NCT #: NCT03569631

STATISTICAL ANALYSIS PLAN (SAP)

Protocol Number:	BPN14770-CNS-203
Protocol Title:	A Randomized, Double-blind, Placebo-Controlled, 2-period Crossover Study of BPN14770 in Adult Males with Fragile X Syndrome
Product Name or Number:	BPN14770
Sponsor:	Tetra Discovery Partners, Inc. 38 Fulton Street West, Suite 303 Grand Rapids, MI 49503-2684
SAP Prepared by:	[redacted]
SAP Version Number (Date):	Version 1.0 (16 July 2020)

TABLE OF CONTENTS

1 LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation or Term	Definition
ABC-FX	aberrant behavior checklist (fragile X-specific factoring system)
ADAMS	anxiety, depression, and mood scale
AE	adverse event
API	active pharmaceutical ingredient
bid, BID	twice daily
BMI	body mass index
BP	blood pressure
BPM	beats per minute
CGI-I	clinical global impression improvement – investigator rated
CGI-S	clinical global impression severity – investigator rated
CNS	central nervous system
CP	completer population
CRF	case report form
ECG	electrocardiogram
ERP	event related potential
FDA	food and drug administration
FXS	fragile X syndrome
IRB	institutional review board
ITI	inter-trial interval
ITT	intent to treat population
KiTAP	test of attentional performance
MEDRA	Medical Dictionary for Regulatory Activities
mg	milligram
n, N	number
NIH-TCB	NIH toolbox cognitive battery for intellectual disabilities
PI	principal investigator
PK	pharmacokinetic
PO	by mouth (per os)
qd, QD	once daily (every 24 hours)
SAE	serious adverse event
SAP	statistical analysis plan
SD	standard deviation
SOC	system organ class
SOP	standard operating procedure

Abbreviation or Term	Definition
VAS	visual analog scale

2 STUDY OVERVIEW

This is a single center, Phase 2, randomized, double-blind, placebo-controlled, 2-period crossover study to obtain preliminary assessment of the effects of BPN14770 in subjects with Fragile X Syndrome. As schematic display of the study design is shown in Figure 1.

The study will consist of a Screening period of up to 28 days prior to initial study drug administration, followed by two 12-week double-blind treatment periods. The screening and the baseline visit for Period 1 may occur at the same time, provided the results of screening safety labs can be rapidly obtained. Subjects will be allowed to combine Period 1 Week 12 and Period 2 Day 1 visits. No washout period will be utilized between double-blind treatment periods; instead, efficacy and biomarker assessments during the second double-blind period (Period 2) will be obtained after a minimum of 6 weeks to allow for drug washout. A final follow-up phone contact for safety is planned one week after the conclusion of Period 2.

Eligible patients will be randomized in a blinded, balanced (1:1) fashion to receive either BPN14770 25 mg bid or matching placebo during Period 1, followed by the opposite treatment during Period 2.

Brief cognitive and behavioral assessments will be performed during each clinic visit. Safety and tolerability assessments throughout the study will include adverse event monitoring, ECGs, vital signs, blood chemistry, hematology, and urinalysis.

Pharmacokinetic samples will be collected at screening and at end of Periods 1 and 2 (Week 12) to confirm that study drug is present when expected and to estimate plasma exposure. Biomarker samples will be also collected at the time the pharmacokinetics samples are obtained.

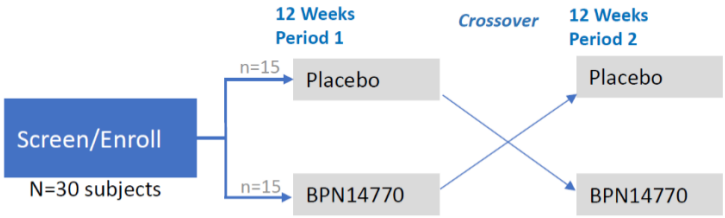

Figure 1: Study Design Schematic

Study Duration

The total duration of the study of each subject will be up to 29 weeks, including a maximum of a 4-week screening period, two 12-week double-blind treatment periods, and a follow-up call approximately 1 week after last treatment.

Study Procedures

The Screening Visit will occur up to 28 days prior to the first study drug administration on Day 1. During screening, subjects and their parent/legal authorized guardian, if indicated, will review and sign an Informed Consent/Assent form prior to any study procedures being performed. Following confirmation of a prior diagnosis for Fragile X, subjects will have information collected regarding their neurological and medical/surgical history, race/ethnicity, social history (tobacco, alcohol, and/or drug use), and use of prescription and over-the-counter medications. Subjects will undergo a full physical exam, have vital signs measured, and 12-lead ECG performed. Height, weight, and BMI will also be collected. Fasting blood samples will be collected for chemistry and hematology, as well as biomarkers and pharmacokinetics. Urine will be collected for urinalysis. The Stanford-Binet test will also be administered at screening, as will an assessment of suicidality risk.

At the Baseline visit (Period 1, Day 1), prior to randomization, patients will receive an abbreviated physical examination including vital signs. However, these will not be repeated if the Screening and Baseline visits occur on the same day. Cognitive and behavioral evaluations will be performed, and these measurements will be used as a common set of baseline measurements to which post-treatment assessments will be compared for both treatment periods. During both double-blind periods, subjects will receive twice-daily treatment with blinded study medication as randomly assigned at baseline. Doses of study medication should be taken in the morning and at night, at least 6 hours apart and at least 30 minutes prior to or 1 hour after meals. Subjects will return to the clinic at the end of Weeks 2, 6, and 12 for each period. Cognitive and behavioral evaluations will be repeated at Weeks 6 and 12 for each period, with the exception of event-related potentials and eye tracking which will be repeated only at Week 12 for each period. Additionally, subjects will be monitored for adverse events, changes in concurrent medications and suicidality via a telephone call at the end of Week 1 of each period, and one week following completion of Period 2 (concurrent medications and AEs only). During clinic visits, adverse effects will be assessed, and laboratory measures, vital signs, and ECGs will be repeated per protocol. Suicidality risk with also be evaluated during the treatment periods per protocol; if a concern is detected, the subject will be referred for further evaluation and treatment. If a patient discontinues the study before Week 12 of a period, the clinical site will endeavor to collect all assessments to be completed at Week 12.

3 STUDY OBJECTIVES

3.1 Study Objectives

The objectives in this study are:

- To obtain preliminary assessment of the efficacy of BPN14770 25 mg bid,
- To evaluate the safety and tolerability of BPN14770 25 mg bid, and
- To obtain pharmacokinetic, pharmacodynamic, and biomarker data on BPN14770.

3.2 Exploratory Efficacy Endpoints

The following instruments will be used to assess the exploratory efficacy endpoints:

- NIH Toolbox Cognitive Battery Modified for Intellectual Disabilities (NIH-TCB)
- Test of Attentional Performance (KiTAP)
- Clinical Global Impression Severity – Investigator rated (CGI-S)
- Clinical Global Impression Improvement – Investigator rated (CGI-I)
- Visual Analog Scale (VAS) Rating Scale using patient-specific behavioral anchors
- Aberrant Behavior Checklist (ABC)
- Anxiety, Depression, and Mood Scale (ADAMS)
- Vineland-3 Rating Scale
- Event-Related Potentials (ERP)
- Eye Tracking

3.3 Safety and Tolerability Endpoints

The following safety assessments will be conducted during the study:

- Treatment-emergent adverse events
- Changes in vital signs
- Clinical laboratory evaluations (chemistry, hematology, urinalysis)
- Electrocardiogram (ECG) measurements

4 GENERAL METHODS

4.1 Analysis Populations

Safety Population: All subjects who sign the study-specific informed consent documents and receive at least one dose of study medication.

Intent to Treat (ITT) Population: All randomized subjects who receive at least one dose of treatment and return for at least one follow-up visit.

Completer Population (CP): All randomized subjects who complete both treatment periods with no significant protocol deviations.

4.2 Summarization of Data

Study results will be summarized by treatment group (Placebo or 25 mg BID) or treatment sequence (A-B, B-A), where appropriate.

No imputation of missing data will be performed. No windowing of visits will be performed unless otherwise specified.

4.3 Sample Size Justification

While no definitive sample size calculations were performed due to the exploratory nature of the study, a common standard deviation of 3 units and an effect size of 2.25 units on the Aberrant Behavior Checklist should provide 80% power to detect a significant difference at alpha = 0.05 with 30 crossover subjects. If carryover effects or excessive dropouts allow only for a first Period analysis, a sample size of 15 per group during the first period would allow for detection of an effect size of 3.19 units.

4.4 Multiplicity

Since this is an exploratory study of BPN14770, no multiplicity adjustments will be made to any alpha levels used for statistical testing for efficacy endpoints. P-values <0.05 will be used as an indication of treatment effect for all efficacy endpoints.

4.5 Output Production and Validation

All analyses will be performed using SAS V 9.3 or higher (SAS Institute, Inc, Cary, North Carolina, USA). Validation and quality control of the tables and listings, which display the results of the statistical analysis of the data from this study, will follow the appropriate [redacted] standard operating procedures (SOPs).

5 SUBJECT DISPOSITION

The number of subjects who are in the safety population, who completed the study, and who discontinued from the study, along with reason of study discontinuation, will be summarized in tabular format by treatment sequence and all subjects for the safety population. Subject disposition information will be displayed for the safety population in a subject listing.

6 DEMOGRAPHIC CHARACTERISTICS

For quantitative variables (e.g., age, height, weight, body mass index [BMI], and Stanford-Binet Full Intelligence scale), summary statistics (number [n], mean, standard deviation [SD], minimum, median, and maximum) will be presented by treatment sequence and all subjects for the safety population. For the qualitative variables (e.g., sex, race, and ethnicity), results will be summarized by treatment group as counts and percentages for the safety population. Individual demographic information for the safety population will be displayed in subject listings.

7 MEDICAL HISTORY & PHYSICAL EXAMINATION

Medical history, physical examination findings and neurological examination findings will be displayed for the safety population in a subject listing.

8 COGNITIVE TESTING ANALYSIS

8.1 NIH-Toolbox Cognition Battery (NIH-TCB)

The NIH-TCB is a battery of extensively validated computer-administered cognitive tests with utility across childhood and adolescence, early adulthood, and old age. The NIH-TCB assessments were designed to minimize floor and ceiling effects which often are present in testing batteries designed for the general population. Downward extensions of many of the NIH-TCB tests have been created to allow feasibility down to a mental age of 3. Therefore, there is good reason to believe that the assessments are appropriate for individuals with intellectual disabilities.

NIH-TCB includes the following evaluations/composite scores:

- Picture Vocabulary
- Oral Reading Recognition
- List Sorting Working Memory
- Pattern Comparison Processing Speed
- Picture Sequence Memory
- Flanker Inhibitory Control and Attention
- Dimensional Change Card Set
- Cognition Crystallized Composite Score
- Cognition Fluid Composite Score
- Cognition Total Composite Score

Summary statistics (n, mean, standard deviation, minimum, median, and maximum) will be presented for the observed values and the change-from-baseline values to each posttreatment time point for the uncorrected standard scores, the dext raw scores, and the dext raw percentages if available for each of the above evaluations/composite scores for the ITT population and the completer population. In these displays, baseline will be defined as the last non-missing value obtained prior to first dose of study medication for Period 1, Day 1 predose

The primary uncorrected standard scores are the uncorrected standard scores for cognition crystallized composite, picture vocabulary, oral reading recognition, pattern comparison processing speed, and picture sequence memory. The primary uncorrected standard scores will be analyzed using a mixed effect model. The change from baseline will be used in the model as the dependent variable. The model will include sequence, period, treatment, visit, predose baseline NIH-TCB, baseline Stanford-Binet Full IQ score, and visit by treatment interaction as

fixed effects, with subject within sequence as a random effect. This analysis will be done for both the ITT population and the completer population.

In addition, for Period 1 data only, the change-from-baseline to Week 12 will be analyzed using an analysis of covariance (ANCOVA) model with fixed terms for treatment, baseline NIH-TCB and baseline Stanford-Binet Full IQ covariates for the ITT population for the primary uncorrected standard scores.

If a subject could not complete the uncorrected standard flanker control and attention test then the Downward extension was to be used. The same method was to be used for the uncorrected standard dimensional change card sort test. Since the Downward extension scores are not the same as the uncorrected standard scores, they could not be used in the calculations for the cognition fluid composite score or the cognition total composite score. Therefore, the above tests will be denoted as secondary uncorrected standard scores and only summary statistics will be provided.

For the uncorrected standard list sorting working memory score, it was observed that more than 50% of the subjects had missing data at baseline and/or at week 12 after approximately 2/3 of the subjects had completed Period 1. If this trend of sparsity of data remains upon completion of the study, the uncorrected standard list sorting working memory score will also be denoted as a secondary uncorrected standard score and only summary statistics will be provided.

The NIH-TCB results will be displayed in subject listings.

8.2 KiTAP – Test of Attention Performance

KiTAP consists of eight nonverbal subtests measuring different basal as well as higher-order components of attention and executive functioning. Four subtests (i.e., alertness, distractibility, go-nogo, and flexibility) will be utilized in the study. Within each subtest, mean reaction time, standard deviation time, number correct, number error, and number omissions were reported as appropriate.

The primary endpoints are the alertness mean reaction time, distractibility correct (distractor), distractibility errors (distractor), flexibility correct, flexibility errors, go-nogo correct, and go-nogo errors. Summary statistics (n, mean, standard deviation, minimum, median, and maximum) will be presented for the observed values and the change-from-baseline values to each posttreatment time point for the primary endpoints for the ITT population and the completer population. In these displays, baseline will be defined as the last non-missing value obtained prior to first dose of study medication for Period 1, Day 1 predose.

The primary endpoints will be analyzed using a mixed effect model. The change from baseline will be used in the model as the dependent variable. The model will include sequence, period, treatment, visit, predose baseline KiTAP score, baseline Stanford-Binet Full IQ score, and visit by treatment interaction as fixed effects, with subject within sequence as a random effect. This analysis will be done for both the ITT population and the completer population.

In addition, for Period 1 data only, the change-from-baseline to Week 12 will be analyzed using an analysis of covariance (ANCOVA) model with fixed terms for treatment, baseline KiTAP and baseline Stanford-Binet full IQ covariates for the ITT population for the primary endpoints.

All KiTAP assessments will be displayed in subject listings.

8.3 Clinical Global Impression Severity (CGI-S)

CGI-S is a global measure to provide a clinical judgment of a subject's overall condition based on a trained clinician's assessment of cognition, behavior and activities of daily living. The assessment of severity will be made with a 7-point scale: 1, not ill; 2, very mild; 3, mild; 4, moderate; 5, marked; 6, severe; 7, extremely severe.

CGI-S will be summarized by treatment group as counts and percentages at each visit for the ITT population and the completer population. In addition, CGI-S will be summarized by treatment group as counts and percentages at each visit for the Period 1 data only.

Summary statistics (n, mean, standard deviation, minimum, median, and maximum) will be presented for the observed values and the change-from-baseline values to each posttreatment time point for CGI-S for the ITT population and the completer population. In these displays, baseline will be defined as the last non-missing value obtained prior to first dose of study medication for Period 1, Day 1 predose.

CGI-S will be analyzed using a mixed effect model. The change from baseline will be used in the model as the dependent variable. The model will include sequence, period, treatment, visit, predose baseline CGI-S score, baseline Stanford-Binet Full IQ score, and visit by treatment interaction as fixed effects, with subject within sequence as a random effect. This analysis will be done for both the ITT population and the completer population.

In addition, for Period 1 data only, the change-from-baseline to Week 12 will be analyzed using an analysis of covariance (ANCOVA) model with fixed terms for treatment, baseline CGI-S and baseline Stanford-Binet full IQ covariates for the ITT population for CGI-S.

The CGI-S results will be displayed in subject listings.

8.4 Clinical Global Impression Improvement (CGI-I)

CGI-I assessment is a 7-point scale that requires the clinician to assess how much the subject's illness has improved or worsened relative to a baseline state at the beginning of the intervention, and rated as: 1, very much improved; 2, much improved; 3, minimally improved; 4, no change; 5, minimally worse; 6, much worse; or 7, very much worse.

CGI-I will be summarized by treatment group as counts and percentages at each visit for the ITT population and the completer population. In addition, CGI-I will be summarized by treatment group as counts and percentages at each visit for the Period 1 data only.

Summary statistics (n, mean, standard deviation, minimum, median, and maximum) will be presented for the observed values at each posttreatment time point for CGI-I for the ITT population and the completer population.

CGI-I will be analyzed using a mixed effect model and will be used in the model as the dependent variable. The model will include sequence, period, treatment, visit, baseline Stanford-Binet Full IQ score, and visit by treatment interaction as fixed effects, with subject within sequence as a random effect. This analysis will be done for both the ITT population and the completer population.

In addition, for Period 1 data only, CGI-I will be analyzed using an analysis of covariance (ANCOVA) model with fixed terms for treatment, and baseline Stanford-Binet full IQ covariates for the ITT population at Week 12.

The CGI-I results will be displayed in subject listings.

8.5 Visual Analog Scale (VAS)

VAS consists of three domains: daily functioning, anxiety/irritability, and irritability/language. For each of the domains, the parents/caregiver will mark on a visual line measuring 10 cm with one side marked "worst behavior" and the other side marked "best behavior." The parents/caregiver marks are measured in millimeter distance so that improvements or worsening of behavior over the treatment period can be evaluated.

Summary statistics (n, mean, standard deviation, minimum, median, and maximum) will be presented for the observed values and the change-from-baseline values to each posttreatment time point for each of the three domains for the ITT population and the completer population. In these displays, baseline will be defined as the last non-missing value obtained prior to first dose of study medication for Period 1, Day 1 predose.

The three domains of the VAS will each be analyzed using a mixed effect model. The change from baseline will be used in the model as the dependent variable. The model will include sequence, period, treatment, visit, predose baseline VAS score, baseline Stanford-Binet Full IQ score, and visit by treatment interaction as fixed effects, with subject within sequence as a random effect. This analysis will be done for both the ITT population and the completer population.

In addition, for Period 1 data only, the change-from-baseline to Week 12 will be analyzed using an analysis of covariance (ANCOVA) model with fixed terms for treatment, baseline VAS and baseline Stanford-Binet full IQ covariates for the ITT population for each domain.

The VAS results will be displayed in subject listings.

8.6 Aberrant Behavior Checklist (ABC)

ABC is a 58-item parent/caregiver rating scale used to assess behaviors across six dimensions or subscales: irritability, hyperactivity, lethargy/withdrawal, stereotypy, inappropriate speech and social avoidance. Items are evaluated on a four-point Likert scale ranging from 0 (not at all a problem) to 3 (the problem is severe in degree).

Summary statistics (n, mean, standard deviation, minimum, median, and maximum) will be presented for the observed values and the change-from-baseline values to each posttreatment time point for each of the six subscales for the ITT population and the completer population. In these displays, baseline will be defined as the last non-missing value obtained prior to first dose of study medication for Period 1, Day 1 predose.

For each subscale, the change-from-baseline to Week 12 will be analyzed using an analysis of covariance (ANCOVA) model which includes sequence, period, and treatment as fixed factors, subject within sequence as a random factor, and baseline ABC score and baseline Stanford-Binet full IQ covariates for the ITT population and the completer population.

In addition, for Period 1 data only, the change-from-baseline to Week 12 will be analyzed using an analysis of covariance (ANCOVA) model with fixed terms for treatment, baseline ABC and baseline Stanford-Binet full IQ covariates for the ITT population for each subscale.

All ABC assessments will be displayed in subject listings.

8.7 Anxiety Depression and Mood Scale (ADAMS)

ADAMS is a 28-item behavior-based informant instrument rated by the parent/caregiver and designed to assess anxiety, depression, and mood disorders in individuals with intellectual disability. Items are rated on a scale of 0 ("behavior has not occurred or is not a problem") to 3 ("behavior occurs a lot or is a severe problem"). The scale is composed of 5 factors which address: Manic/Hyperactive Behavior, Depressed Mood, Social Avoidance, General Anxiety and Obsessive/Compulsive Behavior.

Summary statistics (n, mean, standard deviation, minimum, median, and maximum) will be presented for the observed values and the change-from-baseline values to each posttreatment time point for each of the five factors for the ITT population and the completer population. In these displays, baseline will be defined as the last non-missing value obtained prior to first dose of study medication for Period 1, Day 1 predose.

For each factor, the change-from-baseline to Week 12 will be analyzed using an analysis of covariance (ANCOVA) model which includes sequence, period, and treatment as fixed factors, subject within sequence as a random factor, and baseline ADAMS score and baseline Stanford-Binet full IQ covariates for the ITT population and the completer population.

In addition, for Period 1 data only, the change-from-baseline to Week 12 will be analyzed using an analysis of covariance (ANCOVA) model with fixed terms for treatment, baseline ADAMS and baseline Stanford-Binet full IQ covariates for the ITT population for each factor.

All ADAMS results will be displayed in subject listings.

8.8 Event-Related Potentials (ERP)

Event-related Potentials (ERPs) enable extraction of neural responses associated with specific sensory, cognitive, or motor events from an overall EEG as cited in the protocol. Auditory stimuli are presented and EEG events assessed in relation to timing of the stimuli. For this study, an ERP protocol developed to measure auditory habituation to trains of repeated stimuli, resting state delta power, and oddball task responses in FXS will be used (Ethridge et al, 2017). This ERP protocol has shown feasibility and clinical validity based on comparison to sensory hypersensitivity on the Sensory Profile in FXS. The EEGs will be read by [redacted] who developed the ERP protocol and has extensive experience in ERP evaluation. [redacted] will reject data if there is too much artifact present (i.e., the amount of usable data is below a 50% threshold determined by the total number of trials).

The following parameters will be accessed:

1. Habituation
    - N1 amplitude to first paired stimulus

- FXS typically show increased N1 amplitude to auditory stimuli relative to typically developing controls. It is expected that the N1 amplitude will decrease as a sign of positive impact.

- N1 amplitude to second paired stimulus
  - FXS typically show increased N1 amplitude to auditory stimuli relative to typically developing controls, and in particular show a lack of reduction of N1 amplitude in response to repeated stimuli in a stimulus pair. It is expected that the N1 amplitude to the repeated stimulus will decrease as a sign of positive impact.

- N1 amplitude percent change between first and second paired stimulus
  - This is the amount of amplitude decrease (i.e. the amount of habituation) of the N1 ERP from first stimulus to second (repeated) stimulus in the pair, calculated as a ratio to control for individual differences in absolute ERP amplitude. It is expected that this ratio will increase as a sign of positive impact.

2. Oddball

- P2 amplitude standard condition
  - FXS typically show increased P2 amplitude to auditory stimuli relative to typically developing controls. It is expected that P2 amplitude will decrease as a sign of positive impact.

- P2 amplitude oddball condition
  - FXS typically show increased P2 amplitude to auditory stimuli relative to typically developing controls, but less differentiation between standard and oddball stimulus in amplitude. It is expected that P2 amplitude will decrease as a sign of positive impact.

- P2 amplitude percent change between oddball and standard condition
  - FXS typically show increased P2 amplitude to auditory stimuli relative to typically developing controls, but less differentiation between standard and oddball stimulus in amplitude. Typically developing individuals show increased P2 amplitude to oddball stimuli relative to standard stimuli. It is expected that the ratio of P2 amplitude between standard and oddball stimuli in FXS will increase as a sign of positive impact.

3. Resting

- Ratio of occipital alpha power during eyes closed condition during resting EEG
  - Typically developing individuals show increased occipital alpha power during eyes closed resting EEG compared to eyes open resting EEG, whereas FXS show less differentiation between the two conditions. It is expected that an increase in the ratio of eyes closed to eyes open alpha power in FXS as a sign of positive impact.

- Resting gamma power across whole head during eyes open condition
    - FXS typically show increased gamma power during rest compared to typically developing controls. It is expected that a decrease in gamma power as a sign of positive impact.

Summary statistics (n, mean, standard deviation, minimum, median, and maximum) will be presented for the above values and the change-from-baseline values to each posttreatment time point for each of the eight parameters for the ITT population and the completer population. In these displays, baseline will be defined as the last non-missing value obtained prior to first dose of study medication for Period 1, Day 1 predose.

For each parameter, the change-from-baseline to Week 12 will be analyzed using an analysis of covariance (ANCOVA) model which includes sequence, period, and treatment as fixed factors, subject within sequence as a random factor, and baseline ERP score and baseline Stanford-Binet full IQ covariates for the ITT population and the completer population.

In addition, for Period 1 data only, the change-from-baseline to Week 12 will be analyzed using an analysis of covariance (ANCOVA) model with fixed terms for treatment, baseline ERP and baseline Stanford-Binet full IQ covariates for the ITT population for each factor and the overall score.

All ERP results will be displayed in subject listings.

8.9 Eye Tracking

Eye tracking will be used to assess social gaze in the subjects. Testing will be conducted in a quiet room with the lights turned off. The eye tracker (Tobii) will be calibrated for each subject at the beginning of each session. Following calibration, subjects will view pictures shown on the screen. Each assessment begins with presentation of a scrambled face image for 1 s followed immediately by its matched face image for 3 s. An inter-trial interval containing a uniform grey screen is shown for 0.5, 1, or 2 s, randomly determined. The order of face presentation is pseudorandomized and each eye tracking session lasts approximately 6 min. Measurements will be the number of fixations and the proportion of looking time (%) for the following regions of interest: eyes, nose, mouth, and other. An increased in the number of fixations within the eye region and an increase in the proportion of looking time (%) within the eye region indicates an improvement within a subject.

Summary statistics (n, mean, standard deviation, minimum, median, and maximum) will be presented for the above values and the change-from-baseline values to each posttreatment time point for each of the above parameters for the ITT population and the completer population. In these displays, baseline will be defined as the last non-missing value obtained prior to first dose of study medication for Period 1, Day 1 predose.

For each parameter, the change-from-baseline to Week 12 will be analyzed using an analysis of covariance (ANCOVA) model which includes sequence, period, and treatment as fixed factors, subject within sequence as a random factor, and baseline eye tracking score and baseline Stanford-Binet full IQ covariates for the ITT population and the completer population.

In addition, for Period 1 data only, the change-from-baseline to Week 12 will be analyzed using an analysis of covariance (ANCOVA) model with fixed terms for treatment, baseline eye

tracking and baseline Stanford-Binet full IQ covariates for the ITT population for each parameter.

The data displays may be modified based on the actual data received for the analysis.

Eye tracking results will be displayed in subject listings.

8.10 Pharmacokinetics

The pharmacokinetic concentrations will be displayed in summary tables as well as in the subject listings.

8.11 Biomarkers (Pharmacodynamics)

A description of the biomarkers analysis is not included in this statistical analysis plan. The biomarker analysis, if conducted, will be exploratory and determined by the clinical results observed in the study.

9 SAFETY ANALYSES

9.1 Adverse Events

Adverse events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) coding system (using MedDRA version stated in Data Management Plan). An overall summary table will be presented by treatment group that summarizes all treatment-emergent adverse events, all serious treatment-emergent adverse events, all study drug-related treatment-emergent adverse events as assessed by PI, all study drug discontinuations due to treatment-emergent adverse events, number of deaths, and all severe treatment-emergent adverse events for the safety population. In addition, a frequency table will be presented by treatment group for all treatment-emergent adverse events by system organ class and preferred term for number and percentage of subjects:

- by maximum severity
- by relationship to study drug

An adverse event will be considered to be treatment-emergent for the treatment given in Period 1 if the event first appears after initial dosing in Period 1 or if it is a pre-existing condition that worsens following initial dosing through Period 2, Day 1. An adverse event will be considered to be treatment-emergent for the treatment given in Period 2 if the event first appears after Period 2, Day 1 or if it is a pre-existing condition worsens after Period 2, Day 1. Because the start time of adverse events is not collected on the case report form, it will be assumed that any TEAEs reported as starting on Period 2 Day 1 will be assigned to the treatment from Period 1. Any medical condition/event already present prior to the subject taking the first dose of study treatment will be reported in the medical history. Subject listings of all treatment-emergent adverse events will be provided.

In all displays, adverse events will be displayed by MedDRA System Organ Class and Preferred Terms, with subjects who have the same adverse event counted only once for that event and with subjects who have more than one adverse event within a System Organ Class counted only once

in that System Organ Class. An adverse event will be considered drug-related if, in the investigator's judgment, the event was related to treatment.

9.2 Laboratory Tests (Hematology, Serum Chemistry, and Urinalysis)

For hematology, chemistry, and urinalysis quantitative tests, summary statistics (n, mean, SD, minimum, median, and maximum) will be presented for the observed values at each time point by treatment group for the safety population. Summary statistics will also be presented for the change-from-baseline values to each post baseline time point. In these displays, baseline will be defined as the last non-missing value obtained prior to first dose of study medication for Period 1, Day 1 predose.

For all tests, results will be displayed in subject listings, with those values falling outside the laboratory reference range flagged. Laboratory reference ranges will be provided by the laboratory site(s) and will be included in an appendix of the clinical study report.

9.3 Vital Signs

Summary statistics (n, mean, SD, minimum, median, and maximum) will be presented for the observed values for sitting systolic and diastolic blood pressure, pulse, respiration rate, and oral temperature at baseline and each post dose time point by treatment group for the safety population. Summary statistics will also be presented for the change-from-baseline values to each post baseline time point. In these displays, baseline will be defined as the last non-missing value obtained prior to first dose of study medication for Period 1, Day 1 predose.

All vital signs results will be displayed in subject listings.

9.4 Electrocardiograms

Single ECG results will be collected at Baseline/Screening and at Week 6 and Week 12 of each study period. For the quantitative variables of QRS interval, PR interval, QT interval, QTcB interval, and Heart Rate, summary statistics (n, mean, SD, minimum, median, and maximum) will be presented for the observed values at Baseline/Screening and each post dose time point. Summary statistics will also be presented for change-from-baseline values to each post baseline time point. In these displays, baseline will be defined as the last non-missing value obtained prior to first dose of study medication for Period 1, Day 1 predose.

All ECG results will be displayed in subject listings.

10 SUBJECT LISTINGS

Data that are collected and entered into the study database but that are not displayed in the summary tables specified in the preceding sections will be presented in subject listings. These will include (but will not be limited to) data from the following data collection modules:

Protocol Deviations

Drug Accountability/Dosing Information

Inclusion/Exclusion Criteria

Caregiver Information

Concomitant Medications

Concomitant Treatments/Procedures

Stanford-Binet Non-Verbal and Verbal Intelligence Scale

Substance Abuse

Suicidality Assessment

11 INTERIM ANALYSES

No statistical interim analyses are planned.

12 REFERENCES

Ethridge et al. Neural synchronization deficits linked to cortical hyper-excitability and auditory hypersensitivity in fragile X syndrome. Molecular Autism (2017) 8:22 DOI 10.1186/s13229-017-0140-1.

13 FINAL SIGN-OFF FOR TETRA DISCOVERY PARTNERS, PROTOCOL BPN14770-CNS-203 STATISTICAL ANALYSIS PLAN

[redacted]

13-Aug-2020 | 9:08 AM EDT

Date

[redacted]

[redacted]

14-Aug-2020 | 1:15 PM PDT

Date

[redacted]
Tetra Discovery Partners, Inc.
Grand Rapids, MI 49503-2684

[redacted]

14-Aug-2020 | 10:40 AM PDT

Date

[redacted]
Tetra Discovery Partners, Inc.
Grand Rapids, MI 49503-2684

14 REVISIONS TO STATISTICAL ANALYSIS PLAN

Date	Revision	Statistician's Signature
•	•	•
•	•	•
•	•	•
•	•	•
•	•	•
•	•	•

15 APPENDIX

15.1 Conventions for Statistical Tables and Subject Listings

The following conventions are used in the mockups for the statistical tables and subject listings:

[Sponsor] = [Tetra Discovery Partners]

[Protocol] = [BPN14770-CNS-203]

[Drug name] = [BPN14770]

Subject number in listings will represent the subject's screening number.

In general, font size will be at least 9-point and all margins will be 1 inch.