[redacted]

Sponsor	Bionomics Limited
Protocol Title:	A Phase 2, Randomized, Double-blind Study to Evaluate the Efficacy and Safety of BNC210 Compared to Placebo for the Acute Treatment of Social Anxiety Disorder
Protocol Number:	BNC210.013
[redacted]	[redacted]
Document Version:	v2.0
Document Date:	18Nov2022

Approvals

Role	Signatures
Biostatistician	Print Name: [redacted]
	Sign Name:
[redacted]	Print Name: [redacted]
	Sign Name:
Bionomics Limited Representative	Print Name: [redacted]
	Sign Name:

[redacted]

Document History

Final V1.0	16Sep2022 – Initial Version
Final V2.0	11Nov2022 - [redacted]

[redacted]

[redacted]

Table of Contents

[redacted]

[redacted]

List of Abbreviations

Abbreviation	Definition
AE	adverse event
AIC	Akaike information criterion
ANCOVA	analysis of covariance
ATC	anatomical therapeutic chemical
[redacted]	
[redacted]	
BMI	body mass index
CI	confidence interval
CRF	case report form
CSR	clinical study report
C-SSRS	Columbia-Suicide Severity Rating Scale
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
ECG	electrocardiogram
EMA	European medicines agency
FAS	full analysis set
FDA	food and drug administration
HAM-D	Hamilton Depression Rating Scale
HIV	Human immunodeficiency virus

[redacted]

Abbreviation	Definition
HBsAg	Hepatitis B surface antigen
HCV	Hepatitis C virus
HR	heart rate
ICH	International Conference on Harmonization
ITT	intent-to-treat
IWRS	interactive web response system
LLOQ	lower limit of quantification
LS	least squares
LSAS	Liebowitz Social Anxiety Scale
MAR	missing at random
MCMC	Markov-Chain Monte-Carlo
MedDRA	medical dictionary for regulatory activities
MI	multiple imputation
MMRM	mixed model for repeated measurements
MNAR	missing not at random
N	number
[redacted]	
PP	per-protocol
PT	preferred term

[redacted]

Abbreviation	Definition
REML	restricted maximum likelihood
RR	respiratory rate or relative rate
RSS	Royal Statistical Society
SAD	social anxiety disorder
SAE	serious adverse event
SAF	safety population
SAP	statistical analysis plan
SBP	systolic blood pressure
SCID-5-CT	Structured Clinical Interview for DSM-5 Disorders – Clinical Trials version
SD	standard deviation
SE	standard error
SI	International System of Units
SOC	system organ class
SSPS-N	Self-Statements of Public Speaking (Negative Self-Statements Subscale)
STAI	Spielberger State-Trait Anxiety Inventory
SUDS	Subjective Units of Distress Scale
TEAE	treatment-emergent adverse event
ULOQ	upper limit of quantitation
WHO-DD	world health organization drug dictionary

[redacted]

1. Overview

This statistical analysis plan (SAP) describes the planned analysis and reporting for Bionomics Limited protocol number BNC210.013 (A Phase 2, Randomized, Double-blind Study to Evaluate the Efficacy and Safety of BNC210 Compared to Placebo for the Acute Treatment of Social Anxiety Disorder) dated 17-Nov-2021 Version 1.1. Reference materials for this statistical plan include the protocol and the accompanying sample data collection documents. Operational aspects related to collection and timing of planned clinical assessments are not repeated in this SAP unless relevant to the planned analysis.

The structure and content of this SAP provides sufficient detail to meet the requirements identified by the Food and Drug Administration (FDA), European Medicines Agency (EMA), and International Conference on Harmonization (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use: Guidance on Statistical Principles in Clinical Trials (ICH, 1998) All work planned and reported for this SAP will follow internationally accepted guidelines, published by the American Statistical Association (ASA, 2018) and the Royal Statistical Society (RSS, 2014), for statistical practice.

The planned analyses identified in this SAP may be included in clinical study reports (CSRs), regulatory submissions, or future manuscripts. Also, post-hoc exploratory analyses not necessarily identified in this SAP may be performed to further examine study data. Any post-hoc or unplanned, exploratory analysis performed will be clearly identified as such in the final CSR.

The statistical plan described hereafter is an *a priori* plan. It will be approved before any unblinded inferential or descriptive analysis of data pertaining to Bionomics Limited's study BNC210.013.

2. Study Objectives and Endpoints

2.1. Study Objectives

2.1.1. Primary Objective

The primary objective is to assess the effects of BNC210 compared to placebo in reducing self-reported anxiety severity as measured by the Subjective Units of Distress Scale (SUDS) in the performance phase of a public speaking challenge.

2.1.2. Secondary Objectives

The secondary objectives are to:

- Assess the effects of BNC210 compared to placebo in reducing self-reported anxiety severity as measured by the SUDS in a public speaking challenge;

[redacted]

- Assess the effects of BNC210 compared to placebo in reducing self-reported symptoms of anxiety as measured by the Spielberger State-Trait Anxiety Inventory – State component (STAI-State) in a public speaking challenge;
- Assess the effects of BNC210 compared to placebo in reducing self-reported negative cognitions during speaking as measured by the Self-Statements During Public Speaking Scale (Negative Self-Statements subscale; SSPS-N) in a public speaking challenge;
- To assess the safety and tolerability of BNC210;
- [redacted]

[redacted]

[redacted]

- [redacted]
- [redacted]
- [redacted]

2.2. Study Endpoints

2.2.1. Efficacy Endpoints

2.2.1.1. Primary Efficacy Endpoint

The primary efficacy endpoint of this study is the mean change between Baseline to the average of the performance phase of a public speaking challenge in the SUDS for participants receiving a single dose of BNC210 (225 mg or 675 mg) compared to participants receiving placebo, where the average of the performance phase is defined as the mean of up to six timepoints (0, 1, 2, 3, 4, 5 minutes) measured during a public speaking challenge.

2.2.1.1.1 Primary Estimand

The primary question of interest for this Phase 2 study is: What is the intervention difference in self-reported anxiety severity, measured by the mean of the SUDS in up to six timepoints (0, 1, 2, 3, 4, 5 minutes) measured during a public speaking challenge after approximately 1 hour post dose in participants with social anxiety disorder (SAD), regardless of task discontinuation or staff intervention for any reason?

Although all efficacy variables will be measured during 1 day reducing the likelihood that intercurrent events may impact the research question, participants may have experienced possible known or unknown intercurrent events prior to arriving at the study site. Participants may discontinue the study for any reason and may differentially experience initiation of task

[redacted]

discontinuation or staff intervention. The "Treatment Policy Strategy" will be adopted for handling all known or unknown intercurrent events. To that end, the Intent-To-Treat (ITT) principle will serve as the analytical basis for interpreting the estimand. In other words, the difference in a single dose of BNC210 or placebo in the mean change from baseline to the performance phase in SUDS scores will be evaluated regardless of the occurrence of any such intercurrent event.

2.2.1.2. Secondary Efficacy Endpoint(s)

The secondary efficacy endpoints of this study include the following comparisons of participants receiving a single dose of BNC210 (225 mg or 675 mg) and participants receiving placebo:

- Mean change between Baseline to the average of the anticipation phase of a public speaking challenge in the SUDS for participants receiving a single dose of BNC210 (225 mg or 675 mg) compared to participants receiving placebo, where the average of the anticipation phase is defined as the mean of up to three timepoints (0, 1, 2 minutes) measured during speech preparation;
- [redacted]
- Mean change between Baseline to the end of the performance phase of a public speaking challenge in the STAI-State scores for participants receiving a single dose of BNC210 (225 mg or 675 mg) compared to participants receiving placebo;
- Mean change between Baseline to the end of the anticipation phase of a public speaking challenge in the STAI-State scores for participants receiving a single dose of BNC210 (225 mg or 675 mg) compared to participants receiving placebo;
- Mean change between Baseline to the end of the performance phase of a public speaking challenge in the SSPS-N scores for participants receiving a single dose of BNC210 (225 mg or 675 mg) compared to participants receiving placebo;
- Mean change between Baseline to the end of the anticipation phase of a public speaking challenge in the SSPS-N scores for participants receiving a single dose of BNC210 (225 mg or 675 mg) compared to participants receiving placebo;

[redacted]

2.2.2. Safety Endpoints

The safety endpoints of this study include the following from Baseline to the 7-day Follow-Up Visit. In the event that a measurement is not recorded at Baseline (i.e., ECG), the Screening visit will be used instead:

- Continuous adverse events (AEs) and serious adverse events (SAEs);

[redacted]

- Electrocardiograms (ECGs);
- Vital signs (body temperature, pulse rate, respiratory rate, systolic and diastolic blood pressure);
- Clinical laboratory results (hematology, blood chemistry, and urinalysis).

[redacted]

3. Overall Study Design and Plan

3.1. Overall Design

This is a randomized, double-blind, placebo-controlled, three arm, parallel group, multi-center Phase 2 study in participants with SAD, who are between the ages of 18 to 65, with a single day treatment period and seven day follow up period.

Approximately 150 participants will be enrolled and randomized using a 1:1:1 ratio to receive either BNC210 (225 or 675 mg) or matched placebo. Participants will complete a single dose of treatment with their allocated study intervention and will then engage in a public speaking task. Participants will also receive a follow up phone/video call 7 days after their study day for AE recording.

[redacted]

3.3. Study Population

The study population is comprised of male and female participants who are 18 to 65 years of age, with a current diagnosis of SAD as defined in the Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM-5) and confirmed by the Structured Clinical Interview for DSM-5 Disorders – Clinical Trials version (SCID-5-CT). Furthermore, the participants must receive a total score of ≥70 on the Leibowitz Social Anxiety Scale (LSAS) and a sub-score of 3 on Question 6 of the LSAS (Fear and Anxiety subscale) at screening.

[redacted]

[redacted]

3.4. Treatments Administered

Participants will be randomized to one of three treatment groups:

- BNC210 225 mg
- BNC210 675 mg
- Placebo

[redacted]

3.5. Method of Assigning Participants to Treatment Groups

Participants who meet inclusion criteria and none of the exclusion criteria will be randomly assigned in a 1:1:1 ratio to receive BNC210 225 mg, BNC210 675 mg, or placebo.

3.6. Blinding and Unblinding

This is a double-blind study in which participants, investigators, raters, and study personnel involved in the conduct of the study are blinded to study intervention.

For emergency unblinding, study personnel will use the interactive web response system (IWRS). The IWRS will be programmed with blind-breaking instructions. If a participant's intervention assignment is unblinded, the Sponsor must be notified within 24 hours of this occurrence. The date and reason for the unblinding must be recorded. Overall unblinding will take place at the end of the study only after database lock.

[redacted]

3.7. Schedule of Events

Detailed schedules of events for the study are provided in Tables 1 and 2.

Table 1 Schedule of Activities - Overview

ASSESSMENTS[1]	Screening	Treatment Day				Follow-up
	≤ 14 days (+3 days)	Baseline / Pre-Dose2	Dosing	Public Speaking Challenge	Post-Challenge	7 days post treatment (±3 days)
Informed consent	X					
Inclusion and exclusion criteria3	X					
Demography	X					
Medical history including current & prior medications	X					
Structured Clinical Interview for DSM-5 Disorders – Clinical Trials version (SCID-5-CT)	X					
Leibowitz Social Anxiety Scale (LSAS)	X					
Hamilton Depression Rating Scale (HAM-D)	X					
Columbia Suicide Severity Rating Scale (C-SSRS)4	X	X				
Human immunodeficiency virus (HIV), Hepatitis B surface antigen (HBsAg) and Hepatitis C virus (HCV) antibody screen	X					
Serum pregnancy test (if applicable)	X					
Urine drug screen[5]	X	X				
Physical Examination[6]	X	X				
Vital signs[7]	X	X			X[8]	

[redacted]

[redacted]

ASSESSMENTS[1]	Screening	Treatment Day				Follow-up
	≤ 14 days (+3 days)	Baseline / Pre-Dose2	Dosing	Public Speaking Challenge	Post-Challenge	7 days post treatment (±3 days)
12-lead electrocardiogram (ECG)[9]	X				X[10]	
Clinical laboratory tests (hematology, biochemistry, urinalysis)	X	X			X[11]	
Concomitant medications recording	X	X				
Urine pregnancy test (if applicable)[12]		X				
Randomization[13]		X				
Subjective Units of Distress Scale (SUDS)[14]		X		X	X	
Spielberger State-Trait Anxiety Inventory (STAI)[15]		X		X	X	
Self-Statements During Public Speaking (SSPS)[16]		X		X	X	
Dosing with Study Intervention[17]			X			
Blood sample collection for BNC210 concentration[18]					X	
Assessment of participant's readiness for discharge[19]					X	
Adverse Event (AE) recording				X	X[20]	X

[redacted]

[redacted]

Table 2. Schedule of Activities (SoA) - Efficacy Assessments on Treatment Day

	Baseline / Pre-Dose (~5 mins prior to dosing)	**Dosing**	**Resting / Post-Dose** (~55 mins post dose)	**Instructions on Challenge**	**Public Speaking Challenge**								
					Speech Preparation - Anticipation Phase (2 mins duration)			**Performance Phase** (5 mins duration)					
					0 min	1 min	2 min	0 min	1 min	2 min	3 min	4 min	5 min
Subjective Units of Distress Scale (SUDS)	X		X		X	X	X	X	X	X	X	X	X
Spielberger State-Trait Anxiety Inventory – State component (STAI-State)	X		X				X						X
Self-Statements During Public Speaking (SSPS)[1]	X		X				X						X

[redacted]

[redacted]

4. Statistical Analysis and Reporting

4.1. Introduction

Data processing, tabulation of descriptive statistics, calculation of inferential statistics, and graphical representations will primarily use SAS (release 9.4 or higher). If the use of other software is warranted, the final statistical methodology report will detail what software was used for what purposes.

Continuous (quantitative) variable summaries will include the number of participants (n) with non-missing values, mean, standard deviation (SD) median, minimum, and maximum.

Categorical (qualitative) variable summaries will include the frequency and percentage of participants who are in the particular category or each possible value. In general, the denominator for the percentage calculation will be based upon the total number of participants in the study population for the treatment groups, unless otherwise specified. The denominator for by-visit displays will be the number of participants in the relevant study population with non-missing data at each visit.

The minimum and maximum will be reported with the same degree of precision (i.e., the same number of decimal places) as the observed data. Measures of location (mean and median) will be reported to 1 degree of precision more than the observed data and measures of spread (SD) will be reported to 2 degrees of precision more than the observed data.

Percentages will be presented to 1 decimal place, unless otherwise specified.

Unless otherwise indicated, all statistical tests will be conducted at the ≤0.05 significance level using 2-tailed tests. p values and corresponding 95% confidence intervals (CIs) will be presented for statistical tests.

Statistical testing will be performed at the 0.05 level using two-tailed tests. A p value of ≤ 0.10 but > 0.05 will be considered evidence of a trend.

4.2. Interim Analysis and Data Monitoring

No interim analyses are planned.

5. Analysis Populations

The following analysis populations are planned for this study:

- **Safety Population (SAF):** The Safety Population includes all participants who receive any amount of the study intervention. The Safety Population will be used for the analysis of the safety endpoints. Participants will be analyzed by actual treatment received.
- **Intent-to-Treat Population (ITT):** The ITT population includes all randomized participants. Participants will be analyzed by randomized treatments.

[redacted]

- **Full Analysis Set (FAS):** The FAS population includes all randomized participants who receive any amount of the study intervention. Participants will be analyzed by randomized treatment.
- **Per Protocol (PP)**: The per-protocol population will include participants from the FAS population who have no major protocol deviations. [redacted] Participants will be analyzed by randomized treatments.
- [redacted]

6. General Issues for Statistical Analysis

6.1. Statistical Definitions and Algorithms

6.1.1. Baseline

The last non-missing observation recorded before receiving the first dose of study drug will be used as the baseline observation for all calculations of change from baseline.

6.1.2. Adjustments for Covariates

The primary endpoint analyzing the change in SUDS Scores from Baseline to the average of the public speaking challenge will be assessed with a Mixed Model for Repeated Measures (MMRM). The model will include fixed effects for treatment, interaction between treatment and timepoint, sex, use of a mask, and a covariate for baseline score. See Section 8 for more details.

6.1.3. Multiple Comparisons

Because this is a Phase 2 study, the level of significance will be set at 0.05 for primary and secondary endpoints with no adjustment for multiple comparisons.

6.1.4. Handling of Dropouts or Missing Data

All possible efforts will be made to ensure that participants stay in the study and all data are collected as scheduled however, the occurrence of missing data cannot be completely eliminated. The primary endpoint will be analyzed using MMRM.

[redacted]

[redacted]

[redacted]

6.1.4.1. Multiple Imputation Methods

Data are considered MAR (missing at random) if, conditional upon the independent variables in the analytic model, the missingness depends on the observed outcomes of the variable being analyzed but does not depend on the unobserved outcomes of the variable being analyzed. This assumption implies that the behavior of the post dropout observations can be predicted from the observed variables, and therefore that treatment effect can be estimated without bias using the observed data (European Medicines Agency, 2010). For studies of missing data in a controlled clinical trial setting, MAR is usually considered as a plausible underlying missing mechanism (Molenberghs and Kenward, 2007; Siddiqui et al., 2009; Mallinckrodt et al., 2008; Mallinckrodt et al., 2013). The assumption of MAR is often reasonable because, particularly in longitudinal studies wherein the evolution of treatment effects is assessed by design over time, the observed data and the models used to analyze them can explain much of the missingness (Little and Rubin, 1987; Verbeke and Molenberghs, 2000). This point may be especially relevant in well-controlled studies, in which extensive efforts are made to observe all outcomes and factors that influence them while participants are following protocol-defined procedures. Thus, longitudinal clinical trials by their very design aim to reduce the amount of MNAR (missing not at random) data (missingness explained by unobserved responses), thereby increasing the plausibility of MAR (Mallinckrodt et al., 2008).

Although the assumption of MAR is often reasonable in clinical trials, the possibility of MNAR data cannot be ruled out. Therefore, analysis valid under MNAR will be performed as the primary/main analysis.

Any participants who do not begin or who discontinue the public speaking challenge early will have these efficacy assessments analyzed as imputed using MI techniques. This analysis will be presented as the main analysis.

Multiple imputation (MI) is a simulation-based approach where missing values are replaced using an appropriate stochastic model given the observed data and covariates, creating multiple completed data sets. These completed datasets are then analyzed using standard analysis methods (MMRM for this study), and the different parameter estimates across the datasets are then combined to produce unique point estimates, standard errors, and confidence intervals taking into account the uncertainty of the imputation process.

In most randomized clinical trials that collect data over time, the great majority of missing data follow a monotone pattern. That is, once a participant has a missing data for some timepoint,

[redacted]

data will be missing for all subsequent timepoints. In this study, if a subject does not begin the public speaking challenge or discontinues the public speaking challenge early, then all subsequent measures for the public speaking challenge will be missing. Typically, there is also a small amount of non-monotone missing data (i.e., some participants have missing values for intermediate timepoints, but have non-missing data at subsequent timepoints).

MI with Placebo-Based Imputation

A placebo-based (jump to control) multiple imputation for missing primary and some secondary endpoints will be carried out for participants who discontinue the public speaking challenge early, as indicated previously. The imputation has three broad components; i) the multiple imputation process for the placebo data; ii) the multiple imputation process for the BNC210 data; and iii) the analysis model that will be used to draw inference regarding the primary causal estimands along with the method for combining the results across the multiply-imputed datasets.

Multiple imputation based on a standard MAR imputation approach will be performed in SAS using a general three-step approach. Each step will be conducted separately for the primary and a subset of secondary endpoints:

Step 1: If the data has a non-monotone pattern of missingness, then a monotone data augmentation method using Markov-Chain Monte-Carlo (MCMC) will be used to impute data that is missing and required to make the missing data pattern monotone. Fifty datasets with a monotonic missing pattern will be generated. This method will use a non-informative Jeffreys prior to derive the posterior mode from the expectation-maximization (EM) algorithm as the starting values for the MCMC method. Intermittent missing values will be imputed using the MCMC method assuming a multivariate normal distribution over all variables included in the imputation model (i.e., treatment group, baseline, and each post-baseline timepoint). The MCMC statement of the MI procedure in SAS (PROC MI) will specify the CHAIN=MULTIPLE option, so that the procedure uses multiple chains and completes the default 200 burn-in iterations before each imputation, and the IMPUTE=MONOTONE option to create the 50 partially imputed datasets with a monotone missing pattern. The seed of the pseudorandom number generator used to randomly generate imputations for the missing values in Step 1 is 6411776.

Assumptions underlying the partial imputation step are such that participants with missing data follow the same model as other participants in their respective treatment arm that have complete data.

If the raw data has a monotone pattern of missingness, then the same procedures described above can be followed to create 50 identical datasets that will be used as an input dataset for the next step including both treatments. This step will re-impute the BNC210 group with placebo-based data.

Step 2: The second stage will impute missing post-informative participant withdrawal monotone data in a sequential manner, using a method proposed by Ratitch and O'Kelly. The post-informative participant withdrawal values in each of the imputed datasets for the BNC210 arm will be set to missing, thus leaving a dataset with multiply-imputed values and monotone missing

[redacted]

data patterns due to early termination as a result of AEs, lack of efficacy, i.e. these data will contain the complete data for all placebo participants (observed and imputed under MAR) and the BNC210 participants (intermittent missing data imputed under MAR and monotone missing data for non-informative withdrawal imputed under MAR), with monotone missing data due to informative withdrawal left as missing (to be imputed under MNAR). The second stage of the imputation procedure will handle the MNAR imputation for the monotone missing post-informative participant withdrawal data in the BNC210 participants. A new dataset will then be created with all placebo data. Once the MNAR data are imputed, these data (containing all placebo data and imputed MNAR data for BNC210) will then be combined with non-missing and imputed MAR data from BNC210 participants to create the datasets to be used for analysis. The second state imputation will be done sequentially by timepoint for each first stage imputation. The seed number of 1592065 will be used for the imputation procedure described in Step 2.

For both Steps 1 and 2, minimum and maximum values for the SUDS (0 and 100) will be specified (separately) in the MI procedure to avoid imputed values outside the possible range of values. When an intended imputed value is less than the minimum or greater than the maximum value specified, the MI procedure in SAS will redraw another value for imputation.

Step 3: MMRM analyses will be performed separately for each of the 50 complete analysis datasets and the results will be combined into one multiple imputation inference (estimated treatment effect, standard error, p-value and associated 95% CI) using the SAS MIANALYZE procedure. The treatment difference will be tested at the 2-sided 0.05 level and corresponding 95% CIs will be calculated. In the case that there are no missing data for a particular timepoint, p-values and 95% CIs will come from the MMRM analysis on the observed data.

6.1.4.2. Supportive Analysis for Primary and Secondary Efficacy Endpoints

Supportive analyses will be performed on the primary and a subset of secondary efficacy endpoints to demonstrate the robustness of the conclusions. As a supportive analysis, the primary and select secondary efficacy endpoints will be analyzed in the FAS population using MMRM with no MI, described below. The primary efficacy endpoint and select secondary endpoints will also be analyzed in the PP population using MMRM with no MI.

6.1.4.3. MMRM Analysis Method

Primary and select secondary efficacy variables will be assessed with a MMRM with MI, and with a MMRM with no imputation as supportive analyses. The other secondary endpoints will be assessed with a MMRM with no MI. The model will include fixed effects for treatment, interaction between treatment time, sex, use of a mask, and covariate for Baseline score. The MMRM method has been demonstrated extensively as an appropriate choice for the primary analysis in longitudinal confirmatory clinical trials with continuous endpoints (Mallinckrodt et al., 2008). This analysis method, which is from a broader class of direct-likelihood analyses methods, makes use of fully and partially observed data sequences from individual participants by estimating the covariance between data from different time points (Molenberghs and Kenward, 2007). Further, it is often useful to implement MMRM using an unstructured approach

[redacted]

to modeling both the treatment-by-time means and the (co)variances, leading to what is essentially a multivariate normal model wherein treatment group means at the primary time point are adjusted to reflect both the actual observed data and the projected outcomes from the participants with missing data (Cnaan et al., 1997; Molenbergs et al., 2004; Molenbergs and Kenward, 2007).

Despite careful planning and study conduct, the occurrence of missing data cannot be completely eliminated. As a direct likelihood method, the MMRM is a preferred approach for handling missing data in such designs. MMRM is a full multivariate model in nature, which avoids potential bias as a predetermined model and operates in a more general MAR framework (Mallinckrodt et al., 2001).

6.1.5. Pooling of Sites

This is a multicenter study and there will be no pooling of a site factor.

6.1.6. Derived Variables

- **Change from baseline** = value at current time point – value at Baseline.
- **Body Mass Index (BMI) kg/m^2** =

$$\left(\frac{weight\ (kg)}{\frac{height(cm)}{100} * \frac{height(cm)}{100}}\right)$$

- **Treatment Emergent Adverse Event (TEAE)** = any AE that starts or deteriorates on or after the first dose of study medication and through Follow-up Day 7 visit.

[redacted]

[redacted]

[redacted]

[redacted]

6.1.8. Data Adjustments/Handling/Conventions

All collected data will be presented in listings. Data not subject to analysis according to this plan will not appear in any tables or graphs but will be included only in the data listings.

All p values will be displayed in four decimals and rounded using standard scientific notation (e.g., 0.XXXX). If a p value less than 0.0001 occurs it will be shown in tables as < 0.0001.

Adverse Events and Concomitant Medication Coding

An AE is any untoward medical occurrence in a clinical study participant, temporally associated with the use of study intervention, whether or not considered related to study intervention. All SAEs will be collected from the signing of the ICF until the Follow-up Visit at the timepoints specified in Table 2. All AEs will be collected from the start of intervention until the Follow-up Visit timepoints specified in Table 2. Any medical occurrence that begins before the start of study intervention, but after obtaining informed consent will be recorded as medical history/current medical conditions and not as an AE.

Adverse events will be coded using the MedDRA version 23.0 or higher thesaurus. Concomitant medications will be coded using the World Health Organization Drug Dictionary (WHO-DD Global B3 version March 2020).

Concomitant Treatment Definition and Handling of Data

[redacted]

A concomitant treatment refers to all treatment, including concomitant therapies, vaccines, herbal treatments, vitamins, recreational drugs, and non-pharmacological or behavioral treatments such as psychotherapy, taken from time of Screening or during the study.

Partial Date Imputation

If partial dates occur, the convention for replacing missing dates for the purpose of statistical analysis is as follows: if just day is missing then the day assigned is the first day of the month or the date of first dose (if in the same month), whichever is later; if just month is missing then the month assigned is the month of the first dose, unless that results in a date prior to the first dose in which case the date of first dose is assigned; and if both month and day are missing then the month assigned is the month of the first dose and the day assigned is either the first day of the month or the first dose date, whichever is later.

If partial times occur, the convention is as follows: if the missing time occurs on the day of the first dose and both the hour and minute are missing then the time assigned is the time of the first dose, otherwise if both the hour and minute are missing and the date is not the date of first dose the time assigned is 12:00; if the date is the same as the date of the first dose and only hour is missing the hour assigned is 12 or the hour of first dose, whichever is later, and if the date is the same as the date of first dose and only the minute is missing the minute assigned is 30 or the minute of first dose, whichever is later. Otherwise the hour assigned is 12 if the hour is missing and the date is not the same as the date of first dose and the minute assigned is 30 is the the date is not the same as the date of first dose.

These conventions will be applied only to adverse event onset and concomitant medication dates and times with the following precaution: if the missing date and time reflect the date and time of onset of an adverse event, the modified date and time will be constructed to match the first documented date/time post drug administration while preserving the order in which the AE was reported in the CRF.

Lower and Upper Limit of Quantitation

In general, for quantitative laboratory values reported as "<" or "≤" the lower limit of quantitation (LLOQ), one-half of the reported value (i.e., LLOQ/2) will be used for analysis. The exception to this data treatment is for plasma concentrations of BNC210 that are reported as <LLOQ, where a value of zero will be used in calculating summary statistics.

For quantitative laboratory values reported as ">" or "≥" the upper limit of quantitation (ULOQ), the reported value (i.e., ULOQ) will be used for analysis.

Laboratory Test Results

For analysis purposes, repeat laboratory test results will not be used unless the original laboratory value is missing or indicated as invalid, in which case the first non-missing repeat laboratory value will be used for data analysis.

[redacted]

The International System of Units (SI) will be used in reporting all efficacy and safety laboratory values.

Treatment Duration and Exposure

Treatment Duration and Exposure will not be calculated given this is a single dose study with study drug given at the site. Participants who are not dosed will not be included in the safety or efficacy analyses.

6.2. Special Handling for COVID-19 Disruptions

Study visits may be delayed/not performed as a result of COVID-19 disruptions (e.g., sites were closed or participants were under stay-at-home orders). Participants who must delay their study visit past the screening period will be re-screened.

7. Study Patients/Participants and Demographics

7.1. Disposition of Patients/Participants and Withdrawals

The number of participants randomized, completing the treatment day visit, completing all visits through follow-up, and withdrawing, along with reasons for withdrawal, will be tabulated overall and by treatment group. The number of participants in each analysis population will be reported.

7.2. Protocol Violations and Deviations

Protocol deviations will be listed. Protocol deviations will be classified as "Important" or "Non-Important". An important deviation poses a possible safety issue to the participant or it has a potential impact on the statistical analysis of the clinical data. A non-important deviation is identified as any protocol deviation that does not meet the criteria for an important deviation.

[redacted] Protocol deviations which are deemed to be "Important" and "Non-Evaluable" (i.e., a deviation that has a potential impact on the efficacy analysis), will be classified as major protocol deviations. Participants with at least 1 major deviation will be excluded from the PP analysis population.

Important protocol deviations may include:

- Significant and/or persistent dosing error
- Participant did not meet eligibility criteria and did not have a waiver or dispensation by medical monitor
- Error in randomization (i.e., received wrong drug)
- Use of prohibited concomitant treatment during participation in the trial

The final decision regarding inclusion and exclusion of participants from the analysis populations will be based on a final listing of protocol deviations. This will be determined during

[redacted]

a (blinded) review meeting before any unblinding occurs or database freeze/lock, with input from the Clinical and Biostatistics team members and approval from the Sponsor.

Protocol deviations will be summarized by type and by treatment group for the SAF population.

7.3. Demographics and Other Baseline Characteristics

Demographic variables will include age, sex, height, and weight. Information on race and ethnicity will be collected for any eventual analysis of differences in response to the study intervention in accordance with local regulatory requirements. Baseline participant characteristics will include whether the participant wore a mask during the performance phase of the public speaking challenge, medical history including confirmation of SAD diagnosis and LSAS total score at Screening. Baseline evaluation of the SUDS score, STAI-State score, and SPSS-N score will also be reported. These analyses will be presented for the SAF, FAS, and PP populations.

8. Efficacy Analysis

Efficacy analyses on the primary efficacy endpoint using MI will also evaluate select secondary endpoints, including evaluations of the change from baseline and [redacted], the average of the anticipation phase, and the average of the performance phase and will be performed on the FAS population. The primary efficacy analysis using observed data will also evaluate select secondary endpoints, including evaluations of the change from baseline and [redacted], the average of the anticipation phase, and the average of the performance phase and will be performed on the FAS and PP populations. Efficacy analyses on the other secondary endpoints will be performed on the FAS population, [redacted]. The FAS will be considered the primary population for efficacy analyses.

For all efficacy analyses, participants will be analyzed by the randomized treatment group assignment (BNC210 225 mg, BNC210 675 mg or placebo). All efficacy data will be presented in participant listings.

8.1. Primary Efficacy Analysis

The primary efficacy endpoint for this study is the mean change between Baseline and the average of the performance phase of a public speaking challenge in the SUDS for participants receiving a single dose of BNC210 (225 mg or 675 mg) compared to participants receiving placebo, where the performance phase is defined as up to six timepoints (0, 1, 2, 3, 4, 5 minutes) measured during a public speaking challenge and the average is defined as the mean of all measured timepoints during that phase.

A restricted maximum likelihood (REML)-based MMRM model with MI will be used as the primary analysis method. The repeated measures include post-baseline timepoints (i.e., the resting period and the average of up to 0, 1, 2 minutes in the anticipation phase) with change from baseline SUDS scores during a public speaking challenge as the dependent variable. Each change from baseline score will be derived as the difference from that timepoint (i.e., the resting

[redacted]

period, the anticipation phase, and the performance phase) with averages utilized in the anticipation and performance phases. The average for the anticipation and performance phases will thereby be used as each respective timepoint. The MMRM model will include fixed effects for treatment, timepoint, interaction between treatment and timepoint average, sex, use of a mask, and a covariate for baseline score, using an unstructured covariance structure. If the unstructured covariance model will not converge using the Newton-Raphson algorithm, the following alternatives will be tried in the order presented, stopping if convergence occurs:

1) Fisher Scoring Algorithm

2) Factor Analytic Structure

3) Successive univariate analytic regression method as proposed in Lu and Mehrotra (2010)

If still the mentioned methods fail, a more specific covariance structure, first the Toeplitz covariance structure (alternative 4) then the auto-regressive order 1 (AR[1]) (alternative 5) structures will be tested. A sandwich estimator will be used with structured covariance models to control Type I error. If still the model does not converge, additional covariance structures will be explored and the final decision will be documented in the CSR.

Two comparisons (BNC210 225 mg versus placebo and BNC210 675 mg versus placebo) will be made and the least squares (LS) mean difference of averages along with the corresponding 95% confidence interval (CI) will be reported. For each treatment arm, the LS mean difference along with the corresponding p value will be weighted equally across the 2 BNC210 arms.

The null hypothesis for the primary efficacy endpoint of the equality of BNC210 and placebo is:

> H_{01}: Mean change in SUDS scores between baseline and the average of the performance phase in a public speaking challenge between either single dose (225 mg or 675 mg) of BNC210 and placebo are equal.

The null hypothesis of equal treatment effect will be rejected if the statistical analysis results in a 2-sided p value for treatment through the average of the SUDS scores during the performance phase in a public speaking challenge of the study of less than or equal to 0.05. Least squares (LS) means will be calculated for each treatment group for each post-baseline timepoint in the model. The difference between both levels of BNC210 (225 mg and 675 mg) and placebo change from baseline in the average of the SUDS score in the performance phase will be estimated, with the corresponding 2-sided 95% CI constructed for each time point. The change from baseline LS means with standard error, 95% CI for the LS means, p value for testing if the LS mean is zero, LS mean difference between treatment groups (BNC210 225 mg minus placebo, BNC210 675 mg minus placebo) with standard error, 95% CI for the LS mean difference, and p value for testing if the treatment LS means are equal will be presented.

The trial will be claimed successful if the hypothesis of no treatment effect on the primary efficacy endpoint of the FAS population is rejected at the 0.05 (2-sided) significance level.

As a sensitivity analysis, the primary analysis MMRM without MI will be performed in the FAS and PP populations.

[redacted]

Descriptive summaries will also be provided for the primary endpoint. Data from all performance phase timepoints will be included in the MMRM analysis.

Graphical displays of the means and LS mean change from baseline in the SUDS Scores over time will be presented for each treatment group for all timepoints, with error bars denoting the standard error.

8.2. Secondary Efficacy Analysis

Analyses similar to that described for the primary efficacy endpoint will be performed for each of the following secondary endpoints: Mean changes in SUDS scores between baseline and the average of the anticipation phase in a public speaking challenge between either single dose (225 mg or 675 mg) of BNC210 and placebo, [redacted] will be analyzed using MMRM with MI in the FAS population and MMRM without MI in the FAS and PP populations as part of the primary efficacy models.

[redacted]

The null hypotheses for the secondary efficacy endpoints of the equality of a single dose of BNC210 and placebo are:

H_{02}: Mean changes in SUDS scores between baseline and the average of the anticipation phase in a public speaking challenge between either single dose (225 mg or 675 mg) of BNC210 and placebo are equal.

[redacted]

H_{06}: Mean changes in STAI-State scores between baseline and the performance phase in a public speaking challenge between either single dose (225 mg or 675 mg) of BNC210 and placebo are equal.

H_{07}: Mean changes in STAI-State scores between baseline and the anticipation phase in a public speaking challenge between either single dose (225 mg or 675 mg) of BNC210

[redacted]

and placebo are equal.

H_{08}: Mean changes in SSPS-N scores between baseline and the performance phase in a public speaking challenge between either single dose (225 mg or 675 mg) of BNC210 and placebo are equal.

H_{09}: Mean changes in SSPS-N scores between baseline and the anticipation phase in a public speaking challenge between either single dose (225 mg or 675 mg) of BNC210 and placebo are equal.

[redacted]

For evaluations of SSPS-N, the SSPS-Trait version is used at Baseline/Pre-Dose and Resting/Minute 55 Post-Dose, and the SSPS-State version is used for the anticipation phase, performance phase and post-challenge phase.

Graphical displays of the means and LS mean change from baseline in the STAI-State Scores and SSPS-N Scores over time will be presented for each treatment group for all timepoints, with error bars denoting the standard error.

[redacted]

[redacted]

9. Safety and Tolerability Analysis

All safety analyses will be performed on the SAF population, defined as all participants who were randomized and received 1 dose of study medication.

For all safety and tolerability analyses, participants will be analyzed by the treatment received (BNC210 225 mg, BNC210 675 mg, or placebo) and, if applicable, overall for all participants receiving the study medication.

Safety data collected at the Baseline visit (treatment day) or the last preceding visit if not collected at Baseline will be used as the baseline value for safety analyses.

Safety measures including AEs, C-SSRS, physical examination findings, ECGs, vital signs, and clinical laboratory test results (hematology, biochemistry, urinalysis) will be summarized descriptively. For quantitative variables, descriptive statistics including number of observations, mean, median, standard deviation, minimum, and maximum will be presented for observed and change from baseline values at each study visit. Qualitative variables will be summarized using counts and percentages. No formal inferential analyses will be conducted for safety variables unless otherwise noted.

All safety and tolerability data will be presented in participant listings.

9.1. Adverse Events

In general, treatment emergent adverse events (TEAEs) are defined as AEs that are newly occurring or worsening after the first dose of study drug. In this study they are defined as AEs that start or deteriorate on or after the administration of study medication and no later than 7 days following the dose of study medication or reported through the Follow-up Day 7. For any participants who die during the study and the date of death is between the date of study medication administration and the date of study discontinuation (as entered by the site), inclusive, all AEs (including those resulting in death) that occur during the study will be considered as TEAEs and will be included in the TEAE summaries. All summaries of AEs will be based on TEAEs unless specified otherwise.

All AEs, TEAEs, and SAEs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA), Version 23.0 or higher.

The incidence of TEAEs will be summarized by treatment group, by system organ class (SOC) and preferred term (PT), by severity, and by relationship to study intervention.

The number and percent of participants reporting TEAEs, grouped by MedDRA SOC and PT (coded using MedDRA v. 23.0 or higher), will be tabulated by severity and treatment group. In the case of multiple occurrences of the same TEAE within the same participant, each participant will only be counted once for each preferred term.

In the summaries showing severity and relationship to study medication the event with the maximum severity (severe) or strongest relationship (definitely related) will be reported. If a particular event is missing the severity and/or relationship, then the strongest possible severity or relationship will be assumed for analysis (severity = severe, relationship = definitely related).

[redacted]

In the AE data listings, all AEs will be displayed.

9.1.1. Adverse Events Leading to Withdrawal

A data listing of AEs leading to withdrawal from study will also be provided, displaying details of the event(s) captured on the CRF.

9.1.2. Deaths and Serious Adverse Events

Any deaths that occur during the study will be listed.

Serious adverse events (SAEs) will be listed.

An SAE is an AE defined as any untoward medical occurrence that, at any dose, meets one or more of the criteria listed:

1) Results in death.
2) Is life-threatening.
3) It requires inpatient hospitalization or prolongation of existing hospitalization.
4) Results in a congenital abnormality or birth defect.
5) It is an important medical event that may jeopardize the participant or may require medical intervention to prevent one of the outcomes listed above.
6) Results in persistent or significant disability/incapacity

9.2. Clinical Laboratory Evaluations

Central laboratory tests (hematology, biochemistry, urinalysis) will be performed at Screening, Baseline/Pre-Dose, and Post-Challenge.

A serum pregnancy test (if applicable), HIV, HBsAg, and HCV tests will be performed at Screening. A urine pregnancy test (if applicable) will be conducted at Baseline/Pre-Dose. Urine drug screening tests will be conducted at Screening and Baseline/Pre-Dose. These results will not be reported unless applicable for enrollment/discontinuation reasons.

Laboratory test results will be summarized descriptively by treatment and time point as both observed values and change from baseline values.

The number of participants with clinical laboratory values below, within, or above the normal range by time point and in relation to Baseline will be tabulated for each clinical laboratory analyte by treatment group.

A listing of all laboratory analytes by category (hematology, biochemistry, urinalysis) will be provided.

9.3. Vital Signs

Vital signs will be collected at Screening, Baseline/Pre-Dose, and Post-Challenge. Descriptive summaries of actual values and changes from baseline will be calculated for body temperature,

[redacted]

pulse rate, respiratory rate, and seated systolic and diastolic blood pressure (at 5 minutes rest). Vital sign measurements, including body temperature, will also be listed.

9.4. Electrocardiograms

Triplicate 12-Lead ECGs will be collected at Screening and Post-Challenge. Descriptive summaries and change from screening will be presented for ECG measures of PR interval, QRS interval, QT interval, QT_CF interval, and HR. These summaries will be presented by study visit and treatment group.

The number and percentage of patients/participants with normal and abnormal ECG results will be summarized for the SAF population by treatment group at each visit. Summary statistics for all collected visits will be displayed for QT and both QT_C correction methods by treatment group. Fridericia's correction for QT_C will be applied as follows:

Fridericia's correction (QT_Cf) will be derived as follows:

$$\text{Fridericia's Correction } (QT_Cf) \quad QTc_f = \frac{QT_{msec}}{\sqrt[3]{RR}}$$

where: Relative Rate: RR = 60 / HR. HR = Heart Rate obtained from the ECG.

9.5. Physical Examinations

A full physical examination will be done at Screening and an abbreviated physical exam will be completed at Baseline/Pre-Dose. A listing of physical examination findings will be provided for complete and brief examinations.

Complete physical examinations will include assessments of the general appearance, skin and lymphatics, eyes, ears, nose, throat, cardiovascular system, respiratory system, abdomen/gastrointestinal system, musculoskeletal and neurological systems. Other body systems may also be examined as required. Height and weight will also be measured at Screening only.

Abbreviated physical examinations will include, at a minimum, assessments of the eyes, ears, nose, throat, cardiovascular system, respiratory system, and abdomen/gastrointestinal system. Other body systems may also be examined as required.

9.6. Further Safety Evaluations

9.6.1 Columbia Suicide Severity Rating Scale (C-SSRS)

The C-SSRS is an instrument to assess both suicidal behavior and ideation. The Screening version will be administered at Screening and the Since Last Visit version of the scale will be administered at Baseline/Pre-Dose. C-SSRS data will be listed.

[redacted]

9.7. Concomitant Medication

Prior and concomitant treatments are coded using WHO-DD Global B3 version March 2020 and will be listed.

Prior treatments will be flagged separately from concomitant treatments. Treatments that started before administration of the study medication will be considered prior treatments, whether or not they were stopped before study medication. Any treatments continuing or starting after study medication will be considered to be concomitant. If a treatment starts before the study medication and continues after the study medication, it will be considered both prior and concomitant.

[redacted]

11. Changes from Planned Analysis

Due to the nature of the repeated measures models for the primary analysis, secondary endpoints that evaluate CFB at the [redacted] anticipation phases are included in all primary efficacy analyses and are denoted as select secondary endpoints. [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

12. References

1. US Federal Register. (1988) International Conference on Harmonization; Guidance on Statistical Principles for Clinical Trials. Department of Health and Human Services: Food and Drug Administration [Docket No. 97D-0174]. Federal Register Volume 63, Number 179, pages 79583-49598. September 16, 1998.
2. ASA. (2018) Ethical Guidelines for Statistical Practice. Prepared by the Committee on Professional Ethics, April 2018. http://www.amstat.org/about/ethicalguidelines.cfm
3. ICH (1998). ICH Harmonized Tripartite Guideline. Statistical Principles for Clinical Trials E9; 1998. https://database.ich.org/sites/default/files/E9_Guideline.pdf
4. RSS. (2014) The Royal Statistical Society: Code of Conduct, 2014. https://rss.org.uk/about/policy-and-guidelines/code-of-conduct/.
5. Liebowitz, M.R., Salman, E., Nicolini, H., Rosenthal, N., Hanover, R., & Monti, L. (2014). Effect of an Acute Intranasal Aerosol Dose of PH94B on Social and Performance Anxiety in Women with Social Anxiety Disorder. *American Journal of Psychiatry,* 171, 675-682.
6. Mallinckrodt C.H., Lane P.W., Schnell D., Peng Y., Mancuso J.P. 2008. Recommendations for the Primary Analysis of Continuous Endpoints in Longitudinal Clinical Trials. *Drug Information Journal*, 42, 303-319.
7. Molenberghs G., Kenward M.G. 2007. *Missing Data in Clinical Studies*. New York: Wiley.
8. Cnaan, A., Laird, N.M., & Slasor, P. (1997). Using the General Linear Mixed Model to Analyse Unbalanced Repeated Measures and Longitudinal Data. *Statistics in Medicine*. 16, 2349-2380.
9. Moelnberghs, G., Thijs, H., Jansen, I., Beunckens, C., Kenward, M.G., Mallinckrodt, C.,

& Carrol, R.J. (2004). Analyzing incomplete longitudinal clinical trial data. *Biostatistics.* 5, 445-464.

10. Mallinckrodt C. H., Clark W. S., David S. R. (2001). Accounting for dropout bias using mixed-effects models. *Journal of Biopharmaceuticals Statistics*. 11, 9-21.
11. European Medicines Agency. 2010. Guideline on Missing Data in Confirmatory Clinical Trials.
12. Siddiqui O., Hung H.M., O'Neil R. 2009. MMRM vs. LOCF a comprehensive comparison based on simulation study and 25 NDA datasets. *Journal of Biopharmaceuticals Statistics*, 19(2), 227-246.
13. Mallinckrodt C.H., Roger J., Chuang-Stein C., Molenberghs G., O'Kelly M., Ratitch B., Jannsens M., Bunouf P. 2013. Recent Developments in the Prevention and Treatment of Missing Data. *Drug Information Journal,* 48, 68-80.
14. Little R., Rubin D. 1987. *Statistical Analysis with Missing Data*. New York: John Wiley.
15. Verbeke G., Molenberghs, G. 2000. *Linear Mixed Models for Longitudinal Data*. New York: Springer.
16. Lu, K., Mehrotra, D.V. 2010. Specification of covariance structure in longitudinal data analysis for randomized clinical trials. *Statistics in Medicine,* 29, 474-488.
17. Ratitch, B., O'Kelly, M., Tosiello, R. 2013. Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. *Pharmaceutical Statistics.* 21(6), 337-347.