NCT03817684

Statistical Analysis Plan

Sponsor Name: Tetra Discovery Partners, Inc.

Protocol Number: BPN14770-CNS-201

Protocol Title: A Randomized, Double-blind, Placebo-controlled, 3-Arm Parallel Design Study to Evaluate the Effects of BPN14770 in Patients with Early Stage Alzheimer's Disease

Protocol Version and Date: Amendment 3.0, 24 September 2019

Syneos Health Project Code: 7001704

Author: [redacted], Biostatistician

Revision History

Version #	Date (DD-Mmm-YYYY)	Document Owner	Revision Summary
0.1	07-May-2019	[redacted]	Initial Release Version
0.2	01-Nov-2019	[redacted]	Updated per Tetra review comments on first draft SAP
1.0	09-Dec-2019	[redacted]	Updated per Tetra review comments on second draft SAP
1.1	14-Jan-2020	[redacted]	Add section 8.3 Multiple Testing Strategy
2.0	16-Jan-2020	[redacted]	Updated per Tetra review comments and modified analysis for CGI-I
3.0	02-Mar-2020	[redacted]	A number of clarifications were made to the text to ensure clarity of methods. Updated primary analysis to accommodate the study design and the repeated measures collected for each endpoint. Added detail regarding each efficacy endpoint, eg, what the scale is for each endpoint, interpretation (eg, are lower values 'better' or 'worse'), how missing individual questions/items for each scale affect derivation of the total and/or domain scores. For safety, additional specification was added on assessment of laboratory shifts (low/normal/high) as well as ECG assessment of QTc. Patient [redacted] was the same subject, randomized at 3 different treatment centers. Due to the multiple treatments and overlapping treatment periods, this patient (and all 3 patient numbers associated with the patient) is excluded from the Safety, mITT, PP, Completers, and PK populations. This patient will be reported in the CSR via a detailed patient narrative.

I confirm that I have reviewed this document and agree with the content.

Approvals
Syneos Health Approval
[redacted] Lead Biostatistician
[redacted] Senior Reviewing Biostatistician
Tetra Discovery Partners, Inc. Approval
[redacted] Clinical Project Manager Sponsor Contact
[redacted] Consulting Statistician Sponsor Contact
[redacted] Sponsor Contact

Table of Contents

Revision History 2
Approvals 3
1. Glossary of Abbreviations 7
2. Purpose 9
   2.1. Responsibilities 9
   2.2. Timings of Analyses 9
3. Study Introduction 10
   3.1. Study Objective 10
   3.2. Brief Description 10
   3.3. Subject Selection 10
   3.4. Determination of Sample Size 11
   3.5. Treatment Assignment & Blinding 12
   3.6. Administration of Study Medication 12
   3.7. Study Procedures and Flowchart 13
4. Endpoints 15
   4.1. Primary Efficacy Endpoint 15
   4.2. Secondary Efficacy Endpoints 15
   4.3. Safety Endpoints 15
5. Analysis Populations 16
   5.1. Screened Population 16
   5.2. Randomized Population 16
   5.3. Safety Population 16
   5.4. Modified Intent-to-Treat Population 16
   5.5. Per Protocol Population 16
   5.6. PK Population 16
   5.7. Completers Population 16
   5.8. Protocol Deviations 17
6. General Aspects for Statistical Analysis 18
   6.1. General Methods 18
   6.2. Key Definitions 18

6.3. Missing Data 19
6.4. Visit Windows 19
6.5. Pooling of Centers 19
6.6. Subgroups 19
6.7. Data listings 19
7. Demographic, Other Baseline Characteristics and Medication 20
7.1. Subject Disposition and Withdrawals 20
7.2. Demographic and Other Baseline Characteristics 20
7.3. Medical/Surgical History 21
7.4. Medication 21
8. Efficacy 23
8.1. Primary Efficacy Endpoint and Analysis – RBANS DMI 23
8.1.1. Sensitivity Analysis using Multiple Imputation (MI) Approach for the Primary Efficacy Endpoint of RBANS DMI 24
8.1.2. Additional Sensitivity Analysis for the Primary Efficacy Endpoint 26
8.2. Secondary Efficacy Endpoints and Analyses 26
8.2.1. RBANS total Score and other RBANS index scores 27
8.2.2. ADCS-ADL Total Score 27
8.2.3. Mini-Mental State Evaluation (MMSE) Score 28
8.2.4. Clinical Dementia Rating Sum of Boxes (CDR-SB) Score 28
8.2.5. Clinical Global Impression- Improvement (CGI-I) Score 29
8.2.6. Composite Endpoint 30
8.3. Multiplicity and Experiment-wise Alpha Control Considerations 30
9. Analysis Of Pharmacokinetics 32
10. Analysis Of Biomarker Data 33
11. Safety 34
11.1. Extent of Exposure 34
11.2. Treatment Compliance 34
11.3. Adverse Events / Adverse Drug Reactions 35
11.4. Laboratory Evaluations 36
11.5. Vital Signs 37
11.6. ECG 37

11.7. Physical Examination 37
11.8. Columbia Suicide Severity Rating Scale (C-SSRS) 37
12. Changes from Analysis Planned in Protocol 40
13. Reference List 41
14. Programming Considerations 42
14.1. General Considerations 42
14.2. Table, Listing, and Figure Format 42
14.2.1. General 42
14.2.2. Headers 42
14.2.3. Display Titles 43
14.2.4. Body of the Data Display 43
14.2.5. Footnotes 44
15. Quality Control 46

1. Glossary of Abbreviations

Abbreviation	Description
AD	Alzheimer's disease
ADCS-ADL	Alzheimer's Disease Cooperative Study Activities of Daily Living Inventory
AE	Adverse event
ANCOVA	Analysis of covariance
ApoE4	Apolipoprotein E4
ATC	Anatomical Therapeutic Chemical
BMI	Body Mass index
CDR	Clinical Dementia Rating
CDR-SB	Clinical Dementia Rating – Sum of Boxes
CGI-I	Clinical Global Impression – Improvement
CI	Confidence Interval
CP	Completers
C-SSRS	Columbia-Suicide Severity Rating Scale
CTMS	Clinical Trial Management System
eCRF	Electronic case report form
ECG	Electrocardiogram
GLM	Generalized Linear Models
ICH	International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
ITT	Intent-to-Treat
LSMeans	Least Square Mean
MMRM	Mixed model for repeated measures
MAR	Missing at Random
MNAR	Missing Not at Random
Max	Maximum
MCI	Mild Cognitive Impairment
MedDRA	Medical Dictionary for Regulatory Activities
MHIS	Modified Hachinski Ischemia Score
Min	Minimum

Abbreviation	Description
MMSE	Mini Mental State Examination
N/A	Not Applicable
PP	Per Protocol
PT	Preferred Term
QTc	Corrected QT Interval
RBANS (DMI)	Repeatable Battery for the Assessment of Neurological Status (Delayed Memory Index)
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SD	Standard deviation
SE	Standard Error
SI	Standard International System of Units
SMD	Standardized Mean Difference
SOC	System Organ Class
SOP	Standard operating procedure
TEAE	Treatment-eEmergent adverse event
TFL	Table, Figure, and Listing
WHO	World Health Organization

2. Purpose

The purpose of this statistical analysis plan (SAP) is to ensure that the data listings, summary tables, and figures that will be produced, and the statistical methodologies that will be used, are complete and appropriate to allow valid conclusions regarding the study objectives.

2.1. Responsibilities

Syneos Health will perform the statistical analyses and are responsible for the production and quality control of all tables, figures, and listings.

2.2. Timings of Analyses

The final analysis is planned after all subjects complete the final study visit or terminate early from the main phase of the study and the data processing has been completed to lock the study database. Prior to unblinding the treatment allocations, the per-protocol population will be identified and documented.

Unless otherwise specified, the analysis includes all data collected in the database through the time of the database lock.

Note, there are no planned interim analyses in this study.

3. Study Introduction

3.1. Study Objective

In subjects with a clinical diagnosis of early stage Alzheimer's disease (AD) receiving cholinesterase inhibitor therapy:

- To evaluate the efficacy of two dose levels (10 mg bid and 25 mg bid) of BPN14770 as compared to placebo over a 13-week treatment period
- To evaluate the safety and tolerability of BPN14770 10 and 25 mg bid as compared to placebo over a 13-week treatment period
- To obtain pharmacokinetic data on BPN14770 in this subject population.

The outcomes from this study will be used to determine the dose(s) to be used in subsequent Phase 3 studies.

3.2. Brief Description

This is a Phase 2, randomized, double-blind, placebo-controlled, 3-arm, parallel-design study to evaluate the effects of BPN14770 in approximately 255 subjects with a clinical diagnosis of early stage AD. The study will consist of a Screening period of 28 days or less. The Screening period will be followed by a 13-week double-blind treatment period, and a final follow-up visit for safety one week after treatment is concluded.

Randomization will be stratified based on presence or absence of an ApoE4 allele in the subject's genetic profile. Within each stratum, subjects will be randomized in a blinded, balanced fashion to receive either BPN14770 10mg bid, BPN14770 25 mg bid, or matching placebo. A target of 80 evaluable subjects per group (240 total) is planned (see the sample size determination in Section 3.4). Approximately 85 subjects per group will be enrolled (255 total) in order to account for potential dropouts. Subjects will have a total of eight clinic visits: Screening; Baseline (Day 1); Weeks 1, 2, 4, 8, 13, and the final Follow-up Visit. The screening visit will consist of consenting the subject, obtaining a medical history, full physical exam, height, weight, BMI, ECG, serology and drug testing, chemistry, hematology, urinalysis, ApoE4 testing, pregnancy testing, and cognitive assessments. If the subject fully qualifies for the study, randomization will occur at the Baseline visit following verification of continued eligibitly, at which time the subject will receive study medication.

Cognitive testing will occur at each visit, Week 1 through Week 13, but notat the final Follow-up Visit (note, not all cognitive assessments are performed at each visit). Safety assessments will also be performed at each visit, to include adverse event monitoring, vital signs, ECG, and chemistry and hematology. Urinalysis will also be performed at Screening and Baseline and Weeks 1, 4 and 13. Lipid testing will be performed at Baseline and Week 13. Risk of suicidality will be assessed at Screening, Baseline and Weeks 1, 4, and 13.

Pharmacokinetic sampling will be performed at the clinic visits at Weeks 1, 4, 8 and 13.

3.3. Subject Selection

Individuals are eligible for the study if they meet all of the inclusion and none of the exclusion criteria. The inclusion/ exclusion criteria will be assessed during Screening, which will be up to 28 days prior to first

study drug administration. Continued subject eligibility will be verified on Baseline at Baseline, prior to randomization. Specifically, subjects must meet the following AD criteria:

- Clinical Dementia Rating (CDR) score of 0.5 or 1 (on a scale from 0 to 3, with higher scores indicating worse outcome), with Memory Box score of 0.5 or greater.
- MMSE score of 20 or greater (range 0 to 30, with lower scores indicating worse outcome)
- RBANS DMI score < 85 (range 40 to 160, with lower scores indicating worse outcome)

See Sections 6.2 and 6.3 in the study protocol for the complete list of all eligibility criteria.

3.4. Determination of Sample Size

The BPN14770-CNS-201 study will enroll 255 subjects (85 per treatment arm), in anticipation of evaluating at least 240 subjects (80 per treatment). The Primary Efficacy Outcome measure will be change from baseline at Week 13 for the Repeatable Battery for the Assessment of Neurological Status - Delayed Memory Index (RBANS DMI).

The normalized mean of the RBANS DMI is 100, with an SD of 15, and the test-retest reliability is 0.80 with no practice effects using alternate versions of the test [Randolph 1998]. RBANS DMI was used in a recent natural history study of Mild Cognitive Impairment (MCI) due to probable AD [Karantzoulis 2013]. In the large amnestic MCI sample (81 subjects), the mean was 73.4 with a SD of 16.5 in a normal distribution. Performance of healthy, age-matched controls was a mean of 101.4 with a SD of 14.4 (81 subjects) yielding estimates of the standard deviation for the DMI from 14.4-16.5. Therefore, the power analysis is based on the assumption that the standard deviation of the response variable is 16.

The power of the proposed study is driven by the assumed difference between BPN14770 25mg (higher dose) and placebo. Based on randomization of 80 subjects per group into the parallel-design study, the probability is 80 percent that the study will detect a treatment difference at a two-sided 0.05 significance level, if the true difference between BPN14770 25 mg and placebo is 7.15 units. Using the standardized mean difference (SMD) to estimate effect size [Faraone 2008], a change of 7.15 units on the DMI score is an effect size of 0.45.

Table 1 provides sample size estimates for 80 percent and 90 percent probability that the study will detect a treatment difference at a two-sided 0.05 significance level assuming the standard deviation of the DMI is 16.

Table 1 Clinical Sample Size (1:1 Active + Placebo) Based on Effect Size (Change From Baseline) And Desired Power

	Number of Subjects Per Group Based on Desired Effect Size (SMD)					
Standardized Mean Difference (SMD)	0.3	0.35	0.4	0.45	0.5	0.75
RBANS DMI change (M1-M2)	4.8	5.6	6.4	7.2	8	12
Number of Subjects (Power 0.8)	176	130	100	79	64	29
Number of Subjects (Power 0.9)	235	173	133	105	86	39

For comparison, a donepezil study reported by Seltzer et al., 2004, randomized a total of 153 subjects (2:1 active vs placebo) and reported treatment benefit at a significance level of 0.03 after 12 weeks of

treatment [Seltzer 2004]. In a larger study of mild-to-moderate AD subjects, the effect size of donepezil reported by Rogers et al. 1998 was 0.4 [Rogers 1998],so the assumption of an effect size of 0.45 for BPN14770 in a milder, early AD population, using a sensitive outcome measure, is reasonable.

Therefore, based on clinical trials of donepezil in a similar population, the proposed sample size will provide adequate power to detect a clinically and statistically significant difference of BPN14770 treatment in early AD.

3.5. Treatment Assignment & Blinding

In this randomized, double-blind, placebo controlled, 3-arm, parallel group study, 255 subjects will be assigned to receive one of three treatments via balanced randomization (1:1:1). Randomization will occur separately within each ApoE4 stratum (presence/absence of ApoE4 isoform) to ensure similar numbers of ApoE4 positive subjects are present in each treatment group. The three treatments are:

A. BPN14770 10mg, twice-daily
B. BPN14770 25mg, twice-daily
C. Matched placebo, twice daily

A randomization schedule will be generated by a statistician unassociated with the study execution prior to the start of the study. Subjects will be randomized on Day 1 and receive their bottle of medication that same day.

This is a randomized, double-blind study, meaning that neither site staff nor subject will know what the subject is receiving. Sponsor staff are also blinded to treatment assignments.

Randomization codes will be provided to the Investigator or designee for use in case of an emergency code break requirement. Confirmation of receipt of the randomization code will be required by the Sponsor. The PI or designee will be responsible for maintaining the blind, consistent with protocol design, throughout the study, except in the case of an emergency code break.

The Investigator will be required to notify the Sponsor in the event of any breaking of the blind for any reason. Blinding is not to be broken during the study unless considered necessary by the Investigator for emergency situations for reasons of subject safety. The Investigator should contact the Medical Monitor before breaking the blind. When the blind is broken, the reason must be fully documented.

Except in the event of an emergency code break for an individual subject, the subjects, Investigator, and all other study personnel involved with subject assessments will remain blinded to the actual treatment assignments of the subjects. The Investigator will be ultimately responsible for ensuring that the integrity of the blind is maintained throughout the study.

3.6. Administration of Study Medication

Study drug will be taken by mouth (PO) with water (at least 120 mL = 4 ounces) in the morning and before bed. Subjects should attempt to take study drug at least 30 minutes prior to breakfast and at least one hour after dinner. The first dose should be taken one hour after dinner on Day 1.

3.7. Study Procedures and Flowchart

Table 2 shows the schedule of planned study procedures. Every effort should be made to adhere to this procedure schedule.

Table 2 Schedule of Assessments

	Screening	Baseline	Treatment					Follow-Up
	Days -28 thru Day 0	Day 1	Week 1	Week 2	Week 4	Week 8	Week 13[e]	Week 14
Visit Number	1	2	3	4	5	6	7	8
Target Study Day	---	1	7	14	28	56	91	98
Visit Window			±2	±2	±2	±3	±3	±3
Informed Consent	X							
Eligibility Criteria Confirmed	X	X						
Medical/Surgical History	X	X						
Demographics and Social History	X							
Full Physical Exam	X						X	
Abbreviated Physical Exam		X						X
Height, Weight, BMI	X	X[a]					X[a]	
Vital Signs [b]	X	X	X	X	X	X	X	X
ECGs (single)	X	X	X	X	X	X	X	
Serology (HbsAg, HCV)	X							
Chemistry and Hematology	X	X[c]	X	X	X	X	X[c]	
Lipids		X[c]					X[c]	
ApoE4 testing	X							
Serum Pregnancy [d]	X						X	
Urine Pregnancy [d]		X						
Urinalysis	X	X	X		X		X	
C-SSRS	X	X	X		X		X	
RBANS	X	X	X	X	X	X	X	
MMSE	X	X	X	X	X	X	X	
CGI-I			X	X	X	X	X	
ADCS-ADL		X				X	X	
CDR	X						X	
Hachinski (MHIS-NACC)	X							
PK and Biomarker Sampling		X (biomarkers only)	X		X	X	X	
Randomization		X						
Dispense Medication		X		X	X	X		
First Dose of Medication		X						

		(PM dose following dinner)						
Daily Medication Administration BID			**X----------X**					
Medication accountability			**X**	**X**	**X**	**X**	**X**	
Prior/Concomitant Medication Recording	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
Adverse Event Recording		**X**	**X**	**X**	**X**	**X**	**X**	**X**
Final Disposition								**X**

[a] Weight only
[b] Vital signs include pulse rate, blood pressure, respiration rate, and temperature
[c] Fasting for a minimum of 8 hours prior to blood draw is required at Baseline (Day 1) and End of Treatment (Week 13)
[d] All females of childbearing potential, even if using birth control)
[e] If a subject terminates from the study prior to Week 13, evaluations intended for Week 13 should be completed at discontinuation, if possible.

4. Endpoints

4.1. Primary Efficacy Endpoint

The primary efficacy outcome measure is the change from baseline to Week 13 in Repeatable Battery for the Assessment of Neurological Status - Delayed Memory Index (RBANS DMI).

4.2. Secondary Efficacy Endpoints

There are six secondary efficacy outcomes of interest. Change from baseline will be evaluated for each of the outcome measures below:

- RBANS total score
- Alzheimer's Disease Cooperative Study Activities of Daily Living Inventory - Mild Cognitive Impairment version (ADCS-ADL) total score
- Mini-Mental State Evaluation (MMSE) score
- Clinical Dementia Rating sum of boxes (CDR-SB) score
- Modified ADCOMS (m-ADCOMS) Composite Endpoint

The observed values will be assessed for the endpoint below:

- Clinical Global Impression-Improvement (CGI-I) score

4.3. Safety Endpoints

The following safety endpoints will be collectedduring the study.

- Treatment-emergent adverse events
- Vital signs
- Clinical laboratory evaluations (chemistry, hematology, lipids, urinalysis)
- Electrocardiogram (ECG) measurements

5. Analysis Populations

5.1. Screened Population

The Screened Population will include all screened subjects who signed the informed consent form. This population will be used for summaries of subject disposition, and listings unless otherwise specified.

5.2. Randomized Population

The Randomized Population will include the subset of subjects from the Screened Population who have been randomized. Subjects will be classified according to randomized treatment.

However: Patient [redacted] was the same subject, randomized at 3 different treatment centers. Due to the multiple treatments and overlapping treatment periods, this patient (and all 3 patient numbers associated with the patient) is excluded from the Safety, modified intent-to-treat (mITT), Per Protocol (PP), Completers, and PK populations. This patient will be reported in the CSR via a detailed patient narrative. This patient is therefore excluded from ALL analysis populations.

5.3. Safety Population

The Safety Population will include all randomized subjects who received at least one dose of study treatment. Subjects will be analyzed according to treatment received. Specifically, if a subject randomized to placebo takes one (or more) doses of active treatment (e.g., if the subject is dispensed the wrong drug kit at the site), the subject will be analyzed in that active treatment arm for all safety assessments. The Safety Population will be used for all analyses of safety endpoints.

5.4. Modified Intent-to-Treat Population

The modified Intent-to-Treat (mITT) Population will include all randomized subjects who received at least one dose of treatment and returned for at least one follow-up visit with the primary endpoint collected (non-missing). Subjects will be analyzed according to randomized treatment. The mITT Population will be used for all analyses of efficacy endpoints.

5.5. Per Protocol Population

The Per Protocol (PP) Population includes all mITT subjects who received at least one dose of treatment and returned for at least one follow-up visit with no significant protocol violations. Subjects will be analyzed according to randomized treatment. The PP population will be used to evaluate the robustness of the mITT results.

5.6. PK Population

The PK Population will include all subjects in the Safety Population who receive BPN14770, with at least one post-dose blood sample value with a non-missing (valid) plasma concentration of BPN14770.

5.7. Completers Population

The Completers Population (CP) includes all mITT subjects who completed all 13 weeks of treatment with no significant protocol violations. The CP population will be used to evaluate the robustness of the ITT results.

5.8. Protocol Deviations

A protocol deviation is any change or alteration from the procedures stated in the study protocol, consent document, recruitment process, or study materials (e.g., questionnaires).

A major protocol deviation is a deviation that has an impact on subject safety, may substantially alter risks to subjects, may have an effect on the integrity of the study data, or may affect the subject's willingness to participate in the study. Major protocol violations may include (but are not limited to): deviation from inclusion/exclusion criteria, withdrawal criteria met during the study but subject was not withdrawn, prohibited concomitant medications, treatment with incorrect study treatment (e.g., in the event a subject is accidentally given the wrong drug kit by the site), and substantial deviations from the dosing schedule.

A minor protocol deviation is a deviation that does not impact subject safety, compromise the integrity of the study data, or affect the subject's willingness to participate in the study.

For this project, the protocol deviation collection will be managed in Clinical Trial Management System (CTMS) by the Site Monitors during the monitoring visits. The project team reviews the protocol deviation categories and classifications, and confirms these with the Sponsor on an ongoing basis. Protocol deviations will be summarized by classifications and categories on the Randomized Population. All protocol deviations will be listed.

Protocol violations that result in exclusion from the PP population will be identified prior to unblinding.

6. General Aspects for Statistical Analysis

6.1. General Methods

Unless otherwise specified, continuous endpoints will be descriptively summarized with n, mean, standard deviation (SD), median, minimum (min) and maximum (max). The minimum and maximum will be reported with the same degree of precision (ie, the same number of decimal places) as the raw data. Mean and median will be reported to one degree of precision more than the raw data and standard deviation will be reported to two degrees of precision more than the raw data. In addition, change from baseline values will be calculated at each time point and summarized descriptively. For categorical endpoints, descriptive summaries will include counts and percentages." See the programmings specifications at the end of this SAP (Section14) for more details.

For tabulations, treatments will be sorted, left to right, as Placebo, BPN14770 10 mg, and BPN14770 25 mg. In those cases where a combined column is needed, this will be at the far right as 'Pooled BPN14770' (if the 2 active arms) or 'Total' (if all 3 treatment arms). All analyses and summary outputs will be generated using SAS® version 9.4 (or higher).

All subject data, including those derived as applicable, will be presented in the subject data listings; listings will display all subjects who were screened and have the related data collected. In general, the subject data listings will be sorted by treatment group, site (if site ID is not part of subject number), subject number and assessment visit and date (and time, if applicable). The summary tables will be presented descriptively by treatment groups and overall, if applicable.

Unscheduled or repeat assessments will not be included in summary tables, but will be included in listings. Unless otherwise specified, if there are multiple results at a given visit/time point, the latest non-missing value within the visit/time point window will be used.

Where statistical models (e.g., mixed models for repeated measures (MMRM), analysis of covariance (ANCOVA)) are used to determine treatment effect estimates, least square means (LSmeans) and the standard error (SE) of the LSMeans will be presented. Plots of changes over time in efficacy endpoints will present these LSMeans (SE) values. Each of these statistical models will include all 3 treatment groups, and thus pairwise comparsions will be made with all treatments contributing to the error estimates.

6.2. Key Definitions

For the purposes of this study, the term "study drug" or "treatment" refers to BPN14770 or placebo.

The study day is the day relative to the date of first dose of study drug. If the assessment date is on or after the date of first dose of study drug, the study day is calculated as date of assessment - date of first dose of study drug + 1 (and thus the day of first dose is designated as Study Day 1). If the assessment date is prior to the date of first dose of study drug, the study day is calculated as date of assessment - date of first dose of study drug (and thus the day prior to first dose is designated as Study Day -1).

Unless otherwise specified, baseline is the last non-missing measurement before the first dose of the study drug, which is expected to be the baseline visit at Day 1, or screening if the Day 1 data are not available.

The change from baseline will be calculated for each Post-baseline assessment as:

- Change from Baseline = Post-baseline value – Baseline value.

6.3. Missing Data

For the primary efficacy determination, observed cases data will be utilized.

A supportive analyses will be performed for the primary efficacy endpoint to account for data missing at random (MAR) and for data missing not at random (MNAR). MNAR data are often due subjects discontinuing due to lack of efficacy or to due treatment-emergent AE. This sensitivity analysis will include the use of multiple imputation (MI). Multiple imputation provides a useful strategy for analyzing data sets with missing values. Instead of filling in a single value for each missing value, Rubin's (Rubin 1987) multiple imputation strategy replaces each missing value with a set of plausible values that represent the uncertainty about the correct value to impute. This analysis is described in the Primary efficacy endpoint section (Section 8.1.1).

6.4. Visit Windows

The visit windows are specified in the Schedule of Assessment in Section 3.7. The visits recorded in database will be used for the analyses, and there is no plan to re-assign visits based on actual visit dates, except the Early Termination (ET) visit. For ET visit, the next scheduled visit closest to the termination date will be assigned for the efficacy analysis specified in Section 8.

6.5. Pooling of Centers

Thirty eight (38) sites enrolled subjects into the study, with nearly 75% having enrolled < 10 subjects. As such, sites will be pooled based geographical location (based in US time zones: Eastern, Central/Mountain, Pacific) into pooled final site designations of at least 30 subjects per site.

6.6. Subgroups

Subgroup analysis of the primary efficacy endpoint will be performed for the following subgroups:

- Presence or absence of the ApoE4 allele
- By gender (male and female)
- Ethnicity (Hispanic vs non-Hispanic)
- MMSE (dichotomized as the baseline median across the pooled treatment groups)

Tabulation of the incidence of TEAEs will also be presented for those subgroups.

6.7. Data listings

Data listings will be presented for all data collected. The derived Study Day will be presented along with the date in these data listings. Specifically for TEAEs and medications, the Study Start Day and Study Stop Day will be included. Listings will be sorted by subject number, and then study visit or study day, accordingly.

7. Demographic, Other Baseline Characteristics and Medication

7.1. Subject Disposition and Withdrawals

Subject disposition will be summarized for the Screened Population.

For subject study status, the number and percentage of subjects for each one of the following categories will be presented by treatment groups and overall.

- Screened subjects (Total only)
- Screen failure subjects (Total Only)
- Randomized Population
- Safety Population
- Modified Intent-to-Treat Population
- Per Protocol Population
- Completers Population
- Subjects who completed the study
- Subjects who discontinued the study prematurely by primary reason

For all categories of subjects (except for the screened subjects and screen failures), percentages will be calculated using the number of randomized subjects as the denominator, unless otherwise specified.

Subject disposition and analysis populations data will be listed with clear indication of which populations in which each subject has been included.

7.2. Demographic and Other Baseline Characteristics

The following demographic and baseline variables will be summarized using the Safety Population and the mITT population, by treatment group and overall:

- Age (years)
- Age categorized (as <65, >=65 to <75, and >=75)
- Gender
- Race
- Ethnicity
- Body mass index (BMI) (kg/m^2) at baseline
- BMI categorized (dichotomized at <30 kg/m^2 vs >= 30 kg/m^2)
- APOE4 test results (presence / absence of the allele)
- Pre-treatment MMSE

- MMSE categorized as >=26, 23 to 25, <23
- RBANS DMI
- RBANS DMI by median overall (pooled across all treatment groups) <median vs >=median

Modified Hachinski Ischemia Score (MHIS), smoking history, alcohol consumption, drug use, serology, and pregnancy test results will be collected and listed but not summarized.

7.3. Medical/Surgical History

Medical/surgical history as recorded at screening will be summarized using the Safety Population by treatment group using the number and percentage of subjects reporting each system organ class (SOC) and preferred term (PT). Medical/surgical history will be sorted by descending overall total by SOC and PT using the Medical Dictionary for Regulatory Activities (MedDRA) coding dictionary version 22.0 in the summary table.

Medical/surgical history data will be listed and sorted by treatment group, site, subject number, onset date, SOC, and PT.

7.4. Medication

All concomitant medications taken 14-days prior to the first dose of study drug (Day 1) and during the study through the Week 14 Follow-up Visit should be recorded with indication, daily dose, and start and stop dates of administration.

- Prior medications are those medications taken and stopped before the start of study drug dosing.
- Concomitant medications are medications that are ongoing at or taken on or after the start of study drug dosing.
- New-onset concomitant medications are defined as those medications starting between the first dose of double-blind study medication and 1 day (24 hours) after the start of the last dose of double-blind study medication. Medications with no start time but starting on the day of double-blind study medication will be assumed to be new-onset.

Prior, concomitant, and new-onset medications will be summarized by the number and percentage of subjects taking a particular medication. All the medications will be listed.

A summary of concomitant medications taken during the course of the study will be presented in tabular form using Anatomical Therapeutic Chemical (ATC) classification level 2 and PT via the World Health Organization Drug (WHODrug) Global B3 Mar 2019. The medications will be summarized using the Safety Population by treatment group and sorted alphabetically by ATC level 2 class and PT. If a subject has taken a medication more than once, the subject will be counted only once in the same ATC class or PT.

Medications with incomplete end dates will be considered a concomitant medication if:

- Day and month are missing and the year is equal to or after the year of the first dose date;
- Day is missing and the year is after the year of the first dose;

- Day is missing and the year is equal to the year of the first dose date and the month is equal to or after the month of the first dose date;
- Year is missing; or
- Entire date is missing.

The medication data will be listed, and sorted by treatment group, site, subject number, prior and concomitant status, onset date, ATC level, and PT.

8. Efficacy

8.1. Primary Efficacy Endpoint and Analysis – RBANS DMI

The Repeatable Battery for the Assessment of Neurological Status (RBANS) [Randolph 1998] is a brief, individually administered test measuring attention, language, visuospatial/constructional abilities, and immediate and delayed memory. It consists of 12 subtests, which yield five Index scores and a Total Scale score.

Total scores are calculated by adding up the scores for each of the 12 individual subtests and ranges between 0 and 311, where a low score indicates greater impairment. The range of possible scores for the RBANS DMI is 40 to 160.

The primary efficacy outcome measure is the change from baseline to Week 13 in the delayed memory index of the RBANS [Repeatable Battery for the Assessment of Neurological Status - Delayed Memory Index (RBANS DMI)]. Observed cases data (with no imputation for missing values) will be used for primary efficacy analysis.

The original RBANS DMI score and change from baseline will be summarized by treatment group at each time point.

The primary method of analysis will be an MMRM with restricted maximum likelihood estimation, an unstructured covariance matrix, and the Kenward-Roger approximation for denominator degrees of freedom. The model will include fixed effects for treatment, visit (time), the treatment-by-time interaction, ApoE4, pooled study site, the baseline RBANS DMI score as covariate, and a random effect for subject. [Similar models will be defined for the other continuous secondary efficacy endpoints.] Primary inferences will be drawn from treatment differences in the LSmean changes from baseline derived from the MMRM models at Week 13. As additional supportive information, treatment differences for Weeks 1, 4, and 8 will also be derived using the MMRM model. Least Squares Means for each visit and the SE of the LSMeans will be displayed graphically.

In an MMRM model, missing data are not explicitly imputed. Instead, all available postbaseline assessments up to endpoint for the subjects in the mITT population are utilized and, via modelling of the within subject correlation structure, endpoint treatment differences which are adjusted to take account of missing data are derived. The estimation of treatment differences using the MMRM model is based upon an assumption that the missing data follows a missing-at-random mechanism (i.e., the missingness of the observations may be dependent on the observed outcomes or covariates, but not on unobserved outcomes). For this primary analysis, no imputation of the missing values will be made or required and the data used in the analysis will be the actual observed responses at each visit.

Exmaple SAS code:

```
Proc Mixed data=indat ;
 class trtp SITEGR1 ApoE4 avisitn subjid;
 model chg=  base SITEGR1 ApoE4 trtp avisitn trtp*avisitn /ddfm=kr;
 Random intercept / subject=subjid;
 repeated avisitn / type=un subject=subjid;
 lsmeans trtp*avisitn /cl diff ;
run;
```

In the unlikely circumstance that there are convergence problems with the MMRM analysis, this will be explored. For example, the SCORING=4 option could be used in the MIXED statement, which makes SAS use Fisher scoring for the first 4 iterations. If the convergence problem cannot be resolved, the unstructured covariance matrix will be replaced by the AR(1) covariance structure in combination with a random subject effect.

8.1.1. Sensitivity Analysis using Multiple Imputation (MI) Approach for the Primary Efficacy Endpoint of RBANS DMI

A listing of the number of subjects with each unique missing pattern for the RBANS DMI will be presented. This will be generated using PROC MI in SAS (see example, Table *3*). These data will be examined in order to determine if there was any general difference in the pattern of missing data between the treatment groups, as well as to explore the assumption of missingness at random.

Table 3 Examples of Patterns of Missingness

Pattern	Treatment Group	Number (%) of Subjects	Baseline	Week 1	Week 2	Week 4	Week 8	Week 13
1	25 mg	xx (xx.x%)	X	X	X	X	X	X
	10 mg	xx (xx.x%)	X	X	X	X	X	X
	Placebo	xx (xx.x%)	X	X	X	X	X	X
2	25 mg	xx (xx.x%)	X	X	X	X	X	.
	10 mg	xx (xx.x%)	X	X	X	X	X	.
	Placebo	xx (xx.x%)	X	X	X	X	X	.
Etc.								

For the primary endpoint (RBANS-DMI), 100 imputed datasets will be created, with data imputation for missing values due to missingness at random as well as due to discontinuation due to AE or due to lack of efficacy The 100 imputed datasets were analyzed using an analysis of covariance (ANCOVA) model with contrasts to test the primary efficacy endpoint. The model used treatment as the main effect and stratification factor ApoE4, pooled study site and baseline RBANS-DMI as covariates. Data from all three treatment groups were included in the same ANCOVA model for purposes of the testing procedures.

The 100 resulting treatment effect parameters and standard errors (SEs) from these ANCOVA were presented individually and then combined to provide a distribution of parameters (and SEs) upon which the primary hypothesis of treatment effect was determined. The results of the analyses were presented in summary tables with standard summary statistics as well as least squares (LS) means, active versus placebo LS mean differences (treated group – placebo), SEs, confidence intervals, and paired-comparison p-values. The mean RBANS-DMI values and standard errors of the mean (SEM) for each treatment group were also presented graphically. This analysis was referred to as the multiple imputation analysis.

Imputation Methodology:

For subjects with arbitrary missing data patterns (missing at random [MAR]), here meant to represent RBANS-DMI scores that were missing for reasons other than the subject discontinuing (and specifically, ‘intermediate time points’, eg, a subject has a valid non-missing Week 4 RBANS-DMI, but no Week 8, the

Week 8 score is considered MAR), a Markov chain Monte Carlo method (Schafer 1997) that assumed multivariate normality to impute all such MAR missing values was used to create preliminary imputed (resulting) data sets that have strictly monotone missing patterns. A monotone missing pattern implies that all observations after a certain time point are missing, but that there are no intermediate missing values. This pattern occurs when missing values are usually caused by subject discontinuation.

After this first step, any subjects who had monotone missing values were then divided into two groups. For subjects that did not discontinue due to an AE, a second imputation using regression-based method for continuous variables was then run to complete the imputed datasets by imputing the remaining monotone missing RBANS-DMI scores assuming MAR.

For subjects with missing data as a result of discontinuation due to an AE the data was considered as missing not at random (MNAR). The same pattern-mixture approach was also taken to the imputation (as for the monotone MAR data). However, while a similar regression-based method for continuous variables was applied, a scaled (constant) adjustment to the imputed values was made such that the outcomes that were imputed were distributionally 'worse' than those imputed for the monotone MAR data. This scaled adjustment increased each imputed value (from the regression) for these MNAR data by a constant of '10' (ten). For example, if the regression provided an imputed value of 70 for a given missing value for a subject at a given time point, the MNAR imputed value was set to 70-10 = 60 for that subject at that time point. This would, in effect, guarantee that the overall distribution of missing values for discontinuations due to an AE was WORSE than that for other reasons.

Exapple SAS codes:

Create the missing patterns of the data:

```
proc mi data=dat nimpute=0;
by trt01pn
var Base Week1 Week2 Week4 Week8 Week13;
ods output missPattern=pattern;
run;
```

Create monotone missing patterns under MAR assumption:

```
proc mi data = non_mono out = monotone nimpute = 100 seed = xxx
var trt01pn Base Week1 Week2 Week4 Week8 Week13;
mcmc chain = multiple impute = monotone;
run;
```

Imputing the remaining monotone missing RBANS-DMI scores

```
proc mi data=monotone out=imputed1 seed=xxx nimpute=1
by _imputation_;
class trt01pn ApoE4 Site;
var trt01pn ApoE4 Site Base Week1 Week2 Week4 Week8 Week13;
monotone reg(/details);
run;

Note: for data considered as MNAR, the imputed value will scaled via
subtraction of '10'in order to ensure a 'worse' distribution than for
data that is MAR.
```

Perform ANCOVA using imputed data:

```
proc mixed data=imputed;
by _imputation_;
class trt01pn(ref='0') ApoE4 Site;
model CHG = ApoE4 Site trt01pn base;
lsmeans trt01pn /pdiff cl;
ods output lsmeans=lsmeans diffs=diffs;
run;

Note: CHG is the change from baseline to Week 13 in RBANS-DMI.
```

Combine the analysis results into a single estimation with standard error

```
proc mianalyze parms=lsmeans;
modeleffects trt01pn;
ods output ParameterEstimates=lsm;
by trt01pn;
run;
```

Sample SAS code presented above is for illustrative purposes, and the actual code may be adjusted as needed. Note that all imputed values will be 'forced' to fall into the range of 'possible' values for the RBANS-DMI (which ranges from 40 to 160).

8.1.2. Additional Sensitivity Analysis for the Primary Efficacy Endpoint

In addition, pooling of the two active BPN14770 arms will be performed and a comparison to the placebo arm will be performed, using similar MMRM models as for the primary assessment, for the mITT population.

The primary efficacy population will be the mITT population. The PP population and the CP population outcomes will be used to evaluate the robustness of the mITT results, using the models described above.

A separate sensitivity analysis using a non-parametric (distribution-free) Wilcoxon Rank Sum Test of the pairwise treatment differences will be run for the Week 13 change from baseline data. SAS code similar to the below will be used. A test of the normality of the residuals will be performed to assess the validity of the non-parametric vs parametric assumptions.The results will be presented in a summary table with standard summary statistics as well as the Wilcoxon rank sum means, active vs. placebo rank-sum mean differences and p-values.

```
Proc npar1way data=dat;
  Class treat;
  Var change;
  ods select WilcoxonScores;
Run;
```

8.2. Secondary Efficacy Endpoints and Analyses

Change from baseline will be evaluated for each of the outcome measures below, using similar MMRM models as used for the primary efficacy endpoint.

- RBANS total score, as well as immediate memory, visuospatial/constructional, attention, and language indices
- Alzheimer's Disease Cooperative Study Activities of Daily Living Inventory - Mild Cognitive Impairment version (ADCS-ADL) total score
- Mini-Mental State Evaluation (MMSE) score
- Clinical Dementia Rating sum of boxes (CDR-SB) score
- Modified ADCOMS (m-ADCOMS) Composite Endpoint.

For the analysis of secondary efficacy endpoints, the primary efficacy population will be the mITT population. The PP population and the CP population may be used to evaluate the robustness of the mITT results if the mITT population is materially/substantially greater in sample size (ie, 10% or more) than those two populations. Otherwise, analyses for those two sensitivity populations will not be performed.

An analysis of the Clinical Global Impression- Improvement (CGI-I) score will also be provided.

8.2.1. RBANS total Score and other RBANS index scores

The RBANS total score will be analyzed similarly as for the primary analysis for the primary efficacy endpoint in Section 8.1. The original RBANS total score and change from baseline will be summarized by treatment group at each time point. The LS Mean and SE of change from baseline in RBANS total score over time will also be presented graphically by treatment group.

The pairwise differences in change from baseline between the treatment groups will be assessed at each time point.

Age-based index scores are generated for each RBANS domain as well as a total scale, also expressed as an index score (normal mean=100; SD=15) that ranges from 40-160, calculated from the RBANS stimulus document. Change from baseline in the other RBANS index scores (immediate memory, visuospatial/constructional, attention, language) will be summarized by time point, and analyzed similarly as the primany efficacy endpoint.

8.2.2. ADCS-ADL Total Score

The Alzheimer's Disease Cooperative Study scale for ADL (ADCS-ADL) is administered by an independent rater. The ADCS-ADL is an interviewer-administered informant-based scale where the informant (caregiver) responds to 23 activities of daily living questions about the subject. The questions range from basic to instrumental activities of daily living and take approximately 20 minutes to complete. The total score ranges from 0-78 and a higher score signifies greater functional ability.

Functional ability as assessed in any disease state, including AD, is typically described in terms of activities of daily living (ADLs) that are divided into broad conceptual categories of basic ADLs (BADLs) and instrumental ADLs (IADLs). The ADCS-ADL is a 23 item scale that includes 6 BADL items and 17 IADL items that provide a total score from 0-78, with a lower score indicating greater severity. Caregivers are asked to rate the degree to which their family member or loved one can perform a variety of tasks.

Basic score will also be calculated as the sum of questions 1-6b and ranges from 0-22 (activities included in the basic score are: eating, walking, using the toilet, bathing, grooming and dressing) and Instrumental

score, sum of questions 7-23, ranges from 0-56. (activities included in the instrumental score are: using the telephone, watching television, conversations, clearing dishes, personal belongings, making drinks, making snacks, taking rubbish out, getting out and about, shopping, keeping appointments, being left alone, current events, reading, writing, pastimes/hobbies, household chores).

The ADCS-ADL scores (total, as well as basic and instrumental sub-scores) will be analyzed similarly as for the primary analysis for the primary efficacy endpoint in Section 8.1. The ADCS-ADL scores and change from baseline will be summarized by treatment group at each time point. The LS Mean and SE of change from baseline in ADCS-ADL scores over time will also be presented graphically by treatment group.

The pairwise differences in change from baseline between the treatment groups (will be assessed at each time point.

8.2.3. Mini-Mental State Evaluation (MMSE) Score

The Mini–Mental State Examination (MMSE) is a 30-point questionnaire that is used extensively in clinical and research settings to measure cognitive impairment. It is commonly used to screen for dementia. It consists of a series of questions and tests, each of which scores points if answered correctly. The MMSE tests a number of different mental abilities, including a person's memory, attention and language.

The MMSE [Folstein 1975] consists of 11 items covering orientation, memory (recent and immediate), concentration, language and praxis. Scores range from 0 to 30, with lower scores indicating greater cognitive impairment. Scores for each of the 11 individual tests are not recorded on the CRF, therefore if any item is missing then the total score will be set to missing.

The maximum MMSE score is 30 points, with lower scores indicating worse disease. A score of 20 to 24 suggests mild dementia, 13 to 20 suggests moderate dementia, and less than 12 indicates severe dementia. On average, the MMSE score of a person with Alzheimer's declines about two to four points each year. (source: *https://www.alz.org/alzheimers-dementia/diagnosis/medical_tests*)

The MMSE scores will be analyzed similarly as for the primary analysis for the primary efficacy endpoint in Section 8.1, using an MMRM. The MMSE scores and change from baseline will be summarized by treatment group at each time point. The LS Mean and SE of change from baseline in MMSE scores over time will also be presented graphically by treatment group.

The pairwise differences in LS Mean change from baseline between the treatment groups will be assessed at each time point.

8.2.4. Clinical Dementia Rating Sum of Boxes (CDR-SB) Score

The Washington University Clinical Dementia Rating scale [Morris 1993] is a structured, clinician-rated interview that collects information on cognitive capacity from both the collateral source and subject for the evaluation of staging severity of dementia. The CDR was developed primarily to assess severity level in persons with dementia but it can be used to stage dementia in other illnesses as well (eg, Parkinson's disease). Six domains are assessed and then synthesized to assign a Global CDR score. The domains are memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care. Impairment is defined only when caused by cognitive loss rather than by physical disability or other non-cognitive factors. Severity ratings range along a 5-point scale (except for the personal care domain):

- CDR-0: no cognitive impairment
- CDR-0.5: questionable or very mild dementia
- CDR-1: mild
- CDR-2: moderate
- CDR-3: severe

The CDR is frequently utilized in clinical and research settings to describe stage-dependent features of dementia and for use as a clinically-meaningful outcome measure in clinical drug trials.

CDR is an ordinal parameter (scale 0-3) and is only collected at baseline and Week 13. The CDR sum of boxes (CDR-SB) score is obtained by summing each of the domain box scores, with scores ranging from 0 (least impairment) to 18 (severe impairment).

The CDR-SB score (and each individual domain) will be analyzed similarly as for the primary analysis for the primary efficacy endpoint in Section 8.1. The CDR-SB score and change from baseline will be summarized by treatment group. The LS Mean and SE of change from baseline in CDR-SB scores will also be presented graphically by treatment group.

The pairwise differences in change from baseline between the treatment groups will be assessed at Week 13.

8.2.5. Clinical Global Impression- Improvement (CGI-I) Score

The CGI-I employs a 7-point Likert scale to measure the overall improvement in the subject's condition posttreatment. The Investigator will rate the subject's total improvement whether or not it is due entirely to drug treatment. Response choices include: 0=not assessed, 1=very much improved, 2=much improved, 3=minimally improved, 4=no change, 5=minimally worse, 6=much worse, and 7=very much worse.

The CGI-I does not, by it's very definition, have a baseline, and thus the model for this variable will not include a baseline covariate. The CGI-I will be analyzed similarly as for the primary analysis for the primary efficacy endpoint in Section 8.1, with modification to the MMRM model accordingly.

A table of the number (%) of subjects by each CGI-I response category will be provided for each study visit. A CMH row means scores test (stratified by ApoE allele) will be provided at each time point. Further, at each time point, subjects will dichotomized as 'improved' or 'not improved', and p-values will be derived to test the treatment group comparibility accordingly (using a CMH test).

The percentage of subjects with each score at each visit will also be presented graphically by treatment (via bar chart).

An analysis will be performed by using PROC GLIMIX in SAS, and the cumulative odds ratios will be obtained by using the ESTIMATE statement. The difference between treatment groups will be evaluated by a generalized linear model (GLM) at each time point. The model will include fixed effects of ApoE4 stratification factor, study site (pooled), and treatment group. The cumulative odds ratios will be presented along with the SE, a 95% CI, and the p-value.

8.2.6. Composite Endpoint

The composite endpoint will be based on a modification to the AD Composite Score (ADCOMS), a composite clinical outcome for prodromal Alzheimer's disease trials [Wang 2016]. Since the assessments collected in this study were not the same as those collected in the published study, the statistician cited on the pubilication (S. Hendrix) assisted in the development of a modified ADCOMS (m-ADCOMS) composite model for this study.

A partial least square regression is performed and validated using data from the Alzheimer's Disease Cooperative Study (ADCS) and from the Alzheimer's Disease Neuroimaging Initiative (ADNI). The proposed model was derived using the original ADCOMS items [Wang 2016] but leaving out the ADAS-COG Word Finding Difficulty and the ADAS-COG Orientation, and substituting the RBANS immediate word recall (LRL) for the ADAS-COG Immediate Word Recall item (after dividing by the total points possible and multiplying by 10) and also the RBANS list recognition (LRN) for the ADAS-COG Recognition item (after dividing by the total points possible and multiplying by 12). The proposed model has an estimated mean square deviation ratio (MSDR) that is more than 95% of the original ADCOMS MSDR.

The same weights as used for the original ADCOMS are used to calculate the composite score for this study:

Composite =LRL*0.00847483 + LRN *0.003732761 + (5-M1)*0.041567 + (1-M7)*0.038238 + C1*0.054321 + C2*0.109100 + C3*0.089039 + C4*0.069493 + C5*0.058724 + C6*0.078152

Scale	Item ID	Item Name
RBANS	LRL	List Recall (LRL)
	LRN	List Recognition
MMSE	M1	Orientation Time
	M7	Drawing
CDR-SB	C1	Personal Care
	C2	Community Affairs
	C3	Home and Hobbies
	C4	Judgment and Problem Solving
	C5	Memory
	C6	Orientation

As the calculation of the composite scores needs to use the CDR-SB scores, which are only collected at baseline and Week 13, the composite scores will be analyzed similarly as for the primary efficacy endpoint at Week 13 only. The composite scores and change from baseline will be summarized by treatment group. The LS Mean and SE of change from baseline in composite scores will also be presented graphically by treatment group.

The pairwise differences in change from baseline between the treatment groups will be assessed and presented, using similar methology as for the primary efficacy endpoint.

8.3. Multiplicity and Experiment-wise Alpha Control Considerations

This is a Phase 2 dose-finding study, powered only for the primary efficacy endpoint; outcomes from this study are expected to be used to determine the dose to carry forward into Phase 3. Secondary efficacy

endpoints are expected to contribute data providing evidence of benefit from BPN14770 treatment and to be used as part of the preponderance of evidence in the dose-selection decision-making. Due to the limited sample size in this Phase 2 dose-finding study, experiment-wise alpha control will be limited to control for the two pairwise tests for the primary efficacy endpoint only.

- BPN14770 25 mg vs placebo
- BPN14770 10 mg vs placebo

A hierarchical approach will be taken, in which the 25 mg vs placebo comparison will be tested first, and only if statistically significant (p-value $\leq$0.05) will testing of the 10 mg vs placebo comparison be performed. This will ensure alpha control of the primary endpoint for the study as well as a gated procedure for the secondary endpoints for each dose. Only if the primary efficacy endpoint for a given dose is found to be statistically significant will the other efficacy endpoints will be tested at their nominal p-values, with no control for multiple testing.

9. Analysis Of Pharmacokinetics

Plasma BPN14770 concentrations will be obtained during the Week 1, 4, 8 and 13 clinic visits. Samples will be drawn according to the time of the clinic visit, with documentation of time of day sampling was performed, and time of day of most recent dose of study medication.

Plasma BPN14770 concentrations will be descriptively summarized by dose and visit for the PK Population. Concentration values below the lower limit of quantification (BLQ) will be treated as zero.

A listing for Plasma BPN14770 concentrations with dose, visit, and collection time will be produced.

10. Analysis Of Biomarker Data

Biomarker data will be analyzed post-study in an exploratory fashion. This SAP does not address those analyses.

11. Safety

The population used for safety analyses will be the Safety Population. Safety will be assessed on the basis of treatment-emergent adverse events (TEAEs), clinical laboratory values, vital signs, physical exams, the Columbia-Suicide Severity Rating Scale (C-SSRS), and ECGs.

11.1. Extent of Exposure

The following information will be summarized using descriptive statistics by the treatment group:

- Duration of exposure (days)
- Total number of capsules taken
- Cumulative dose (mg)

Duration of exposure will be calculated as:

Date of last dose – date of first dose + 1.

Total number of capsules taken will be calculated as:

Sum of (Number of capsules dispensed at previous visit - Number of capsules returned) over all the visits.

Cumulative dose will be calculated as:

Total number of capsules taken X dose level

The listing for the exposure data will be provided.

11.2. Treatment Compliance

Treatment compliance will be calculated for each visit and overall for the treatment period.

Compliance (%) for each visit will be calculated as:

(Number of capsules actually taken / Number of capsules planned) * 100%

Number of capsules actually taken = Number of capsules dispensed at previous visit - Number of capsules returned/remained.

Number of capsules planned = (current visit date – previous visit date)*2 – 1 (if last dose is in AM) + 1 (if last dose at previous visit is in AM). If last dose information is missing, last dose will be treated as AM dose.

Compliance (%) for overall will be calculated as:

(Total number of capsules taken / Total number of capsules planned) * 100%

Total number of capsules taken is calculated as in Section 11.1; Total number of capsules planned is calculated as (Date of last dose – date of first dose) * 2.

Compliance to study medication will be summarized using descriptive statistics for each visit and overall by treatment group.

Compliance will also be listed along with the study medication accountability data.

11.3. Adverse Events / Adverse Drug Reactions

Adverse events (AEs), which include abnormal and clinically significant clinical laboratory test variables, will be monitored and documented from the time of Informed Consent until study participation is complete (the Follow-up Visit).

All AEs will be coded using Medical Dictionary for Regulatory Activities (MedDRA) version 22.0. The focus of the analysis will be on TEAEs, which will in general be summarized by treatment group and for the combined BPN14770 groups using preferred term (PT) and primary system organ class (SOC). Pre-treatment AEs will be listed (along with all the TEAEs, in the same listing), sorted by start day/date, but pre-treatment AEs will not be tabulated.

TEAEs are defined as any AE that start on or after the first dose of the study drug, or any event that is present at baseline, continues after the first dose of the study drug, and worsens in severity or in frequency. See Section 9.2.3 of the protocol for definition of serious adverse events (SAEs).

Adverse events with incomplete start dates will be considered a TEAE if:

- Day and month are missing and the year is equal to or after the year of the first dose date;
- Day is missing and the year is after the year of the first dose;
- Day is missing and the year is equal to the year of the first dose date and the month is equal to or after the month of the first dose date
- Year is missing; or
- Entire date is missing.

If the relationship to study treatment or the severity (intensity) of the AE is missing, these will be left as missing, and the event will not be included in the tabulations for relatedness and intensity.

For summaries by SOC and PT, a subject will be counted once at the SOC level and once at each PT within the SOC level. For summaries by SOC, PT, and maximum severity, a subject will be counted once at the highest severity level for which the event occurred at the SOC level and the highest severity level for each unique PT within that SOC level. Subjects will be counted only once within each SOC and PT at the strongest relationship to study drug in the following order: Related (possibly related, probably related, or related); Not Related (defined as unlikely related, unrelated, not related, no relation) to study drug.

The summaries presenting frequency of TEAEs will be ordered by decreasing frequency of the combined BPN14770 groups by SOC and PT.

Summary tables of TEAEs will summarize the number and percentage of subjects for the following:

- An overall summary of the number and percentage of subjects reporting TEAEs, study drug-related TEAEs, severe TEAEs, serious TEAEs, TEAEs leading to study drug discontinuation, and TEAEs leading to death.
- TEAEs overall and by SOC and PT
- Study drug-related TEAEs overall and by SOC and PT
- TEAEs by maximum relationship to study treatment, overall and by SOC and PT
- TEAEs by maximum severity, overall and by SOC and PT
- Serious TEAEs, overall and by SOC and PT
- TEAEs leading to study drug discontinuation, overall and SOC and PT

Only TEAEs will be included in the summary tables, but all AEs will be listed. Separate listings for SAEs, TEAEs leading to study drug discontinuation, and TEAEs leading to death will also be provided . A separate listing of death events will also be provided. All AE listings will include Study Day (start and stop) for the start dates and stop dates and will be sorted by Start Day.

11.4. Laboratory Evaluations

Blood samples for chemistry, hematology and lipid testing and urine samples will be collected based on the collection schedule provided in the Schedule of Assessments in Section 3.7 . A list of the clinical laboratory tests that will be performed can be found in Appendix 16.2 of the protocol. A central laboratory will be used to analyze the clinical laboratory samples.

If laboratory values were considered clnically significant by the investigator, the value was to be reported as a TEAE, and thus clinically significant lab values are reported accordingly in the adverse events.

Summary tables will include descriptive statistics for the actual values and changes from baseline by scheduled time point in hematology, chemistry, lipid testing, and quantitative urinalysis test results by treatment group. For qualitative urinalysis parameters, frequency counts and percentages will be presented over time for these categorical data by treatment.

A shift table based on reported lab values classified as low/normal/high will be created for each lab parameter to assess shift from baseline to each subsequent visit. Similarly, a shift table comparing baseline lab value classification to shifts occurring anytime following initial dose by highest classification and lowest clarification will be provided.

All parameters will be converted to consistent units according to the Standard International System of Units (SI) before summarization.

All clinical laboratory results will be listed by site, subject and timing of collection. Out-of-range values will be flagged as low or high where applicable based on lab normal ranges provided by the lab(s). Listings of abnormal hematology and chemistry test results will be provided.

All clinical laboratory parameters will be plotted over time (with mean and SD) by study visit.

11.5. Vital Signs

Vital signs (pulse rate, blood pressure, respiration rate, and temperature) will be measured at Screening, during the Treatment Visits, and Follow-Up as specified in the Schedule of Assessments. Blood pressure at the Screening Visit should be obtained in both arms and the arm with the higher value should be used for ongoing monitoring throughout the rest of the study. For each subject, measurement on the same arm (right or left) using the same method (either automated or manual) should occur throughout the study.

Summary table of actual values and changes from baseline by visit for each vital sign parameter will be provided by treatment groups.

All vital signs measurement data will be listed by subject and time of measurement. Weight collected at Day 1 and Week 13 will be listed with vital signs.

All vital sign parameters will be plotted over time (with mean and SD) by study visit.

11.6. ECG

Single 12-lead ECGs will be performed at clinic visits as specified in Table 2 Schedule of Assessments.

Summary tables by visit and change from baseline to visit for each ECG parameter (Heart Rate, PR interval, QRS interval, QT interval, RR interval, and QTc interval) will be provided.

The overall results of the ECGs recorded as “normal”, “abnormal, not clinically significant (NCS)” or “abnormal, clinically significant (CS)” will be summarized by treatment group at each visit.

12-lead ECG data will be listed by subject and visit.

The ECG analysis will include a careful review of QTcF values. As part of this review, a summary of the number (percent) of subjects with QTcF values (at pre-treatment and then at any time post-first dose) in the following ranges will be provided: ≤450 msec, >450 to ≤480 msec, >480 to ≤500 msec, and >500 msec. This will be performed by visit as well as at ANY time during the double-blind treatment period (where the highest QTcF value will be used for that assessment).

11.7. Physical Examination

The full physical examination will be conducting at the Screening and Week 13 (or Early Termination) Visits as specified in the Schedule of Assessments. An abbreviated physical examination will be conducted at on Day 1 (Baseline) and Week 14 Follow-Up.

Significant findings prior to the start of dosing will be recorded on the Medical History page of the electronic case report form (eCRF). Only changes from baseline physical examination findings that meet the definition of a treatment-emergent AE will be recorded as an AE.

Physical examination data will be listed only.

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

The Columbia Suicide Severity Rating Scale (C-SSRS) is a prospective assessment instrument that directly classifies suicidal ideation and behaviour into 11 preferred categories [Posner 2011]. The C-SSRS involves a series of probing questions to inquire about possible suicidal thinking and behavior.

The C-SSRS will be conducted at Screening, Baseline, Week 1 and Week 4, as specified in the Table 2 Schedule of Assessments.

The C-SSRS data will be listed for all subjects. Listings will include behavior type and/or category for Suicidal Ideation and Suicidal Behavior of the C-SSRS.

The following outcomes are C-SSRS categories and have binary responses (yes/no). The categories have been re-ordered from the actual scale to facilitate the definitions of the composite and comparative endpoints, and to enable clarity in the presentation of the results.

- Category 1 – Wish to be Dead
- Category 2 – Non-specific Active Suicidal Thoughts
- Category 3 – Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act
- Category 4 – Active Suicidal Ideation with Some Intent to Act, without Specific Plan
- Category 5 – Active Suicidal Ideation with Specific Plan and Intent
- Category 6 – Preparatory Acts or Behavior
- Category 7 – Aborted Attempt
- Category 8 – Interrupted Attempt
- Category 9 – Actual Attempt (non-fatal)
- Category 10 – Completed Suicide
- Self-injurious behavior without suicidal intent is also a C-SSRS outcome (although not suicide-related) and has a binary response (yes/no).

Endpoints based on the above categories are defined below.

- Suicidal **ideation**: A “yes” answer at any time during treatment to any one of the five suicidal ideation questions (Categories 1-5) on the C-SSRS.
- Suicidal **behavior**: A “yes” answer at any time during treatment to any one of the five suicidal behavior questions (Categories 6-10) on the C-SSRS.
- 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) on the C-SSRS.
- Self-injurious behavior without suicidal intent.

A shift table to present changes from pre-treatment to any time post-first-dose will be assessed, with the number (and percent) of subjects showing a worsening from baseline on any of the four endpoints.

The C-SSRS was collected at screening (and the assessment was based on subject recall from the past year), while on-treatment collection was based on outcomes ‘since the prior visit’. Thus, the analysis will be done in two ways:

1. Shifts from screening, where C-SSRS is relative to subject recall of the past year (using Screening value C-SSRS, and shifts to each post-screening time point).

2. Shifts from baseline, where C-SSRS is collected relative to subject recall since the PREVIOUS visit, using Baseline value C-SSRS and shifts to each post-baseline time point).

Table 4 presents the general manner in which these data will be tabulated.

Table 4 Sample Table Demonstrating How C-SSRS Outcomes will be Presented

Treatment	Baseline Category	Treatment Category		
		No suicidal ideation or behavior n (%)	Suicidal ideation n (%)	Suicidal behavior n (%)
Drug Name (N=xxx)	No suicidal ideation or behavior	x (%)	x (%)	x (%)
	Suicidal Ideation	x (%)	x (%)	x (%)
	Suicidal Behavior	x (%)	x (%)	x (%)
Comparator Name (N=xxx)	No suicidal ideation or behavior	x (%)	x (%)	x (%)
	Suicidal Ideation	x (%)	x (%)	x (%)
	Suicidal Behavior	x (%)	x (%)	x (%)

Notes: N = number of patients with a baseline and post-baseline C-SSRS assessment, n = number of patients in category. % = 100*n/N.
Baseline refers to [specify definition]
Suicidal Ideation includes any one of the five suicidal ideation events (Categories 1-5). Suicidal behavior includes any one of the five suicidal behavior events (Categories 6-10).
Each patient is counted in one cell only. Patients with both Suicidal Ideation and Suicidal Behavior are included in the Suicidal Behavior category.

A listing of all C-SSRS data will be provided, while a listing of subjects who demonstrated a worsening on any one of the four endpoints will also be provided. These listings will include the Study Day (relative to first dose of double-blind treatment) the C-SSRS was collected.

12. Changes from Analysis Planned in Protocol

There were no meaningful changes from the protocol described analysis.

13. Reference List

1. Faraone, S.V. Interpreting estimates of treatment effects: implications for managed care. P T 33, 700-711 (2008).
2. Folstein M.F., Folstein S.E., McHugh P.R.. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975 Nov;12(3):189-98.
3. Karantzoulis, S., Novitski, J., Gold, M. & Randolph, C. The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): Utility in detection and characterization of mild cognitive impairment due to Alzheimer's disease. Arch Clin Neuropsychol 28, 837-844 (2013).
4. Morris J.C. The Clinical Dementia Rating (CDR): Current version and scoring rules. Neurology, November 01, 1993; 43 (11). doi: https://doi.org/10.1212/WNL.43.11.2412-a
5. Posner, K, GK Brown, B Stanley et al., 2011, The Columbia-Suicide Severity Rating Scale (C448 SSRS): Internal Validity and Internal Consistency Findings From Three Multi-Site Studies With 449 Adolescents and Adults, Am J Psychiatry, 168:1266-1277.
6. Randolph C., Tierney M. C., Mohr E., Chase T. N. The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): Preliminary clinical validity. Journal of Clinical and Experimental Neuropsychology. 1998;20(3):310–319. doi:10.1076/jcen.20.3.310.823. [PubMed]
7. Rogers, S.L., Farlow, M.R., Doody, R.S., Mohs, R. & Friedhoff, L.T. A 24-week, double-blind, placebo-controlled trial of donepezil in patients with Alzheimer's disease. Donepezil Study Group. Neurology 50, 136-145 (1998).
8. Rubin D.B.: Multiple Imputation for Nonresponse in Surveys. New York: John Wiley and Sons, 1987
9. Seltzer, B., et al. Efficacy of donepezil in early-stage Alzheimer disease: a randomized placebo-controlled trial. Arch Neurol 61, 1852-1856 (2004).
10. Wang J, Logovinsky V, Hendrix SB, et al. ADCOMS: a composite clinical outcome for prodromal Alzheimer's disease trials. J Neurol Neurosurg Psychiatry. 2016;87(9):993–999. doi:10.1136/jnnp-2015-312383.

14. Programming Considerations

14.1. General Considerations

- Each output will be stored in a separate file.
- Output files will be delivered in Word format.
- Numbering of TLFs will follow International Conference of Harmonization (ICH) E3 guidance.

14.2. Table, Listing, and Figure Format

14.2.1. General

- All TLFs will be produced in landscape format, unless otherwise specified.
- Treatments will be sorted, left to right, as Placebo, 10 mg, and 25 mg. In those cases where a combined column is needed, this will be at the far right.
- All TLFs will be produced using the Courier New font, size 8.
- The data displays for all TLFs will have a minimum 1-inch margin on all 4 sides.
- Legends will be used for all figures with more than 1 variable, group, or item displayed.
- TLFs will be in black and white (no color), unless otherwise specified.
- Specialized text styles, such as bolding, italics, borders, shading, and superscripted and subscripted text, will not be used in the TLFs, unless otherwise specified. On some occasions, superscripts 1, 2, or 3 may be used (see below).
- Only standard keyboard characters will be used in the TLFs. Special characters, such as non-printable control characters, printer-specific, or font-specific characters, will not be used. Hexadecimal-derived characters will be used, where possible, if they are appropriate to help display math symbols (eg, µ). Certain subscripts and superscripts (e.g., cm^2) will be employed on a case-by-case basis.
- Mixed case will be used for all titles, footnotes, column headers, and programmer-supplied formats, as appropriate.

14.2.2. Headers

- All output should have the following header at the top left of each page:

  Tetra Discovery Partners, Inc.
  Protocol No. BPN14770-CNS-201

- All output should have Page n of N at the top or bottom right corner of each page. TFLs are internally paginated in relation to the total length (ie, the page number should appear sequentially as page n of N, where N is the total number of pages in the table).
- The date output was generated should appear along with the program name as a footer on each page.

14.2.3. Display Titles

- Each TLF should be identified by the designation and a numeral (ie, Table 14.1.1). ICH E3 numbering will be followed. A decimal system (x.y and x.y.z) should be used to identify TLFs with related contents. The title is centered. The analysis set should be identified on the line immediately following the title. The title and table designation are single spaced. A solid line spanning the margins will separate the display titles from the column headers. There will be 1 blank line between the last title and the solid line.

Table x.y.z
First Line of Title
Second Line of Title if Needed
Safety Population

14.2.4. Body of the Data Display

14.2.4.1. *General Conventions*

Data in columns of a table or listing are formatted as follows:

- Alphanumeric values are left-justified;
- Whole numbers (eg, counts) are right-justified; and
- Numbers containing fractional portions are decimal aligned.

14.2.4.2. *Table Conventions*

- Units will be included where available
- If the categories of a parameter are ordered, then all categories between the maximum and minimum category are presented in the table, even if n=0 for all treatment groups in a given category that is between the minimum and maximum level for that parameter. For example, the frequency distribution for symptom severity would appear as:

Severity Rating	N
severe	0
moderate	8
mild	3

  Where percentages are presented in these tables, zero percentages will not be presented and so counts of 0 will be presented as 0 and not as 0 (0%).

- If the categories are not ordered (eg, Medical History, Reasons for Discontinuation from the Study), then only those categories for which there is at least 1 subject represented in 1 or more groups are included.
- An Unknown or Missing category is added to each parameter for which information is not available for 1 or more subjects.

- Unless otherwise specified, the estimated mean and median for a set of values are printed out to 1 more significant digit than the original values, and standard deviations are printed out to 2 more significant digits than the original values. The minimum and maximum should report the same significant digits as the original values. For example, for systolic blood pressure:

n	XX
Mean (SD)	XXX.X (XXX.XX)
Median	XXX.X
Min, Max	XXX, XXX

- P-values are output in the format: “0.xxx”, where xxx is the value rounded to 3 decimal places. Every p-value less than 0.001 will be presented as <0.001. If the p-value are less than 0.0001, then present as <0.0001. If the p-value is returned as >0.999, then present as >0.999

- Percentage values are printed to one decimal place, in parentheses with no spaces, one space after the count (eg, 7 (12.8%), 13 (5.4%)). Values that round down to 0.0. will be displayed as '<0.1', or as appropriate with additional decimal places. Unless otherwise noted, for all percentages, the number of subjects in the analysis set for the treatment group who have an observation will be the denominator. Percentages after zero counts should not be displayed and percentages equating to 100% are presented as 100%, without decimal places.

- Tabular display of data for medical history, prior/concomitant medications, and all tabular displays of adverse event data are presented by the body system, treatment class, or SOC with the highest occurrence in the active treatment group in decreasing order, assuming all terms are coded. Within the body system, drug class and SOC, medical history (by preferred term), drugs (by ATC1 code), and adverse events (by preferred term) are displayed in decreasing order. If incidence for more than 1 term is identical, they should then be sorted alphabetically. Missing descriptive statistics or p-values which cannot be estimated are reported as “-”.

14.2.4.3. *Listing Conventions*

- Listings will be sorted for presentation in order of treatment groups as above, subject number, visit/collection day, and visit/collection time.

- Dates are printed in SAS DATE9.format (“ddMMMyyyy”: 01JUL2000). Missing portions of dates are represented on subject listings as dashes (--JUL2000). Dates that are missing because they are not applicable for the subject are output as “N/A”, unless otherwise specified.

- All observed time values are to be presented using a 24-hour clock HH:MM or HH:MM:SS format (eg, 11:26:45, or 11:26). Time will only be reported if it was measured as part of the study.

- Units will be included where available.

14.2.5. Footnotes

- A solid line spanning the margins will separate the body of the data display from the footnotes.

- All footnotes will be left justified with single-line spacing immediately below the solid line underneath the data display.

- The last line of the footnote section will be a standard source line that indicates the name of the program used to produce the data display, date the program was run, and the listing source (ie, 'Program : myprogram.sas Listing source: 16.x.y.z').

15. Quality Control

SAS programs are developed to produce output such as analysis data sets, summary tables, data listings, figures or statistical analyses. An overview of the development of programs is detailed in Syneos Health SOP Developing Statistical Programs (3907).

Syneos Health SOPs Developing Statistical Programs (3907) and Conducting the Transfer of Biostatistical Deliverables (3908) describe the quality control procedures that are performed for all SAS programs and output. Quality control is defined as the operational techniques and activities undertaken to verify that the SAS programs produce the output by checking for their logic, efficiency and commenting and by review of the produced output.