A Randomized, Placebo-Controlled, Translational Study of ATH-1017 in Subjects with Mild to Moderate Alzheimer's Disease

Pentara Corporation
2261 East 3300 South
Millcreek, UT 84109

Protocol:
ATH-1017-AD-0202 v3.00 dated 25-May-2021

SAP Version: 1.1, 02-Jun-2022

NCT Number: NCT04491006
This NCT number has been applied to the document for
purposes of posting on Clinicaltrials.gov

A Randomized, Placebo-Controlled, Translational Study of ATH-1017 in Subjects with Mild to Moderate Alzheimer's Disease

By signing below, all parties accept that the analysis methods and data presentations are acceptable and that this document is final.

Prepared By:

DocuSigned by:
[redacted]
Signer Name: [redacted]
Signing Reason: I am the author of this document
Signing Time: 6/2/2022 | 2:55:44 PM PDT
03BFBF2959B94D25882657D5383A86F0

6/2/2022

[redacted], PhD — Date

Vice President of Statistics, Pentara Corporation

Reviewed By:

DocuSigned by:
[redacted]
Signer Name: [redacted]
Signing Reason: I approve this document
Signing Time: 6/2/2022 | 3:17:35 PM PDT
4702CD2DFA584E0688AE96D6F0925B7F

6/2/2022

[redacted], PhD — Date

CEO, Pentara Corporation

Reviewed/Approved By:

DocuSigned by:
[redacted]
Signer Name: [redacted]
Signing Reason: I approve this document
Signing Time: 6/2/2022 | 3:40:41 PM PDT
8AA0B90F82EF4D47A8CF[redacted]234

6/2/2022

[redacted], MD, PhD — Date

Chief Medical Officer,

Athira Pharma, Inc.

Modification History

Unique Identifier for this Version	Date of the Document Version	Author	Significant Changes from Previous Authorized Version
0.1	05-APR-2021	[redacted]	not applicable—first version
0.2	12-APR-2022	[redacted]	- specifying pooled treatment arm vs placebo as primary - removing hierarchy for secondary - primary outcome as change from baseline to all post baseline, post-dose P300 latencies - alpha control only for primary endpoint - specify acute and sustained analysis of p300 latency as additional endpoints - covariance structure for MMRM model updated - update covariates list for primary and secondary models - specify [redacted] (subgroup analyses) of +/- ACHEI and for disease severity (mild/moderate) - clarify that a Pearson correlation will be used for the secondary outcome (between change in P300 latency and cognitive measures - specify outcomes of all GST combinations of interest - removed all sensitivity analysis as outcomes of interest are specified as secondaries - rename [redacted] to additional outcomes -7.5.1.2 changed to use the same model as 7.4.1 except with time and time interactions included for the correlations. -sections describing primary analysis changed to prioritize pooled active treatment vs. placebo.

0.3	28-APR-2022	[redacted]	- define ROI (region of interest) as average of CZ, FC1, FC2, CP1, and CP2 electrodes as primary efficacy endpoint - update language in SAP to define primary efficacy endpoint as the average of pre-dose and post-dose records within an analysis window - update language regarding baseline definition, pre-dose (sustained) and pre-to-post (acute) to be more explicit
1.0	11-MAY-2022	[redacted]	-Clarified that all visits are used in the covariance modeling but that visits at week 12, 16, 20, and 26 are averaged for the primary endpoint and key secondary. -Changed the key secondary to match the paragraph text of the protocol, which does not include COWAT, so the GST is ADASCog and CGIC instead. -Ordering of secondary objectives is changed and some are moved to exploratory. -Pooled treatment changed to active and a paragraph added to general section that describes the pooling of active treatments. -Allowed for keeping data if there is only pre or post, eliminating sentence that stated must have both a pre and post for a visit to be included. -Key secondary now uses neither the model of the primary nor a Pearson correlation, instead has its own model. Region of interest defined in general section. -Removed duplicated descriptions by moving as many as possible to the general section. -COWAT and ADASCog version of

			the GST removed since it is only mentioned in the summary and heading of the protocol but that contradicts the text of the protocol. -Concurrent medication coding described. -Clarified primary analysis not using weeks 2 and 6 in the endpoint. -Clarified averaging of primary endpoint visits. -Removed unneeded explanation of overall effect of treatment. -Added treatment descriptions in the general section. -Significant differences defined as p < .05 in the general section.
1.1	02-Jun-2022	[redacted]	-added GST of (COWAT + ADCS-ADL) as secondary analysis

Table of Contents

1 Abbreviations

AChEI	Acetylcholinesterase inhibitor
AD	Alzheimer's disease
AE	Adverse event
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
AKT	Protein kinase B
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
ApoE	Apolipoprotein E
aPTT	Activated partial thromboplastin time
AST	Aspartate aminotransferase
CBC	Complete blood count
CBD	Cannabidiol
CDR	Clinical dementia rating scale
CFB	Change from baseline
CPK	Creatine phosphokinase
CRO	Contract research organization
CYP3A4	Cytochrome P450 3A4
C_{max}	Maximum concentration
CNS	Central nervous system
COWAT	Controlled oral word association test
C-SSRS	Columbia-suicide severity rating scale
CT	Computerized tomography
DSMB	Data safety monitoring board
ECG	Electrocardiogram
eCRF	Electronic case report form
EDC	Electronic data capture
EEG	Electroencephalogram
[redacted]	[redacted]
ERP	Event-related potentials

ET	Early termination
FSH	Follicle-stimulating hormone
fT3	Free tri-iodothyronine
fT4	Free thyroxine
FWER	Family-wise error rate
GCP	Good clinical practice
GDS	Geriatric depression scale
GGT	Gamma-glutamyl transferase
GLP	Good laboratory practice
GST	Global statistical test
HBsAg	Hepatitis B surface antigen
HCV	Hepatitis C virus
HGF	Hepatocyte growth factor
HIV	Human immunodeficiency virus
HR	Heart rate
ICF	Informed consent form
ICH	International council for harmonisation
IEC	Independent ethics committee
IMP	Investigational medicinal product
INR	International normalized ratio
IRB	Institutional review board
IRT	Interactive response technology
LAR	Legally authorized representative
LTP	Long-term potentiation
MAPK	Mitogen-activated protein kinase
MCT	Medium-chain triglyceride
MET	MET receptor tyrosine kinase
mITT	Modified intent-to-treat
MMRM	Mixed model with repeated measures
MMSE	Mini-mental state examination
MRI	Magnetic resonance imaging

NMDA	N-methyl D-aspartate
[redacted]	[redacted]
OD	Once daily
P	Phosphorylated
PD	Pharmacodynamic(s)
PI3K	Phosphoinositide 3-kinase
PK	Pharmacokinetic(s)
PKC	Protein kinase C
PLCγ	Phospholipase C-gamma
PM	Plasma membrane
PRN	As needed
PSP	Post-synaptic potential
PT	Prothrombin time
[redacted]	[redacted]
QTcF	Corrected QT interval using Fridericia's formula
RAC1	Ras-related C3 botulinum toxin substrate 1
RAF	Rapidly accelerated fibrosarcoma (protein)
RAS	Rat sarcoma (protein)
RBC	Red blood cells
ROI	Region of interest
[redacted]	[redacted]
SAE	Serious adverse event
SAP	Statistical analysis plan
SBP	Systolic blood pressure
SC	Subcutaneous
SE	Standard error
SOP	Standard operating procedure
STAT3	Signal transducer and activator of transcription 3
THC	Tetrahydrocannabinol
TSH	thyroid-stimulating hormone
ULN	Upper limit of normal

US(A)	United States (of America)
VAS	Visual analog scale
WBC	White blood cells
[redacted]	[redacted]

Table 1 Schedule of Assessments

				Double-blind placebo-controlled treatment period (26-week)							Safety follow-up [r]
		Screening [a]	Pre-baseline	Baseline							
	Visit:	1	2a	2b	3	4	5	6	7	8/ET [q]	9
	Week:	-4 to -2	-1	1	2	6	12	16	20	26	30
Assessment	Day:	-28 to -6	-5 to -3	1	14 (±7)	42 (±7)	84 (±7)	112 (±7)	140 (±7)	182 (±7)	210 (±7)
Inclusion/ Exclusion		X	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	X								
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	X
Conmeds [l]		X	X	X	X	X	X	X	X	X	X
Hearing Test [m]		X									
ADAS-Cog11 *				X	X	X	X		X	X	X
COWAT *				X	X	X	X		X	X	X
ADCS-CGIC *				X			X			X	
ADCS-ADL23 *				X			X			X	
[redacted]				[redacted]			[redacted]			[redacted]	

				Double-blind placebo-controlled treatment period (26-week)							Safety follow-up [r]
		Screening [a]	Pre-baseline	Baseline							
	Visit:	1	2a	2b	3	4	5	6	7	8/ET [q]	9
	Week:	-4 to -2	-1	1	2	6	12	16	20	26	30
Assessment	Day:	-28 to -6	-5 to -3	1	14 (±7)	42 (±7)	84 (±7)	112 (±7)	140 (±7)	182 (±7)	210 (±7)
EEG (ERP P300 and [redacted]) [n]			X (no dosing)								X (no dosing)
Pre-dose				X	X	X	X	X	X	X	
Post-dose				X	X	X	X	X	X	X	
PK (plasma) [o]				X			X			X	
Biobanking Plasma [p]				X						X	
[redacted]				[redacted]						[redacted]	
[redacted]				[redacted]						[redacted]	
[redacted]				[redacted]						[redacted]	

ADCS-ADL23 = Alzheimer's Disease Cooperative Study-Activities of Daily Living, 23-item version; ADAS-Cog11 = Alzheimer's Disease Assessment Scale-Cognitive Subscale; ADCS-CGIC = Alzheimer's Disease Cooperative Study-Clinical Global Impression of Change; AE = adverse event; CDR = Clinical Dementia Rating Scale; COWAT = Controlled Oral Word Association test; C-SSRS = Columbia-Suicide Severity Rating Scale; ECG = electrocardiogram; ERP = event-related potential; [redacted] GDS = Geriatric Depression Scale; IMP = investigational medicinal product; MMSE = Mini-Mental State Examination; MRI = magnetic resonance imaging; [redacted] PK = pharmacokinetic; [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 2b), Week 12 (Visit 5), Week 26 (Visit 8), and Safety follow-up (Visit 9).

c Blood collection for FSH levels (to confirm post-menopausal state in females), serology, ApoE genotyping, folate, Vitamin B_{12}, fT3, fT4, and 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 subject or caregiver will be assessed at Visits 2b through 7, inclusive.

f First SC injection of IMP will be performed at site under supervision. Subjects will remain at site for 2 hours ± 15 minutes for safety observation follow up. They may be discharged from the clinic at this time absent 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. Subjects 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 subject/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 CT for subjects 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 post-dose on Day 1(Visit 2b) and 30 (± 15) minutes post-dose at all other visits. All ECG assessments will be performed in triplicate approximately 1 minute apart.
j. Vital signs will be performed pre-dose on all visits. Supine BP and HR recordings will be made after the subject 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 subject is supine for 5 minutes; the second blood pressure will be recorded after the subject stood for up to 3 minutes.
k. Safety labs include chemistry, hematology, and urinalysis.
l. Prior or concurrent medications.
m. Subject hearing will be tested to establish suitability for auditory ERP assessment, i.e., ability to hear and differentiate two different tones, using the centrally provided EEG equipment; hearing aid must be removed during the screening hearing test and during EEG recordings.
n. At Pre-baseline visit (Visit 2a, Day -5 to Day -3, no dosing), EEG assessments (ERP P300 and [redacted]) will be performed twice approximately 2 hours apart. EEG data should be uploaded for quality check immediately after the completion of the Pre-baseline visit (Visit 2a). If both EEG assessments failed quality check at pre-baseline, the visit will be repeated once and the screening window of 28 days can be extended upon approval by Medical Monitor.
At Baseline/Day 1 (Visit 2b), EEG assessments (ERP P300 and [redacted]) will be performed at pre-dose following the completion of baseline assessments of ADAS-Cog$_{11}$ and COWAT, and before the ADCS-CGIC assessment, up to 1.5 hour before dose in clinic. EEG will be assessed post-dose at approximately 2 (±1) hours after IMP dosing.
At Visits 3, 4, 5, 6, 7, and 8, EEG assessments (ERP P300 and [redacted]) will be performed at pre-dose up to 1 hour before dose in clinic. EEG will be assessed post-dose following the completion of ADAS-Cog$_{11}$ and COWAT assessments, and before the ADCS-CGIC assessment, at approximately 2 (±1) hours after IMP dosing.
At safety follow up (Visit 9, no dosing), EEG assessments (ERP P300 and [redacted]) will be performed following the completion of ADAS-Cog$_{11}$ and COWAT.
o. PK plasma samples will be collected at post-dose on Baseline/Day 1 (Visit 2b); pre-dose and post-dose at Week 12 (Visit 5) and Week 26 (Visit 8). The pre-dose PK sample is collected anytime 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.
p. Plasma sample will be collected and banked for biomarker analysis (only for subjects who provided consent for plasma biobanking). Ten aliquots will be created from the plasma sample.
q. Subjects who terminate prior to Visit 9 are to complete same assessments as Visit 8/ET (early termination). 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.
r. Safety follow-up visit to be performed for subjects who do not roll over into the optional open-label extension (OLEX) study; subjects who roll over into the OLEX study will complete the safety follow-up visit at the end of the OLEX study.

* At Pre-baseline visit (Visit 2a, Day -5 to Day -3), MMSE should be done first before all other assessments.
At Baseline/Day 1 (Visit 2b), ADAS-Cog$_{11}$, COWAT, and ADCS-CGIC 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, with ADAS-Cog$_{11}$ and COWAT assessments performed first at approximately 1 hour (± 30 minutes) post-dose. ADCS-CGIC assessments, when applicable, will be organized at adjacent times shortly after the individual EEG assessments.

2 Introduction

ATH-1017 is an experimental Alzheimer's disease (AD) treatment. ATH-1017 is a prodrug, which is rapidly converted to the active metabolite ATH-1001 in the plasma after SC injection. ATH-1017 was developed as a water-soluble prodrug of ATH-1001 to allow SC dosing in aqueous vehicles. The active metabolite ATH-1001 acts as a positive modulator at the hepatocyte growth factor (HGF) and its tyrosine kinase, MET, receptor system. Central nervous system (CNS) MET expression is crucial in maintaining the healthy adult brain (Hawrylycz, 2015), and is reduced in AD particularly in the hippocampus and frontal cortex (Hamasaki, 2014). The HGF/MET system presents a new therapeutic target to treat neurodegeneration and restore cognitive function in AD and other neurodegenerative disorders.

3 Trial Overview

3.1 Design

This is a Phase 2 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 subjects with a clinical diagnosis of mild to moderate AD, diagnosed on a 'probable' level according to McKhann, 2011.

The study will be conducted at a total of 6 centers in Australia and 8 centers in the US. Subjects 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; all eligible subjects will be tested for ApoE genotype.

Subjects who meet all inclusion/exclusion criteria will undergo baseline EEG assessments (ERP P300 and [redacted]) at 2 separate baseline visits. At the first baseline visit (Visit 2a, Pre-baseline, Day -5 to Day -3), EEG assessments (ERP P300 and [redacted]) will be performed at 2 separate timepoints approximately 2 hours apart (no dosing). EEG data should be uploaded for quality check immediately after the completion of the Pre-baseline visit (Visit 2a). At the second baseline visit (Visit 2b, Baseline, Day 1), no more than 6 days after the Pre-baseline visit, subjects will be randomized in a ratio of 1:1:1 to 3 parallel arms, either to active treatment (ATH-1017 40 mg/day or ATH-1017 70 mg/day) or placebo. During randomization, subjects will be stratified by baseline MMSE severity: mild (MMSE: 20-24) versus moderate (MMSE: 14-19). At this Baseline visit (Visit 2b), subjects will undergo pre-dose baseline and post-dose EEG assessments (ERP P300 and [redacted]).

Study drugs will be administered by SC injection once daily (OD) preferably during daytime. It is not recommended to take more than one dose within 8 hours. At the baseline visit, the first SC injection of study drug will be performed at site under supervision. The subject should withhold study drug administration on the day of subsequent clinic visits; study drug administration will be done on site under supervision of site staff at these visits.

Each subject is required to have a primary caregiver willing to accept responsibility for supervising or, if required, administering study drug, and assessing the condition of the subject throughout the study in accordance with all protocol requirements. During the double-blind treatment period, clinic visits will take place on Day 1 and thereafter at Weeks 2, 6, 12, 16, 20, and 26, with a safety follow-up visit scheduled 4 weeks after completion of the double-blind period at Week 30 (see Table 1 for schedule of assessments).

Subjects will undergo EEG assessments (ERP P300 and [redacted]) at each post-baseline clinic visit (pre- and post-dose timepoints) through Week 26, plus the safety follow-up visit at Week 30 (see Table 1 for timing of assessments). On Day 1, after completion of the first dose, subjects will remain onsite 2 hours for post-treatment safety observation. As marked circadian fluctuations of cognitive performance have been observed in AD (Hilt, 2015), ADAS-Cog_{11} and COWAT assessments shall occur at clinic visits in the morning at approximately the same time they were performed during the initial Baseline assessment.

Similarly, ADCS-CGIC assessments will be organized at adjacent times to the individual EEG assessment times. Subjects may live at home, in a senior residential setting, or an institutional setting without the need for continuous nursing care, and should not be likely to experience a change in living conditions (e.g., institutionalization, moving to a different city, etc.), or change in primary caregiver, during participation in the trial period. The end of the study is defined as the date of the safety follow-up visit, Visit 9/Week 30. Subjects who terminate prior to Visit 8 are to complete same assessments as Visit 8/early termination (ET).

An independent Data Safety Monitoring Board (DSMB) will conduct periodic review and assessments of unblinded safety data (AEs, labs, ECG, etc.) throughout the study to ensure the safety of study subjects.

Blood draws will take place at scheduled clinic visits (Day 1, Week 12 and Week 26) for analysis of plasma concentrations of ATH-1017 and ATH-1001 (see Table 1 for schedule of assessments).

An open label extension will be offered at participating sites.

3.2 Study Objectives

3.2.1 Primary Objectives

1. To evaluate the effects of ATH-1017 on event-related potential (ERP) P300 latency; change from baseline to post-baseline measurements compared to placebo.
2. To determine the safety and tolerability of ATH-1017.

3.2.2 Key Secondary Objectives

1. To evaluate the correlation of ERP P300 latency and cognition measured by Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog_{11}) and/or executive memory function measured by Controlled Oral Word Association test (COWAT); changes from baseline to post-baseline measurements comparing ATH-1017 to placebo.

3.2.3 Other Secondary Objectives

1. To evaluate the clinical efficacy of ATH-1017; Global Statistical Tests (GSTs) (O'Brien, 1984) that combine the scores from various clinical assessments into a single measure of overall symptomology. The various combinations of clinical outcomes combined for the GSTs are described in Section 7.

2. To evaluate the efficacy of ATH-1017 compared to placebo based on changes from baseline in:
    a) clinical efficacy, measured separately by the ADAS-Cog_{11} (cognition), ADCS-CGIC (global impression of change), and COWAT (executive memory function) questionnaires
    b) Functional ability, measured by ADCS-ADL23

[redacted]

3. To further evaluate the effects of ATH-1017 on ERP P300 latency; (i) evaluation of the acute effects of ATH-1017 on ERP P300 latency change from pre-dose to within-visit post-dose measurement,(ii) evaluation of the sustained effects of ATH-1017 on ERP P300 latency change from pre-dose baseline to pre-dose measurement across visits, (iii) evaluation of persistence of change in ERP P300 latency at Week 30 compared to placebo.

[redacted]

5. To determine the plasma PK profile of ATH-1017 and ATH-1001.

3.3 Sample Size Determination

A total sample size of 60 evaluable subjects (20 per treatment arm) is based on the results of the Phase 1 study, NDX-1017-0101, which demonstrated significant effects of ATH-1017 on ERP P300 in n=7 ATH-1017-treated AD subjects versus n=4 placebo-treated AD subjects, with 8 days of daily SC injection.

3.4 Blinding and Unblinding

The clinical study will be performed in a double-blind manner.

The study blind should not be broken except in a medical emergency (where knowledge of the IMP administered would affect the treatment of the emergency). The decision to break the blind will be made on a case-by-case basis, at the discretion of the Principal Investigator in collaboration with the Sponsor and Medical Monitor. The applicable standard operating procedures (SOPs) at the applicable vendors will be followed for blind breaking procedures.

After database lock, the overall randomization code will be broken only for reporting purposes.

4 Patient Chronology

Study days will be numbered from Day 1, the first day of IMP administration. The duration of events, such as the duration of treatment, will be taken as the end date minus the start date (Day 1), plus 1.

5 General Conditions for the Final Analysis

5.1 Definitions

Safety endpoints representing the change from baseline at a particular visit will be defined as the result at the visit minus the last non-missing result prior to the first dose of study medication. Should a baseline value be missing, then the corresponding screening value, if available, may be used.

Efficacy endpoints representing the change from baseline at a particular visit will be defined as the result at the visit minus the baseline value. Baseline for the primary analysis is defined as the average of all non-missing ERP P300 values measured prior to the first dose. Unless otherwise specified, similar definitions for baseline will be used for other efficacy analyses. Change from baseline will be missing for data collected prior to the first dose of study medication. Patients with a missing score for either baseline or the applicable visit will have a missing value for the change from baseline at that visit.

Whether an improvement is represented by a positive or negative change from baseline will depend on the endpoint.

5.2 General Considerations for the Final Analysis

All analyses will be carried out in SAS version 9.3 and R version 3.4 or higher.

Visits will be designated as Visit 1 (Screening), Visit 2a (Pre-baseline), Visit 2b (Baseline/Week 1 (Day 1)), Visit 3 (Week 2), Visit 4 (Week 6), Visit 5 (Week 12), visit 6 (Week 16), Visit 7 (Week 20), Visit 8/ET (Week 26), and Visit 9 (Safety Follow-up; Week 30) as defined in Table 1.

For all efficacy analyses, unless otherwise specified, active treatment refers to pooled treatment arms (40mg arm and 70mg arm). After this pooled comparison, the 70 mg versus placebo comparison (first) and the 40 mg versus placebo comparison (second) will be performed. A one-sided alpha of 0.05 will be used for all comparisons.

Regarding ERP P300 endpoint, 18 electrodes in the 10-20 system are collected for this endpoint, but not all of them will be used in the analysis. Because the design of the auditory task requires reference and ground electrodes to be in the front and back of the head, respectively, the average of the central electrodes available, CZ, FC1, FC2, CP1, and CP2, known as the region of interest (ROI) offers the best signal to noise ratio for identifying a potential ERP P300 effect and will be used for this endpoint analysis.

Summaries will be shown with columns for each treatment group as well as a column for all patients in the study and applicable population. The number of subjects in the treatment group and population will be shown in the column header. Continuous data will be summarized with the number of non-missing values and their mean, standard deviation, median, minimum and

maximum. Categorical data will be summarized as the count and percentage of distinct patients with each value. Unless noted otherwise and for a given analysis population, the denominator will be the number of subjects with non-missing assessments in this analysis population who belong to the treatment group. Summaries by visit will include both unscheduled and regularly scheduled visits.

All data collected in the clinical database will be listed. Listings will include all subjects with analysis population flags included to designate which population each subject is included with. Listings will generally be sorted by subject ID, parameter and visit date/study day. Unscheduled visits will be included in the listings.

If it happens that no patients qualify for a display, the display will be produced with a note that "No patients qualify for the display." For example, if there are no deaths then the applicable displays will be so noted.

For safety summaries, the last pre-dose measurement is defined as the baseline value. For MMSE, the mean of the pre-dose measurements (obtained at Screening and Pre-baseline) will be used as the baseline value . ADCS-CGIC is a change score, so no further calculation is required. For all other efficacy masures, baseline is the average of all measurements prior to first dosing, or if only one measurement exists that value will be used.

Visit windowing will be applied for analyses which use visit categories. For categorical visit summaries, all visits including early termination assessments and unscheduled visits will be included with the closest scheduled post-baseline visit that includes the efficacy or safety assessment, based on number of days since Day 1. For the primary variable of ERP P300 latency effect, if multiple assessments fall within the same analysis window, pre- and post-dose measurement will be averaged within the analysis window for this variable. For all other efficacy assessments, if multiple efficacy assessments fall within the same analysis window (other than baseline), any non-missing efficacy assessments will be averaged. For all non-efficacy assessments, a last-within-window approach will be used.

Assessments will be considered on-treatment after first dose of study drug and up to the last day of dosing +35 days for all safety assessments. For ERP and efficacy assessments, on treatment will include all assessments after the first dose of study drug and up to the last day of dosing +3 days.

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.

Days will be converted to weeks by dividing by seven. Days will be converted to months by dividing by 30.417. Days will be converted to years by dividing by 365.25. All data collected during the study will be analyzed and reported unless stated otherwise.

5.3 Visit Windowing

Analysis Windows for ADAS-Cog_{11} and COWAT

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

Analysis Windows for EEG (ERP P300 and [redacted]), CSSRS, Physical and Neurological Exams, EG, Vital Signs, Labs

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

Analysis Windows for GDS, Height, Weight

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

Analysis Windows for ADCS-CGIC, ADCS-ADL23, [redacted], PK

Analysis Visit	**Protocol-Specified Day**	**Analysis-Specified Interval**
Week 12	Day 84	Day 2-133

Week 26	Day 182	Day 134-217

Analysis Windows for MMSE

Analysis Visit	**Protocol-Specified Day**	**Analysis-Specified Interval**
Week 20	Day 140	Day 2-161
Week 26	Day 182	Day 162-217

5.4 Procedures for Handling Missing Data

Patients who drop out will have all available post-baseline data included in the analysis, unless otherwise specified. The mixed model is based on an assumption of Missing at Random (MAR) and is designed to handle right censored data for subjects who drop out of the study.

5.5 Interim Analysis

No interim analysis of the data is planned.

5.6 Changes to Statistical Analysis

The statistical analysis plan clarifies that comparison between treatment and placebo will be performed first by evaluating differences between pooled active treatment (40 mg and 70 mg ATH-1017 arms combined) followed by comparisons of individual treatment groups (40 mg ATH-1017 versus placebo and 70 mg ATH-1017 versus placebo).

The primary esstimand and analysis were amended to reflect use of a treatment treatment contrast that averages the difference between treatments in mean CFB across Weeks 12, 16, 20, and 26, whereas the original, protocol-specified approach was the difference at Week 26.

5.7 Analysis Populations/Sets

5.7.1 Modified Intent-to-Treat (mITT) Population

The modified intent-to-treat (mITT) population will include all randomized subjects who took at least one dose of the study medication and who completed at least one ERP P300 baseline assessment and one post-baseline ERP P300. Subjects will be analyzed according to the dose they were randomized to.

5.7.2 Per Protocol Population

The per protocol population will include all mITT subjects who took the assigned medication during the 26 weeks of treatment, completed at least one ERP P300 plus ADAS-Cog_{11} or COWAT post-baseline assessment, and did not have any major protocol deviations. Subjects will be analyzed based on actual treatment received.

5.7.3 Safety Population

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

5.8 *Primary Estimand*

The primary estimand for this study is the mean difference between drug and placebo based on post-baseline changes in ERP P300 latency using a hypothetical strategy to assess the results expected if all patients had completed the treatment period.

The rationale for this as the primary estimand is based on the aim of this study to evaluate a measure suggestive of treatment benefit in support of establishing proof of concept. Specific details of this estimand are:

Treatment regimen: ATH-1017 40 mg/day and ATH-1017 70 mg/day pooled compared versus placebo.

Population: Patients in the mITT analysis set (see Section 5.7.1) with mild to moderate AD as specifically definded by the protocol inclusion / exclusion criteria.

Variable: ERP P300 latency as defined in Section 7.1.

Population summary measure: The LS mean difference in CFB between pooled ATH-1017 treatment arms and placebo averaged across Weeks 12, 16, 20, and 26 (see Section 8.1).

Handling of intercurrent events. The intercurrent events (ICEs) of missed or modified doses, drug discontinuation and concurrent treatment will be handled using a treatment policy strategy which ignores these ICEs. Deaths are not expected during the 26-week double-blind phase follow-up in the enrolled subject population. A hypothetical strategy will be used to handle the ICE of early study discontinuation in order to estimate the effect anticipated if all patients had completed treatment.

5.9 Data Review

Classification of deviations from the protocol as minor or major, and decisions regarding patient population assignments will be decided on a case-by-case basis without knowledge of the treatment assigned and before the database lock in a blinded data review meeting.

After database lock, the responsible statistician will request the treatment codes, the study will be unblinded, and the statistical analysis will be conducted.

5.10 Missing Data

Patients that lack information at a visit due to early termination, etc. will not be imputed for the primary analysis. In each of the following listings, patients missing the applicable data will have the following entry listed for that data point "No Data": Analysis Population Exclusions (Listing 16.1.4), General Comments (Listing 16.1.7), Serious AEs (Listing 16.3.1.2), AEs Leading to Discontinuation or Death (Listing 16.3.2), and Clinically Significant Laboratory Abnormalities (Listing 16.3.4).

In the event that an ADAS-Cog_{11} total score is partially collected, but missing some individual items, a last z-score carried forward (LZCF) or straight line imputation will be performed at the item level to fill in missing values prior to calculating total scores.

6 Study Conduct

6.1 Treatment Assignments

Upon locking the database the study will be unblinded. The treatment assignment will be given to each patient and 100% verified.

6.2 Enrollment, Demographics, and Baseline Characteristics

A summary table that includes all patients will be provided to show the study timelines: the earliest and latest date of screening visit, of study drug intake, and study participation as well as the duration in days of screening, treatment and study participation.

A summary table will be provided for the Safety Population to summarize demographics, i.e. age, sex, race, height and weight at baseline. The table will be repeated for the mITT and PP Populations. The equality between treatment groups for each of these parameters will be tested with a t-test or Fisher's exact test, as appropriate.

The supportive listings will present all data on the Informed Consent / Demographics CRF page. The listings of disposition, IMP dispensing and return, and vital signs also partly support these tables.

6.3 Patient Disposition and Protocol Deviations

A summary table will be provided for all patients to display the number of subjects randomized, Safety Population, mITT 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 displayed in the eCRF.

A summary table will be provided for all randomized patients to include those with one or more major protocol deviations. Major deviations as reported on corresponding eCRF will be classifed by category and summarized; within each classification the count and percent of each category that appears on the CRF will be displayed.

Supportive listings will present all information on the Study Completion / Early Termination eCRF page and the Protocol Deviations page. Another supportive listing will display patients excluded from any analysis population along with the reason(s) for exclusion.

6.4 Extent of Exposure to Study Drug

A summary table will be provided for the Safety Population stratified by treatment group to summarize the duration of treatment, total dose received, average dose per week and compliance. The count and percentage of subjects with compliance below 80% and above 110% will be provided as an assessment of under/overdosing.

The supportive listing will present all data on the Drug Accountability CRF page.

6.5 Medical History / Concomitant Medication

Concomitant medications will be coded according to WHODDrug (WHO Drug dictionary). Medical history findings will be coded with the Medical Dictionary for Regulatory Affairs (MedDRA). The version number will be shown as a footnote in the

displays. A listing will be provided of these data and will include all patients in the study. A summary table will be provided to show the count and percentage of patients with any medical history by system organ class and preferred term. Patients will be counted at most once per term.

6.6 General Comments

A listing will be provided of all general comments recorded on the corresponding CRF page.

7 Efficacy Endpoints

7.1 Primary Variable

The primary variable is change from baseline in ERP P300 latency. Each post-baseline assessment of ERP P300 latency will be calculated as the average across the pre-dose and post-dose values at each respective visit. The primary estimand and analysis are based on combining results across Weeks 12, 16, 20, and 26 via a contrast statement that in essence averages the LS mean treatment group differences across these assessment times (see Section 8.1).

ERP P300 is a method of recording brain activity elicited by external stimuli, e.g., an oddball auditory stimulus, and is a well-established functional biomarker, particularly of working memory access (Ally, 2006). ERP P300 is characterized by a stereotyped series of voltage deflections occurring after the respective odd tone to be counted, with early features (< 100 msec) corresponding to unconscious sensory transmission (auditory cortex, N100), and later features produced by cognitive processing in the ventral attentional network, i.e., P300, referring to the large positive deflection at roughly 300 msec in healthy adults (young or elderly). The P300 latency is sensitive to detecting reduced synaptic transmission related to cognitive decline in AD patients and other dementias (Olichney, 2011).

To assess the P300 component (latency and amplitude), the subject has to perform a task related to auditory stimuli. The stimulus consists of an oddball paradigm with 2 sound stimuli. Stimuli are presented through headphones and auditory stimulation for P300 will be assessed in a recording lasting up to 10 minutes.

7.2 Secondary Efficacy Endpoints

7.2.1 Global Statistical Tests

Global statistical tests (GST) facilitate assessment of a overall change in disease status/trajectory into a single measure by standardizing and then combining individual clinical and/or functional measures. Several combinations of clinical measures will be used to create multiple GSTs. Each GST is a specific combination of outcomes that are measured at then standardized at each visit. Standardization is based on using the mean score from baseline (across all arms) as the mean value and standard deviation is calculated on all residual values (across all arms) per visit, where the residuals are calculated as the distance between inidivual outcomes and the mean outcome within that treatment arm. The GST will be determined for each subject by calculating the *z*-score on the individual level for each of the component endpoints and taking the mean *z*-score across GST component measures for each individual at each time point. This individual patient mean z-scores will be the efficacy outcome variable used in the analyses.

GST scores will be computed using the following combinations (See definitions of GST components below):

Protocol-specified GST

ADAS-Cog$_{11}$ and COWAT

ADAS-Cog$_{11}$ and ADCS-CGIC

Additional exploratory GSTs

ADCS-ADL23 and COWAT

ADAS-Cog$_{11}$ and ADCS-ADL23

ADAS-Cog$_{11}$ and ERP P300 (no correlation analysis)

ADAS-Cog$_{11}$, ERP P300, and ADCS-CGIC (no correlation analysis)

For ex-US purposes, the activities of daily living endpoint, comparison of active to placebo for the first GST may include the ADCS-ADL23 instead of ADCS-CGIC.

7.2.2 Alzheimer's Disease Assessment Scale – Cognitive Subscale (ADAS-Cog$_{11}$)

The ADAS-Cog$_{11}$ is designed to measure cognitive symptom change in subjects 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.

Due to known circadian fluctuations of cognitive capacity (Hilt, 2015), ADAS-Cog$_{11}$ will be assessed in the morning at approximately the same time of day as the baseline assessment for all applicable visits.

ADAS-Cog$_{11}$ assessments will be performed pre-dose at Visit 2 (Baseline/Day 1), and post-dose at approximately 1 hour (± 30 minutes) at Visit 3 (Week 2), Visit 4 (Week 6), Visit 5 (Week 12), Visit 7 (Week 20), Visit 8/ET (Week 26), and Visit 9 (Safety follow-up; no dosing).

7.2.3 Alzheimer's Disease Cooperative Study Clinical Global Impression of Change (ADCS-CGIC)

The Alzheimer's Disease Cooperative Study Clinical Global Impression of Change (ADCS-CGIC) scale is a 7-point scale that requires the clinician to assess how much the subject's illness has improved or worsened relative to a baseline state at the beginning of the intervention and rated as: 1, markedly improved; 2, moderately improved; 3, minimally improved; 4, no change; 5, minimally worse; 6, moderately worse; or 7, markedly worse. The ADCS-CGIC consists of 3 parts: a guided baseline interview administered to the subject and caregiver/support person, a follow-up interview administered to the subject and

caregiver/ support person, and a clinician's rating review (Schneider, 1997). At study start, using the ADCS-CGIC baseline form as a guideline, the ADCS-CGIC rater will obtain an integral clinical impression of the subject's status, can apply any formal testing at his/her discretion, and take personal notes regarding the subject's condition; these will serve as a reference for future change ratings. The ADCS-CGIC will be administered by an experienced clinician who will remain independent of the subject's safety, cognitive and functional outcomes, and will be trained and certified for this study. The ADCS-CGIC rater will ideally remain the same individual for all ADCS-CGIC ratings.

ADCS-CGIC assessments will be performed adjacent to ADAS-Cog_{11} and COWAT assessments pre-dose at Visit 2 (Baseline/Day 1), post-dose at Visit 5 (Week 12), and post-dose at Visit 8/ET (Week 26).

7.2.4 Controlled Oral Word Association Test (COWAT)

The Controlled Oral Word Association Test (COWAT) is an oral verbal fluency test in which the subject is required to make verbal associations to different letters of the alphabet by saying all the words which they can think of beginning with a given letter. Individuals are given 1 minute to name as many words as possible beginning with each of the letters. The procedure is then repeated for the remaining two letters (Benton, 1994; Strauss, 2006). The test score is the total number of different words produced for all 3 letters.

The COWAT will be performed adjacent to the ADAS-Cog_{11} assessment, i.e., pre-dose at Visit 2 (Baseline/Day 1), and post-dose at approximately 1 hour (± 30 minutes) at Visit 3 (Week 2), Visit 4 (Week 6), Visit 5 (Week 12), Visit 7 (Week 20), Visit 8/ET (Week 26), and Visit 9 (Safety follow-up; no dosing).

7.2.5 Alzheimer's Disease Cooperative Study – Activities of Daily Living, 23-item Version (ADCS-ADL23)

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 subject'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 to 78, with lower scores indicating greater functional impairment. ADCS-ADL23 assessments will be performed pre-dose at Visit 2 (Baseline/Day 1), post-dose at Visit 5 (Week 12), and post-dose at Visit 8/ET (Week 26).

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

7.2.10 Pharmacokinetic Variables

Blood will be collected for analysis of ATH-1017 and ATH-1001 in plasma at the timepoints detailed in Table 1. The actual time of blood sampling will be recorded.

Blood sample collection, processing, and shipping details will be outlined in a separate laboratory manual. In brief, blood will be processed and plasma analyzed by a validated method for concentrations of ATH-1017 and ATH-1001. Concentrations of ATH-1017 and ATH-1001 will be listed by treatment, subject, day, and time of sample. Plasma concentration summary statistics will be presented.

8 Statistical Analyses

Efficacy analyses will be performed in the mITT and PP populations, based on pooling of active doses and with by individual dose.

8.1 Primary Analysis

The primary analysis will use a likelihood-based mixed model for repeated measures (MMRM) to compare the estimated changes across post-baseline visits (CFB) between pooled active treatment and placebo in ERP P300 latency. These analyses will assess the magnitude and statistical significance of mean CFB between placebo and the pooled active treatment groups. All post baseline visits will be included in the model. Because the difference between treatments in ERP P300 latency is not expected to vary significantly over time, the primary analysis will be based on the contrast between LS means averaged across Weeks 12, 16, 20, and 26.

Baseline for the primary analysis is defined as the average of all non-missing ERP P300 values measured prior to the first dose.

An ERP risk score will be calculated and used as a covariate in the model, combining known contributors to AD severity. The ERP risk score will be calculated on a per-person basis and is equal to the predicted value from OLS regression of the average of all post-baseline ERP P300 latenecy measurements as the dependent variable. The following covariates are used as explanatory variables: age, sex, screening (stratifying) MMSE score, AChEI treatment status, APOE4 genotype (E4/E4 versus non-E4/E4), baseline ADAS-Cog_{11} total score, baseline ADCS-ADL23 total score, baseline NPI total score, and baseline COWAT total score.

The MMRM analysis of the primary outcome will have CFB in ERP P300 latency as the response variable and will include the following covariates and fixed effects:

- Baseline Test Score (covariate)
- ERP Risk Score (covariate)
- Time (aka visit, categorical)
- Site (random effect)
- Subject (random effect)
- Baseline Test Score by time interaction

Visitwise LS means, the main effect LS mean, and contrasts between LS means are interpreted as the expected mean difference between treatment groups (pooled/separate active doses versus placebo) in CFB. In addition to treatment group differences, output will include p-values, 95% confidence intervals for the difference, effect size, 95% confidence interval for the effect size, and an effect size based upon Cohen's D.

Effect size will be calculated by taking the difference between LS means and dividing by the standard deviation (i.e. the standard error of the estimated difference multiplied by the squared degrees of freedom). The equations below show how effect size and Cohen's D

effect size will be calculated. In the following formulas *p* stands for placebo and *t* stands for treatment, and *SE* for standard error:

$$Effect\ Size = \frac{LSMEAN_P - LSMEAN_t}{LSMEAN_P}$$

For assessments where a higher score indicates worse performance (i.e. ADAS-Cog$_{11}$) Cohen's d will be calculated using the following equation:

$$Cohen's\ D = \frac{LSMEAN_P - LSMEAN_t}{(pooled\ SD)}$$

and for assessments where a higher score indicates better performance (i.e. COWAT) Cohen's d will be calculated using the following equation:

$Cohen's\ D = \frac{LSMEAN_t - LSMEAN_P}{(pooled\ SD)}$ where the pooled standard deviation (pooled SD) is defined as follows:

$$Pooled\ SD = \sqrt{\frac{(n_t - 1)\left(SE_{t\sqrt{n_t}}\right)^2 + (n_p - 1)\left(SE_{p\sqrt{n_p}}\right)^2}{(n_t + n_p - 1)}}.$$

The number of subjects with an observed efficacy outcome, mean, standard deviation, median, 25th percentile (Q1), 75th percentile (Q3), minimum and maximum will all be reported and accompany the estimates from the MMRM outlined in this section. A compound symmetry (CS) covariance structure will be used for this analysis.

8.2 Key Secondary Analysis

The correlation (association) between CFB in ERP P300 with the GST that includes ADAS-Cog$_{11}$ and ADCS-COWAT will be established using an MMRM analysis. In these analyses CFB in GST will be the response variable of interest and CFB to each post-baseline visit in ERP P300 latency will be the primary independent (explanatory) variables of interest, with other covariates fit to isolate the association between ERP P300 latency and GST. Treatment will not be fit in the model. A clinical risk score will be created using the same method described for the primary analysis in section 8.1, but using the average of all post-baseline ADAS-Cog$_{11}$ measurements as the dependent variable instead of ERP P300 Latency.

The covariates for the MMRM analysis will include:

- Baseline value ERP P300 (covariate)
- Clinical Risk Score (covariate)
- Time (aka visit, categorical)
- Site (random effect)
- Subject (random effect)
- Baseline value ERP P300 by time interaction

The covariance structure for the repeated measures in this model will be unstructured (UN). If UN does not converge for the model, the MMRM model will be simplified to allow convergence as described in the following paragraph. Variance components will be used as the covariance structure for the random site effect in the model.

If the MMRM does not converge, using the model described above, the analysis will be rerun using a first-order heterogeneous autoregressive (ARH[1]) covariance structure, and then a compound symmetric (CS) structure. If convergence is still not achieved then the model will be simplified to exclude the baseline by time interaction and rerun first with the UN and then with ARH[1], followed by CS. If convergence is still not achieved, the following terms will be excluded from the model: age, first with the UN and then ARH[1], followed by CS. If convergence is still not achieved, then these models will have all covariates removed except center/site and visitwise CFBs in ERP P300 latency.

A partial Pearson correlation that controls for treatment will also be calculated by including data from all treatment arms but will account for variance due to treatment (e.g., using the partial statement in SAS PROC CORR).

8.3 Other Secondary Analyses

8.3.1 Global statistical tests

Each of the GSTs defined in Section 7.2.1 will be analyzed as follows:

The MMRM analysis will include the respective CFB to each post-baseline visit in GSTs as the repeatedly measured response variable, with treatment as an explanatory variable along with other covariates, fixed, and random effects, including:

- Baseline Test Scores of the Efficacy Parameter (covariate)
- AD Risk Score (covariate)
- Time (aka visit, categorical)
- Site (random effect)
- Subject (random effect)
- Time by Baseline Test Score

The same covariance structure process will be used as described for the primary analysis. A Clinical risk score will be created using the same method described for the secondary analysis in section 8.2. For GSTs that involve the CGIC, which is inherently a measure of change and hence no baseline value exists, the baseline value to be used as a covariate in that analysis will be the baseline values of the other component(s) of that GST. When more than one component of a GST exists besides the CGIC, the baseline covariate will be a combined

statistical test of those components defined as the mean of baseline z-scores for those components, where each z-score is defined based on the mean and standard deviation for that outcome at baseline irrespective of treatment group.

8.3.2 Individual clinical, functional, health economic, and biomarker assessments

The ADCS-ADL23, ADAS- Cog_{11}, ADCS-CGIC, [redacted], COWAT, [redacted] will be analyzed using the MMRM analysis as described for the primary analysis, with the exception that for the ADCS-CGIC, which is inherently a measure of change, the baseline ADAS-Cog_{11} score will be used as the baseline covariate. These analyses will use the clinical risk score in place of the ERP risk score as a covariate. Output from these analyses will focus on visitwise contrast at Week 26 (the primary ERP P300 latency analysis was based on a contrast combining results across post-baseline visits).

[redacted]

An additional analysis will be performed on the change in ERP P300 latency from pre-dose to within-visit post-dose measurements averaged across all post-dose measurements within post-baseline analysis windows to estimate the acute effect estimate. In addition, change in ERP P300 latency from pre-dose baseline to pre-dose measurements averaged across all pre-dose measurements within post-baseline analysis windows across visits will be analyzed to estimate the sustained effect.

8.3.3 Subgroup analyses

The following subgroups will be assessed for ERP P300 latency:

- AChEI (subjects taking anti-cholinesterase inhibitors / not taking anti-cholinesterase inhibitors);
- MMSE stratification variable (mild/moderate).
- APoE status (carrier/non-carrier)

Subgroup analyses will be implemented by using the MMRM model as described for the primary analysis with the addition of the treatment-by-subgroup and treatment-by-visit-by subgroup interaction terms.

8.3.4 Additional Analyses

The persistence of the treatment effect on ERP P300 latency will be evaluated by repeating the primary analysis with the addition of the Week 30 (off treatment) data.

Simple correlations will also be evaluated between visitwise post-dose ERP P300 latency and the following outcomes: ADAS-Cog_{11}, ADCS-ADL23, COWAT, GST, and ADCS-CGIC. These correlation will be assessed using partial Pearson correlations that control for treatment. That is, the analyses will include data from all treatment arms but will account for variance due to treatment (e.g., using the partial statement in SAS PROC CORR).

[redacted]

8.3.5 PK Analyses

PK will be analyzed using the mITT population. Descriptive statistics will be provided for the each treatment group. Additionally, correlations of exposure values to clinical outcomes may be assessed, if possible, by correlation to concentration data as well as summarization of outcomes data in the upper and lower PK tertiles.

The PK concentration data collected prior to each dose will be used in a concentration-response relationship analysis to assess correlation with each outcome which will mirror the primary analysis but replace the treatment variable with PK concentration. Analysis will be performed separately for each PK collection time as well as using the highest of the PK timepoints.

9 Safety

9.1 Safety Endpoints

Primary endpoint

- incidence of treatment-emergent adverse events

Additional endpoints

- Incidence and severity of treatment-emergent adverse events (AEs)
- Clinical laboratory tests
- Vital signs
- Physical/neurological examinations
- ECGs
- Use of concomitant medications for treatment of AEs
- C-SSRS
- GDS

9.2 Analysis Populations Evaluated for Safety

All analyses of safety will be performed on the Safety Population.

9.3 Statistical Methods for Safety Endpoints

A summary table will be provided for the Safety Population of the count and percentage of subjects with any safety endpoint. No hypothesis testing will be performed for safety variables.

Adverse Events

AEs reported on CRFs will be coded into system organ classes and preferred terms using the Medical Dictionary for Regulatory Activities (MedDRA v23.1). A treatment-emergent adverse event (TEAE) is defined as an AE with an onset date on or after the start of dosing. The adverse event summary will include only TEAEs. Any AEs that are not considered treatment-emergent will be provided in data listings only.

The incidence of AEs will be summarized for the safety population. Although a preferred term or system organ class may be reported more than once for a subject, each subject will only be counted once in the incidence count for each category. If a subject has the same AE on multiple occasions, the highest severity (severe > moderate > mild) or drug relationship (definite > probable > possible > unlikely > not related) recorded for the event will be presented.

Severity levels include: mild, moderate and severe. Relationships will be grouped into two categories for analysis: related and unrelated. Not related and unlikely will be categorized as “unrelated.” Possible, probable and definite will be categorized as “related.” If severity or drug relationship is missing no data imputation will be performed and no category of missing will be presented.

Summary tables showing the number of subjects and percent within each category will be generated for each of the following types of adverse events:

- All AEs;
- Serious Adverse Events;
- Fatal Adverse Events;
- AEs by causality
- AEs by severity
- AEs for Subjects who Died.

These summaries will present the number and percentage of subjects reporting an adverse event for each classification level. The denominators for calculating the percentages overall will be based on the number of subjects in the safety population. The denominators for calculating the percentages by treatment will be based on the number of subjects exposed to each treatment in the safety population. In addition to these summaries, all AEs will be summarized by action taken, seriousness, severity, and relationship to study drug.

All AEs that occurred in 5% or more of all subjects (active and placebo) will be tabulated for the safety population. These results will be analyzed descriptively and their incidence rate will be summarized.

All SAEs, AEs, AEs leading to premature discontinuation from the study, and AEs with fatal outcome will also be provided in data listings by subject, treatment group, verbatim term and preferred term.

9.2 Vital Signs

Each vital sign will be summarized by treatment and by visit, using descriptive statistics (mean, median, SD, minimum, maximum, and number of subjects) 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 subject, treatment group, visit, and parameter.

9.3 Electrocardiogram

ECG values and change from baseline values will be summarized by visit using descriptive statistics for PR interval, QT interval, QTcB interval, QTcF interval, RR interval, mean heart rate, and QRS duration. ECG abnormalities will be summarized as the count and percentage of subjects 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

subjects 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 and their incidence rate and two-sided 95% confidence intervals will be summarized.

9.4 Clinical Laboratory Evaluations

Continuous blood clinical laboratory analytes absolute values and change from baseline values will be summarized by analyte and visit using descriptive statistics (mean, median, SD, minimum, maximum, and number of subjects). Mean line plots over time will be displayed for each analyte with separate lines for each treatment. Categorical laboratory analytes, classified as normal or abnormal, will be summarized by analyte and visit using the number and percentage of subjects in each category and in each treatment group?. The denominators for calculating the percentages will be based on the number of subjects 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 subjects 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 subjects with non-missing assessments for a particular analyte.

Clinical laboratory results will be provided in data listings by subject, visit and analyte. Abnormal lab results (either below LLN or above ULN) will be provided in a separate listing by subject, center, analyte, treatment group and visit.

9.5 Physical and Neurological Exams

Physical and neurological examination findings will be summarized as the count and percentage of subjects in each treatment group.

9.6 C-SSRS

The C-SSRS responses will be tabulated by visit, treatment group, question and response. All C-SSRS responses will also be provided in a data listing.

9.7 Geriatric Depression Scale (GDS)

The GDS is a self-report measure of depression in older adults with a “Yes/No” response format. The GDS was originally developed as a 30-item instrument. It has since been validated in a shortened form comprising 15 items (Sheikh, 1986). The total score range is 0 to 15, with a higher score indicating more severity. A GDS score of ≤ 7 is required at Screening. In discussion with the Medical Monitor, subjects with a GDS score between 8 and 10 inclusive can be considered for study participation if the increased score is driven by specific domains related to the pandemic and its restrictions, rather than by major depression.

GDS assessments will be performed at Screening, Baseline/Day 1, Visit 5 (Week 12), Visit 8/ET (Week 26), and Visit 9 (Safety follow-up). A listing will be provided, showing responses across all outcomes and timepoints.

9 OTHER LISTINGS

The following additional listings will be provided:

- Subjects excluded from the safety, mITT, and PP populations;
- Clinical laboratory results for hematology, blood chemistry and urinalysis;
- Abnormal laboratory results;
- Physical examination assessments;
- Neurological examination assessments;
- Concomitant medications;
- Dose administration dates and times.

•

10 Bibliography

Aisen, P.S., et al. (2008). High-dose B vitamin supplementation and cognitive decline in Alzheimer disease: a randomized controlled trial. JAMA 300(15), 1774-1783.

Aisen, P.S., et al. (2003). Effects of rofecoxib or naproxen vs placebo on Alzheimer disease progression: a randomized controlled trial. JAMA 289(21), 2819-2826.

Aricept® (donepezil hydrochloride) tablets prescribing information. Revised: July 2015. © 2012 Eisai Inc.

Benton, A. L., Hamsher, K., Rey, G. L., & Sivan, A. B. (1994). Multilingual Aphasia Examination (3rd ed.). Iowa City, IA: AJA Associates.

Citron, M. (2010). Alzheimer's disease: strategies for disease modification. Nature Reviews Drug Discovery 9, 387-398.

Cummings JL, Mega M, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J. The Neuropsychiatric Inventory: comprehensive assessment of psychopathology in dementia. Neurology 1994;44: 2308-2314.

Ebens, A., et al. (1996). Hepatocyte growth factor/scatter factor is an axonal chemoattractant and a neurotrophic factor for spinal motor neurons. Neuron 17:1157-1172.

EuroQol Group (1990). "EuroQol - a new facility for the measurement of health-related quality of life". Health Policy (Amsterdam, Netherlands). 16 (3):199–208.

FDA Guidance for Industry, Drug-Induced Liver Injury: Premarketing Clinical Evaluation, July 2009.

Folstein MF, Folstein SE, McHugh PR. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 12(3), 189-198.

Funakshi, H., and Nakamura, T. (2011). Hepatocyte Growth Factor (HGF): Neurotrophic functions and therapeutic implications for neuronal injury/diseases. Current Signal Transduction Therapy 6:156–167.

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.

Hamasaki, H., et al. (2014). Down‐regulation of MET in hippocampal neurons of Alzheimer's disease brains. Neuropathology 34, 284–290. PMID 24444253.

Hawrylycz, M., et al. (2015). Canonical genetic signatures of the adult human brain. Nature Neuroscience 18, 1832–1844. PMID 26571460.

Hyman, B.T., et al. (2012). National Institute on Aging – Alzheimer's Association guidelines on neuropathologic assessment of Alzheimer's disease. Alzheimer's & Dementia: The Journal

of the Alzheimer's Association 8, 1 – 13.

Hilt DC, Dgetluck N, Sainati S, Moebius HJ. Effect of circadian rhythm on cognition testing in a phase 2B mild/moderate Alzheimer's Disease (AD) study. J Prev Alzheimers Dis. 2015:2(4):P1-41.

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.

Kind P. The EuroQoL instrument: an index of health-related quality of life. In: Spilker B, ed. Quality of Life and Pharmacoeconomics in Clinical Trials, 2nd edn. Philadelphia, PA: Lippincott-Raven Publishers, 1996:191-201.

Maina, F., and Klein, R. (1999). Hepatocyte growth factor, a versatile signal for developing neurons. Nature Neuroscience 2(3):213-217.

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.

Molnarfi, N., Benkhoucha, M., Funakoshi, H., Nakamura, T., and Lalive, P.H. (2015). Hepatocyte growth factor: A regulator of inflammation and autoimmunity. Autoimmun Rev 14, 293–303.

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

Organ, S.L, Tsao, M-S. (2011). An overview of the c-MET signaling pathway. Ther Adv Med Oncol 3(1):S7-S19.

Piguet et al. (2017) Clinical Gene Therapy for Neurodegenerative Diseases: Past, Present, and Future. Human Gene Therapy 28(11):988-1003.

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

Sano, M., et al. (1997). A controlled trial of selegiline, alpha-tocopherol, or both as treatment for Alzheimer's disease. New England Journal of Medicine 336(17), 1216-1222.

Schneider LS, Olin JT, Doody RS, Clark CM, Morris JC, Resiberg B, et al. Validity and reliability of the Alzheimer's Disease Cooperative Study - Clinical Global Impression of Change. The Alzheimer's Disease Cooperative Study. Alzheimer Dis Assoc Disord. 1997;11(2):S22-S32.

Shang, J., et al. (2011). Strong neurogenesis, angiogenesis, synaptogenesis, and antifibrosis of hepatocyte growth factor in rats brain after transient middle cerebral artery occlusion. J. Neurosci. Res. 89, 86–95.

Sheikh JI, Yesavage JA. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. Clinical Gerontologist: The Journal of Aging and Mental Health, 5(1-2), 165–173.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). Compendium of neuropsychological tests: Administration, norms, and commentary (3rd ed., p. 502). Oxford: Oxford University Press.

Tyndall, S.J et al. (2007). Hepatocyte growth factor-induced enhancement of dendritic branching is blocked by inhibitors of N-methyl-D-aspartate receptors and calcium/calmodulin-dependent kinases. J. Neurosci. Res. 85, 2343–2351.

Wimo A, Wetterholm AL, Mastey V, Winblad B. Evaluation of the resource utilization and caregiver time in Anti-dementia drug trials – a quantitative battery. In: Wimo A, Karlsson G, Jönsson B, Winblad B (eds). The Health Economics of Dementia, 1998. Wiley's, London, UK.

Zarit, SH, Reever, KE, Back-Peterson, J. Relatives of the impaired elderly: correlates of feelings of burden. The Gerontologist. 1980;20:649-655.