$_{11}$ and CFB ADCS-ADL23 for creating the z-scores.

Note that a participant is only included in the overall mean and pooled SD at a particular VISIT if that participant has reported values for both CFB ADAS-Cog$_{11}$ and CFB ADCS-ADL23 at that particular VISIT.

In brief, for each VISIT, we calculate an overall mean and a pooled SD, where we include each participant who has values for both CFB ADAS-Cog$_{11}$ and CFB ADCS-ADL23 at that VISIT; z-scores are calculated for each participant at that VISIT for both CFB ADAS-Cog$_{11}$ and CFB ADCS-ADL23 by subtracting the respective overall mean and then dividing by the respective overall pooled SD.

The remainder of this Appendix is an exposition which details how this is done for the case of only two pools, noting that our specific use case does involve nine pools, and so for this study the natural extension of this example is required.

The following notation will be used to compute the GST score. Let x_{ijk} and y_{ijk} denote the changes from baseline to the kth visit for the jth patient in the ith trial arm ($i = 1$ corresponds to placebo and $i = 2$ corresponds to the selected ATH-1017 arm, i.e., ATH-1017 40 mg or ATH-1017 70 mg) for CFB ADAS-Cog$_{11}$ and CFB ADCS-ADL23, respectively. To simplify notation, it will be assumed that the number of patients in a trial arm does not depend on the visit, i.e., n_1 patients in the placebo arm and n_2 patients in the selected ATH-1017 arm at each visit.

The GST score will be computed by applying the following algorithm to the ADAS-Cog$_{11}$ and ADCS-ADL23 change scores. Beginning with ADAS-Cog$_{11}$, the overall mean will be computed at the kth visit, i.e.,

$$m_k = \frac{n_1\bar{x}_{1\cdot k}+n_2\bar{x}_{2\cdot k}}{n_1+n_2},$$

where $\bar{x}_{1\cdot k}$ and $\bar{x}_{2\cdot k}$ are the arm-specific sample means. Next, the pooled SD will be computed at the kth visit, i.e.,

$$s_k = \sqrt{\frac{(n_1-1)s_{1k}^2 + (n_2-1)s_{2k}^2}{n_1 + n_2 - 2}}$$

where s_{1k}^2 and s_{2k}^2 are the arm-specific sample variances. Note that the pooled SDs will be derived using an appropriate pooling algorithm. Finally, the standardized change scores will be defined as follows:

$$x_{ijk}^* = \frac{x_{ijk} - m_k}{s_k}$$

The standardized change scores for ADCS-ADL23 will be computed in a similar way and will be denoted by y_{ijk}^*.

The GST score for the CFB to the kth visit for the jth patient in the ith trial arm is defined as

$$\frac{x_{ijk}^* - y_{ijk}^*}{2}$$

Note that a negative sign has been applied to y_{ijk}^* since an increase in the ADCS-ADL23 score indicates a favorable response. This algorithm will be applied to each post-baseline visit. It is important to note that this detailed example is for the situation of two pools, while for this trial there are nine pools, and the algorithm will be implemented for nine, rather than two pools, using the natural extension of pooled estimation of SD.

A similar algorithm will be applied to compute the GST score at baseline. Specifically, let x_{ij} and y_{ij} denote the baseline values for the jth patient in the ith trial arm ($i = 1$ corresponds to placebo and $i = 2$ corresponds to ATH-1017) for ADAS-Cog$_{11}$ and ADCS-ADL23, respectively. After that, the algorithm defined above needs to be applied to each endpoint, i.e., compute the overall mean at baseline and the pooled SD at baseline to derive the standardized score at baseline for each endpoint (note that a negative sign needs to be again applied to the standardized baseline score for ADCS-ADL23). The GST score at baseline is then defined as the average of the endpoint-specific standardized scores at baseline.

Appendix 2 Statistical Methodology at the Final Analysis

The following notation will be used to set up the decision rules for the final analysis. First, two trial stages are defined:

- Stage 1 is comprised of all patients included in the interim analysis database.
- Stage 2 is comprised of the remaining patients.

Let p_1 through p_4 denote the two-sided p-values for the hypotheses of interest computed from the Stage 1 data and, similarly, let q_1 through q_4 denote the two-sided p-values for the hypotheses computed from the Stage 2 data.

The final decision rules for the four null hypotheses will be defined in terms of the combination p-values denoted by $r(I)$. The combination p-value for the intersection hypothesis $H(I)$ is derived using CHW method, i.e.,

$$Z_{CHW} = \sqrt{w} * Z_1 + \sqrt{1-w} * Z_2$$

where $w = 0.45$ is the pre-specified weight, Z_1 is the t test statistic using the pre-IA patients, Z_2 is the t test statistic using the post-IA patients, and Z_{CHW} is the weighted Z-score for the CHW method. A p-value is calculated from Z_{CHW} based on a standard normal distribution.

The resulting set of decision rules protects the overall Type I error rate at a two-sided 0.05 level with respect to both sources of multiplicity in the trial.

[redacted]

Appendix 3 Schedule of Assessments and Procedures

Assessment		Screening [a]	Randomized placebo-controlled treatment period (26-week)							Safety follow-up [p]
	Visit:	1	2	3	4	5	6	7	8/ET [o]	9
	Week:	-4 to -1	1	2	6	12	16	20	26	30
	Day:	-28 to -1	1	14 (± 7)	42 (± 7)	84 (± 7)	112 (± 7)	140 (± 7)	182 (± 7)	210 (± 7)
Inclusion/Exclusion		X	X							
Informed Consent		X								
Demographics		X								
Medical History		X								
Height and Weight		X	X[b]			X[b]			X[b]	X[b]
Blood [c]		X								
C-SSRS [d]*		X	X	X	X	X	X	X	X	X
GDS		X	X			X			X	X
MMSE *		X	[redacted]							
CDR		X								
Randomization			X							
Drug Dispensing [e]			X	X	X	X	X	X		
Dose of IMP in-clinic [f]			X	X	X	X	X	X	X	
Drug Accountability				X	X	X	X	X	X	
Physical and Neurological Exam [g]		X	X	X	X	X	X	X	X	X
MRI [h]		X								
12-Lead ECG [i]		X	X	X	X	X	X	X	X	X
Vital signs [j]		X	X	X	X	X	X	X	X	X
Safety Labs [k]		X	X	X	X	X	X	X	X	X
AE		X	X	X	X	X	X	X	X	X
Conmeds [l]		X	X	X	X	X	X	X	X	X
ADAS-Cog$_{11}$ *		X	X	X	X	X		X	X	X
[redacted]			[redacted]	[redacted]	[redacted]	[redacted]		[redacted]	[redacted]	[redacted]
[redacted]			[redacted]			[redacted]			[redacted]	
ADCS-ADL23 *			X			X		X	X	
[redacted]			[redacted]			[redacted]			[redacted]	
PK (plasma) [m] *			X			X			X	
Plasma Sample Collection (biomarker analyses) [n]			X			X			X	
[redacted]			[redacted]						[redacted]	
[redacted]			[redacted]						[redacted]	
[redacted]			[redacted]						[redacted]	

ADASCog11, Alzheimer's Disease Assessment Scale-Cognitive Subscale; ADCS-ADL23 , Alzheimer's Disease Cooperative Study-Activities of Daily Living, 23-item version; [redacted] AE, adverse event; ApoE, apolipoprotein E; CDR, Clinical Dementia Rating Scale; [redacted] C-SSRS, Columbia-Suicide Severity Rating Scale; ECG, 12-lead electrocardiogram; [redacted] GDS, Geriatric Depression Scale; IMP, investigational medicinal product; MMSE, Mini-Mental State Examination; MRI, magnetic resonance imaging; [redacted] PK, pharmacokinetic; [redacted] SC, subcutaneous; [redacted]

a. If 28 days is not sufficient to complete the Screening period, the possibility of an extension can be discussed with the Medical Monitor.
b. Only weight collected at Baseline/Day 1 (Visit 2), Week 12 (Visit 5), Week 26 (Visit 8), and Safety follow-up (Visit 9).
c. Blood collection for follicle-stimulating hormone (FSH) levels (to confirm post-menopausal state in females), serology, ApoE genotyping, folate, Vitamin B12, free triiodothyronine (fT3), free thyroxine (fT4), and thyroid-stimulating hormone (TSH).
d. 'C-SSRS Baseline/Screening' version will be administered at Screening and 'C-SSRS Since Last Visit' version will be administered at all post-Screening visits.
e. Dispensing of kits containing study drug will occur every 2 weeks, at the study site or as needed by direct-to-patient shipment; larger provision of study drug is permitted to accommodate personal need, e.g., vacation; drug returns will be recorded, and compliance calculated. IP administration by participant or caregiver will be assessed at Visits 2 through 7, inclusive.
f. First SC injection of IMP will be performed at site under supervision. Participants will remain at site for 2 hours ± 15 minutes for safety observation follow up. They may be discharged from the clinic at this time in the absence of any systemic AEs (but not for local site reaction). Should they have systemic AEs, they should remain under observation for an additional 2 hours and may be discharged at that time with investigator's clearance. Participants should withhold IMP dose on the day of subsequent clinic visits whereupon IMP administration will be done on site under supervision of site staff. There is no specified in-clinic observation period for these subsequent visits but will require investigator discharge from the clinic. It is recommended to contact the participant/caregiver by phone at appropriate intervals to support dosing compliance and injection techniques.
g. Physical and neurological exam should be done post-dose on all visits when dosing applies
h. MRI (or computed tomography (CT) for participants with non-MRI-safe cardiac pacemaker, or other relevant medical reason, with Medical Monitor approval) scan must have been performed within 12 months before Screening. If such scan is unavailable or older than 12 months, it should be repeated to ascertain the diagnosis before randomization.
i. 12-lead ECGs will be performed pre-dose and 30 (± 15) minutes postdose on Day 1 and 30 (± 15) minutes postdose at all other visits. All ECG assessments will be performed in triplicate sequentially.
j. Vital signs will be performed pre-dose on all visits. Supine blood pressure (BP) and heart rate (HR) recordings will be made after the participant has been supine for at least 5 minutes. Orthostatic BP will be recorded as follows: the first blood pressure will be the average of 3 measurements recorded after the participant is supine for 5 minutes; the second blood pressure will be recorded after the participant stood for up to 3 minutes.
k. Safety labs include chemistry, hematology, and urinalysis.
l. Prior or concurrent medications.
m. PK plasma samples will be collected at pre-dose and post-dose on Baseline/Day 1 (Visit 2); pre-dose and post-dose at Week 12 (Visit 5) and Week 26 (Visit 8). The pre-dose PK sample is collected any time before dosing. The post-dose PK sample is collected anytime between 30 minutes and 120 minutes after dosing as practical. The actual time of dosing and of PK sampling will be recorded.
n. Plasma samples will be collected for biomarker analysis. Prior to protocol amendment v7 (03 May 2023) this was optional for participants.
o. Participants who terminate prior to Visit 9 are to complete same assessments as Visit 8/early termination (ET). For clinical outcome assessments if completed within 4 weeks of the ET visit, they do not need to be repeated; all safety outcomes and drug accountability should be performed regardless of interval.

p. Safety follow-up visit to be performed for participants who do not roll over into the optional open-label extension (OLEX) study; participants who roll over into the OLEX study will complete the safety follow-up visit at the end of the OLEX study.

* At Baseline/Day 1 (Visit 2), MMSE should be done first before all other assessments pre-dose. ADAS-Cog_{11}, [redacted] will be performed pre-dose; ADCS-ADL23, [redacted] C-SSRS, [redacted] will be performed anytime during the visit.

For visits after the baseline (except for safety follow-up when dosing is not applicable), all clinical outcome assessments will be performed post-dose. [redacted]

Appendix 4 Imputation of Partial Dates

Table 3 Imputation Rules for Partial Dates – AEs and Prior/Concomitant Medications

Parameter	Missing	Additional Condition	Imputation
Start Date	D only	M and Y same as M and Y of first study drug dosing	Date of first study drug dose
		M and/or Y not the same as M and Y of first study drug dosing	First day of indicated month
	M and D	Y same as Y of first study drug dosing	Date of first study drug dose
		Y not the same as Y of first study drug dosing	01 Jan of indicated year
	M, D, and Y	none – date completely missing	Date of first study drug dose
End Date	D only	M and Y same as M and Y of last study date	Date of last study date
		M and/or Y not the same as M and Y of last study drug dosing	Last day of indicated month
	M and D	Y same as Y of last study date	Date of last study date
		Y not the same as Y of last study drug dosing	31 Dec of indicated year
	M, D, and Y	none – date completely missing	AE or medication is ongoing

D, day; AE, adverse event; M, month; Y, year

Note: If the end date is complete and the imputed start date is after the end date, then the start data will be imputed using the end date. If the imputed end date is before the start date, then the imputed end date will be equal to the start date.