[redacted]

STATISTICAL ANALYSIS PLAN

[redacted]

PROTOCOL NUMBER: BNC210.007

TITLE: A Randomized, Double-blind, Placebo-controlled Phase II Study of BNC210 in Adults with Post-Traumatic Stress Disorder (PTSD)

STUDY DRUG: BNC210

AUTHOR: [redacted]

SPONSOR:
Bionomics Limited
31 Dalgleish Street
Thebarton SA 5031
AUSTRALIA

DATE FINAL: 27th September 2018

DOCUMENT VERSION: FINAL V1.6

[redacted]

TITLE: A Randomized, Double-blind, Placebo-controlled Phase II Study of BNC210 in Adults with Post-Traumatic Stress Disorder (PTSD)

AUTHORS: [redacted]

NAME (TITLE):	**DATE**	**SIGNATURE**
[redacted]	_____	_____
[redacted]		

REVIEWERS: [redacted]

NAME (TITLE):	**DATE**	**SIGNATURE**
[redacted]	_____	_____
[redacted]		
NAME (TITLE):	**DATE**	**SIGNATURE**
[redacted]	_____	_____
[redacted]		

[redacted]

VERSION HISTORY

Document	Summary of Change	Date of Issue
Draft Version 1.0		31 May 2018
Final Version 1.1	Revised following comments from Bionomics Ltd [redacted]	15th July 2018
Final Version 1.2	Revised following comments from Bionomics Ltd [redacted]	10th August 2018
Final Version 1.3	Included repeating Demographic outputs for mITT Population	21st August 2018
Final Version 1.4	Included additional subgroup analysis (Section 6.4), and clarified how non-numeric laboratory results will be reported (Section 11.4)	23rd August 2018
Final Version 1.5	Include analysis of PGI-I and CGI-I results (Sections 9.1.3, 9.1.4 and 10.2)	25th September 2018
Final Version 1.6	Include Subgroup analyses for MADRS and HAM-A (Section 6.4), and the analysis of the CAPS-5 Criterion B – Question B5 score (Sections 9.1.1, 10.1.2 and 10.2)	27th September 2018

[redacted]

LIST OF ABBREVIATIONS

Abbreviation	Definition
AE	Adverse Event
AIC	Akaike's Information Criterion
ALP	Alkaline Phosphatase
ALT	Alanine Aminotransferase (SGPT)
AQoL	Assessment of Quality of Life
AST	Aspartate Aminotransferase (SGOT)
ASL	Arterial Spin Labeling
AUC	Area Under Curve
BIC	Bayesian Information Criterion
b.i.d.	Twice Daily
BMI	Body Mass Index
BOCF	Baseline Observation Carried Forward
BUN	Blood Urea Nitrogen
C-SSRS	Columbia Suicide Severity Rating Scale
CAPS-5	Clinician-Administered PTSD Scale for the DSM-5
CGI-I	Clinical Global Impressions – Improvement
CGI-S	Clinical Global Impressions – Severity
CHRNA7	Nicotinic Acetylcholine Receptor Subunit Alpha-7
CI	Confidence Interval
Cmax	Maximal Observed Plasma Concentration
CRP	C-reactive Protein
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, 5th Edition
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
HAM-A	Hamilton Anxiety Rating Scale
IP	Investigational Product
IRB	Institutional Review Board
ITT	Intent-to-treat
IWRS	Interactive Web Response System
LEC-5	Life Events Checklist for DSM-5 including interview with abuse ratings.
LOCF	Last Observation Carried Forward
LSMeans	Least Squares Means
MADRS	Montgomery-Åsberg Depression Rating Scale
MAR	Missing At Random
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple Imputation
mITT	Modified Intent-to-Treat

[redacted]

MMRM	Mixed Model for Repeated Measurements
MSI-BPD	McLean Screening Instrument for Borderline Personality Disorder
PAL	Paired Associates Learning
PCL-5	Standard PTSD Checklist for DSM-5
PD	Pharmacodynamic
PGI-I	Patient Global Impressions – Improvement
PGI-S	Patient Global Impressions – Severity
PK	Pharmacokinetics
PRN	As Needed
PSQI	Pittsburgh Sleep Quality Index
PT	Preferred Term
PTSD	Post-traumatic Stress Disorder
QTcF	Fridericia-corrected QT Interval
RVP	Rapid Visual Information Processing
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Standard Deviation
SDS	Sheehan Disability Scale
SOC	System Organ Class
SOP	Standard Operating Procedure
SSID	Study Subject Identification Number
SSRI	Selective Serotonin Reuptake Inhibitor
SWM	Spatial Working Memory
TEAE	Treatment-emergent Adverse Event
ULN	Upper Limit of Normal

[redacted]

CONTENTS

[redacted]

1. INTRODUCTION

This document describes the planned statistical analyses for Protocol BNC210.007, a randomized, double-blind, placebo-controlled Phase II study of BNC210 in adults with Post-Traumatic Stress Disorder (PTSD).

The planned statistical analyses of Protocol BNC210.007 shall be described in this statistical analysis plan (SAP).

This SAP supplements the study protocol BNC210.007 Final Version 4.0 dated 30th May 2018. If the protocol is amended, this SAP will be revised as required. The plan will be finalized before the study database is locked.

2. STUDY OBJECTIVES

2.1 Primary Objective

- To assess the effects of BNC210 on Investigator-rated symptoms of PTSD measured by Clinician-Administered PTSD Scale for the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5) (CAPS-5) scores.

2.2 Secondary Objectives

- To assess the effects of BNC210 on Investigator-rated symptoms of PTSD measured by each of the four symptom clusters (Intrusion; Avoidance, Negative Alterations in Cognition and Mood; Arousal and Reactivity) of the CAPS-5.
- To assess the effects of BNC210 on other psychiatric outcomes in subjects with PTSD including anxiety and depression.
- To assess the effects of BNC210 on global functioning and Quality of Life in subjects with PTSD.
- To assess the effects of BNC210 on patient-reported outcomes in subjects with PTSD.
- To assess the safety and tolerability of BNC210 in subjects with PTSD.

[redacted]

[redacted]

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

3. STUDY DESIGN

3.1 General Description

This is a randomized, double-blind, parallel, placebo-controlled, multi-centre study with a 12-week, 4-arm treatment phase (placebo and BNC210). Subjects will attend a Screening visit within 3 weeks before randomization to confirm eligibility. Approximately 192 subjects who fulfil the inclusion criteria and none of the exclusion criteria will be randomized on Day 1 using a 1:1:1:1 ratio. Subjects will then complete 12 weeks of their allocated treatment. The Investigational Product (IP) (total daily dose 300 mg, 600 mg or 1200 mg) or Placebo will be taken twice daily (b.i.d.) with food. Subjects will return to the clinical site at weekly or 2-weekly intervals to receive study medication, complete safety assessments and questionnaires. Subjects are then requested to attend a post-study visit at 15 weeks (that is 3 weeks after the last treatment).

3.2 Length of Study

Screening evaluation within 3 weeks of randomization. Treatment phase: 12 weeks of treatment, b.i.d. (weekly or fortnightly visits to the clinic). Post-study visit: 3 weeks after

[redacted]

the last treatment (i.e. 15 weeks after the first treatment). Recruitment timeframe estimate is 2 years.

3.3 Schedule of Events

The schedule of assessments can be found on pages 9-10 of the final protocol.

3.4 Treatment Assignment

An Interactive Web Response System (IWRS) will allocate a randomization number at the baseline visit but the number will not be used for subject identification purposes. Data associated with subjects who fail screening will not be entered into the eCRF. [redacted] subjects will be randomly assigned to the four treatment groups in a 1:1:1:1 ratio until the close of randomization. The IWRS vendor will be supplied with a validated randomization list by an Independent Statistician.

3.5 Determination of Sample Size

Sample size estimation is based on the following. [redacted]

[redacted] With the level of significance set to 0.05 and a 2-sided test, a sample size of 36 subjects per arm is required to achieve a power of 80%. For PTSD patients a relatively high withdrawal rate is anticipated, therefore the sample size was increased by 33% to 48 per arm (=1.33x36) (Hoskins et al. 2015; Marshall et al. 2001; Marshall et al. 2007; Tucker et al. 2001). Because this is a Phase II study the level of significance will be 0.05 with no adjustment for multiple comparisons.

3.6 Internal safety review

The safety data shall be monitored throughout by a subject matter expert.

3.7 Independent Safety Review Committee

An independent Safety Review Committee will be established to monitor subject safety throughout the study. The meetings will occur approximately every 6 months and will review all safety data including vital signs, electrocardiogram (ECG), pathology results, physical examinations, Columbia Suicide Severity Rating Scale (CSSRS) and continuous adverse event (AE) reporting.

3.8 Interim Analysis

There are no planned interim analyses.

3.9 Final Analysis

The final analysis will be conducted once all enrolled subjects have completed the study and the database lock has occurred.

[redacted]

4. ANALYSIS POPULATIONS

4.1 All Enrolled Population

The All Enrolled population will include all subjects who provided informed consent.

4.2 ITT/mITT Populations

The Intent-to-treat (ITT) population will comprise all randomised subjects; analysed as per randomised treatment. This population will be used for subject disposition summaries and all listings unless otherwise specified. The Modified ITT (mITT) population will be those subjects from the ITT population who received at least one dose of study medication. The mITT population will use treatment groups as randomised and will be used for all efficacy analyses, tabulations, and listings.

4.3 Safety Population

The safety population will comprise all randomised subjects who receive at least one dose of study medication and will be analyzed as per actual treatment received. The safety population will be used for all safety analyses and listings.

The analysis sets will be validated during a blind review meeting.

4.4 Description of the Populations

A summary table with the description of the number of subjects in each analysis set, the number of subjects who completed the study, and the number of subjects who discontinued classified by reason of withdrawal, will be prepared. Corresponding individual listings will be prepared.

Listings with end of study status and visit dates will also be prepared.

[redacted]

5. DATA HANDLING CONVENTIONS

5.1 Treatment Groups

Data will be grouped by treatment group, using the following treatment names.

- Placebo
- BNC210 150 mg b.i.d.
- BNC210 300 mg b.i.d.
- BNC210 600 mg b.i.d.

Subjects will take 150 mg or 300 mg or 600 mg BNC210 or placebo b.i.d., with food, as oral suspensions from single-use [redacted] bottles for 12 weeks. Doses should be administered approximately 12 hours apart (±2 hours), with breakfast, and with an evening meal.

5.2 Data Presentation

In the Schedule of Assessments in the Protocol the visit representing 12 weeks after the start of treatment is labelled Week 13. The outputs presenting the results from this study will label this visit as Week 12. Similarly, other visits will be labelled as in the table below:

Protocol - Schedule of Assessments		Outputs for Reporting
Day	Week	Week
1	1	Baseline
8	2	1
15	3	2
29	5	4
43	7	6
57	9	8
71	11	10
85	13	12
106	16	15

:

[redacted]

Descriptive analyses will be completed for this study. For continuous parameters, these will include the number of observations, mean, median, standard deviation, minimum and maximum. The minimum and maximum values will be presented to the same number of decimal places as recorded in the database. Mean, median and standard deviation will be presented to 1 more decimal place than the raw data. Categorical parameters will be summarised as frequency counts and percentages. Percentages will be rounded to 1 decimal place, with the denominator being the number of subjects in the relevant population, unless otherwise stated. All confidence intervals reported will be two-sided 95% confidence levels, unless specified otherwise.

Summary tables, and listings for baseline data, drug exposure, efficacy and safety will be presented as four treatment groups as described in Section 5.1.

The baseline assessment at Day 1 will be used as a reference point for all four treatment groups when summarising the change from baseline.

The expected number of subjects in each treatment group will be as follows:

- Placebo (N=48)
- BNC210 150 mg b.i.d. (N=48)
- BNC210 300 mg b.i.d. (N=48)
- BNC210 600 mg b.i.d. (N=48)

5.3 Premature Withdrawal and Missing Data

The primary analysis of the CAPS-5 measure will be conducted based on observed and imputed data. All observed data for subjects who withdraw early will be used for the

analysis up to the point of their discontinuation. The Multiple Imputation (MI) approach used to impute values for missing data is described in Section 10.1.1.

5.4 Common Derivations

5.4.1 Baseline

For all parameters presented by visit in the summary tables, baseline will be defined as the last, valid, non-missing assessment prior to the first study drug administration.

5.5 Visit Window Conventions

Visit windows will not be applied to the data except for deriving baseline assessments (Section 5.4.1). Visit numbers (Weeks post Baseline) recorded on the eCRF will be used for the analysis and reporting of the data (except for BNC210 plasma concentration data where the recorded dates will be used to determine time (days) since Baseline).

5.6 Software

All analyses will be performed using SAS version 9.4 or later. Some figures will be created using R 3.2.2. The data processing, analysis, and validation for the figures created in R will be conducted using SAS version 9.4 or later.

6. STATISTICAL CONSIDERATIONS

6.1 Multiplicity Adjustments

There will be no adjustment for multiplicity.

6.2 Multi-center Studies

Given the small sample sizes at some centers there will be no adjustment for center.

6.3 Other Strata and Covariates

Gender, Age, Country (Australia/USA), Nicotine Use at Baseline (Current/Former/Never) and Concomitant Antidepressant Use at Baseline (Yes/No) will be included as covariates in the primary efficacy analysis model (two subjects reported to be transgender will be

[redacted]

excluded from this analysis). Concomitant Antidepressant Use will be based on data recorded on the Concomitant Medication page on the eCRF and where the Start to Stop date period includes the Date of First Dose of Study Medication. The list of medications, and associated indication, will be reviewed by a Clinician (this will be completed prior to database lock).

6.4 Sub-Group Analyses

The primary endpoint (CAPS-5 Total Score) and key secondary endpoints (MADRS and HAM-A) will be repeated for the following sub-group:

- Country: Australia, USA
- Baseline Nicotine use (Current/Former/Never)
- Age (≤median, > median)
- Concomitant Antidepressant Use at Baseline (Yes/No)
- Gender (Male, Female) (subjects reported to be transgender will not be included).
- Time (Years) since initiating PTSD event (≤10 years, > 10 years).

7. OUTPUT SHELLS

Template shells for the tables, listings, and figures are presented in [redacted]

8. STUDY POPULATION

8.1 Subject Disposition

The number and percentage of subject who were included in each analysis population, completed the study, and completed each visit will be presented by treatment group. The reason for early withdrawal will be summarized and listed by treatment group. In addition, the number of days from Randomisation to Withdrawal/Last Visit will be determined for each subject and the treatment group results summarized and presented using Kaplan-Meier methodology.

[redacted]

8.2 Protocol Deviations

A protocol deviation is defined as any intentional or unintentional change to, or noncompliance with, the approved protocol, procedures, or requirements. Deviations may result from the action or inaction of the subject, investigator, or site staff. All deviations will be tracked and should be reported to Institutional Review Boards (IRBs) in accordance with their reporting policy.

Possible protocol deviations will be documented in the source documents and tracked by the Sponsor.

A final list of protocol deviations will be approved by the Sponsor, the Medical Monitor, and others as appropriate, and after database lock provided to McCloud Consulting Group. The final list of Important protocol deviations will include date of deviation, deviation, reason for deviation, action taken, and comments, and will be included in the final study report.

8.3 Demographics and Baseline Characteristics

Baseline data will be used to describe the safety population (and repeated for the mITT population). This information will include the following:

- Demographics including gender (and female reproductive status), age, race and country (Australia or USA)
- Medical History
- Screening weight, height, and Body Mass Index (BMI)
- Time (Years) since initiating PTSD event
- Screening Nicotine Use Status
- Baseline Antidepressant Use

Baseline measurements will be summarised, by treatment group, as number of observations, mean, standard deviation, minimum, median, and maximum for continuous variables, and number of observations, frequencies and percentages for categorical

variables. Age will be derived as (date of initial informed consent–date of birth+1)/365.25. All baseline data will be summarized descriptively in tables and listed by treatment group.

9. EFFICACY VARIABLES

9.1 Definitions of Variables

All the assessments/questionnaires below will be assessed at time-points as captured in the Schedule of Assessments found in the protocol. Imputation of missing Domain or Total Scores will be performed using MI as described in Section 10.1.1. The imputation will be at the level of the Domain or Total Scores. In other words the imputation of the Domain or Total Scores will not be done for the individual questions comprising each Domain or Total Score.

9.1.1 Clinician-Administered PTSD Scale for the DSM-5 (CAPS-5)

The CAPS-5 is a 30-item structured interview used to diagnose PTSD and assess PTSD symptoms over the past week. Information about the frequency and intensity of each item is combined into a severity rating, and the CAPS-5 total severity is calculated by adding the severity scores for the 20 PTSD symptoms in the DSM-5 (National Centre for PTSD). The severity of PTSD is rated from 0 (absent) to 4 (extreme/incapacitating).

In addition to the total CAPS-5 score, the cluster scores for:

- Criterion B: Intrusion (the sum of the individual severity scores for *Intrusion symptoms* items 1-5);
- Criterion B: Intrusion – Question B5
- Criterion C: Avoidance (the sum of *Avoidance symptoms* items 6 and 7);
- Criterion D: Negative Alterations in Cognition and Mood (the sum of *Cognitions and mood symptoms* items 8-14); and

- Criterion E: Arousal and Reactivity (the sum of *Arousal and reactivity symptoms* items 15-20)

will be analysed and reported separately as secondary measures of efficacy.

9.1.2 PTSD Checklist for DSM-5 (PCL-5)

The PCL-5 is a 20-item self-report assessment of the 20 DSM-5 symptoms of PTSD (National Centre for PTSD). Each symptom is rated on a scale from 0 (not at all) to 4 (extremely). The Total Score is the sum of the responses to the 20 items. If some responses are missing then the following approach will be adopted - if no more than 5 responses are missing then calculate a Total score (and assume missing responses have a value of zero); if more than 5 responses are missing then Total score is missing.

9.1.3 Clinical Global Impression scale–Improvement and Severity (CGI-I/CGI-S)

The CGI-S is a rating scale designed to assess the severity of the subject's symptoms, and the CGI-I is designed to assess the change in the subject's condition since the initial assessment. Severity is rated on the CGI-S from 1 (normal, not at all ill) to 7 (among the most extremely ill of subjects). The changes in the subject's condition are rated on the CGI-I from 1 (very much improved) to 7 (very much worse). Improvement will also summarised as binary responses (1) (very much improved/much improved OR other) and (2) (very much improved OR other).

9.1.4 Patient Global Impression Scales – Improvement and Severity (PGI-I/PGI-S)

The PGI-S is a self-rating scale designed to assess the severity of the subject's symptoms, and the PGI-I is designed to assess the change in the subject's condition since the initial assessment. Severity is rated on the PGI-S from 1 (normal) to 4 (severe). The changes in the subject's condition are rated on the PGI-I from 1 (very much improved) to 7 (very

much worse). Improvement will also summarised as binary responses (1) (very much improved/much improved OR other) and (2) (very much improved OR other).

9.1.5 Assessment of Quality of Life (AQoL-8D)

The AQoL-8D is a 35 question self-rating scale designed to assess a subject's quality of life on the following eight dimensions; Independent Living, Happiness, Mental Health, Coping, Relationships, Self-Worth, Pain and the Senses. The scale takes approximately 5 minutes to complete. The Total Score is the sum of the responses to the 35 items. If some responses are missing then the following approach will be adopted - if no more than 5 responses are missing then calculate a Total score (and assume missing responses have a value of zero); if more than 5 responses are missing then Total score is missing.

9.1.6 Hamilton Anxiety Rating Scale (HAM-A)

The HAM-A is an interview questionnaire that measures severity of anxiety symptoms based on 14 parameters, including anxious mood, tension, fears, insomnia, somatic complaints and behaviour during the interview (Hamilton, 1959). It takes approximately 15-20 minutes to complete and each parameter is rated on a scale of 0 (not present) to 4 (very severe).

9.1.7 Montgomery-Asberg Depression Rating Scale (MADRS)

The MADRS is a 10-item self-diagnostic and clinician-rated questionnaire to measure the presence and severity of depressive episodes and includes the following symptoms: 1) apparent sadness; 2) reported sadness; 3) inner tension; 4) reduced sleep; 5) reduced appetite; 6) concentration difficulties; 7) lassitude; 8) inability to feel; 9) pessimistic thoughts; and 10) suicidal thoughts (Williams and Kobak, 2008). The rater must decide whether the rating lies on defined scale steps (0, 2, 4, 6) or between them (1, 3, 5).

[redacted]

9.1.8 CANTAB Cognitive Assessment

The CANTAB cognitive assessment is a battery of tests designed to detect both cognitive impairment and enhancement. The battery will measure cognitive function through assessing sustained attention, episodic memory, working memory and executive function. All tests will be completed electronically using iPads. The test battery will take approximately 20-25 minutes to complete. The CANTAB cognitive assessment must be completed at the same time during each visit (+ 1hr of baseline assessment start time).

A Paired Associates Learning (PAL) task will be used to assess episodic memory. During the test, boxes are displayed on the screen and open in turn to reveal a number of patterns. Participants are instructed to try to remember the location in which each pattern was shown. After all the boxes have been opened, each pattern is then shown in the centre of the screen in a randomized order, and the participant touches the box in which the pattern was located. If an error is made, all the patterns are re-presented to remind the participant of their locations. As the test progresses so the stages become more difficult as the number of patterns to be remembered increases, up to a maximum of 8 patterns. For participants who fail to complete all levels, an adjusted total is calculated that allows for errors predicted in the stages that were not attempted. The key outcome measures for this task are total errors and first trial memory score.

A Spatial Working Memory (SWM) task will be used to assess the ability of the subject to retain spatial information and manipulate it in working memory. It is a self-ordered task that also gives a measure of strategy. A number of coloured boxes are presented on the screen, and the computer hides a token in these boxes one at a time. The participant is instructed to touch the boxes in turn to search for the token that has been hidden. When a token is found it should be placed in a home area on the right side of the screen. The participant then searches for more tokens until the same number of tokens as the number of coloured boxes has been found. The key task instruction is that the computer will never hide a token in the same coloured box twice in the same problem. As the test progresses,

[redacted]

so it becomes more difficult. The key outcome measures for this task are between errors and strategy.

A Rapid Visual Information Processing (RVP) task will be used to assess sustained attention, outputting measures of response accuracy, target sensitivity and reaction times. For the RVP task, single digits appear in a pseudo-random order at a rate of 100 digits per minute in box in the centre of the screen. Subjects must detect a series of 3-digit target sequences (e.g. 3-5-7; 2-4-6; 4-6-8) and respond by touching the button at the bottom of the screen when they see the final number of the sequence. Nine target sequences appear every minute. Key outcome measures include target sensitivity and reaction times.

Factors that could influence the cognitive function assessments must be captured within the CANTAB system. This includes, but is not limited to, drug and medication use within 12 hours of assessment, lack of sleep, and failure to understand the test requirements.

In addition to the total composite score, the six z-scores:

- Z-Score (ZPALFAMS) for PAL first attempt memory score (PALFAMS)
- Z-Score (ZPALTEA) for PAL total errors adjusted (PALTEA)
- Z-Score (ZSWMBE) for SWM between errors (SWMBE)
- Z-Score (ZSWMS) for SWM strategy (SWMS)
- Z-Score (ZRVPA) for RVP A' prime (RVPA)
- Z-Score (ZRVPMDL) for RVP median latency (RVPMDL)

will be analysed and reported separately as secondary measures of efficacy.

[redacted]

[redacted]

9.1.9 Sheehan Disability Scale (SDS)

The SDS is a rating scale designed to measure impairment in three domains: work/school, social life, and family life/home responsibilities. Each item is rated on a scale of 0 (not at all) to 10 (extremely), for a total score of 0 to 30.

In addition to the total SDS score, the three domain scores:

- Work/School
- Social Life
- Family Life/Home Responsibilities

will be analysed and reported separately as secondary measures of efficacy.

9.1.10 Pittsburgh Sleep Quality Index (PSQI)

The PSQI is a questionnaire consisting of 19 questions designed to assess the following 7 components of sleep quality: subjective sleep quality, sleep latency (i.e., how long it takes to fall asleep), sleep duration, habitual sleep efficiency, sleep disturbances, use of sleeping medication, and daytime dysfunction over the previous month. Each question is weighted on a scale from 0 to 3 and a global score is obtained, with lower scores correlating with better sleep quality. [redacted]

10. EFFICACY ANALYSIS

10.1 Primary Efficacy Endpoint

The primary efficacy endpoint is CAPS-5 total symptom severity score and will be analysed with a Mixed Model for Repeated Measurements (MMRM) with sequential multiple imputation used to estimate the missing observations (as described below in Section 10.1.1). The repeated measurements will range from baseline to Week 12 (Day 85) for the primary endpoint.

[redacted]

The primary model will include the endpoint from days 8 to 85 as the dependent variable, the baseline (day 1) value as a baseline covariate, day, and treatment group as factors, and treatment group by day as an interaction term. Additional covariates in the model will include Gender (two subjects reported to be transgender will be excluded from this analysis), Age, Country (Australia/USA), Nicotine Use at Baseline (Current/Former/Never) and Concomitant Antidepressant Use at Baseline (Yes/No).

The following variance-covariance structures will be investigated (in the model using the observed CAPS-5 data): Compound Symmetry, Ante-Dependence Order 1, Unstructured, Autoregressive Order1, Toeplitz, and Variance Component. The most appropriate structure will be chosen for the final model by comparing the Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC) statistics, as well as using statistical judgement. If appropriate, the same variance-covariance structure will be used for all variables.

The primary comparisons of interest will be the difference in Least Squares means (LSMeans, means adjusted for the covariates in the model) at Week 12 between each of the three active study medication groups (separately) and placebo. The results will be presented as LSMeans for each of the four treatment groups at Week 12, and the LSMeans of treatment differences (between each of the three active medication dose groups and placebo at Week 12) and their associated 95% confidence intervals. The treatment group LSMeans overall, and the difference between the groups will also be presented [redacted].

Because this is a Phase II study, the level of significance will be 0.05 with no adjustment for multiple comparisons.

In addition to the analysis results, the summary statistics for the actual values at each of the visits (up to Week 15), and the change from Baseline, will be presented for each of the treatment groups. Plots of the observed treatment group means (and 95% confidence

[redacted]

intervals) over time, as well as the LSMeans (and 95%CI) over time from the MMRM for Imputed and Observed data will be produced [redacted].

10.1.1 Multiple Imputation of Missing Data in Primary Analysis

For the primary analysis, missing CAPS-5 values will be assumed to be missing at random (MAR). Under the MAR assumption, it is assumed that missing values are conditional on the independent variables in the missing data model and the observed dependent variable. In other words, subject dropout can be modelled by the observed covariates and outcomes. It is assumed that the statistical behaviour of the unobserved data, conditional on the observed covariates and outcomes, is what it would have been, had it been observed. The MAR mechanism assumes that missing data is not related to unobserved covariates.

This clinical trial was designed to observe all relevant information, therefore, it is reasonable to assume that the missing data follows the MAR mechanism. As such, missing values in the primary analysis will be imputed using multiple imputation with 100 samples drawn, which is an appropriate method for handling data which is MAR.

The analysis will be conducted using all subjects in the mITT population. For subjects with a missing baseline CAPS-5 score the screening assessment score will be used as the baseline score. The analysis will include subjects who were randomised but did not go on to complete any post baseline efficacy assessments.

There are 3 stages to the multiple imputation process, as described below (Mallinckrodt C, and Lipkovich I, 2017).

1. Imputation Phase

During this phase, 100 datasets will be created where missing CAPS-5 values will be imputed by randomly selecting data from an appropriate distribution.

First, only intermittent missing data will be imputed, so that the remaining missing data has a monotone missing data pattern. For the imputation of the intermittent missing data, all variables in the imputation model will be assumed to have a multivariate joint normal distribution and will be modelled as continuous variables. The variables in the imputation model will include the variables in the primary analysis MMRM model.

A Markov Chain Monte Carlo (MCMC) procedure will be used, where missing data will be selected from a multivariate joint normal distribution conditional on the observed data.

Following the imputation of the intermittent missing data, the remaining monotone missing data will be imputed using stepwise sequential Bayesian regression.

The multiple imputation procedures will be implemented using PROC MI in SAS with a random seed of 9023476. Imputed data that falls outside the plausible range for the CAPS-5 (i.e. 0 to 80) will not be truncated and will be used in the analysis as imputed. This is because post-imputation truncation of data can lead to biased estimates (Rodwell, Lee, Romaniuk, & Carlin, 2014).

2. Analysis Phase

The 100 imputed datasets will each be analysed using the planned statistical model described in Section 10.1.

3. Pooling Phase

The parameter estimates from the 100 analyses conducted in step 2 will be combined into a single set of statistics that reflect the uncertainty associated with the imputed values using PROC MIANALYZE in SAS. Using the parameter estimates and standard errors from the 100 imputed datasets, this procedure applies Rubin's (1987) rules to make valid univariate statistical inferences.

[redacted]

10.1.2 Sensitivity Analyses

As a sensitivity analysis of the CAPS-5 Total Score and four cluster scores, and the CAPS-5 Criterion B – Question B5, the MMRM will be implemented on:

- The observed data (no imputation will be applied for the missing data). Mixed models with repeated measures are an alternative method that can be used in the presence of missing data, when the data is assumed to be MAR.

Additional sensitivity analysis of the CAPS-5 Total Score the MMRM will be implemented on:

- The data where the missing values are substituted using the Last Observation Carried Forward (LOCF) approach
- The data where the missing values are substituted using the Baseline Observation Carried Forward (BOCF) approach.

10.2 Secondary Efficacy Measures

For each continuous secondary efficacy variables listed in Section 9, including the four CAPS-5 Cluster scores, the analysis methodology described, and outputs proposed, in Section 10.1 will be followed. These continuous secondary efficacy measures include:

- CAPS-5 Criterion B: Intrusion (the sum of the individual severity scores for *Intrusion symptoms* items 1-5);
- Criterion B: Intrusion – Question B5
- CAPS-5 Criterion C: Avoidance (the sum of *Avoidance symptoms* items 6 and 7);
- CAPS-5 Criterion D: Negative Alterations in Cognition and Mood (the sum of *Cognitions and mood symptoms* items 8-14); and
- CAPS-5 Criterion E: Arousal and Reactivity (the sum of *Arousal and reactivity symptoms* items 15-20).

[redacted]

- MADRS
- HAM-A
- PCL-5
- PSQI
- AQoL-8D
- Social functioning: SDS Total (and three domain scores separately)
- CANTAB Cognitive Assessment Composite Score (and separately the six z-scores used in the composite score formula)
- Sleep monitoring: PSQI

The categorical endpoints listed below will be summarised at each time point by treatment group for the observed data (i.e. no imputation for missing data). The number and percentage of subjects with each response will be presented [redacted].

- CGI-S/CGI-I
- PGI-S/PGI-I

Overall treatment group differences in the binary response rates for Improvement, using the response classifications described in Sections 9.1.3 and 9.1.4 respectively, will be performed at each post-baseline timepoint with a Chi-Square test.

[redacted]

[redacted]

Individual estimates of area under the plasma concentration-time curve (AUC) will be made using appropriate pharmacokinetic modelling techniques (by third party supplier). Relationship between exposure (as measured by AUC) and efficacy will be explored initially by correlation analysis.

10.4 Compliance Analysis

Plasma levels of BNC210 taken at Screening, Week 2, Week 6 and Week 12 will be used as a dichotomic compliance marker. A participant with 3 detectable levels will be determined to be a good complier; with one or 2 detectable levels an intermittent complier and with 3 non-detectable levels will be determined to be a non-complier. Individual estimates of area under the plasma concentration-time curve will be made (by a third party supplier).

11. SAFETY

11.1 Exposure

Study medication compliance will be calculated as the planned versus actual dose received (based on the number of [redacted] bottles dispensed and the number of used [redacted] bottles returned) over the duration of the treatment period (first to last treatment

[redacted]

dates). Treatment compliance and treatment duration will be summarized with descriptive statistics (n, mean, SD, median, minimum and maximum). A listing will display the date study medication was consumed, medication interruptions, duration of study drug and percentage treatment compliance.

11.2 Columbia Suicide Severity Rating Scale (C-SSRS)

The C-SSRS is a suicidal ideation rating scale designed to identify behaviours that may be indicative of a patient's intent to commit suicide. The scale is administered via a semi-structured interview and measures both passive and active suicidal ideation and the intensity and duration of the ideation. Both suicidal and non-suicidal self-injurious behaviour is also assessed.

The approach described in document *Columbia–Suicide Rating Scale Scoring and Data Analysis Guide* (Version 2.0, Finalized February 2013, Contributors: Mary E. Nilsson, et.al.) [redacted] will be followed (descriptive summary statistics, excluding the optional formal hypothesis testing but producing outputs consistent with the Tables 1,2 and 4 and the Listing, [redacted]).

11.3 Adverse Events (AEs)

The Principal Investigator and clinical facility staff are responsible for detection, recording, and reporting of events that meet the criteria and definition of AEs as listed below. AEs will be recorded from first administration of IP through Week 15 (Post Study Visit). The Principal Investigator and clinical facility staff are responsible for detection, recording and reporting of pregnancy and appropriate follow up.

An AE is any untoward medical occurrence in a patient or clinical investigational subject administered a pharmaceutical product and which does not necessarily have to have a causal relationship with this treatment. An AE can therefore be any unfavorable and unintended sign (including an abnormal laboratory finding, for example), symptom, or

disease temporally associated with the use of a medicinal product, whether or not related to the medicinal product.

AEs with an onset date on or after the date of administration of study drug will be recorded. If the onset date is missing, the AE will still be recorded. Each subject will have their adverse event assigned to their treatment group.

AEs will be summarized using the latest version of the Medical Dictionary for Regulatory Activities by System Organ Class (SOC) and Preferred Term (PT). The incidence and percentage of subjects with at least 1 occurrence of a PT will be included, according to the most severe grade using a 3-point scale (mild, moderate, severe). The number of events per PT will also be summarized. Causality (relationship to study treatment) will be summarized separately.

The incidence and frequency of AEs, serious AEs (SAEs), related AEs, and AEs leading to treatment discontinuation will be summarized by treatment group according to SOC and PT. Adverse events will be listed. The duration of AEs will be determined and included in listings, along with the action taken and outcome.

Duration of Adverse Events in days will be calculated as the AE stop date minus the AE onset date + 1. Duration will not be calculated for subjects who do not have both a full AE onset date and a full AE stop date.

11.4 Laboratory Test and other Test Abnormalities as Adverse Events

Abnormal assessments (e.g., ECGs and vital signs) that are judged by the Principal Investigator as clinically significant or if the abnormal assessment results in clinical sequelae will be recorded as AEs.

Laboratory abnormalities will be reported by the Investigator as AEs if the abnormality is considered clinically significant or if the abnormality results in clinical sequelae. Clinically significant abnormal laboratory findings or other abnormal assessments that are detected

during the study or are present at baseline and significantly worsen following the first administration of IP will be reported as AEs.

The Principal Investigator and Co-Investigators will exercise their medical and scientific judgment in deciding whether an abnormal laboratory finding or other abnormal assessment is clinically significant.

The following clinical laboratory tests will be performed:

- Haematology
- Biochemistry
- C-reactive protein
- Urinalysis

Complete blood count and clinical chemistry will be summarized by descriptive statistics and the associated changes from baseline will be calculated and presented by treatment group and scheduled visit. Where results were recorded as '<value' or '>value' then the value used for the descriptive statistics will be (value-0.1) or (value+0.1) respectively. For example, the numeric value 3.9 will used were a result was recorded as '<4.00'. The last available assessment prior to the first drug administration will be defined as baseline value.

Data listings will include actual laboratory values, the relevant normal ranges for that subject's site, and whether or not the measurement was normal or abnormal according to the relevant laboratory normal ranges.

11.5 Pregnancy Test

Female subjects will have a serum pregnancy test performed at screening. The results will be only listed.

[redacted]

11.6 Vital Signs

Vital Signs (systolic blood pressure, diastolic blood pressure, pulse rate, temperature, and respiration rate) will be assessed throughout the study and weight and height at the Screening assessment only.

Vital sign results and change from baseline will be summarized over time by treatment group, and listed by treatment and subject [redacted]).

11.7 12-lead ECG

12-lead ECG measurements include QT and QTc intervals.

Where three ECG measurements are recorded at an assessment, then the average of those three measurements will be used as the measure at that visit in the summary table. The observed ECG results, and change from Screening, will be summarized at each scheduled visit and listed by treatment group [redacted].

11.8 Physical Examination

A complete physical examination will be completed at screening, Weeks 2, 6, and 12/Early Termination, and the Post Study Visit. The results will be classified as "Normal", "Abnormal" or "Not done" and subjects with any abnormalities will have the reason listed. Physical Examination results will be listed by Treatment [redacted].

11.9 Concomitant Medications/Treatments

Concomitant medications will by coded using the WHO drug dictionary (latest release). The start and stop dates will be used to identify when a concomitant medication was taken during the study.

Concomitant medications, Benzodiazepine Use, Psychotherapy and Nicotine Use will summarized separately by treatment group and the details included in listings.

Concomitant medications prescribed due to an AE will also be listed separately.

[redacted]

12. REFERENCES

1. SAS/STAT Software. 9.4 ed. Cary, NC: SAS Institute
2. Rodwell, L., Lee, K., Romaniuk, H., & Carlin, J. (2014). Comparison of methods for imputing limited-range variables: a simulation study. BMC Medical Research, 14-57.
3. Mallinckrodt, C. Lipkovich I, (2017). Analysing Longitudinal Clinical Trial Data, A Practical Guide, Chapman and Hall
4. Liang, K.-Y., and Zeger, S. L. (1986). "Longitudinal Data Analysis Using Generalized Linear Models." Biometrika 73:13–22.
5. Stokes, M. E., Davis, C. S., and Koch, G. G. (2012). Categorical Data Analysis Using SAS. 3rd ed. Cary, NC: SAS Institute Inc.