Cover Page for ClinicalTrials.gov

Sponsor name:	Annovis Bio, Inc.
NCT number:	NCT05357989
Sponsor trial ID:	ANVS-22001
Official title of study:	A 6-month prospective, randomized, double-blind, placebo-controlled clinical trial investigating the efficacy, safety, and tolerability of two different doses of buntanetap or placebo in patients with early Parkinson's disease
Document type:	Statistical Analysis Plan
Document date:	January 11, 2024

Statistical Analysis Plan

for

A 6-month prospective, randomized, double-blind, placebo-controlled clinical trial investigating the efficacy, safety, and tolerability of two different doses of buntanetap or placebo in patients with early Parkinson's disease

Sponsor ANNOVIS BIO, INC., USA

Product/Compound ANVS-22001

Phase of the study III

EudraCT number 2022-001542-38

	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
ADL	Activities of daily living
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
CGIS	Clinical Global Impression of Severity
CI	Confidence interval
CL	Clearance
CM	Concomitant medication
CP	Conditional Power
CSF	Cerebrospinal Fluid
CSP	Clinical Study Protocol
C-SSRS	Columbia Suicide Severity Rating Scale
CV	Coefficients of variation
DSMB	Data Safety Monitoring Board
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
ES	Enrolled set
FDA	Food and Drug Administration
GFAP	Glial Fibrillary Acidic Protein
IC	Informed Consent
IDMB	Independent DSMB
IMP	Investigational Medicinal Product
ITT	Intent-to-Treat set
LSM	Least Square Mean
MAR	Missing at Random
MCMC	Markov Chain Monte Carlo
MDS-UPDRS	Movement Disorder Society-United Parkinson's Disease Rating Scale
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
PD	Parkinson's disease
PGIC	Participant Global Impression of Change
PK	Pharmacokinetic
PPS	Per-protocol set
PT	Preferred Term
QD	Once a Day
RCTC	Rheumatology Common Toxicity Criteria
SAE	Serious Adverse Event
SAF	Safety analysis set
SAP	Statistical Analysis Plan
SD	Standard Deviation
SOC	System Organ Class / Standard of care
$T_{½}$	Half-life
tau	Tau protein
TDP-43	TAR DNA-binding Protein 43
TEAE	Treatment-Emergent Adverse Event
t_{max}	Time to peak drug concentration
TOEPH	Heterogeneous Toeplitz
UN	Unstructured
WAIS-IV	Wechsler Adult Intelligence Scales, 4th edition
WHO	World Health Organization

2 Document Version History

Version Number	Version date	Section(s) Updated	Change since previous version (with reason)
Draft 0.1	24-OCT-2022	All	First version
Draft 0.2	03-NOV-2022	All	Internal review comments incorporated
Draft 0.3	17-NOV-2022	All	Sponsor first review comments incorporated
Final 1.1	09-DEC-2022	All	Final version
Final 1.1 Amendment	25-APR-2023	All	Bonferroni method was updated to Hochberg method. Revised the randomized sample size and evaluable sample size as needed in the Interim Analysis section for consistency with the protocol.
Final 2.0 Amendment	07-DEC-2023	All	FDA comments incorporated
Final 2.1 Amendment	11-JAN-2024	Section 7 Section 8.5	Study drug exposure was included in defining Per-protocol set. A separate PK SAP is not needed.

3 Introduction

This Statistical Analysis Plan (SAP) is based on Clinical Study Protocol (CSP) Global-Version 5.0, dated August 21, 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 patients, 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 development of combination therapies that target 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. All three proteins have been implicated in the pathogenesis of PD. Furthermore, at older ages, there is a high incidence of PD patients with mixed pathologies, such as PD dementia. Therefore, it is reasonable to hypothesize that inhibiting expression of all three proteins should lead to a better efficacy outcome in PD patients than inhibiting just one.

The objective of this study is to confirm efficacy of buntanetap in early PD subjects. During the dose-finding Phase 2a study, a dose-response curve between 0 and 80 mg once daily (QD) was observed with 10 and 20mg QD showing the best benefit. Therefore, decision has been taken to dose subjects with 10 and 20 mg QD.

4 Study Objectives and Endpoints

Objective	*Endpoint*
Primary	
The primary objective is to assess efficacy and safety of buntanetap in early PD subjects	**1. Change from baseline to Month 6 in MDS-UPDRS Part II (OFF-state)** Change in the Score from the Activities of Daily Living (ADL) Scale in the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) (Part II, a MDS-UPDRS Subscale) from Baseline to Month 6. The MDS-UPDRS is a 50-item rating scale designed to assess Parkinson's disease-related disability and impairment. The scale comprises four parts: Part I evaluates mentation, behavior, and mood symptoms; Part II evaluates ADL; Part III evaluates motor function; and Part IV evaluates complications of dopaminergic therapy. Part II contains 13 questions which are to be completed by the participant (range 0-52). A higher score indicates more severe symptoms of PD. **2. 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 PD subjects' daily living, mobility, and complications	**1. Change from baseline to Month 6 in the sum of MDS-UPDRS Part II+III (OFF-state)** Change in the Sum of the Score from the ADL Scale and Motor Examination in the MDS-UPDRS (Parts II+III, a MDS-UPDRS Subtotal) from Baseline to Month 6. Part II contains 13 questions which are to be completed by the participant (range 0-52). Part III contains 33 scores based on 18 items (range 0-132). MDS-UPDRS Part II and III combined score equals the sum of Parts II and III (range 0-184). A higher score indicates more severe symptoms of PD. **2. Change from baseline to Month 6 in MDS-UPDRS Part III (OFF-state)** Part III (motor function) contains 33 scores based on 18 items (Range 0-132). A higher score indicates more severe symptoms of PD. **3. Change from baseline to Month 6 in the sum of MDS-UPDRS Total score (OFF-state)** The MDS-UPDRS is a 50-item rating scale designed to assess Parkinson's disease-related disability and impairment. The scale comprises four parts: Part I evaluates mentation, behavior, and mood symptoms; Part II evaluates ADL; Part III evaluates motor function; and Part IV evaluates complications of dopaminergic therapy. The Total score is the sum of the subscale scores for Parts I to III and ranges from 0 to 236, with higher scores reflecting greater severity. **4. Percentage of Responders with "Much Improved" or "Very Much Improved" on Participant Global Impression of Change (PGIC) (ON-state)** The PGIC is the participant-reported outcome. The qualitative assessment of meaningful change will be determined by the participant in response to the question, "Compared to your condition at the beginning of treatment, how much has your condition changed?" Scores are: 1=very much improved; 2=much improved; 3=minimally improved; 4=no change; 5=minimally worse; 6=much worse; and 7=very much worse. Percentage of responders with much improved and very much improved on PGIC scale will be assessed. PGIC will be taken at home while the subject is during ON-state (with their Standard of care (SOC) for Parkinson disease).

Objective	*Endpoint*
	5. Change from baseline to Month 6 on Clinical Global Impression of Severity (CGIS) (OFF-state) The CGIS on the severity of movement impairment as assessed by the site rater. Site raters will be asked: Considering your total clinical experience with the Parkinson disease population, how ill is the patient's movement at this time? Answers were based on a 7-point scale, with 1=not assessed, 2= very mild, 3= mild, 4= moderate, 5= moderate severe, 6= severe, 7=extremely severe.
Exploratory	
Investigate change after buntanetap in potential biomarkers and improvement in cognitive functions	**1. Change from baseline to Month 6 in the plasma biomarkers measured** Potential biomarkers to be measured in plasma are Neurofilament Light (NFL), Glial fibrillary acidic protein (GFAP) and TAR DNA-binding protein 43 (TDP43). **2. Change from baseline to Month 6 in Wechsler Adult Intelligence Scale (WAIS IV) coding test (OFF-state)** In the digital symbol coding test 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. The higher score indicates better outcome. **3. Change from baseline to Month 6 in Mini-Mental State Exam (MMSE) (OFF-state)** The MMSE is a 30-point test widely used for the testing of the cognitive function, with a higher score indicating better function. A higher score indicates better outcome. **4. 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 6-month prospective, randomized, double-blinded, placebo-controlled clinical trial investigating the efficacy, safety, and tolerability of two different doses of buntanetap or placebo in patients with early PD.

Qualified subjects will be randomly assigned at a 1:1:1 ratio to one of three treatment arms: buntanetap 10 mg, buntanetap 20 mg, and placebo after a screening period of up to 42 days.

MDS-UPDRS, CGIS, WAIS IV, and MMSE will be assessed in clinical OFF-state during visits, i.e., after subjects having stopped medications of SOC Parkinson's disease 12 hours before clinical visits.

PGIC will be assessed in clinical ON-state during visits, i.e., subjects are allowed to take their SOC Parkinson medication during buntanetap treatment.

450 subjects with early PD will be randomized to 10 mg, 20 mg of buntanetap QD or placebo.

They will undergo a Screening Visit, and if they provide Informed Consent (IC) and are considered eligible per the inclusion and exclusion criteria, will proceed to participate in the treatment period. Randomized subjects will visit the clinic for the first-time dosing in clinic with administration of 10 mg or 20 mg of buntanetap or placebo, followed by an at home dosing period of 6 months, with daily administration of 10 mg or 20 mg of buntanetap or placebo.

Subjects will be required to visit clinics at 1 month, 2 months, 3 months, and 6 months (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), psychometric tests (MDS-UPDRS, WAIS, MMSE), PGIC (ON-state), and CGIS.

The Schedule of Events is shown in Appendix 1.

6 Determination of Sample Size

A sample size of up to 450 subjects (1:1:1 randomized to 10 mg, 20 mg buntanetap and placebo respectively) was determined.

Qualified subjects will be randomized at a 1:1:1 to one of the following three treatment groups: buntanetap 10 mg, 20 mg QD and placebo. A sample size of 360 (120 per treatment group) will have 90% power to detect a treatment difference at a two-sided 0.025 significance level. In the sample size calculation, it was assumed a treatment difference of 2.55 and a common standard deviation (SD) of 5.59 based on historical data of MDS-UPDRS Part II+III. Approximately 450 subjects will be randomized (approximately 150 per arm) to account for 120 subjects calculated from the power analysis plus screen failures and dropout rates.

7 Data Sets to be Analyzed

Subjects in this study must be in accordance with the criteria specified below. Subjects who do not meet all inclusion criteria, disease diagnostic criteria, or who meet any exclusion criteria may not be randomized into the study without prior approval from the Project Directors and Medical Monitor. Membership in the analysis sets will be determined in a Pre-Analysis review meeting (see Section 8.2) before unblinding.

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

- The Enrolled set (ES) will include all subjects 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 subjects who are randomized. The primary efficacy analyses will be based on the ITT.
- The Modified Intent-to-Treat set (mITT) will consist of all subjects who randomized, take at least 1 dose of the study drug, and have at least 1 post-baseline assessment. The mITT will be used to support the primary efficacy analyses.
- The Safety analysis set (SAF) will be the primary safety analysis set and will consist of all randomized subjects who receive at least 1 dose of the study drug. The safety analyses will be based on the SAF.
- The Per-protocol set (PPS) will consist of all mITT subjects who complete at least 145 days treatment exposure (Last dose date - First dose date + 1 >= 145 days) without any major protocol deviations that potentially could impact the primary efficacy endpoint. 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 subjects receiving buntanetap and with at least one sampling profile allowing derivation of at least one PK parameter. Intense PK

subgroup is only in the US. 16-people per group will go through 10hr sampling timepoints at 0, 1hr, 1.5hr, 2hr, 3hr, 4hr,5hr, 6hr, 7hr, 8hr, 9hr, and 10hr (±10 mins for all collection timepoints).

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 during the pre-analysis review when all data on protocol violations/deviations are available and 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 in a separate PK SAP will be based on the PK set.

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 Pre-Analysis Review

A pre-analysis review meeting will be held prior to database lock. During this meeting, the assignment of subjects to the analysis sets will be discussed and decided reviewing

- Screen failures with reason for failure
- Discontinued subjects with reason for discontinuation
- Protocol deviations
- Available efficacy data by visit
- Preliminary subjects' classifications of analysis sets as far as possible

8.3 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.

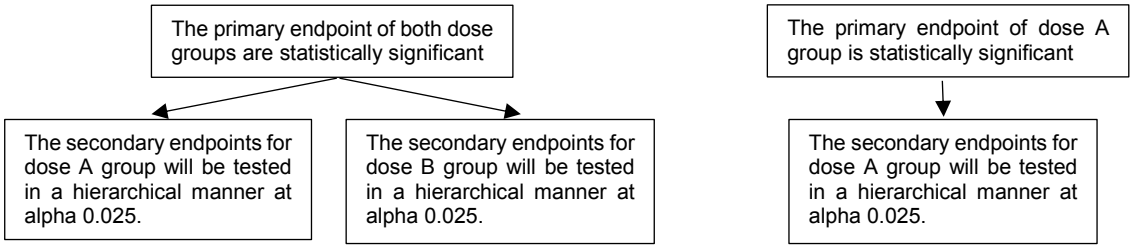
8.3.1 Level of Significance, Multiple Comparisons and Multiplicity

There are two active treatment groups (10 mg and 20 mg) in the study. Each of the two active treatments will be compared to the control (two hypothesis tests for the primary endpoint). The comparison of each of the two treatment groups against control for the primary endpoint will be performed using the Hochberg step up procedure, where if the largest p-value is <0.05, then the null hypothesis will be rejected for both groups. If the largest p-value is >0.05, then the critical value for the second test will be at two-sided 0.025.

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. Change from baseline to Month 6 in the sum of MDS-UPDRS Part II+III (OFF-state)
2. Change from baseline to Month 6 in MDS-UPDRS Part III (OFF-state)
3. Change from baseline to Month 6 in MDS-UPDRS Total score (OFF-state)
4. Percentage of Responders with "Much Improved" or "Very Much Improved" on PGIC (ON-state)
5. Change from baseline to Month 6 on CGIS (OFF-state)

If the primary endpoint of both dose groups are statistically significant, the secondary endpoints for each dose group will be tested in a hierarchical manner at alpha 0.025. If the primary endpoint of a specific dose group is statistically significant, the secondary endpoints for that dose group will be tested in a hierarchical manner at alpha 0.025. If the primary endpoint of both dose groups are not statistically significant, the hypothesis testing for all subsequent endpoints will be considered exploratory. Two flowcharts are also presented in the following.



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

8.3.2 Definitions

MMSE – OFF-state	Mini-Mental State Examination is a 30-point test, widely used test of cognitive function.
PGIC – ON-state	Participant Global Impression of Change is the participant-reported outcome. The qualitative assessment of meaningful change will be determined by the participant in response to the question, "Compared to your condition at the beginning of treatment, how much has your condition changed?" Scores are 1= very much improved; 2= much improved; 3= minimally improved; 4= no change; 5= minimally worse; 6= much worse; 7= very much worse
CGIS – OFF-state	Clinical Global Impression of Severity is a scale on the severity of movement impairment as assessed by the site rater. Site raters will be asked: Considering your total clinical experience with the Parkinson disease population, how ill is the patient's movement at this time? Answers were based on a 7-point scale, with 1= not assessed, 2= very mild, 3= mild, 4= moderate, 5= moderate severe, 6= severe, 7=extremely severe
WAIS-IV – OFF-state	In the digital symbol Wechsler Adult Intelligence Scales, 4th edition, coding test 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.

MDS-UPDRS – OFF-state	Movement Disorder Society-Unified Parkinson's Disease Rating Scale is a 50-item rating scale designed to assess Parkinson's disease-related disability and impairment comprising the four parts: Part I - Non-Motor Aspects of Experiences of Daily Living Part II - Motor Aspects of Experiences of Daily Living Part III - Motor examination Part IV - Motor complications
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

8.3.3 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 efficacy primary 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) 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). 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 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 (Month 1, Month 2, Month 3, Month 6).

Treatment comparisons (10 mg vs. Placebo / 20 mg 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 (10 mg – Placebo / 20 mg – Placebo) will be estimated and resulting 2-sided p-values and associated 95% CI will be presented.

8.3.4 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.

8.3.5 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.

8.3.6 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 (40 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)
    - Ethnicity (Hispanic or Latino, Not Hispanic or Latino, Not Reported, Unknown)

- Substance use – tobacco and alcohol
    - Smoker (Never, Current or Former)
    - Alcohol user (Never, Current or Former)
- Time since diagnosis of Parkinson's disease

8.3.7 Primary Analysis

The primary efficacy endpoint in this clinical trial is the change from baseline to Month 6 in the score for MDS-UPDRS part II (OFF-state) between each of the two regimens of buntanetap and Placebo.

The statistical hypothesis to be tested for primary efficacy is described as:

- Null Hypothesis (H0):

  The difference in change in MDS-UPDRS Part II (OFF-state) scores from Baseline to Month 6, between treatment (active vs placebo) regimen = 0.

- Alternative Hypothesis (H1):

  The difference in change in MDS-UPDRS Part II (OFF-state) scores from Baseline to Month 6, between treatment (active vs placebo) regimen ≠ 0.

The primary endpoint will be analyzed via MMRM. The model will include treatment, timepoint, treatment-by-timepoint interaction as the fixed effects and baseline as the covariate. The variance-covariance matrix will be assumed to be unstructured.

This analysis will be conducted after all subjects have reached Month 6 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.

8.3.8 Secondary Analysis

Secondary efficacy endpoints in this clinical trial are

- Change from baseline to Month 6 in the sum of the score of MDS-UPDRS Parts II+III – OFF-state
- Change from baseline to Month 6 in the score of MDS-UPDRS Part III – OFF-state
- Change from baseline to Month 6 in the total score for MDS-UPDRS – OFF-state
- Percentage of responders with "Much improved" or "Very much improved" on PGIC – ON-state
- Change from baseline to Month 6 on CGIS – OFF-state.

Change from baseline to Month 6 in the sum of MDS-UPDRS Parts II+III – OFF-state

The MDS-UPDRS is a 50-item rating scale designed to assess Parkinson's disease-related disability and impairment. The scale comprises four parts: Part I evaluates mentation, behavior, and mood symptoms; Part II evaluates ADL; Part III evaluates motor function; and Part IV evaluates complications of dopaminergic therapy. Part II contains 13 questions which are to be completed by the participant (range 0-52). Part III contains 33 scores based on 18 items (Range 0-132). MDS-UPDRS Part II and III combined score equals the sum of Parts II and III (Range 0-184). A higher score indicates more severe symptoms of PD.

MDS-UPDRS Parts II+III will be analyzed by MMRM, similar to the primary efficacy endpoint.

Change from baseline to Month 6 in the MDS-UPDRS Part III – OFF-state

Part III contains 33 scores based on 18 items (Range 0-132).

MDS-UPDRS Part III will be analyzed by MMRM, similar to the primary efficacy endpoint.

Change from baseline to Month 6 in the MDS-UPDRS Total score – OFF-state

The UPDRS is a 50-item rating scale designed to assess Parkinson's disease-related disability and impairment. The scale comprises four parts:

Part I evaluates mentation, behavior, and mood symptoms;

Part II evaluates ADL;

Part III evaluates motor function; and

Part IV evaluates complications of dopaminergic therapy.

The MDS-UPDRS Total score is the sum of the subscale scores for Parts I to III and ranges from 0 to 236, with higher scores reflecting greater severity. MDS-UPDRS will be assessed by the investigator at baseline and each visit at the clinic (Month 1, Month 2, Month 3, and Month 6).

MDS-UPDRS Total score will be analyzed by MMRM, similar to the primary efficacy endpoint.

Percentage of responders with “Much improved” or “Very much improved” on PGIC – ON-state

The PGIC is the participant-reported outcome. The qualitative assessment of meaningful change will be determined by the participant in response to the question, "Compared to your condition at the beginning of treatment, how much has your condition changed?" Scores are

1= very much improved

2= much improved

3= minimally improved

4= no change

5= minimally worse

6= much worse

7= very much worse

The assessment is done during visits at the clinic (Month 1, Month 2, Month 3, and Month 6) and ON-state i.e., subjects were allowed to treat the PD with SOC medication.

For PGIC, responders are defined as responses with score 1 (very much improved) and score 2 (much improved). PGIC will be analyzed for responders (scores 1 and 2) using logistic regression method.

Descriptive statistics for each visit will be produced for the PGIC.

Change from baseline to Month 6 on CGIS – OFF-state

The clinical global impressions scale on the severity of movement impairment as assessed by the site rater. Site raters will be asked: Considering your total clinical experience with the Parkinson disease population, how ill is the patient’s movement at this time? Answers were based on a 7-point scale, with 1= not assessed, 2= very mild, 3= mild, 4= moderate, 5= moderate severe, 6= severe, 7= extremely severe.

The assessment is done at baseline visit and during visits at the clinic (Month 1, Month 2, Month 3, and Month 6) and OFF-state i.e., 12 hours after last intake of medication of standard of care for Parkinson’s disease.

For CGIS, responders are defined as a decrease in CGIS score from baseline to Month 6. CGIS will be analyzed for responders using logistic regression method.

Descriptive statistics and change from baseline for each visit will be produced for the CGIS.

8.3.9 Exploratory Analysis

Exploratory endpoints in this clinical trial are

- Change from baseline to Month 6 in the plasma biomarkers measured
- Change from baseline to Month 6 in the WAIS coding test – OFF-state
- Change from baseline to Month 6 in MMSE – OFF-state
- PK analysis.

Changes from baseline to Month 6 in the plasma biomarkers measured

Potential biomarkers to be measured in plasma are Neurofilament Light (NFL), Glial fibrillary acidic protein (GFAP) and TDP43.

Change from baseline to Month 6 in WAIS IV coding test – OFF-state

In the digital symbol coding test 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. The assessment is done by the investigator at baseline, Month 2, and Month 6 and OFF-state i.e., 12 hours after last intake of medication of SOC Parkinson's disease.

Descriptive statistics and change from baseline for each visit will be produced for the WAIS coding test.

Change from baseline to Month 6 in MMSE - OFF-state.

The MMSE is a 30-point test, widely used test of cognitive function. The MMSE will be assessed by the investigator at screening, baseline, and Month 6. The MMSE comprises sections on orientation to place, orientation to time, registration, attention and calculation, recall, naming, repetition, comprehension, reading, writing, and drawing. The assessment is done by the investigator at screening, at baseline and at end-of-treatment visit and OFF-state i.e., 12 hours after last intake of medication of SOC Parkinson's disease.

Descriptive statistics and change from baseline for each visit will be produced for the MMSE.

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 methods. 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.

8.3.10 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 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.

8.3.11 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, 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).

8.3.11.1 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.

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 Project Directors and 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

8.3.11.2 Permitted concomitant medication

Per protocol concomitant treatment with anti-parkinsonian medications at stable doses for 4 weeks or greater prior to screening are allowed.

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

- tricyclic antidepressants
- antipsychotics prescribed for any reason, except ≤50 mg quetiapine daily, risperidone ≤1.5 mg/day, olanzapine ≤5 mg/day, and aripiprazole ≤10 mg/day
- psychostimulants

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

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.

The 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 action and currently available data for these vaccine candidates and for buntanetap, it is not believed that there are specific risks to consider in subjects 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.

8.3.11.3 Coding and imputation of incomplete dates

All concomitant medications/therapies will be classified according to ATC level 3 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 the 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.

8.3.12 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 Subject 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 severity according to Rheumatology Common Toxicity Criteria (RCTC) categories
- TEAEs leading to treatment discontinuation of study drug
- TEAEs leading to withdrawal from study
- TEAEs leading to death

The overall summary presentation will be repeated for:

- Severity categories
- TEAEs related to study drug

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.

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

Deaths, other serious AEs (SAEs), and other significant AEs will be listed, these include the following TEAEs:

- SAEs leading to death
- SAEs not leading to death
- AEs for which action was taken with study drug
- AEs leading to withdrawal from study

This listing will be divided into sub-listings with criteria as a subheading. An AE can appear in more than one sub-listing.

AE listings will include the relative time/day in relation to date and time of first administration of study drug and duration of AEs. If AE start time and time of first study drug administration is available, relative time and duration will be given as HH:MM for AEs with onset/duration less than 24 hours after first administration of study drug and otherwise as days. For AEs with incomplete dates no relative day or duration will be calculated.

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.

8.4 Other Safety Assessment

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.

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.

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 (absolute and percentage): basophils, eosinophils, lymphocytes, monocytes, neutrophils 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).

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.

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.

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. 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 subject’s lifetime, during the double-blind treatment period, and during the safety follow-up period will also be presented by 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.

Supportive listings will be provided and will include the subject 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 subjects who have suicidal ideation or suicidal behavior will also be provided.

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. 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 presented.

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.

8.5 Plasma Biomarkers and Pharmacokinetic Analysis

Biomarkers, including 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).

8.6 Multicenter Studies

No per-center analyses are planned.

8.7 Adjustment for Covariates

Baseline value will be used as a covariate in the MMRM analyses for the primary efficacy endpoint (see Section 8.3.7).

8.8 Examination of Subgroups

No subgroup analyses are planned for this study.

8.9 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.

The MMRM analysis of the primary endpoint 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 primary endpoint analysis. 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. The variables for the imputation model will include MDS-UPDRS Part II score at baseline, Month 1, Month 2, Month 3, and Month 6. Approximately 50 datasets will be imputed and the same MMRM analysis as the one applied to the primary analysis will be fitted. The ROUND, MINIMUM and MAXIMUM options will be utilized to ensure imputed values are clinically plausible integers (0 – 52). 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. A random seed number will be selected for the multiple imputation procedure and will be retained.

8.10 Interim Analysis

An interim sample size review will be performed after in total n_1 = 120 evaluable subjects (40 subjects per treatment group) have primary endpoint data, i.e., MDS-UPDRS part II+III at Month 2. Evaluable subjects are defined as randomized and having received at least one dose of the study drug. The aim of the interim sample size review is to reassess the size of the estimated mean and SD of the primary effectiveness variable (MDS-UPDRS part II+III). The sample size will be adjusted accordingly (see following section).

No interim analyses are planned for the purpose of stopping the study early for futility.

8.10.1 Re-estimation of Sample Size

The initial sample size estimation provides statistical power to demonstrate effectiveness among subjects treated with buntanetap. A potential sample size adjustment is to assure that this trial will have the intended precision. The decision of sample size adjustment will be made based upon the estimates of MDS-UPDRS part II+III at baseline and after two months of treatment calculated for the first 120 evaluable subjects with primary endpoint data. Sample size will not be decreased regardless of the results of this evaluation. To warrant the originally intended >90% power of the study at the 0.025 alpha level (2-sided), the evaluable sample size may be increased from n_2 = 360 (120 per treatment group) to a maximum of n_{max}=540 (180 per treatment group). Additional subjects will be enrolled to account for screen failures and drop-outs. The sample size re-estimation will be evaluated for the efficacy endpoint based on the observed interim mean and SD of treatment difference of the MDS-UPDRS part II+III score.

The sample size re-estimation method is based on evaluation of conditional power in relationship to pre-specified decision rules defined by ranges of attainable conditional power values. Only an increase in sample size is possible under this approach when observed conditional power falls within the 'promising zone' as described below (Mehta and Pocock 2011).

The conditional power at the interim analysis is defined by equation (6) from Mehta and Pocock (2011).

The targeted conditional power at the first interim review is 90%, however we pre-specify a range of conditional power values below 90% that would deem our interim results promising and warrant sample size re-estimation. Table 1 provides cut-off values from Mehta and Pocock (2011) for the lower bound of the promising zone, CP_{min}, under some typical two-stage adaptive designs. CP_{min} is 0.42 (42%) for the parameter choices (n_{max}/n_2 = 1.5, n_1/n_2 = 0.33) in this study.

Table 1. CP_{min} cut-off values for some typical two-stage adaptive designs with no early stopping either for efficacy or futility

Sample size ratios		**CP_{min} values for targeted conditional power**	
Maximum allowed (n_{max} / n_2)	**At interim look (n_1 / n_2)**	**80 per cent**	**90 per cent**
1.5	**0.25**	0.42	**0.42**
1.5	0.5	0.41	0.41
1.5	0.75	0.38	0.38
2	0.25	0.37	0.37
2	0.5	0.36	0.36
2	0.75	0.33	0.33
3	0.25	0.32	0.32

3	0.5	0.31	0.31
3	0.75	0.30	0.27
∞	0.25	0.32	0.28
∞	0.5	0.31	0.27
∞	0.75	0.30	0.25

CP: Conditional Power

Table 2 summarizes the sample size re-estimation decision based on the conditional power observed at interim analysis. The sample size is re-estimated when the interim conditional power falls within the promising zone of 0.42 to 0.90 (42% to 90%).

Table 2. Sample Size Re-Estimation Decision

Conditional power	Decision
Less than 42%	"Unfavorable zone" – No change to sample size
42% to 90%	"Promising zone" – Increase sample size
At least 90%	"Favorable zone" – No change to sample size

The re-estimated cumulative final sample size (n_{2*}) is computed using equations (7-9) from Mehta and Pocock (2011) and will be increased by a factor based on a re-assessed attrition rate as per the initial sample size calculation.

8.10.2 Clarifications Based on Food and Drug Administration (FDA)'s Feedback

The interim analysis for sample size re-estimation has been completed in March 2023. Based on the interim analysis results and sample size re-estimation decision rules, the results of the interim analysis suggested to increase the sample size to 180 randomized subjects per arm. As of June 9, 2023, the study has completed enrollment with a total sample size of 522 (174 per arm) subjects.

The original evaluable sample size was 120 per arm, and a 20% drop out rate was assumed, which rendered to a sample size of 150 randomized subjects per arm (120/0.8=150). Based on the blinded data review, we noticed a low dropout rate of approximately 6%. With this dropout rate, 180 per arm randomized subjects will lead to 170 (180*0.94) evaluable subjects per arm, which is around 42% increase from originally calculated sample size of 120 per arm (170/120=1.42). The sample size increased was considered reasonably small (42%) based on the Metha and Pocock (2011) paper that the overall Type I error rate will not be inflated.

The sample size re-estimation for our 3-arm study was handled by the worst-case scenario rules in Table 3 below. The sample size re-estimation method evaluates the conditional power in relationship to a pre-specified decision rule defined by ranges of attainable conditional power values. Only an increase in sample size was possible under this approach if observed conditional power was within the Promising Zone.

Table 3. Worst-case Scenario Approach to Sample Size Re-estimation for Multiple Dose Comparisons

Outcome	Worst-case Scenario Approach
Both doses Unfavorable	Study proceeds with originally planned sample size
One dose Unfavorable One dose Promising	Study proceeds with sample size re-estimation increase suggested by one dose comparison in Promising Zone
One dose Unfavorable One dose Favorable	Study proceeds with originally planned sample size
Both doses Promising	Study proceeds with larger of the sample size re-estimation increases suggested by two dose comparisons in Promising Zone
One dose Promising One dose Favorable	Study proceeds with sample size re-estimation increase suggested by one dose comparison in Promising Zone
Both doses Favorable	Study proceeds with originally planned sample size

According to Chen, Yuan and Li (2018), the bias and error rate using the original maximum likelihood estimator (MLE) point estimator and confidence interval should be small and appropriate to use in the final analysis. In addition, the interim analysis was done using the original primary endpoint MDS-UPDRS Part II+III. Based on the Agency's feedback, the original primary endpoint is changed to MDS-

UPDRS Part II, and MDS-UPDRS Part II+III is kept as a secondary endpoint. Thus, the results of interim analysis are not relevant to the new primary endpoint. With a sample size of 174 per arm, the study will have 90% power to detect a treatment difference of 3 assuming a common SD of 8, based on a 2-sided t-test with alpha of 0.025.

8.11 Data Monitoring

An independent DSMB (IDMB) will be convened to review the safety information from the study on an ongoing basis. The IDMB, 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 IDMB will initially be provided with data blinded to treatment status, but they may request unblinded data if there is a safety concern.

Additionally, the IDMB 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 IDMB may at any time request additional information from the Coordinating Center.

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

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

Based on the review of safety data, the IDMB 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 IDMB, 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 IDMB charter.

8.12 Reporting Conventions

Any p-values $\geq$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.

8.13 References

1. Mehta, C. R., & Pocock, S. J. (2011). Adaptive increase in sample size when interim results are promising: a practical guide with examples. Statistics in medicine, 30(28), 3267-3284.
2. Chen YHJ, Yuan SS, Li X. Statistical inference following sample size adjustment based on the 50%-conditional-power principle. J Biopharm Stat. 2018; 28(3): 575-587

APPENDIX 1: SCHEDULE OF EVENTS

Study Period	**Screening/Baseline Up to 42 Days**		**Double-Blind Trial Treatment Period 6 Months[e]**						**Unscheduled Visit [h] (USV)**	**Early Discontinuation Visit**
Visit Timing	**Day -42 to 0**		**1 month ± 7 days**	**2 months ± 7 days**	**3 months ± 7 days**	**4.5 months ± 7 days**	**1-3 days prior to end of trial visit**	**6 months ± 7 days (end of trial visit) [J]**		
Procedure	**Screening[I] (up to 42 days before Day 0)**	**Base line [J] (Day 0)**	**Clinic Visit**	**Clinic Visit**	**Clinic Visit**	**Phone**	**Phone**	**Clinic Visit**	**Clinic Visit**	**Clinic Visit**
Informed consent	**X**									
Inclusion and exclusion criteria	**X**	**X**								
Demography information	**X**									
Full physical examination and neurological examination	**X**	**X**	**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**	**X**	**X**
Hoehn & Yahr Stage	**X**									

Study Period	**Screening/Baseline Up to 42 Days**		**Double-Blind Trial Treatment Period 6 Months[e]**						**Unscheduled Visit [h] (USV)**	**Early Discontinuation Visit**
Visit Timing	**Day -42 to 0**		**1 month ± 7 days**	**2 months ± 7 days**	**3 months ± 7 days**	**4.5 months ± 7 days**	**1-3 days prior to end of trial visit**	**6 months ± 7 days (end of trial visit) [J]**		
Procedure	**Screening[I] (up to 42 days before Day 0)**	**Base line [J] (Day 0)**	**Clinic Visit**	**Clinic Visit**	**Clinic Visit**	**Phone**	**Phone**	**Clinic Visit**	**Clinic Visit**	**Clinic Visit**
MMSE	X	X						X		
PGIC			X	X	X			X		
CGIS		X	X	X	X			X		
WAIS-IV		X		X				X		
MDS-UPDRS		X	X	X	X			X		
C-SSRS	X	X	X	X	X	X	X	X	X	X
Clinical Laboratory safety tests[d&g]	X	X	X	X	X			X	X	X
12-lead ECG	X	X	X	X	X			X	X	X
Vital signs[a]	X	X	X	X	X			X	X	X
Randomization		X								
Study intervention dispensed		X	X	X	X			X[f]		
Study intervention compliance review			X	X	X	X	X	X		X
AE/SAE review	X	X	X	X	X	X	X	X	X	X
Concomitant medication review	X	X	X	X	X	X	X	X	X	X

Study Period	**Screening/Baseline Up to 42 Days**		**Double-Blind Trial Treatment Period 6 Months[e]**						**Unscheduled Visit [h] (USV)**	**Early Discontinuation Visit**
Visit Timing	**Day -42 to 0**		**1 month ± 7 days**	**2 months ± 7 days**	**3 months ± 7 days**	**4.5 months ± 7 days**	**1-3 days prior to end of trial visit**	**6 months ± 7 days (end of trial visit) [J]**		
Procedure	**Screening[I] (up to 42 days before Day 0)**	**Base line [J] (Day 0)**	**Clinic Visit**	**Clinic Visit**	**Clinic Visit**	**Phone**	**Phone**	**Clinic Visit**	**Clinic Visit**	**Clinic Visit**
PK sampling[c]		**X**						**X**		
Sampling for blood biomarkers		**X**						**X**		
24-hour phone follow-up[b]		**X**	**X**	**X**	**X**			**X**	**X**	**X**

a=Vital signs include sitting blood pressure, pulse, temperature, respiration rate.
b=Phone follow-up should occur approximately 24 hours after baseline, each clinic visit and after discharge from the end of trial or early discontinuation visits.
c=Sampling Times: 0 (pre-dosing) and 1.5hrs after dosing, except for 16-people per group who will go through 10hr sampling timepoints at 0, 1hr, 1.5hr, 2hr, 3hr, 4hr,5hr, 6hr, 7hr, 8hr, 9hr, and 10hr (±10 mins for all collection timepoints). Intense PK subgroup is only in the US.
d=Safety labs at all clinic visits should be taken fasted.
e=Subjects who drop out after the initial 5 months of study treatment will complete the full End of Trial visit. Subjects who dropped out prior to the initial 5 months of study treatment will complete the Early Discontinuation Visit assessments.
f=The study intervention dispensing on the end of trial visit is only for the one pill given on that day.
g=Blood and urine samples will be obtained for clinical safety lab assessments as described in the Laboratory Manual.
h= Investigator will determine the extra assessment if any is needed.
i= With investigator's permission, lab re-testing is allowed within a month.
j= For the intense PK subgroup participants, baseline visit and end-of-trial visit can be done in 2 days.