Cover Page for ClinicalTrials.gov

Sponsor name:	Annovis Bio, Inc.
NCT number:	NCT05686044
Sponsor trial ID:	ANVS-22002
Official title of study:	A Randomized, Double-Blind, Placebo-Controlled, Dose-Ranging, Multicenter Study of Buntanetap in Participants with Mild to Moderate Alzheimer's Disease
Document type:	Statistical Analysis Plan
Document date:	February 27, 2024

Statistical Analysis Plan

for

A Randomized, Double-Blind, Placebo-Controlled, Dose-Ranging, Multicenter Study of Buntanetap in Participants with Mild to Moderate Alzheimer's Disease

Sponsor ANNOVIS BIO, INC., USA

Product/Compound ANVS-22002

Phase of the study II/III

US IND Number 72,654

	Name, Title, Affiliation	Signature & Date
Written by	Wenxin Lu, PhD Senior Statistician I, Statistics Pharmapace, Inc.	[redacted]
Approved by Sponsor Representative	Maria Luisa Maccecchini President and CEO ANNOVIS BIO, INC.	[redacted]

Contents

1 Abbreviations

αSYN	Alpha-synuclein
Aβ	Amyloid beta
AD	Alzheimer's disease
ADAS-Cog	Alzheimer's Disease Assessment Scale–Cognitive Subscale
ADCS-ADL	Alzheimer's Disease Cooperative Study–Activities of Daily Living Scale
ADCS-CGIC	Alzheimer's Disease Cooperative Study–Clinical Global Impression of Change
AE	Adverse Event
APP	Amyloid β Precursor Protein
AR(1)	First-order Autoregressive
ATC	Anatomical Therapeutic Chemical
AUC	Area under the curve
C_{max}	Maximum plasma and CSF concentration
CI	Confidence interval
CL	Clearance
CM	Concomitant medication
CSF	Cerebrospinal Fluid
CSP	Clinical Study Protocol
C-SSRS	Columbia Suicide Severity Rating Scale
CV	Coefficients of variation
DSMB	Data Safety Monitoring Board
DSST	Digital Symbol Substitution Test
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
ES	Enrolled set
FDA	Food and Drug Administration
GFAP	Glial Fibrillary Acidic Protein
IC	Informed Consent
IMP	Investigational Medicinal Product
iSAP	interim Statistical Analysis Plan
ITT	Intent-to-Treat set
LSM	Least Square Mean
MAR	Missing at Random
MCMC	Markov Chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified Intent-to-Treat set
MMRM	Mixed Model for Repeated Measures
MMSE	Mini-Mental State Examination

MNAR	Missing Not at Random
NFL	Neurofilament Light
p-tau	Phosphorylated tau
PD	Parkinson's disease
PK	Pharmacokinetic
PPS	Per-protocol set
PT	Preferred Term
QD	Once a Day
SAE	Serious Adverse Event
SAF	Safety analysis set
SAP	Statistical Analysis Plan
SD	Standard Deviation
SOC	System Organ Class
TEAE	Treatment-Emergent Adverse Event
$T_{½}$	Half-life
tau	Tau protein
TDP-43	TAR DNA-binding Protein 43
t_{max}	Time to peak drug concentration
TOEPH	Heterogeneous Toeplitz
UN	Unstructured
WHO	World Health Organization

2 Document Version History

Version Number	Version date	Section(s) Updated	Change since previous version (with reason)
Final 1.0	16-JAN-2024	All	Final version
Final 2.0 Amendment	27-FEB-2024	Section 8.8	Missing data algorithm was added on item scores of co-primary endpoint ADAS-Cog11 as a sensitivity analysis

3 Introduction

This Statistical Analysis Plan (SAP) is based on Clinical Study Protocol (CSP) Version 3.1, dated September 25, 2023. In the case where there is a difference between the protocol and the SAP, the SAP will be the ruling document.

Currently, there is no treatment available to stop or reverse the progression of Alzheimer's (AD) and Parkinson's disease (PD). Neurodegenerative diseases such as AD and PD share many common characteristics, including the central role of neurotoxic aggregating proteins in their pathogenesis. Amyloid β (Aβ) and tau aggregates (senile plaques and neurofibrillary tangles, respectively) have been traditionally associated with AD, while alpha-synuclein (αSYN) aggregates (Lewy bodies) have been associated with PD. However, it is becoming increasingly clear that all these proteins are involved in both diseases and that aggregation of one can lead to accumulation of another. Furthermore, in several studies of brains from older AD participants, a high percentage of all AD brains present mixed pathologies, such as Lewy body disease. Several clinical trials targeting just one (often Aβ) of the neurotoxic aggregating proteins have failed. Finally, other fragments of Amyloid β Precursor Protein (APP) have been implicated in AD pathology. Collectively, these facts point to the need for lowering multiple neurotoxic aggregating proteins simultaneously, if we are to have a good chance of at least halting disease progression.

Buntanetap has a unique mechanism of action, in that it inhibits the translation and, therefore, reduces the levels of several neurotoxic aggregating proteins both *in vitro* and *in vivo*, including αSYN, APP, its fragments, and tau. Although APP and its downstream products (Ab oligomers, c-terminus peptide, and amyloid plaques), and tau neurofibrillary tangles have been well documented to be culprits of AD, recent research has shown that other misfolded proteins are also part of the equation. For example, increased levels of αSYN and TAR DNA-binding protein 43 (TDP-43) have been shown to be correlated with deficits in cognitive functions in AD participants. Therefore, it is reasonable to hypothesize that inhibiting expression of all the above neurotoxic proteins should lead to a better efficacy outcome in AD participants than inhibiting just one.

The objective of this study is to assess efficacy and safety of buntanetap in participants with mild to moderate AD. We only tested 80 mg once daily (QD) in Phase 2a study in AD participants. During the dose-finding Phase 2a PD study, a dose-response curve between 0 and 80 mg QD was observed with 10 and 20 mg QD showing the best benefit. Therefore, we want to do a similar dose-ranging trial to test the efficacy of 7.5 mg, 15 mg and 30 mg QD buntanetap.

4 Study Objectives and Endpoints

Objective	*Endpoint*
Primary	
The primary objective is to assess efficacy and safety of buntanetap in participants with mild to moderate AD.	**1. Alzheimer's Disease Assessment Scale-Cognitive Subscale 11 (ADAS-Cog 11)** Change in ADAS-Cog 11 from Baseline to Week 12. ADAS-Cog 11 is an 11-item score measuring cognitive functions and non-cognitive functions such as mood and behavior. Total scores range from 0-70, with higher scores indicating greater cognitive impairment. **2. Alzheimer's Disease Cooperative Study-Clinical Global Impression of Change (ADCS-CGIC)** ADCS-CGIC is a validated categorical measure of change in the participant's clinical condition between baseline and follow-up visits. It measures whether the effects of active treatment are substantial enough to be detected by a skilled and experienced clinician on the basis of a clinical interview and examination. It relies on both direct examination of the participant and an interview of the study partner. A skilled and experienced clinician who is blinded to treatment assignment rates the participant on a 7-point Likert scale, ranging from 1 (marked improvement) to 7 (marked worsening). Lower scores indicate better improvement. **3. Safety and tolerability** • Adverse events (AE) • Severity of AEs • Drug related AEs • AEs leading to study discontinuation • Electrocardiogram (ECG) findings • Clinical laboratory test results • Vital sign measurements • Physical examination findings

Objective	*Endpoint*
Secondary	
Assess improvement on AD participants' daily living and cognitive functions	**1. Alzheimer's Disease Cooperative Study-Instrumental Activities of Daily Living Scale (ADCS-ADL)** The ADCS-ADL is a 23-item inventory developed as a rater-administered questionnaire answered by the participant's study partner. The ADCS-ADL measures both basic and instrumental activities (instrumental activity items 6a, 7-23) of daily living by participants with a lower score indicating greater severity. **2. Mini Mental State Examination (MMSE) Score** MMSE is a brief screening instrument used to assess cognitive function (orientation, memory, attention, ability to name objects, follow verbal/written commands, write a sentence, and copy figures). Total score ranges from 0 to 30 with a lower score indicating greater disease severity. **3. Digital Symbol Substitution Test (DSST)** In the DSST individuals are asked to record associations between different symbols and numbers within time limits. The total score is the sum of all the correctly coded numbers and higher scores indicate better performance.
Exploratory	
Investigate change after buntanetap in potential biomarkers	**1. Plasma biomarkers** Potential biomarkers to be measured in plasma are Aβ42/40 ratio, phosphorylated tau 181 (pTau181), pTau217, Neurofilament Light (NFL), Glial fibrillary acidic protein (GFAP) and TAR DNA-binding protein 43 (TDP43). **2. Pharmacokinetics (PK) Analysis** The following PK parameters will be determined: Area under the curve (AUC), C_{max}, T_{max}, $t_{½}$, and CL.

5 Overall Study Design

This is a 12-week randomized, double-blind, placebo-controlled, dose-ranging, multicenter clinical trial investigating the efficacy, safety, and tolerability of three different doses of buntanetap or placebo in participants with mild to moderate AD.

Qualified participants will be randomly assigned at a 1:1:1:1 ratio to one of the four treatment arms: buntanetap 7.5 mg, buntanetap 15 mg, buntanetap 30 mg, and placebo after a screening period of up to 42 days.

ADAS-Cog 11, ADCS-CGIC, ADCS-ADL, DSST, and MMSE will be assessed by clinicians who have successfully completed the requisite certifications/trainings for each assessment. All efforts will be made to ensure participants will be assessed by the same clinician throughout the study.

320 mild to moderate AD participants will be randomized to 7.5 mg, 15 mg, 30 mg of buntanetap QD or placebo.

Participants who have signed an informed consent and meet screening eligibility requirements will be randomly assigned at a 1:1:1:1 ratio to one of the four treatment arms: buntanetap 7.5 mg, buntanetap 15 mg, buntanetap 30 mg, and placebo, through an Interactive Randomization System, after a screening period of up to 42 days, followed by an at home dosing period of 12 weeks. Participants will be required to visit clinics at Day 0 (baseline), 6 weeks, and 12 weeks (end-of-trial), where they will undergo study procedures that include safety assessments (AE and concomitant medication monitoring, 12-lead ECGs, clinical laboratory testing, vital signs assessments, and physical/neurological examinations), and psychometric tests (ADAS-Cog11, ADCS-CGIC, ADCS-ADL, DSST, MMSE).

The Schedule of Events is shown in Appendix 1.

6 Determination of Sample Size

A sample size of up to 320 participants (1:1:1:1 randomized to 7.5 mg, 15 mg, 30 mg buntanetap and placebo respectively) is planned to be randomized.

Qualified participants will be randomized at a 1:1:1:1 ratio to one of the following four treatment groups: 7.5 mg, 15 mg, 30 mg QD buntanetap and placebo. A sample size of 280 (210 combining all three buntanetap-treatment groups and 70 in control) will have 90% power to detect a treatment difference for the co-primary efficacy variable ADAS-Cog11 at a two-sided significance level of 0.05, assuming a delta of 2.7 between the treatment (pooling all three treatment arms together) and placebo, and a common standard deviation (SD) of 6.0. This sample size will also have an 89% power to detect a significant difference for the co-primary efficacy variable ADCS-CGIC at a two-sided significance level of 0.05, assuming a delta difference 0.23 between the treatment (pooling all three treatment arms together) and placebo, and a common SD of 0.5. Approximately 320 participants will be randomized (approximately 80 per arm) to account for an expected 12.5% dropout rate.

7 Data Sets to be Analyzed

The following analysis sets will be used for the statistical analysis and presentation of data:

- The Enrolled analysis set (ES) will include all participants who signed IC and meet all inclusion and exclusion criteria. All data collected from ES will be presented in listings and if relevant in tables and/or graphs. Screening failures will have selected data be presented in a separate listing if needed.
- The Intent-to-Treat set (ITT) will consist of all participants who are randomized. The primary efficacy analyses will be based on the ITT.
- The Modified Intent-to-Treat set (mITT) will consist of all participants who are randomized, receive at least one dose of study drug, and have both a baseline and at least one post-baseline assessment value for the ADAS-Cog11 or ADCS-CGIC score after taking study drug. The mITT will be used to support the primary efficacy analyses.
- The Safety analysis set (SAF) will consist of all randomized participants who receive at least one dose of study drug. The safety analyses will be based on the SAF.

- The Per-protocol set (PPS) will consist of all participants in the mITT who complete at least 67 days treatment exposure (Last dose date - First dose date + 1 >= 67 days) without any major protocol deviations that potentially could impact the primary efficacy endpoint. The membership of this analysis set will be determined prior to database lock. The PPS will be used to support the primary and secondary endpoints as supportive analyses.
- The PK set will analyze all available data from safety participants receiving buntanetap and with at least one sampling profile allowing derivation of at least one PK parameter.

Protocol deviations leading to exclusions from PPS include but are not limited to the following:

- Non-fulfilment of all inclusion criteria
- Fulfilment of at least one exclusion criterion
- Use of certain concomitant medications (for details see section 8.3.11)

The final criteria for the PPS, regarding which protocol deviations that warrant exclusion, will be determined before breaking the blind.

Baseline presentations will be based on the SAF.

Safety presentations will be based on the SAF.

Exposure to study drug will be given for the SAF.

The ITT and mITT will be analyzed according to randomized (planned) treatment. Safety and PPS will be analyzed on actual treatment.

Plasma concentration-time data for buntanetap will be based on the SAF. PK parameters derivations and statistical analyses are described in a separate PK SAP and will be based on the PK set (see Section 8.5).

8 Statistical and Analytical Plans

8.1 Changes to the Planned Analyses

There have been no changes in the planned analyses in the study protocol.

8.2 Hypothesis and Statistical Methods

For tests of statistical significance on efficacy variables the significance level (p-value) will be provided together with the 95 percent confidence intervals (CI) of the mean, minimum, median, first and third quartiles and maximum.

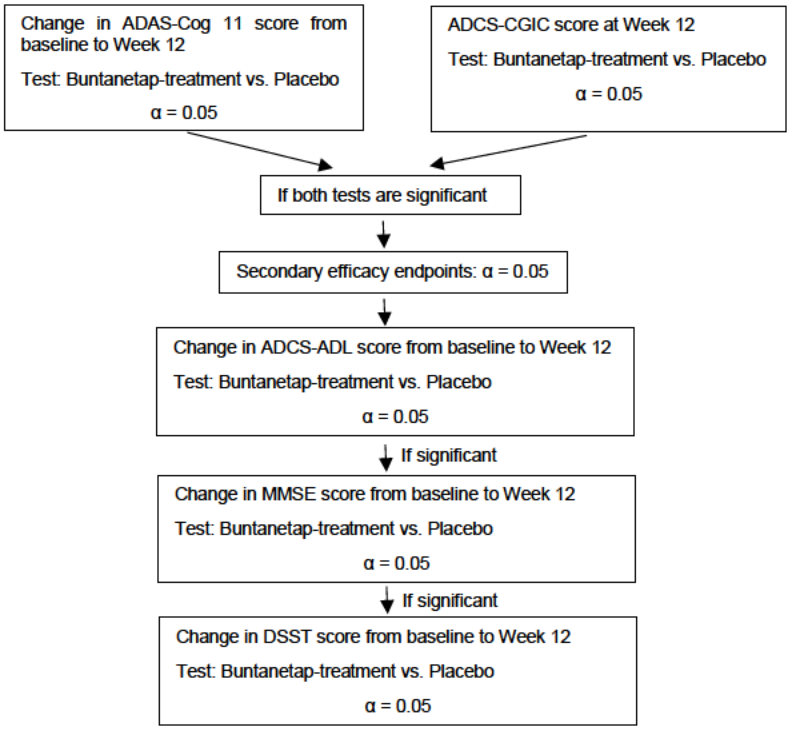
8.2.1 Level of Significance, Multiple Comparisons and Multiplicity

The primary study hypothesis is that buntanetap-treatment (pooling three treatment arms together) is superior to placebo in reducing the cognitive impairment and function impairment of AD participants, as assessed by the two co-primary endpoints: change from baseline in the ADAS-Cog11 score at 12-week and ADCS-CGIC score at 12-week. The two co-primary endpoints will be tested at a two-sided significance level of 5% without multiplicity adjustment. If both test hypotheses are rejected at the 5% level, the test will continue onto secondary endpoints in a hierarchical testing order, at a two-sided significance level of 5%. If one of the two co-primary endpoints is not statistically significant, the hypothesis testing for all subsequent endpoints will be considered exploratory.

To control for study-wise type 1 error rate, a fixed-sequence hierarchical testing procedure will be applied to adjust for multiplicity arising from multiple secondary endpoints. The hierarchical testing order for the secondary endpoints are:

1. ADCS-ADL
2. MMSE score
3. DSST

This hierarchical testing process is presented graphically below:



The study is also designed to explore the effectiveness of three buntanetap-treatment dose level. Nominal p-values will be provided for these test hypotheses of each dose arm vs. placebo, without multiplicity adjustment.

All other exploratory endpoints are considered supportive and no adjustment for multiplicity will be applied.

8.2.2 Definitions

Treatment-Emergent Adverse Event	Treatment emergent adverse events (TEAEs) are defined as undesirable events not present prior to medical treatment, or an already present event that worsens either in intensity or frequency following the treatment. If an AE starts at the same day as the study drug the start time of the AE will be compared with the start time of the administration of study drug. If the start time of the AE is at or after start time of administration of study drug the AE is a TEAE. If no clear assignment is possible due to incomplete or missing start dates, the AE will be counted as TEAE.
Baseline	Baseline will be the last assessment before or up to the first dosing of the study drug. This is valid for vital sign, laboratory data, and safety data, where a difference from baseline is derived. For diary data and questionnaire data, baseline will be the assessment at the baseline visit.
Height [cm]	All values collected in inches will be converted to cm as follows: Height [cm] = height [feet] * 30.48
Weight [kg]	All values collected in pounds will be converted to kg as follows: weight [kg] = weight [pounds] * 0.454

Relative day	The relative day of an event is derived as: Relative day = (Start date) - (Date of first administration of Investigational Medicinal Product (IMP) + 1 For events occurring or starting before the date of first administration of IMP, the relative day is derived as: Relative day = (Start date) - (Date of first administration of IMP) In this way, there will be no Day 0. Day 1 is the same day as the day of first administration of IMP, and Day -1 is the day before.
Date format	All dates in analysis datasets and tables, listings and figures will be in the format YYYY-MM-DD

8.2.3 Summary Statistics

Data will be summarized by means of summary statistics. Summary statistics will be provided for the variables described in the following sections.

Categorical data will be presented as counts and percentages, i.e., n (xx.x%). Percentages will be based on the number of subjects in the analysis data set used and given as integers. Zero frequencies will be given without percentages.

For continuous variables, these statistics will include the number of subjects, mean, SD, median, minimum, and maximum.

All variables will be analyzed descriptively with appropriate statistical methods: categorical variables by frequency tables (absolute and relative frequencies) and continuous variables by sample statistics (i.e., number of observations, mean, SD, 95% CI of the mean, minimum, median, first and third quartiles and maximum). Minimum and maximum values will be displayed as reported.

Since the two co-primary endpoints and secondary endpoints are captured at multiple visits, the method selected to analyze the endpoints, which are treated as continuous, will be a Mixed Model for Repeated Measures (MMRM). For the co-primary endpoint ADAS-Cog11 score and secondary endpoints, MMRM will be applied with change from Baseline score as the response (dependent variable) and with Baseline value as a covariate and treatment, timepoint, treatment-by-timepoint interaction as fixed effects (independent variables). The Model for co-primary endpoint ADCS-CGIC score will include treatment, timepoint, treatment-by-timepoint interaction as the fixed effects. Estimation of parameters will be done by means of Restricted Maximum Likelihood method. The Unstructured (UN) variance-covariance structure will be initially used to explore the variance-covariance matrix across visits. If the algorithm doesn't converge, a heterogeneous Toeplitz (TOEPH) as the covariance matrix will be tried first, and then first-order autoregressive (AR(1)) to achieve convergence. The Kenward-Roger approximation will be used to estimate denominator degrees of freedom. The two-sided 95% CI of the treatment differences will be presented.

The co-primary endpoint ADAS-Cog11 score and secondary endpoints will be analyzed using the model explained above as follows:

<Endpoint>= intercept + Base + T + V + T*V,

with

Base: baseline score;

T: treatment;

V: post-baseline visit (Week 6, Week 12).

The co-primary endpoint ADCS-CGIC score will be analyzed using the model explained above as follows:

<Endpoint>= intercept + T + V + T*V,

with

T: treatment;

V: post-baseline visit (Week 6, Week 12).

Treatment comparisons (pooled Buntanetap-treatment vs. Placebo) will be carried out by means of the contrasts on the treatment factor by visit effect in SAS® Software, using proc mixed. Treatment effects will be estimated by means of Least Square Means (LSM) and 95% CI. Differences between treatments (pooled Buntanetap-treatment – Placebo) will be estimated and resulting 2-sided p-values and associated 95% CI will be presented.

8.2.4 Patient/Subject Data Listings

Data collected in the electronic case report form (eCRF) will generally be listed. eCRF check questions (e.g., reminders) will not be listed.

Listings will be sorted by treatment group, study center, patient/subject id, visit and assessment time as applicable. If center id is part of patient/subject id, then center could be omitted.

In CRF modules where a date is recorded, the relative day may be included in the corresponding listing. In modules where both a start date and stop date are recorded, the duration may be included in the listing.

8.2.5 Subject Disposition

The numbers of subjects screened, enrolled, completed, discontinued, and withdrawn during the study, as well as the reasons for all post-enrolment discontinuations will be listed and/or summarized by treatment group. Disposition and reason for study discontinuation will also be provided as a by-subject listing.

8.2.6 Demographic and Other Baseline Characteristics

Demographics and baseline characteristics will be summarized by treatment group using appropriate descriptive statistics.

The following summaries will be given by treatment and in total:

- Demography
    - Age (years)
    - Age (years), categories (55 to <65, 65 to <75, 75 to <85)
    - Sex (Male or Female)
    - Race (White, Asian, Black or African American, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Other)
    - Ethnicity (Hispanic or Latino, Not Hispanic or Latino, Not Reported, Unknown)
    - Height (cm)
    - Weight (kg)
    - BMI (kg/m^2)

8.2.7 Primary Analysis

The co-primary endpoints in this clinical trial are the change from baseline to Week 12 in ADAS-Cog11 score and the ADCS-CGIC score at Week 12 between buntanetap-treatment (pooling three treatment arms together) and placebo.

The statistical hypothesis to be tested for co-primary endpoint ADAS-Cog11 score is described as:

- Null Hypothesis (H0):

  The difference in change in ADAS-Cog11 score from baseline to Week 12 between buntanetap-treatment (pooling three treatment arms together) and placebo = 0.

- Alternative Hypothesis (H1):

The difference in change in ADAS-Cog11 score from baseline to Week 12 between buntanetap-treatment (pooling three treatment arms together) and placebo ≠ 0.

The statistical hypothesis to be tested for co-primary endpoint ADCS-CGIC score is described as:

- Null Hypothesis (H0):

  The difference in ADCS-CGIC score at Week 12 between buntanetap-treatment (pooling three treatment arms together) and placebo = 0.

- Alternative Hypothesis (H1):

  The difference in ADCS-CGIC score at Week 12 between buntanetap-treatment (pooling three treatment arms together) and placebo ≠ 0.

The two co-primary endpoints will be analyzed via MMRM. The model for the co-primary endpoint ADAS-Cog11 score will include treatment, timepoint, treatment-by-timepoint interaction as the fixed effects and baseline as the covariate. The Model for the co-primary endpoint ADCS-CGIC score will include treatment, timepoint, treatment-by-timepoint interaction as the fixed effects. The variance-covariance matrix will be assumed to be UN.

This analysis will be conducted after all participants have reached Week 12 or early withdrawal, whichever comes first. The primary analysis will be conducted in ITT. The mITT and PPS will be used to support the primary analysis.

Descriptive statistics per visit and change from baseline for each visit will be provided.

As the study is also designed to explore the effectiveness of three buntanetap-treatment dose levels on co-primary endpoints, similar MMRM will be applied comparing each buntanetap-treatment dose arm with placebo.

8.2.8 Secondary Analysis

Secondary efficacy endpoints in this clinical trial are

- Change from baseline to Week 12 in ADCS-ADL
- Change from baseline to Week 12 in MMSE score
- Change from baseline to Week 12 in DSST

Change from Baseline to Week 12 in ADCS-ADL

The ADCS-ADL is a 23-item inventory developed as a rater-administered questionnaire answered by the participant‘s study partner. The ADCS-ADL measures both basic and instrumental activities (instrumental activity items 6a, 7-23) of daily living by participants with a lower score indicating greater severity.

ADCS-ADL score will be analyzed by MMRM, similar to the co-primary endpoint ADAS-Cog11 score.

Change from Baseline to Week 12 in MMSE Score

MMSE is a brief screening instrument used to assess cognitive function (orientation, memory, attention, ability to name objects, follow verbal/written commands, write a sentence, and copy figures). Total score ranges from 0 to 30 with a lower score indicating greater disease severity.

MMSE score will be analyzed by MMRM, similar to the co-primary endpoint ADAS-Cog11 score.

Change from baseline to Week 12 in DSST

In the DSST individuals are asked to record associations between different symbols and numbers within time limits. The total score is the sum of all the correctly coded numbers and higher scores indicate better performance.

DSST score will be analyzed by MMRM, similar to the co-primary endpoint ADAS-Cog11 score.

As the study is also designed to explore the effectiveness of three buntanetap-treatment dose levels on secondary endpoints, similar MMRM will be applied comparing each buntanetap-treatment dose arm with placebo.

8.2.9 Exploratory Analysis

Exploratory endpoints in this clinical trial are

- Plasma biomarkers
- PK analysis.

Plasma Biomarkers

Potential biomarkers to be measured in plasma are Aβ42/40 ratio, pTau181, pTau217, NFL, GFAP and TDP43.

PK analysis

Plasma concentration-time data for buntanetap will be listed and summarized descriptively (number of subjects (N), mean, geometric mean, median, SD, coefficients of variation (CV), minimum, and maximum) for the SAF.

Plasma buntanetap PK parameters will be calculated using a non-compartmental method. Calculations will be based on the actual sampling times recorded during the study. The following PK parameters will minimally be determined, as data permit: Area under the curve (AUC), C_{max}, T_{max}, $t_{½}$, and CL. A separate PK SAP will include details on PK parameters derivations and statistical analyses based on the PK set.

8.2.10 Exposure to Treatment

Duration of exposure will be summarized quantitatively using the number of days on study drug for each patient, displayed by treatment group. Duration will be calculated as (last dose date – first dose date + 1).

Compliance with treatment will be assessed as the percentage of tablets taken in relation to number of tablets expected to be taken.

$$\textit{Compliance with study drug (\%)} = \frac{Total\ number\ of\ tablets\ taken}{Number\ of\ tablets\ expected}$$

Exposure and compliance will be presented using summary statistics by treatment group and the pooled buntanetap-treatment group.

8.2.11 Medical History and Concomitant Medications

Medical history will be coded using the most recent version of the Medical Dictionary for Regulatory Activities (MedDRA). Additionally, a listing of the relevant medical history will be provided.

Prior and concomitant medications will be summarized by treatment group and by the pooled buntanetap-treatment group, by the number and percentage of subjects taking each medication, classified using World Health Organization (WHO) Drug Dictionary Anatomical Therapeutic Chemical (ATC) classes and preferred terms (PT).

8.2.11.1 Prohibited concomitant medication

Investigational agents are prohibited 4 weeks or five half-lives, whichever is greater, prior to entry and for the duration of the trial.

Aducanumab/Aduhelm and Lecanemab/Leqembi are prohibited for 4 weeks or five half-lives, whichever is greater, prior to entry and for the duration of the trial.

Initiation of prohibited medications during the course of the study is discouraged, however, if an excluded medication is initiated after screening, the site should consult with the Medical Monitor for further guidance.

Buntanetap is mainly metabolized by CYP3A4 in vitro. Therefore, we recommend avoiding concomitant use of strong or moderate CYP3A4 inhibitor/inducers, examples see below according to Food and Drug Administration (FDA) guidance on drug development and drug interactions.

CYP3A4 inhibitors	Itraconazole, Ketoconazole, Azamulin, Troleandomycin, Verapamil
CYP3A4 inducers	Rifampicin

8.2.11.2 Permitted concomitant medication

Use of the following medications is allowed only if the participants have been stabilized for at least 4 weeks before screening and will continue to be stable throughout the study:

- tricyclic antidepressants
- antipsychotics prescribed for any reason. However, the dosage must not exceed the following for these medications:
  Quetiapine ≤50 mg/day, risperidone ≤1.5 mg/day, olanzapine ≤5 mg/day, and aripiprazole ≤10 mg/day

If the participant is on other antipsychotics besides the ones mentioned above, please consult the Medical Monitor before screening the participant.

- psychostimulants

Discontinuation and/or change of any of the above medication during the course of the study is discouraged, however, if necessary, the site should consult with the Medical Director and Sponsor for further guidance.

For concomitant treatment with anti-cholinesterase, it is allowed to use for at least 4 weeks prior to screening.

For any anticonvulsant medications used for epilepsy or mood stabilization, neuropathic pain indications, dosing must be stable for at least 4 weeks prior to screening.

For mood-stabilizing psychotropic agents, including but not limited to lithium, dosing must be stable for at least 4 weeks prior to Screening.

Use of short acting benzodiazepines and hypnotics for treatment on an as-needed basis for insomnia or daily dosing as anxiolytics is permitted but should be avoided for 8 hours before administration of cognitive tests. For sedating medication given for any short-term use, all cognitive assessments must be administered at least 24 hours after administration of the sedative.

At this time, there is no controlled clinical data available for anyone receiving SARS-COVID19 vaccination while being treated with buntanetap. Based on the intended mechanism of actions and currently available data for these vaccine candidates and for buntanetap, it is not believed that there are specific risks to consider in participants participating in this study who have received SARS-COVID19 vaccination or who are planning to receive the vaccination. As such, SARS-COVID19 vaccination is not prohibited. If appropriate, SARS-COVID19 vaccination should be considered as a concomitant medication.

8.2.11.3 Coding and imputation of incomplete dates

All concomitant medications/therapies will be classified according to ATC level 2 group text and WHO Drug Dictionary PT. The medications will be classified into categories Prior and Concomitant drug administration based on start date and end date in relation to IMP exposure.

- Prior medications are those for which end dates of the medication/therapy are strictly before date of first administration of study drug.
- Concomitant medications are those for which the period between their start dates and end dates coincide with exposure to study drug and can be further classified into:
  - Concomitant medications starting prior to first exposure to study drug having start dates strictly before first exposure to study drug and end dates on same date or after date of first administration to study drug or are ongoing.
  - Concomitant medications starting on the date of first administration of study drug or after but before or on the date of last administration of study drug

For medication and therapies with partial dates:

- If start date is completely missing, it will be assumed that the medication started before date of first administration of study drug.
- If end date is completely missing and ongoing is not ticked, it will be assumed that the medication ended before date of first administration of study drug.
- If only the year is available, month 6 will be imputed for the purpose of classification.
- For medications/therapies where only year and month (including imputed dates) are available this information will be used in analogy with how complete dates are used for classification, e.g.:

Start date	End date	Ongoing	IMP start date	IMP end date	Classification
2020-07	2020-07		2020-08-14	2020-09-13	Prior
2020-07		No	2020-08-14	2020-09-13	Prior
2020-07		Yes	2020-08-14	2020-09-13	CM before
2020-07	2020-08		2020-08-14	2020-09-13	CM before
2020-08	2020-08		2020-08-14	2020-09-13	CM before
2020-08		No or Yes	2020-08-14	2020-09-13	CM before
2020-08	2020-09		2020-08-14	2020-09-13	CM before
2020-09	2020-09		2020-08-14	2020-09-13	CM after
2020-09		No or Yes	2020-08-14	2020-09-13	CM after

CM: Concomitant Medication

A relative day will not be calculated for medications with incomplete dates.

The concomitant medications will be presented in a summary table broken down on timing in relation to study drug (i.e., Prior and Concomitant). Each subject will only be counted once for each medication and timing category, on a PT level in each period.

One list for each timing period will be presented for the SAF. In addition, a list of anti-parkinsonian medications will be provided.

8.3 Safety Assessment

All safety summary analyses will be performed using the actual treatment for the Safety Analysis by each treatment group and by the pooled buntanetap-treatment group.

8.3.1 Adverse Events

Adverse events include but are not limited to:

- worsening or change in nature, severity, or frequency of conditions or symptoms present at the start of the study;
- subject deterioration due to primary illness;
- intercurrent illness;
- drug interaction.

An abnormal laboratory value will only be reported as an AE if the investigator considers it clinically significant, or if it leads to the participant being withdrawn from the study.

Symptoms and conditions present at the beginning of the study will be characterized, so that AEs can be defined as any new symptom, or any increase in frequency or severity of an existing symptom.

AEs will be coded according to the most recent version of the MedDRA system and will be tabulated by system organ class (SOC) and PT.

A TEAE is an undesirable event that is not present prior to medical treatment, or an already present event that worsens either in intensity or frequency following the treatment. Only TEAEs will be included in summary tables and non-TEAEs will be listed separately.

Relationship to clinical study treatment will be mapped according to the scheme below:

- Potentially related: will include all adverse events with a relationship rating of “definitely”, “probably” or “possibly”.
- Unlikely/not related: will include all adverse events with a relationship rating of “unlikely” or “unrelated”.

An overall summary table will give

- number of events
- number of unique events
- number of subjects with at least one event

for:

- Overall summary of TEAEs
- Serious TEAEs
- Serious TEAE related to study drug
- TEAEs related to study drug
- TEAEs by maximum severity
- TEAEs leading to treatment discontinuation and/or Interruption of study drug
- TEAEs leading to withdrawal from study
- TEAEs leading to death

The total number and percentage of subjects with at least one TEAE and the total number of TEAEs will be presented by SOC and PT. If more than one TEAE occurs with the same PT for the same subject, then the subject will be counted only once for that PT using the most severe and most related occurrence for the summary.

Frequent TEAEs are defined as any PT with ≥5% incidence in any of the treatment arms, which will be summarized by PT.

An AE listing by subject will display all events, and will include the verbatim term in addition to the MedDRA SOC and PT. This listing will also include all relevant eCRF data associated with the event: date of onset, date resolved, date of last dose, severity, frequency, outcome and relationship to study drug and action taken with study drug. Separate listings will be presented for subjects with treatment-emergent serious AEs (SAEs), related TEAEs, TEAEs leading to drug withdrawn, fatal TEAEs (if any).

If AE start date is incomplete, AEs with start month same as month of first administration of study drug will be regarded as TEAEs.

Adverse events will be reported to Data Safety Monitoring Board (DSMB) (see Section 8.11) per DSMB Charter.

8.3.2 Vital Signs

Descriptive summaries (mean, SD, median, minimum, and maximum) of actual values and changes from baseline for all visits at which they were assessed will be calculated for systolic blood pressure, diastolic blood pressure, pulse, respiratory rate, and oral temperature.

Shift tables showing the number of subjects who shifted from normal, abnormal not clinically significant or abnormal clinically significant at baseline to normal, abnormal not clinically significant or abnormal clinically significant at each post-baseline time of assessment will be presented.

8.3.3 Clinical Laboratory Measurements

For continuous laboratory parameters, descriptive statistics will be presented for each visit and for the changes from baseline to each subsequent visit by treatment group, and the pooled buntanetap-treatment group.

CLINICAL SAFETY LAB TESTS		
METABOLIC PANEL	**COMPLETE BLOOD COUNT**	**URINALYSIS**
Sodium (Na) Potassium (K) Chloride (Cl) Carbon Dioxide (CO2) Blood Urea Nitrogen (BUN) Glucose Calcium (Ca) Creatinine (Crn) Bilirubin (direct and total) Albumin Protein Total Glutamic-Oxaloacetic Transferase (AST, SGOT) Glutamic-PyruvateTransferase (ALT, SGPT) Alkaline Phosphatase (ALP) eGFR (estimated Glomerular Filtration Rate)	White Blood Cell Count (WBC) Red Blood Cell Count (RBC) Hemoglobin (Hb) Hematocrit (HCT) Mean Corpuscular Volume (MCV) Mean Corpuscular Hemoglobin (MCH) Mean Corpuscular Hemoglobin Concentration (MCHC) Red Blood Cell Distribution Width (RDW) Mean Platelet Volume (MPV) Platelet Count (PLT) **Screening Only** Hemoglobin A1C (HbA1c)	Color Appearance Specific Gravity pH Blood Glucose Protein Ketones Leukocyte Esterase Nitrite Urobilinogen Bilirubin (total)

Additionally, clinical laboratory parameters will be categorized as low, normal, or high according to laboratory range specifications and the number and percentage of subjects in each category will be presented in shift tables.

Shift tables showing the number of subjects who shifted from normal, abnormal not clinically significant or abnormal clinically significant at baseline to normal, abnormal not clinically significant or abnormal clinically significant at each post-baseline time of assessment will be presented.

Laboratory values that are outside the normal range will also be flagged in the data listings, along with corresponding normal ranges.

If laboratory values are below the limit of quantification, the value corresponding to the limit of quantification will be used when summarizing data (e.g., if the result is <x.x, then the value x.x will be used in the statistical analysis).

8.3.4 Physical Examination

Number and percentage of normal, abnormal but not clinically significant or abnormal and clinically significant findings will be tabulated by visit for the body systems:

- skin,
- head,
- eyes,
- ears,
- nose,
- throat,
- pulmonary,
- cardiovascular,
- abdomen,
- musculoskeletal,
- extremities.

Shift tables showing the number of subjects who shifted from normal, abnormal not clinically significant or abnormal clinically significant at baseline to normal, abnormal not clinically significant or abnormal clinically significant at each post-baseline time of assessment will be presented.

Physical examination data will also be presented in listings.

8.3.5 Neurological Examination

Number and percentage of normal, abnormal but not clinically significant or abnormal and clinically significant findings will be tabulated by visit for the body systems

- cranial nerves,
- strength,
- coordination,
- reflexes,
- sensation,
- tremor,
- gait,
- mental status.

Shift tables showing the number of subjects who shifted from normal, abnormal not clinically significant or abnormal clinically significant at baseline to normal, abnormal not clinically significant or abnormal clinically significant at each post-baseline time of assessment will be presented.

Neurological examination data will also be presented in listings.

8.3.6 Cardiac Assessments

A 12-lead ECG will be performed prior to the administration of study drug at the timepoints of assessments.

The Investigator will enter a signed and dated clinical interpretation of the ECG trace in the eCRF. At each timepoint, absolute values and change from Baseline of ECG numeric variables will be summarised through descriptive statistics by each treatment group, and the pooled buntanetap-treatment group. Observed QTc will also be categorized into the following categories (msec) and the number and percentage of subjects in each category will be summarized at each timepoint and the overall post-baseline maximum:

- Observed: <=450, 451-480, 481-500, and >500; >450; >480
- Change from Baseline: <=10, 11-30, 31-60, and >60; >30

For ECG interpretations, the number of available observations and the number of normal, abnormal clinically significant and abnormal not clinically significant values (absolute and in percentage) will be summarized at each timepoint. The overall post-baseline worst interpretation will also be summarized.

Shift tables showing the number of subjects who shifted from normal, abnormal not clinically significant or abnormal clinically significant at baseline to normal, abnormal not clinically significant or abnormal clinically significant at each post-baseline time of assessment will be presented.

8.3.7 Columbia Suicide Severity Rating Scale (C-SSRS)

The number and percentage of subjects with suicidal ideation or suicidal behavior based on the C-SSRS will be summarized by treatment group and the pooled buntanetap-treatment group. Suicidal ideation or behavior are defined as:

- Suicidal ideation: A “yes” answer at any time during treatment to any one of the five suicidal ideation questions (Categories 1-5)
- Suicidal behavior: A “yes” answer at any time during treatment to any one of the five suicidal behavior questions (Categories 6-10)
- Suicidal ideation or behavior: A “yes” answer at any time during treatment to any one of the ten suicidal ideation and behavior questions (Categories 1-10)

The intensity of most severe suicidal ideation during the participant's lifetime, during the double-blind treatment period, and during the safety follow-up period will also be presented by treatment group and the pooled buntanetap-treatment group for the SAF. Intensity of most severe ideation is a sum of the 5 intensity items (Frequency, Duration, Controllability, Deterrents, and Reasons for Ideation).

Actual lethality and potential lethality on most lethal attempt will also be summarized by treatment group and the pooled buntanetap-treatment group.

Supportive listings will be provided and will include the participant number, study center number, lifetime history, and post-baseline values. Intensity of suicidal ideation, suicidal behavior type, and lethality of suicidal behavior will also be included in these listings.

A listing of all AEs occurring in participants who have suicidal ideation or suicidal behavior will also be provided.

8.4 Plasma Biomarkers and Pharmacokinetic Analysis

Biomarkers, including Aβ42/40 ratio, pTau181, pTau217, GFAP, NFL and potentially TDP43, will be summarized descriptively for SAF.

Plasma concentration-time data for buntanetap will be listed and summarized descriptively (number of subjects (N), mean, geometric mean, median, SD, coefficients of variation (CV), minimum, and maximum) for the SAF.

Non-compartmental analysis of one analyte (buntanetap) in blood matrix concentrations will be performed using validated computer software Phoenix. Results will be tabulated and graphically displayed using Phoenix and appropriately interpreted by a PK Scientist. Calculations will be based on the actual sampling times recorded during the study. The following PK parameters will minimally be determined, as data permit:

- Area under the curve (AUC), including $AUC_{0\text{-}10hr}$ and $AUC_{0\text{-}inf}$
- Maximum concentration (C_{max}) and time of maximum concentration (T_{max})
- Terminal half-life ($t_{½}$)
- Systemic clearance (CL).

8.5 Multicenter Studies

No per-center analyses are planned.

8.6 Adjustment for Covariates

Baseline value will be used as a covariate in the MMRM analyses for the co-primary endpoint ADAS-Cog11 score (see Section 8.3.7).

8.7 Examination of Subgroups

No subgroup analyses are planned for this study.

8.8 Handling of Dropouts, Intercurrent Events, and Missing Data

In case of incomplete dates, the following approach will be used:

- If only the day is missing and the month and year are different from the month and year of the index date, the day will be imputed with 1. In case the month and year are the same as the index day the missing day will be imputed with the last day of the month.
- In case of missing months, they will be imputed
    - with July if the year is not the same as the year of the index date
    - with the following month as the month of the index date if the year is the same as the year of the index date.

If not indicated otherwise in case of completely missing dates no imputation will be done.

For co-primary endpoint ADAS-Cog11, if there are 3 or less missing item scores at a visit, the missing item score will be imputed using the group mean score of that item within the same treatment group at the same visit. If there are more than 3 item scores missing, the total score will be set as missing. The same MMRM analysis as the one applied to the primary analysis will be fitted as a sensitivity analysis.

The MMRM analysis of the two co-primary endpoints assumes that data are missing at random (MAR). The analysis will be conducted in ITT set and subjects will be analyzed as randomized. ITT subjects discontinue from the study for any reason without completing the study are considered intercurrent events. To assess the robustness of this MAR assumption, a conservative control-based multiple imputation method based on missing not at random (MNAR) assumption will be applied to impute missing data arising from intercurrent events for the two co-primary endpoints analyses. In this MNAR approach, subjects who do not complete the study, missing data will be imputed by observed data obtained from the placebo group using multiple imputation with Markov Chain Monte Carlo (MCMC) simulation method for each treatment separately. For the co-primary endpoint ADAS-Cog11 score, the variables for the imputation model will include ADAS-Cog11 score at baseline, Week 6, and Week 12. The ROUND, MINIMUM and MAXIMUM options will be utilized to ensure imputed values are clinically plausible integers (0 – 70). For the co-primary endpoint ADCS-CGIC score, the variables for the imputation model will include ADCS-CGIC score at Week 6 and Week 12. The ROUND, MINIMUM and MAXIMUM options will be utilized to ensure imputed values are clinically plausible integers (1 – 7). Approximately 50 datasets will be imputed and the same MMRM analysis as the one applied to the primary analysis will be fitted. The results from these 50 analyses will be combined using Rubin's rule to construct the treatment estimates from the parameter estimates and associated standard errors. The difference of the treatment means (active minus placebo) will be presented with the associated standard error and two-sided 95% confidence interval. The imputation will be performed for the two co-primary endpoints separately. Two random seed numbers will be selected for the multiple imputation procedure and will be retained.

8.9 Interim Analysis

An interim sample size re-estimation will be performed after a total of n_1 = 90 evaluable participants across four treatment arms have 6 weeks of assessments for the co-primary efficacy endpoints ADAS-Cog11 and ADCS-CGIC. Evaluable participants are defined as ITT participants who received the study drug and have baseline and Week 6 values of both co-primary efficacy endpoints. The aim of the interim sample size review is to re-assess the size of the estimated mean difference between the treatment (pooling three treatment arms together) and placebo, and SDs of the co-primary endpoints ADAS-Cog11 and ADCS-CGIC to ensure sufficient power at the final analyses. No interim analyses are planned for the purpose of stopping the study early for success or futility.

At the time preparing this SAP, the interim sample size re-estimation was done and interim Statistical Analysis Plan (iSAP) was signed off in October, 2023. Please refer to iSAP for more details.

8.10 Data Monitoring

An independent DSMB will review the safety information from the study on an ongoing basis. The DSMB, will identify the study-specific data parameters and format of the information to be reported, as well as the timing of reports based on the enrollment status of the study. The DSMB will initially be provided with data blinded to treatment status, but they may request unblinded data if there is a safety concern.

Additionally, the DSMB will be informed of the occurrence of any SAEs probably related to the drug within 7 days of being reported to the Coordinating Center. The DSMB may at any time request additional information from the Coordinating Center.

The DSMB will consist of at least one statistician and two physicians with experience of DSMBs and clinical trials.

The data to be considered by the DSMB, the process for quality control of reports and data, and the documentation of decisions were described in detail in the mutually agreed DSMB charter prior to the study start. The primary responsibility of the DSMB will be to protect the safety of the study subjects.

Based on the review of safety data, the DSMB will make recommendations regarding the conduct of the study. These may include amending safety monitoring procedures, modifying the protocol or

consent, terminating the study, or continuing the study as designed. Using the Safety Review Process (review of lab data, vitals, and adverse events) and the DSMB, there is substantial oversight and case review to alert the investigators, in a timely manner, to any safety issues that may arise. Further details will be provided in the DSMB charter.

8.11 Reporting Conventions

Any p-values ≥0.001 will be reported to three (3) decimal places; p-values <0.001 will be reported as <0.001. PK parameters will be reported to two decimal places. The mean and median will be displayed to one decimal place greater than the original value and the measure of variability (e.g., SD) will be displayed to two decimal places greater than the original value.

APPENDIX 1: SCHEDULE OF EVENTS

Study Period	**Screening/Baseline**		**At-Home Treatment**		**Unscheduled Visit[k]**	**Early Discontinuation Visit**
Visit Timing	**Day -42 to 0**		**6 weeks ± 7 days**	**12 weeks ± 7 days (end-of-trial) [b]**		
Procedure	**Screening [j] (up to 42 days before Day 0)**	**Baseline (Day 0)**	**Clinic Visit -**	**Clinic Visit -**	**Clinic Visit**	**Clinic Visit**
Informed consent	X					
Inclusion and exclusion criteria	X	X				
Demography information	X					
Full physical and neurological examination	X	X	X	X	X	X
Height, Weight, BMI	X					
Medical and psychiatric history (includes substance use)	X					
Urine pregnancy test (WOCBP only)	X	X	X	X	X	X
MMSE [l]	X	X		X		X
ADCS-CGIC		X[d]	X[e]	X[e]		X[e]
ADAS-Cog11		X	X	X		X
ADCS-ADL		X	X	X		X
DSST		X	X	X		X
C-SSRS	X	X	X	X	X	X
Safety laboratory tests [c]	X	X	X	X	X	X
12-lead ECG	X	X	X	X		X
Vital signs[a]	X	X	X	X	X	X
Randomization		X				
Genetic sample (ApoE)		X				
Study intervention		X[g]	X[h]	X[l]		
Study intervention compliance review			X	X		
AE/SAE review	X	X	X	X	X	X

Study Period	**Screening/Baseline**		**At-Home Treatment**		**Unscheduled Visit[k]**	**Early Discontinuation Visit**
Visit Timing	**Day -42 to 0**		**6 weeks ± 7 days**	**12 weeks ± 7 days (end-of-trial) [b]**		
Procedure	**Screening [j] (up to 42 days before Day 0)**	**Baseline (Day 0)**	**Clinic Visit -**	**Clinic Visit -**	**Clinic Visit**	**Clinic Visit**
Concomitant medication review	X	X	X	X	X	X
PK sampling		X		X		
Sampling for blood biomarkers		X		X		
24-hour phone follow-up		X	X	X	X	X

[a] Vital sign will include sitting blood pressure, pulse, temperature, respiration rate.
[b] Participants that drop out before the initial 9 weeks of study treatment will complete the Early Discontinuation visit and participants that drop out after the initial 9 weeks of study treatment will complete the End-of-Trial visit.
[c] Safety labs should be taken fasted.
[d] ADCS-CGIC evaluation only.
[e] ADCS-CGIC evaluation and scoring the scale.
[g] Participants are given one pill during the visit after taking pre-dose (0h) blood samples and the pills for the remaining 6 weeks at the end of the visit.
[h] Participants are given pills for the remaining 6 weeks at the end of the visit.
[I] Participants are given one pill during the visit after taking pre-dose (0h) blood samples.
[J] At PI's discretion, re-testing can be done within 30 days. If the investigator deems re-screening is justified, they must contact Medical Monitor and sponsor for discussion on a case-by-case basis. If/when rescreening MMSE, the test must be done at least 3 weeks apart from the previous one.
[K] Investigator will determine the extra assessment if any is needed.
[l] To avoid the learning effect, all MMSEs should be done at least 3 weeks apart.