Statistical Analysis Plan

[redacted]

Sponsor	Bionomics Limited
Protocol Title:	A Phase 2b, Randomized, Double Blind, Two Arm Study to Investigate the Effects of BNC210 Tablet Formulation Compared to Placebo in Adults with Post-Traumatic Stress Disorder (PTSD)
Protocol Number:	BNC210.012
[redacted]	[redacted]
Document Version:	Final 3.0
Document Date:	02-Sep-2023

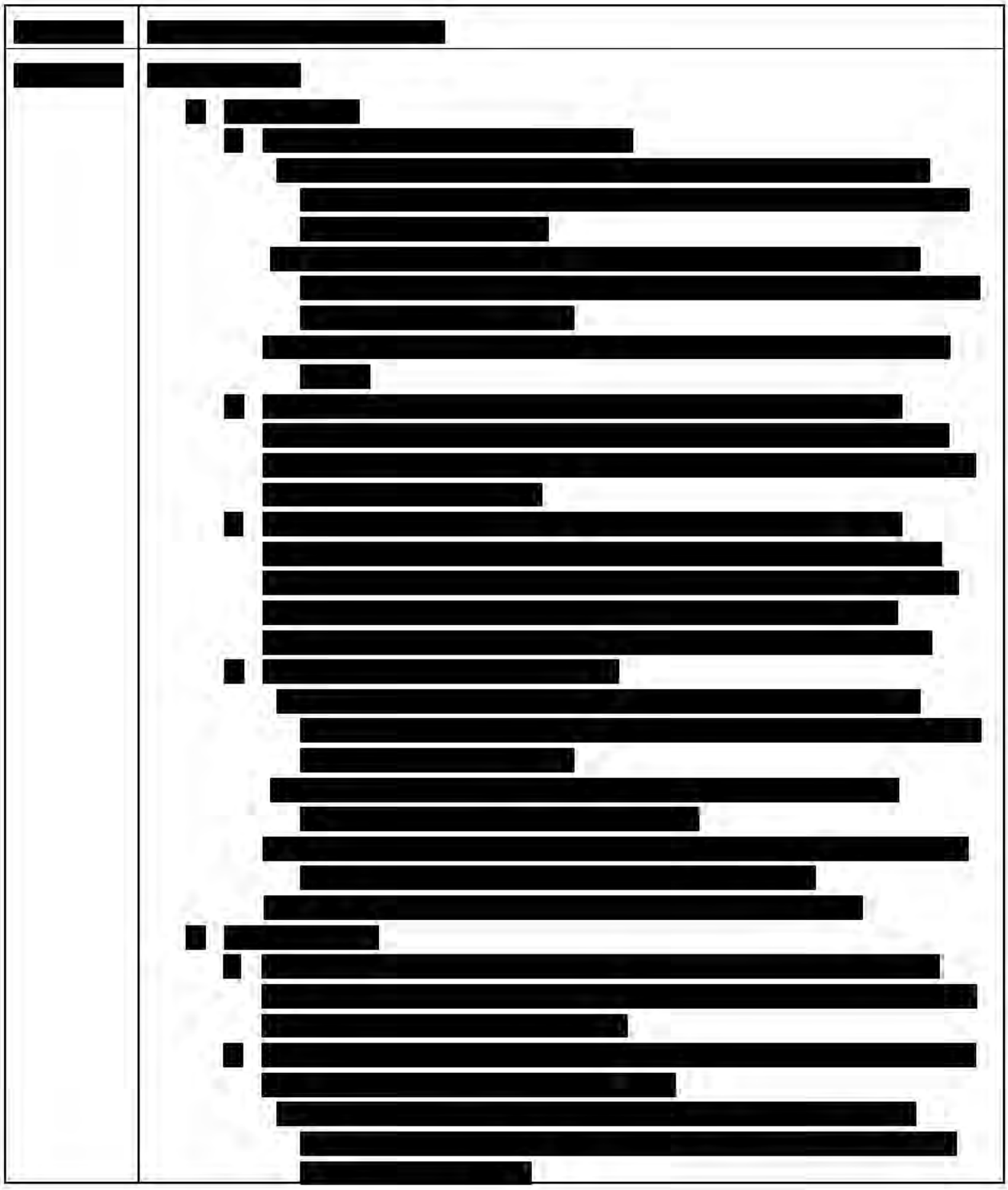
Approvals

Role	Signatures
Biostatistician	Print Name: [redacted] Sign Name:
Premier Senior Statistical Reviewer	Print Name: [redacted] Sign Name:
Bionomics Limited Representative	Print Name: [redacted] Sign Name:

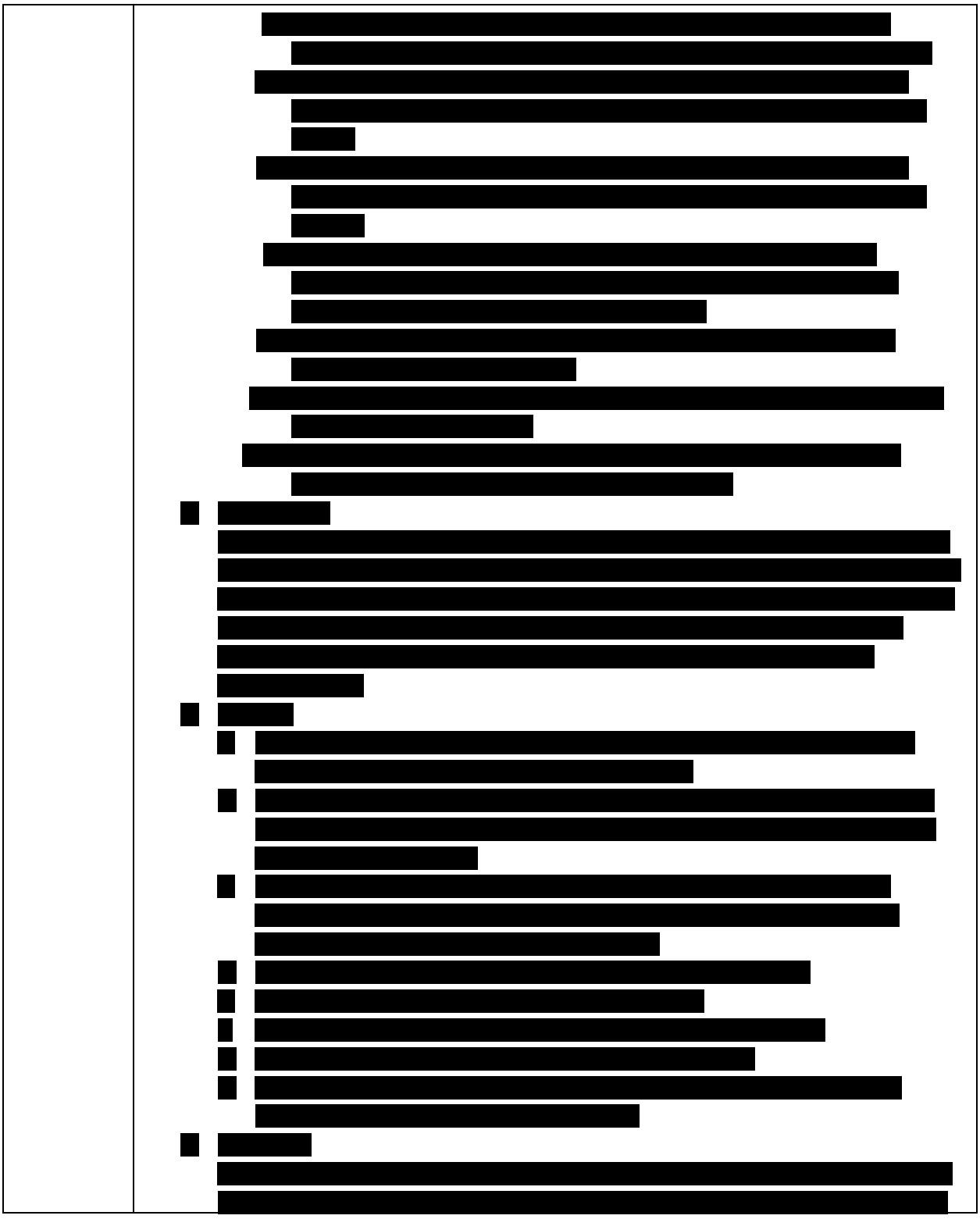
[redacted]

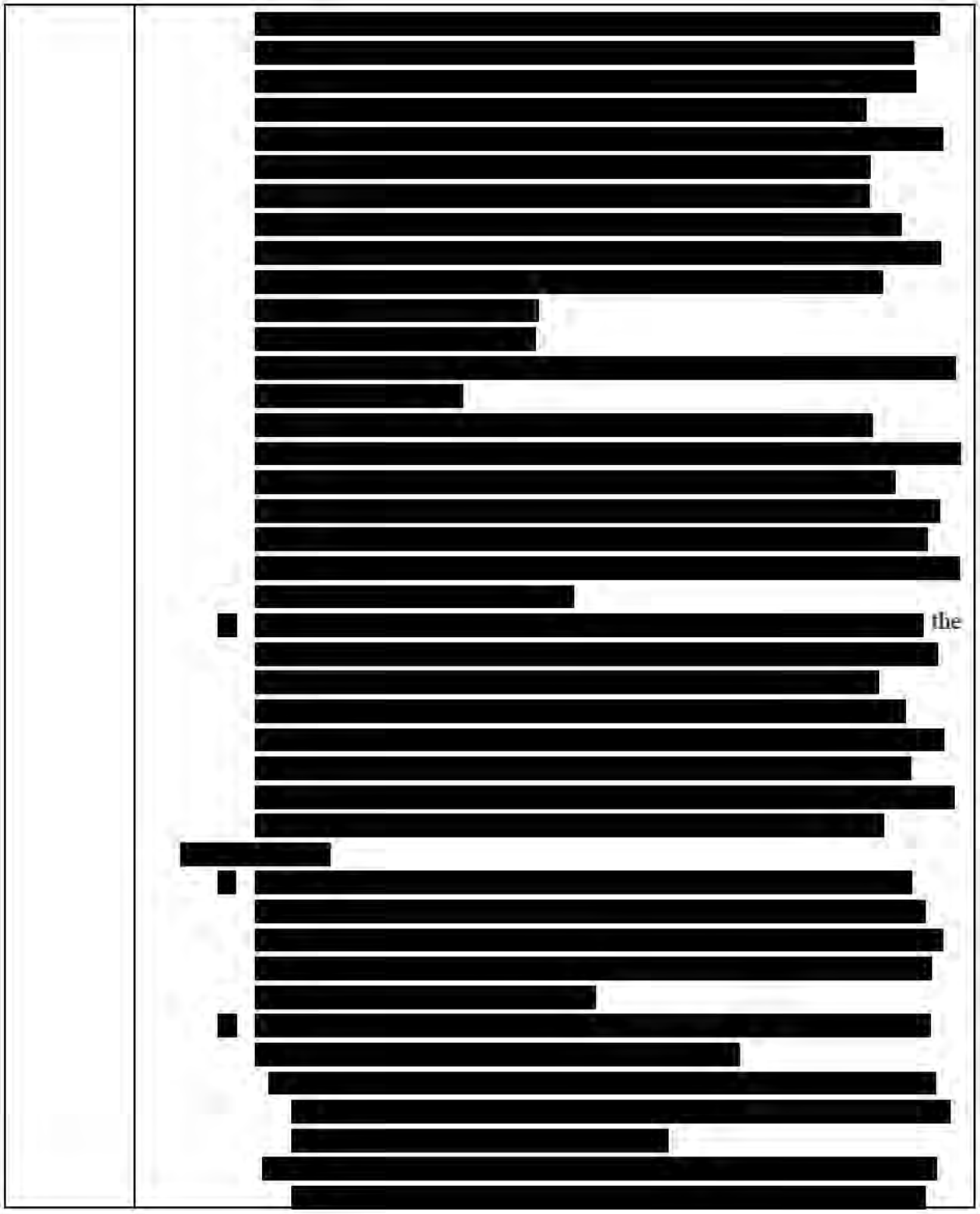
[redacted]

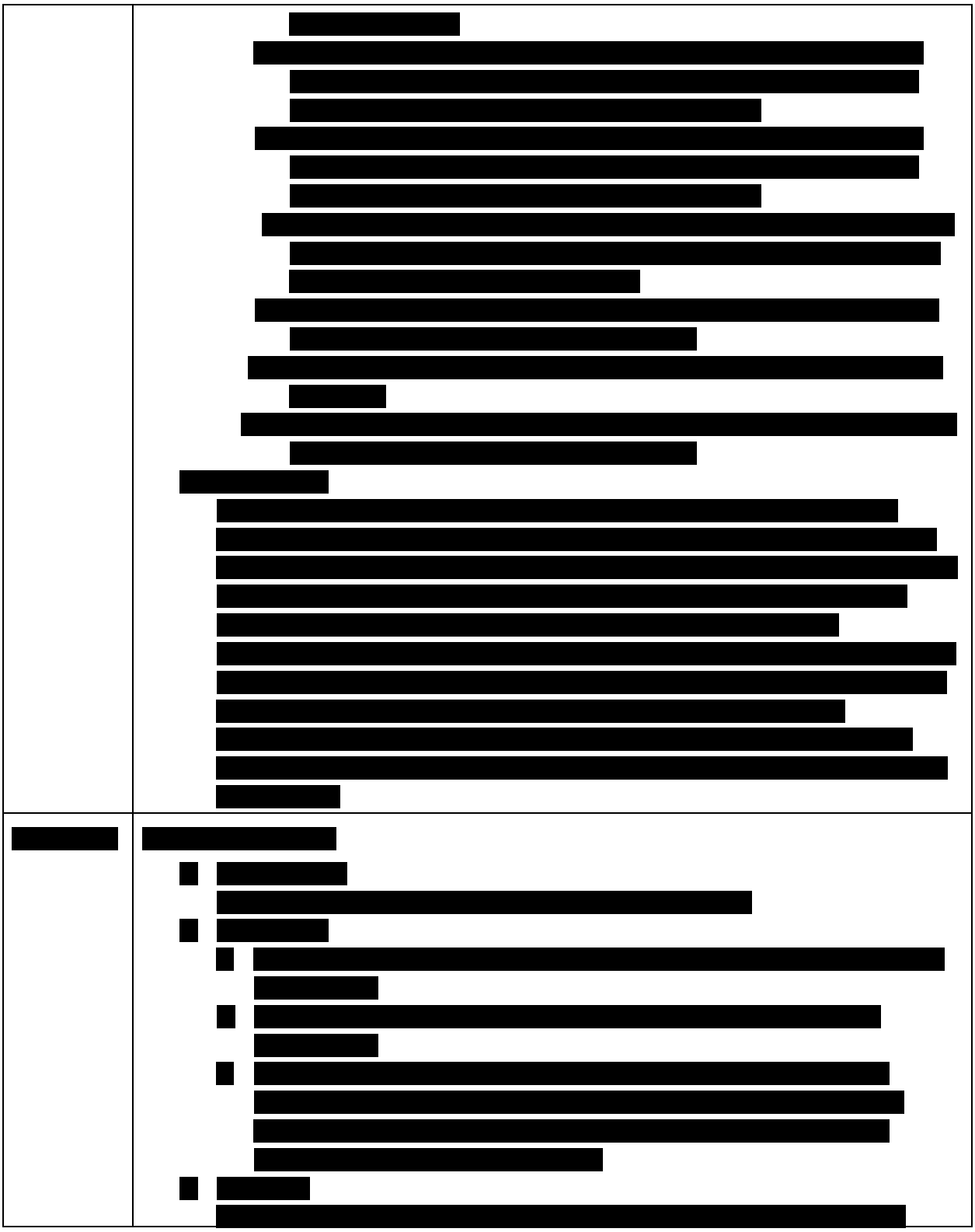
Document History

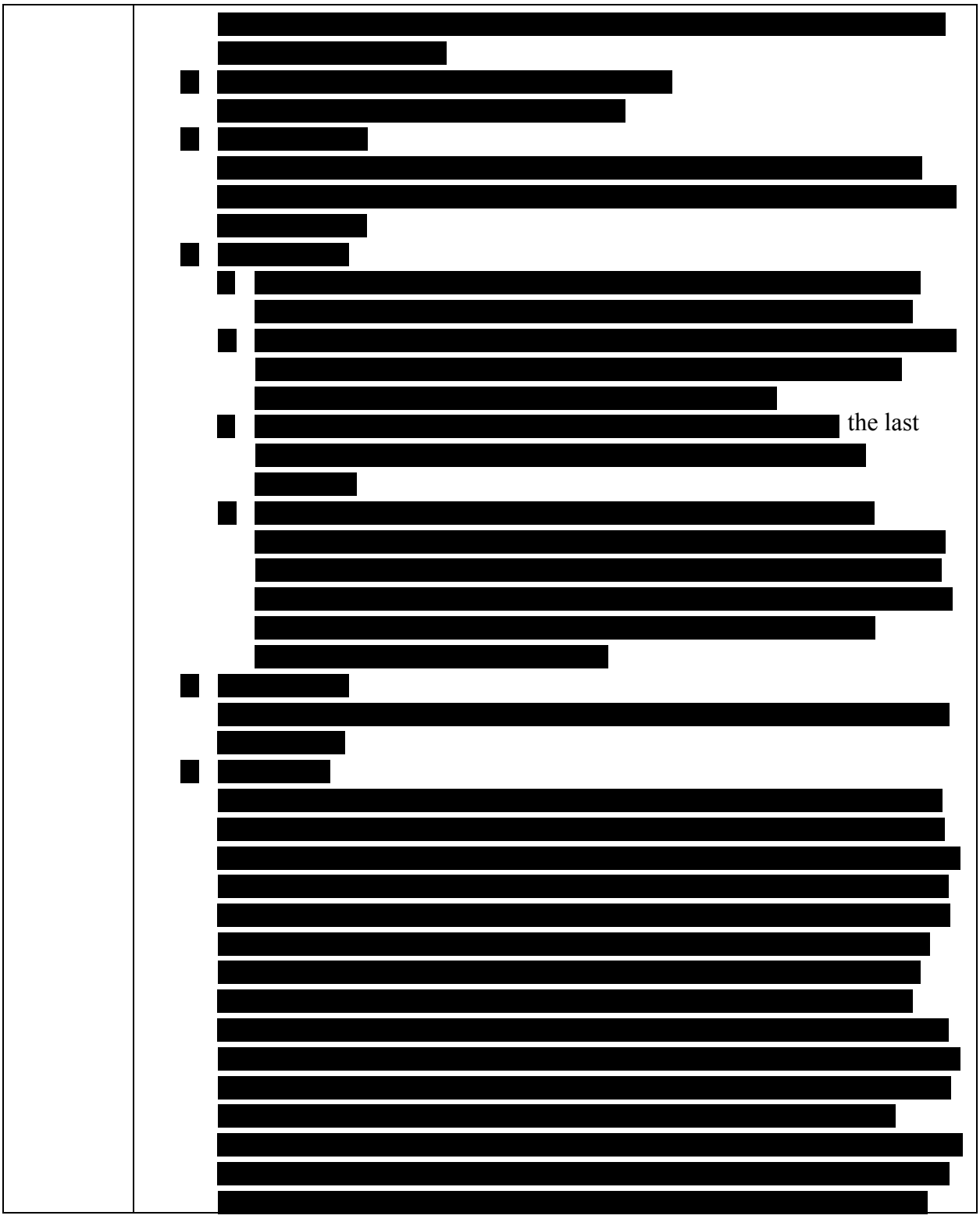


[redacted]



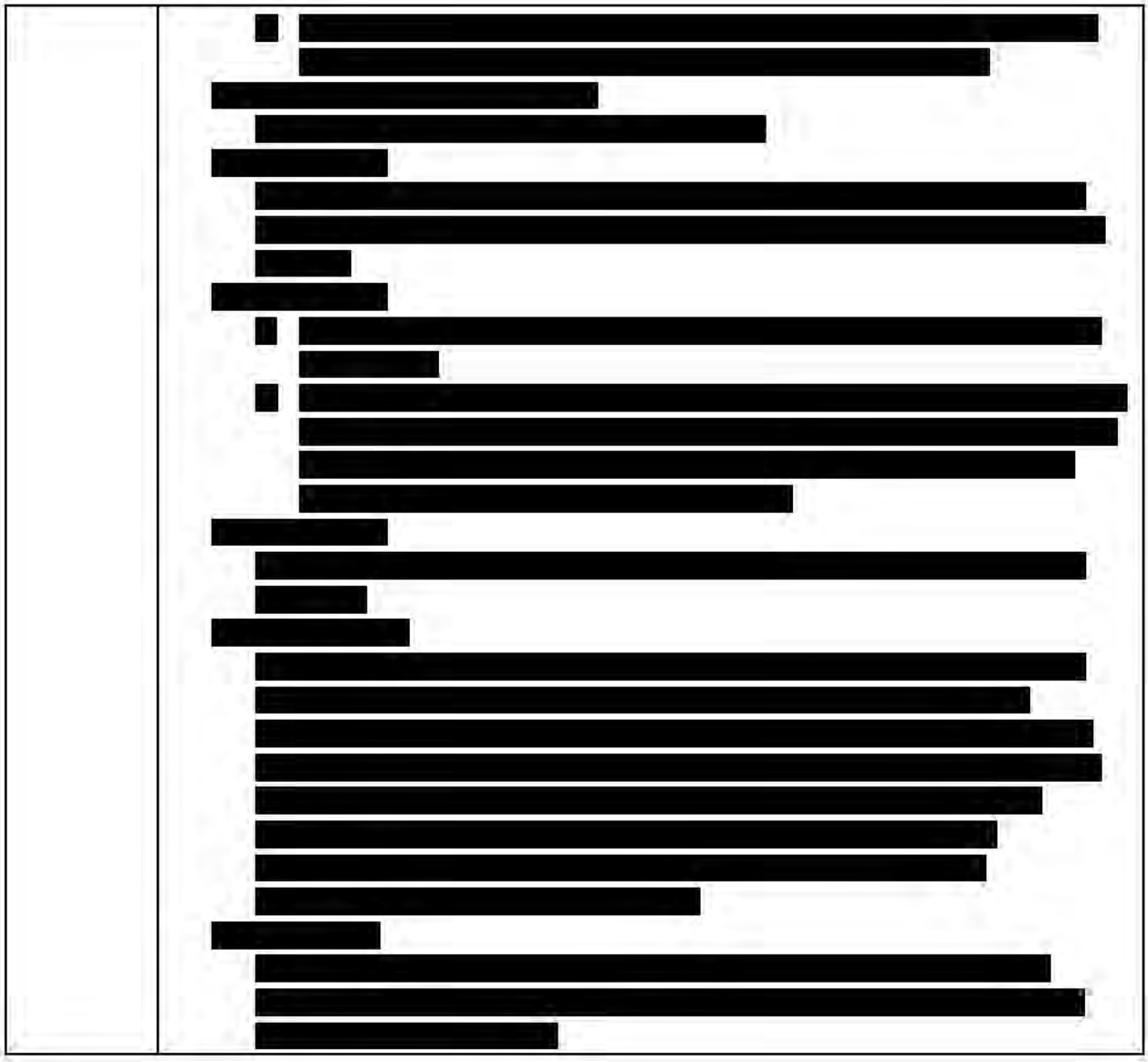





the last

[redacted]

[redacted]



[redacted] [redacted]

Table of Contents

List of Tables

[redacted]

[redacted]

List of Abbreviations

Abbreviation	Definition
AE	adverse event
AIC	Akaike information criterion
ASA	American Statistical Association
ATC	anatomical therapeutic chemical
ANCOVA	analysis of covariance
b.i.d.	twice daily
BMI	body mass index
CAPS-5	Clinician-Administered PTSD Scale for the DSM-5
CGI-I	Clinical Global Impression – Improvement Scale
CGI-S	Clinical Global Impression – Severity Scale
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 – 5th Edition
ECG	Electrocardiogram
EM	expectation maximization
EMA	European Medicines Agency

[redacted]

[redacted]

Abbreviation	Definition
EOT	end of treatment
FDA	Food and Drug Administration
HAM-A	Hamilton Anxiety Rating Scale
HIV	Human immunodeficiency virus
HBsAg	Hepatitis B surface antigen
HCV	Hepatitis C virus
HR	heart rate
ICH	International Conference on Harmonization
ISI	Insomnia Severity Index
ITT	intent-to-treat
IWRS	interactive web response system
LEC-5	Life Events Checklist for DSM-5
LLOQ	lower limit of quantitation
LS	least squares
MADRS	Montgomery-Asberg Depression Rating Scale
MAR	missing at random
MCMC	Markov-Chain Monte-Carlo
MedDRA	Medical Dictionary for Regulatory Activities
MI	multiple imputation

Abbreviation	Definition
mITT	modified intent-to-treat
MMRM	mixed model for repeated measurements
MNAR	missing not at random
N	number
PCL-5	PTSD Checklist for DSM-5
PE	physical examination
PGI-I	Patient Global Impression – Improvement Scale
PGI-S	Patient Global Impression – Severity Scale
PK	Pharmacokinetic
PP	per-protocol (population)
PT	preferred term
PTSD	Post-Traumatic Stress Disorder
REML	restricted maximum likelihood
RR	respiratory rate or relative rate
RSS	Royal Statistical Society
SAE	serious adverse event
SAF	safety (population)
SAP	statistical analysis plan
SCID-5-CT	Structured Clinical Interview for DSM-5 disorders – Clinical Trials Version

[redacted]

[redacted]

Abbreviation	Definition
SD	standard deviation
SDS	Sheehan Disability Scale
SE	standard error
SI	International System of Units
SOC	system organ class
TEAE	treatment-emergent adverse event
ULOQ	upper limit of quantitation
WHO-DD	World Health Organization drug dictionary

[redacted] [redacted]

1. Overview

This statistical analysis plan (SAP) describes the planned analysis and reporting for Bionomics Limited protocol number BNC210.012 (A Phase 2b, Randomized, Double Blind, Two Arm Study to Investigate the Effects of BNC210 Tablet Formulation Compared to Placebo in Adults with Post-Traumatic Stress Disorder (PTSD)), dated 17 Nov 2022 Version 4.0 (UK and US versions). 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 BNC210.012.

2. Study Objectives and Endpoints

2.1. Study Objectives

2.1.1. Primary Objective

The primary objective is to assess the effects of BNC210 on investigator-rated symptoms of PTSD measured by Clinician-Administered PTSD Scale for the DSM-5 (CAPS-5) Total Symptom Severity Scores.

2.1.2. Secondary Objectives

The key secondary objectives are to:

- Assess the effects of BNC210 on investigator-rated global functioning in participants with PTSD;
- Assess the effects of BNC210 on patient-reported social functioning in participants with PTSD.

[redacted]

[redacted]

Other secondary objectives are to:

- Assess the effects of BNC210 on Investigator-rated symptom clusters of PTSD measured by Criterions B, C, D and E of the CAPS-5 symptom cluster scores;
- Evaluate the response rate and remission rate of BNC210 on Investigator-rated symptoms of PTSD measured by CAPS-5 Total Symptom Severity Scores;
- Assess the effects of BNC210 on Investigator-rated global functioning in participants with PTSD;
- Assess the effects of BNC210 on Investigator-rated symptoms of anxiety and depression in participants with PTSD;
- Assess the effects of BNC210 on patient-reported symptoms of PTSD;
- Assess the effects of BNC210 on patient-reported global functioning and sleep quality in participants with PTSD;
- Assess the safety and tolerability of BNC210 in participants with PTSD.

[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 from Baseline to Week 12 in the CAPS-5 Total Symptom Severity Score for participants receiving BNC210 compared to participants receiving placebo.

2.2.1.1.1 Primary Estimand

The primary question of interest for this Phase 2 study is: What is the difference between BNC210 and placebo in the mean change from Baseline to Week 12 in the CAPS-5 Total Symptom Severity Score among participants with PTSD when using randomized treatment.

Over the 12-week randomized treatment period, participants may be exposed to possible known or unknown inter-current events that could possibly impact the estimand, such as treatment discontinuation due to a specific adverse effect or perhaps a lack of effect. The "Treatment Policy Strategy" will be adopted for handling all known or unknown inter-current events in this study. To this end, the Intent-To-Treat (ITT) principle will serve as the analytical basis for

[redacted]

[redacted] [redacted]

interpreting the estimand. In other words, the difference between BNC210 and placebo in the mean change from Baseline to Week 12 in the CAPS-5 Total Symptom Severity Score will be evaluated regardless of the occurrence of any such inter-current event.

2.2.1.2. Secondary Efficacy Endpoint(s)

The key secondary efficacy endpoints of this study include the following comparisons of participants receiving BNC210 and participants receiving placebo:

- Mean change from Baseline to Week 12 in the Clinical Global Impression – Severity Scale (CGI-S);
- Mean change from Baseline and Week 12 in the Sheehan Disability Scale (SDS).

The other secondary efficacy endpoints of this study include the following comparisons of participants receiving BNC210 and participants receiving placebo:

- Mean change from Baseline and Week 12 in the symptom cluster severity score CAPS-5 Criterion B: Intrusion;
- Mean change from Baseline and Week 12 in the symptom cluster severity score CAPS-5 Criterion C: Avoidance;
- Mean change from Baseline to Week 12 in symptom cluster severity scores for CAPS-5 Criterion D: Negative Alterations in Cognitions and Mood;
- Mean change from Baseline to Week 12 in symptom cluster severity scores for CAPS-5 Criterion E: Arousal and Reactivity;
- Proportion of participants who achieve a $\geq$ 30% improvement from Baseline to Week 12 on CAPS-5 Total Symptom Severity Score;
- Proportion of participants who achieve a $\geq$ 50% improvement from Baseline to Week 12 on CAPS-5 Total Symptom Severity Scores;
- Proportion of participants who achieve remission in PTSD symptoms with a score of $\leq$ 11 from Baseline to Week 12 on CAPS-5 Total Symptom Severity Scores;
- Mean scores at Week 12 in the Clinical Global Impression – Improvement Scale (CGI-I);
- Mean change from Baseline to Week 12 in the Hamilton Anxiety Rating Scale (HAM-A);
- Mean change from Baseline to Week 12 in the Montgomery-Åsberg Depression Rating Scale (MADRS);
- Mean change from Baseline to Week 12 in the PTSD Checklist for DSM-5 (PCL-5);
- Mean change from Baseline and Week 12 in the Patient Global Impression – Severity Scale (PGI-S);
- Mean scores at Week 12 in the Patient Global Impression - Improvement Scale (PGI-I);
- Mean change from Baseline and Week 12 in the Insomnia Severity Index (ISI).

[redacted]

2.2.2. Safety Endpoints

The safety endpoints of this study include the following from Baseline to Week 15:

- Continuous adverse events (AEs) and serious adverse events (SAEs);
- Electrocardiograms (ECGs);
- Vital signs (body temperature, pulse rate, respiratory rate, systolic and diastolic blood pressure);
- Clinical laboratory results (hematology, blood chemistry, and urinalysis);
- Physical examination;
- Columbia Suicide Severity Rating Scale (C-SSRS).

2.2.3. Pharmacokinetics Endpoint

A population PK analysis will be done under a separate analysis plan.

3. Overall Study Design and Plan

3.1. Overall Design

This is a randomized, double-blind, placebo-controlled, parallel group, multi-center Phase 2b study in participants with PTSD, who are between the ages of 18 to 75 (inclusive), with a 12 week, 2-arm treatment period.

Approximately 200 participants will be enrolled and randomized using a 1:1 ratio to receive either BNC210 or matched placebo. Participants will complete 12 weeks of treatment with their allocated study intervention. Participants in the US will complete study visits every 2 weeks during the treatment period and a follow-up visit at Week 15 (3 weeks after final intervention). Participants in the UK will complete study visits every week for the first 2 weeks, then every 2 weeks from Week 4 onwards during the treatment period, and a Follow-up visit at Week 15 (3 weeks after final intervention).

3.2. Sample Size and Power

A sample size of approximately 200 participants (100 participants per treatment arm) is

[redacted]

[redacted]

calculated to provide ≥ 80% power to detect a 6-point difference between BNC210 and placebo groups in change from Baseline to Week 12 for the CAPS-5 Total Symptom Severity Scores, with an expected standard deviation (SD) of 12.5. These assumptions are based on data from the previous BNC210 Phase 2 PTSD study and pharmacometric analysis. In a two-sample t-test using a two-sided significance level of 0.05 and power of 80%, 70 participants per treatment group would be needed to demonstrate a 6-point difference between treatment groups, and assuming 30% drop-out over the 12-week treatment period leads to a planned sample size of 100 per group.

3.3. Study Population

The study population is comprised of male and female participants who are 18 to 75 years of age, with a current diagnosis of PTSD as defined by the CAPS-5 and with the index trauma event occurring in adulthood (at least 18 years old), and a CAPS-5 Total Symptom Severity Score of ≥ 30 at Screening and Baseline with no > 25% decrease in CAPS-5 Total Symptom Severity Score between Screening and Baseline.

3.4. Treatments Administered

Participants will be randomized to one of two treatment groups:

- BNC210 900 mg
- Placebo

Participants will self-administer the study intervention twice daily (b.i.d.) for 12 weeks. [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 ratio to receive BNC210 900 mg or matching placebo.

3.6. Blinding and Unblinding

This is a double-blind study in which participants/Investigators/raters, etc. are blinded to study intervention.

The Interactive Web Response System (IWRS) will be programmed with blind-breaking instructions. In case of an emergency, the Investigator has the sole responsibility for determining if unblinding of a participant's intervention assignment is warranted. Participant safety must always be the first consideration in making such a determination. If the Investigator decides that unblinding is warranted, the Investigator should make every effort to contact the Medical Monitor and/or Sponsor prior to unblinding a participant's intervention assignment unless this could delay emergency treatment for the participant. 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.

[redacted]

[redacted]

[redacted]

3.7. Schedule of Events

A detailed schedule of events for participants enrolled in the US is provided in Table 1A.

Table 1A: Schedule of Events (US participants)

Assessment[1]	**Screening** ≤-21 days	**Treatment Period**							**Week 15 / Follow-up** (±5 days)
		Baseline[2] (+3 days)	**Week 2**[3] (±3 days)	**Week 4** (±3 days)	**Week 6** (±3 days)	**Week 8** (±3 days)	**Week 10** (±3 days)	**Week 12 / Early Termination** (±3 days)	
Informed consent	X								
Inclusion and exclusion criteria[4]	X								
Demography	X								
Medical history including current & prior medications	X								
Structured Clinical Interview for DSM-5 disorders – Clinical Trials Version (SCID-5-CT) and Personality Disorders (SCID-5-PD) Borderline Personality Disorder section only	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 pregnancy test (if applicable)		X						X	

1 See Section 8 in the protocol for guidance on the order of assessments to be completed at each visit
2 All baseline assessments to be completed pre-dose
3 Visit can be performed at the study site, or remotely via phone/video call.
4 Recheck eligibility and clinical status before randomization

[redacted]

[redacted]

Assessment[1]	Screening ≤-21 days	Treatment Period							Week 15 / Follow-up (±5 days)
		Baseline[2] (+3 days)	Week 2[3] (±3 days)	Week 4 (±3 days)	Week 6 (±3 days)	Week 8 (±3 days)	Week 10 (±3 days)	Week 12 / Early Termination (±3 days)	
Urine drug screen[5]	X	X		X		X		X	
12-lead electrocardiogram (ECG)[6]	X			X		X		X	
Vital signs[7]	X	X		X		X		X	X
Physical examination (including height only at Screening; weight only at Screening and Week 12 / Early Termination)[8]	X	X		X		X		X	X
Clinical labs (hematology, biochemistry, urinalysis)	X	X		X	X[9]	X	X[9]	X	X
Columbia Suicide Severity Rating Scale (C-SSRS)[10]	X	X		X		X		X	X
Life Events Checklist for DSM-5 (LEC-5)[11]	X								
Clinician-Administered PTSD Scale for DSM-5 (CAPS-5)[12]	X	X		X		X		X	
Montgomery-Åsberg Depression Rating Scale (MADRS)	X	X		X		X		X	

5 Screening assessment to be performed by the central laboratory. Subsequent assessments to be performed locally by study site using urine dip stick test
6 ECGs to be completed in triplicate with an approximate 2-minute break between tests. Participants must be resting in a semi-supine or supine position for at least 5 minutes prior to performing the ECG. The same ECG machine should be used for all recordings from an individual participant, where possible. Single triplicate repeat during Screening allowed in the event of a clinically significant abnormality including prolonged Fridericia corrected QT interval (QTcF).
7 Blood pressure and heart rate will be measured in a sitting position after resting for 5 minutes
8 Full physical examination at Screening & W12 / Early Termination. Abbreviated physical examination at Baseline, Week 4, Week 8, and Follow-up
9 Blood biochemistry testing only for Aspartate Aminotransferase (AST), Alanine Aminotransferase (ALT), Alkaline Phosphatase (ALP), Bilirubin (direct and total)
10 "Screening C-SSRS" to be used at Screening. "C-SSRS Since Last Visit" to be used at every other visit
11 To be completed prior to CAPS-5 assessment at Screening
12 Past month version to be used at every visit and should be audio recorded at every visit using the provided iPhone.

Assessment[1]	Screening ≤-21 days	Treatment Period							Week 15 / Follow-up (±5 days)
		Baseline[2] (+3 days)	Week 2[3] (±3 days)	Week 4 (±3 days)	Week 6 (±3 days)	Week 8 (±3 days)	Week 10 (±3 days)	Week 12 / Early Termination (±3 days)	
Randomization[13]		X							
Hamilton Anxiety Rating Scale (HAM-A)		X		X		X		X	
Clinical Global Impressions Severity and Improvement scales (CGI-S/I)[14]		X		X		X		X	
Patient Global Impression – Severity and Improvement scales (PGI-S/I)[14]		X		X		X		X	
PTSD Checklist (PCL-5)		X						X	
Sheehan Disability Scale (SDS)		X		X		X		X	
Insomnia Severity Index (ISI)		X		X		X		X	
Study Intervention dispensing		X[15]		X		X			
Concomitant medications recording		X	X	X		X		X	X
Adverse Event (AE) recording[16]			X	X		X		X	X
Blood sample collection for concentration of BNC210				X			X	X	

13 Once eligibility has been confirmed
14 Improvement scale at Week 12 only
15 Participants to administer one PM dose on day of completing Baseline assessments and randomization
16 See Appendix 3 in the protocol for definitions

[redacted] [redacted]

A detailed schedule of events for participants enrolled in the UK is provided in Table 1B.

Table 1B: Schedule of Events (UK participants)

Assessment[1]	**Screening** ≤-21 days	**Treatment Period**								**Week 15 / Follow-up** (±5 days)
		Baseline[2] (+3 days)	**Week 1**[3] (±3 days)	**Week 2** (±3 days)	**Week 4** (±3 days)	**Week 6** (±3 days)	**Week 8** (±3 days)	**Week 10** (±3 days)	**Week 12 / Early Termination** (±3 days)	
Informed consent	X									
Inclusion and exclusion criteria[4]	X									
Demography	X									
Medical history including current & prior medications	X									
Structured Clinical Interview for DSM-5 disorders – Clinical Trials Version (SCID-5-CT) and Personality Disorders (SCID-5-PD) Borderline Personality Disorder section only	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 pregnancy test (if applicable)		X							X	
Urine drug screen[5]	X	X			X		X		X	

1 See Section 8 in the protocol for guidance on the order of assessments to be completed at each visit
2 All baseline assessments to be completed pre-dose
3 Visit can be performed at the study site, or remotely via phone/video call.
4 Recheck eligibility and clinical status before randomization
5 Screening assessment to be performed by the central laboratory. Subsequent assessments to be performed locally by study site using urine dip stick test

[redacted]

[redacted]

[redacted]

Assessment[1]	Screening ≤-21 days	Treatment Period								Week 15 / Follow-up (±5 days)
		Baseline[2] (+3 days)	Week 1[3] (±3 days)	Week 2 (±3 days)	Week 4 (±3 days)	Week 6 (±3 days)	Week 8 (±3 days)	Week 10 (±3 days)	Week 12 / Early Termination (±3 days)	
12-lead electrocardiogram (ECG)[6]	X	X[7]		X	X		X		X	
Vital signs[8]	X	X[9]		X	X		X		X	X
Physical examination (including height only at Screening; weight only at Screening and Week 12 / Early Termination)[10]	X	X			X		X		X	X
Clinical labs (hematology, biochemistry, urinalysis)	X	X		X	X	X[11]	X	X[11]	X	X
Columbia Suicide Severity Rating Scale (C-SSRS)[12]	X	X	X	X	X		X		X	X
Life Events Checklist for DSM-5 (LEC-5)[13]	X									
Clinician-Administered PTSD Scale for DSM-5 (CAPS-5)[14]	X	X			X		X		X	
Montgomery-Åsberg Depression Rating Scale	X	X			X		X		X	

6 ECGs to be completed in triplicate with an approximate 2-minute break between tests. Participants must be resting in a semi-supine or supine position for at least 5 minutes prior to performing the ECG. The same ECG machine should be used for all recordings from an individual participant, where possible. Single triplicate repeat during Screening allowed in the event of a clinically significant abnormality including prolonged Fridericia corrected QT interval (QTcF).
7 ECG to be completed 2 hours post initial dose
8 Blood pressure and heart rate will be measured in a sitting position after resting for 5 minutes
9 At Baseline, Vital Signs to be collected pre-dose and 2 hours post initial dose.
10 Full physical examination at Screening & W12 / Early Termination. Abbreviated physical examination at Baseline, Week 4, Week 8, and Follow-up
11 Blood biochemistry testing only for Aspartate Aminotransferase (AST), Alanine Aminotransferase (ALT), Alkaline Phosphatase (ALP), Bilirubin (direct and total)
12 "Screening C-SSRS" to be used at Screening. "C-SSRS Since Last Visit" to be used at every other visit
13 To be completed prior to CAPS-5 assessment at Screening
14 Past month version to be used at every visit and should be audio recorded at every visit using the provided iPhone.

[redacted]

[redacted]

[redacted]

Assessment[1]	Screening ≤-21 days	Treatment Period								Week 15 / Follow-up (±5 days)
		Baseline[2] (+3 days)	Week 1[3] (±3 days)	Week 2 (±3 days)	Week 4 (±3 days)	Week 6 (±3 days)	Week 8 (±3 days)	Week 10 (±3 days)	Week 12 / Early Termination (±3 days)	
(MADRS)										
Randomization[15]		X								
Hamilton Anxiety Rating Scale (HAM-A)		X			X		X		X	
Clinical Global Impressions Severity and Improvement scales (CGI-S/I)[16]		X			X		X		X	
Patient Global Impression – Severity and Improvement scales (PGI-S/I)[16]		X			X		X		X	
PTSD Checklist (PCL-5)		X							X	
Sheehan Disability Scale (SDS)		X			X		X		X	
Insomnia Severity Index (ISI)		X			X		X		X	
Study Intervention dispensing		X[17]			X		X			
Concomitant medications recording		X	X	X	X		X		X	X
Adverse Event (AE) recording[18]			X	X	X		X		X	X
Blood sample collection for concentration of BNC210					X		X		X	

15 Once eligibility has been confirmed
16 Improvement scale at Week 12 only
17 Participants to administer one PM dose on day of completing Baseline assessments and randomization
18 See Appendix 3 in the protocol for definitions

[redacted]

[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, SD, median, minimum, and maximum.

Categorical (qualitative) variable summaries will include the frequency and percentage of participants who are in the particular category for 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.

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 SAF population includes all participants who receive any amount of the study intervention. The SAF population will be used for the analysis of the safety endpoints. Participants will be analyzed by actual treatment received.
- **Intent-to-Treat Population (ITT):** Based on the ITT principle that asserts that the effect of a treatment can be best assessed by evaluating on the basis of the intention to treat a participant (i.e., the planned treatment regimen) rather than the actual treatment given, the ITT population includes all randomized participants. Participants will be analyzed by randomized treatments.

[redacted] [redacted]

- **Modified Intent-to-Treat Population (mITT):** The mITT population includes all participants in the ITT population who receive any amount of the study intervention and have at least one post-baseline primary efficacy assessment (CAPS-5). Participants will be analyzed by randomized treatment. This will be the primary/main population for analysis.
- **Per Protocol Population (PP)**: The PP population will include participants from the mITT population who have no major protocol deviations. Before data are released for statistical analysis, a blinded review of all data will be performed in conjunction with the Sponsor to identify protocol deviations that may potentially affect the results. At this time, it will be determined if participants and/or data should be excluded from the PP population. The list of participants or observations to be excluded from the PP population will be provided in the Clinical Study Report. Participants will be analyzed by randomized treatments.
- **Pharmacokinetic Population (PK):** The PK population will include all participants from the SAF population who have a valid concentration measurement.

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 and key secondary endpoints will each be assessed with a Mixed Model for Repeated Measures (MMRM) with and without multiple imputation (MI). Each model will include fixed effects for treatment, visit, interaction between treatment and visit, center, 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

[redacted]

[redacted]

[redacted]

[redacted]

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 and key secondary endpoints will be analyzed using MMRM with MI using placebo-based (jump to control) imputation. These endpoints will also be analyzed using MMRM without MI as a supportive analysis.

All other secondary efficacy data will be analyzed as observed (with missing data not imputed). Safety data will not be imputed and will be analyzed as observed.

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 participant who withdraws or is discontinued from the study or who misses a scheduled visit or assessments up through Week 12 will have their primary and key secondary efficacy missing data analyzed as imputed using MI techniques. This analysis will be presented as the main analysis.

Multiple imputation 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

[redacted] [redacted]

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 visit, data will be missing for all subsequent visits. Typically, there is also a small amount of non-monotone missing data (i.e., some participants have missing values for intermediate visits, but have non-missing data at subsequent visits).

MI with Placebo-Based Imputation

A placebo-based (jump to control) multiple imputation for missing primary and key secondary endpoints will be carried out for participants who withdraw from the study or have missing data at a scheduled visit through Week 12, 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 key 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 visit). 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 634176.

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.

[redacted] [redacted]

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[14]. Once the data set is made monotone missing, an MI model under a MAR assumption will be used for the remaining monotone missing data. 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 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 study visit 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 CAPS-5 Total Symptom Severity Scores (0 and 80), CGI-S (1 and 7), and SDS (20 and 80) 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 visit, p-values and 95% CIs will come from the MMRM analysis on the observed data.

6.1.4.2. Supportive Analysis for Primary and Key Secondary Efficacy Endpoints

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

6.1.4.3. MMRM Analysis Method

Primary and key secondary efficacy variables will be assessed with a MMRM model with MI. All other continuous secondary and exploratory efficacy variables will be assessed with a MMRM model with no imputation. Primary and key secondary efficacy variables will also be

[redacted]

[redacted]

assessed with a MMRM model with no imputation as supportive analyses. The model will include fixed effects for treatment, visit, interaction between treatment and visit, center, and covariate for Baseline score. The MMRM method has been demonstrated extensively as an appropriate choice for 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 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; Molenberghs et al., 2004; Molenberghs 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 good 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. Analysis Visit Windows

Analyses of all visit-based efficacy variables for this study will be performed using the analysis visit windows defined by study day relative to the first dose of study medication as outlined below

[redacted]

[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

6.1.6. Pooling of Sites

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

6.1.7. Derived Variables

- **Change from baseline** = value at current time point – value at baseline.
- **PCL-5 =** sum of 20 individual items with a score ranging from 0 (not at all) to 4 (extremely) for a total severity score of 0 to 80.
- **SDS** = sum of 3 individual questions scored on a scale of 0 (not at all) to 10 (extremely) for a total score of 0 to 30.
- **ISI =** sum of 7 individual questions scored from 0 (none) to 4 (very much) for a total score of 0 to 28.
- **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 reported through Follow-up Week 15/EOT visit.

- [redacted]

- [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 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 the study medication, whether or not considered related to the study medication. All AEs that occur after any participant has received study treatment or within 21 (±5 days) days following the cessation of treatment, will be recorded.

Adverse events will be coded using the MedDRA version 24.0 or higher thesaurus. Concomitant medications will be coded using the World Health Organization Drug Dictionary (WHO-DD Global B3 version March 2021).

Concomitant Treatment Definition and Handling of Data

A concomitant treatment refers to all treatment, including concomitant medications, vaccines, herbal supplements, vitamins, recreational drugs, and behavioral treatments such as counseling or regular psychotherapy, taken between the dates of the first dose of study medication and the end of the participant's participation in the study (Follow-up Week 15), inclusive.

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 month after the first dose is used; 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

[redacted] [redacted]

and the date is not the same as the date of first dose and the minute assigned is 30 if the date is not the same as the date of first dose.

These conventions will be applied only to AE 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 AE, 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 case report form (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.

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.

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

Treatment Duration and Exposure

For participants who are missing the date of last study drug application, for any reason, the last known contact date will be used in the calculation of treatment duration and study medication exposure. Study drug compliance will not be calculated for those participants whose date of last study drug application is unknown.

6.2. Special Handling for COVID-19 Disruptions

Study visits may be delayed/not performed as a result of COVID-19 disruptions or other natural disasters (e.g., sites were closed or participants were under stay-at-home orders). Where possible, study sites **may** collect data for the following assessments/questionnaires from participants remotely by telephone or video conference; this will be clearly documented in the source:

- Investigator assessments
- Concomitant Medication
- Adverse event monitoring

If possible, sites should adhere to the protocol visit window for remote data collection.

[redacted]

[redacted]

Participant questionnaires normally completed during on-site visits will be provided to participants to be completed at the appropriate time point and returned to the study site upon completion.

The following assessments **cannot** be completed via telemedicine/remotely but may be collected per site COVID-19 protocols or as an unscheduled assessment upon lifting of any restrictions:

- Blood Draw for BNC210 concentration
- ECG
- Laboratory or urine drug screen tests
- Physical examination (PE)
- Vital signs

Study drugs may be delivered to a participant's residence by site personnel who are delegated the responsibility to dispense study drug and coordinate the shipment in the case that restrictions, due to COVID-19 or other natural disaster, prevent in-clinic study visits. Alternatively, a representative of the participant may attend the study site to obtain the study drug on the participant's behalf.

7. Study Patients/Participants and Demographics

7.1. Disposition of Patients/Participants and Withdrawals

The number of participants randomized, completing, 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.

Completers will be recorded as Week 12 completers and Week 15 completers. Week 12 completers will be defined as all participants who complete the randomized treatment period, while Week 15 completers will be defined as those who complete the randomized treatment period as well as the follow-up visit.

7.2. Protocol Violations and Deviations

Protocol deviations will be listed. Assessments that are missed or cannot be performed within the study visit window due to COVID-19 or other natural disaster situations that may preclude an in-clinic study visit will not be considered protocol deviations.

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. Important deviations will be reviewed by the Sponsor and Premier to determine the final classification. 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.

[redacted]

[redacted]

[redacted]

Important protocol deviations may include:

- Significant and/or persistent dosing error
- Participant did not meet eligibility criteria
- 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 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

The following analyses will be conducted for the SAF, mITT and PP populations:

- Demographic variables will include age, sex, height, and weight. Information on race, ethnicity, and country 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 medical history including confirmation of PTSD diagnosis, index trauma event (from CAPS-5 Criterion A), and time since index trauma event (as reported by the participant in months from the life events questionnaire). Baseline evaluation of the CAPS-5 Total Severity Score, PCL-5 total score, CGI-S score, PGI-S score, MADRS total score, HAM-A total score, SDS total score, and ISI total score will also be reported.

A complete participant medical history will be summarized by MedDRA system organ class (SOC) and preferred term (PT), conducted on the SAF population. Prior and concomitant medications will be summarized by treatment group, by the number and percentage of participants taking each medication, classified using WHO-DD Anatomical Therapeutic Chemical (ATC) classes and preferred terms.

7.4. Exposure and Compliance

Treatment compliance (%) will be summarized by randomized treatment group and overall. [redacted]

[redacted]

[redacted]

[redacted]

8. Efficacy Analysis

Efficacy analyses will be performed on the mITT population. MMRM with no MI will be used as a supportive analysis for the primary and key secondary endpoints for the mITT and PP populations according to the planned treatment. The mITT will be considered the primary/main population for efficacy analyses.

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

8.1. Primary Efficacy Analysis

For this study, the primary estimand is the difference between BNC210 and placebo in the mean change from Baseline to Week 12 in the CAPS-5 Total Symptom Severity score. In the course of the 12-week randomized treatment period, participants may be exposed to possible known or unknown inter-current events that could possibly impact the estimand, such as treatment discontinuation due to a specific adverse effect or perhaps lack of effect. The "Treatment Policy Strategy" will be adopted for handling all known or unknown inter-current events in this study. To this end, the ITT principle will serve as the analytical basis for interpreting the estimand. In other words, the difference between BNC210 and placebo in the mean change from Baseline to Week 12 in the CAPS-5 Total Symptom Severity Score will be evaluated regardless of the occurrence of any such inter-current event.

The primary efficacy endpoint is the change from Baseline to Week 12 in the CAPS-5 Total Symptom Severity Score. A restricted maximum likelihood (REML)-based MMRM model with MI will be used as the primary/main analysis method. The repeated measures include post-baseline visits (i.e., weeks 4, 8, 12) with change from baseline in the CAPS-5 Total Symptom Severity Score as the dependent variable. The MMRM model will include fixed effects for treatment, visit, interaction between treatment and visit, center, and 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.

The primary efficacy analysis will compare BNC210 and placebo using the contrast (difference in least squares [LS] means) between treatment groups through to Week 12. Significance tests

[redacted]

[redacted]

will be based on the LS means using a 2-sided significance level (2-sided 95% CIs).

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

> H_{01}: Mean change in CAPS-5 Total Symptom Severity Scores between baseline and Week 12 in the two treatment groups 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 over the last 12 weeks of the study is less than or equal to 0.05. Least squares (LS) means will be calculated for each treatment group for each post-baseline visit in the model. The difference between BNC210 and placebo change from baseline in the CAPS-5 total score will be estimated, with the corresponding 2-sided 95% CI constructed for each visit. 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 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 mITT population is rejected at the 0.05 (2-sided) significance level.

Summary statistics on the CAPS-5 Total Symptom Severity Score by treatment group and visit, including the end of study visit (Week 12), will be presented. Additionally, a figure with the LS mean ± SE of change from baseline in the CAPS-5 Total Symptom Severity Score will be presented by treatment group and visit.

Supportive analyses will be performed on the primary efficacy endpoint to demonstrate the robustness of the conclusions. The analyses will be conducted on the primary efficacy endpoint using no MI on the mITT and PP populations as described above (see Section 6.1.4.2).

8.2. Secondary Efficacy Analysis

Analyses similar to that described above for the primary efficacy endpoint (i.e., MMRM) will be performed for each of the key secondary and other secondary endpoints, including summary statistics, unless otherwise specified. Key secondary, but not other secondary, endpoints will use MI on the mITT population as the main analysis. Supportive analyses for key secondary endpoints will be conducted similarly to above using MMRM with no MI on the mITT and PP populations.

The null hypotheses for the key secondary efficacy endpoints of the equality of BNC210 and placebo are:

> H_{02}: Mean change in CGI-S between Baseline and Week 12 in the two treatment groups are equal.
>
> H_{03}: Mean change in SDS between Baseline and Week 12 in the two treatment groups are equal.

The null hypotheses for the other secondary efficacy endpoints of the equality of BNC210 and

[redacted]

[redacted]

[redacted]

placebo are:

H_{04}: Mean change in CAPS-5 Intrusion Severity Scores (Cluster B) between Baseline and Week 12 in the two treatment groups are equal.

H_{05}: Mean change in CAPS-5 Avoidance Severity Scores (Cluster C) between Baseline and Week 12 in the two treatment groups are equal.

H_{06}: Mean change in CAPS-5 Negative Alterations in Cognitions and Mood Severity Scores (Cluster D) between Baseline and Week 12 in the two treatment groups are equal.

H_{07}: Mean change in CAPS-5 Arousal and Reactivity Severity Scores (Cluster E) between Baseline and Week 12 in the two treatment groups are equal.

H_{08}: Proportion of participants who achieve a ≥30% improvement in CAPS-5 Total Symptom Severity Scores between Baseline and Week 12 in the two treatment groups are equal.

H_{09}: Proportion of participants who achieve a ≥50% improvement in CAPS-5 Total Symptom Severity Scores between Baseline and Week 12 in the two treatment groups are equal.

H_{10}: Proportion of participants who achieve a score ≤11 on the CAPS-5 Total Symptom Severity Score between Baseline and Week 12 in the two treatment groups are equal.

H_{11}: Mean scores of the CGI-I at Week 12 in the two treatment groups are equal.

H_{12}: Mean change in HAM-A Total Scores between Baseline and Week 12 in the two treatment groups are equal.

H_{13}: Mean change in MADRS Total Scores between Baseline and Week 12 in the two treatment groups are equal.

H_{14}: Mean change in PCL-5 Total Scores between Baseline and Week 12 in the two treatment groups are equal.

H_{15}: Mean change in PGI-S between Baseline and Week 12 in the two treatment groups are equal.

H_{16}: Mean scores of the PGI-I at Week 12 in the two treatment groups are equal.

H_{17}: Mean change in ISI between Baseline and Week 12 in the two treatment groups are equal.

An analysis of covariance (ANCOVA) model will be used to analyze the difference in the mean CGI-I and mean PGI-I scores between BNC210 and placebo at Week 12. The model will include treatment as the main effect and baseline score and center as covariates. The LS mean difference along with the corresponding 95% CI will be reported.

The CGI-I and PGI-I will also be summarized descriptively using frequencies and percentages by treatment group for Significant Improvement and All Other Categories. An analysis will be

[redacted]

[redacted]

performed at Week 12 for the P/CGI-I, in which those with a rating of Very Much Improved or Much Improved (i.e., “Significant Improvement"). Odds ratios and corresponding 95% CIs for significant improvement versus all other categories (including missing values at that visit, where appropriate) will be presented. Additionally, for the binary outcomes of the proportion of participants who achieve a) ≥30% improvement; b) ≥50% improvement; c) a score ≤ 11 on the CAPS-5 Total Symptom Severity Score, the number and proportion will be summarized by treatment group and visit. Proportions will be compared between treatments using Pearson’s chi square or Fisher’s exact test as appropriate. All tests will be performed as 2-sided tests at the 0.05 level of significance.

[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 at least 1 dose of study medication.

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

Safety data collected at the baseline visit (Day 0) 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, SD, 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.

[redacted]

[redacted]

9.1. Adverse Events

In general, TEAEs are defined as AEs that start or deteriorate on or after the first dose of study medication and reported through the Follow-up Day Week 15. For any participants who die during the study and the date of death is between the date of first dose of study medication 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 irrespective of the last dose and will be included in the TEAE summaries. All summaries of AEs will be based on TEAEs unless specified otherwise. All SAEs will be collected from the signing of the ICF until the Follow-up visit at the time points specified in the SoA.

All AEs, TEAEs, and SAEs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA), Version 22.1 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. 22.1 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 (probably 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).

In the AE data listings, all AEs will be displayed. AEs that are treatment-emergent will be flagged.

9.1.1. Adverse Events Leading to Withdrawal

A summary of incidence rates (frequencies and percentages) of TEAEs leading to withdrawal of study drug, by treatment group, SOC, and PT will be prepared for the SAF population. No inferential statistical tests will be performed.

A data listing of AEs leading to withdrawal of study drug 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 and also tabulated by SOC and PT and presented by treatment group, severity, and relationship to study drug.

[redacted]

[redacted]

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, Week 4, Week 6, Week 8, Week 10, Week 12 (or early termination), and Follow Up at Week 15. Participants enrolled in the UK will have an additional visit at Week 2.

A serum pregnancy test (if applicable), HIV, HBsAg, and HCV test will be performed at Screening. A urine drug screen will be performed at Screening, Baseline, Week 2 for UK participants only, Week 4, Week 8, and Week 12 and urine pregnancy tests (if applicable) will be conducted at baseline and Week 12. These results will not be reported unless returned positive. Positive screening tests will be summarized descriptively by treatment.

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 (shift table).

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, Week 4, Week 8, Week 12 (or early termination), and Follow-Up at Week 15. Participants enrolled in the UK will have an additional visit at Week 2. Descriptive summaries of actual values and changes from baseline will be calculated for body temperature, pulse rate, respiratory rate, and seated systolic and diastolic blood pressure (at 5 minutes rest). Vital sign measurements will also be included in the listings.

9.4. Electrocardiograms

Triplicate 12-Lead ECGs will be collected at Screening, Week 4, Week 8, and Week 12 (or early termination). Participants enrolled in the UK will have additional visits at Baseline and Week 2.

[redacted]

[redacted]

Descriptive summaries will be presented for ECG measures of PR interval, QRS interval, QT interval, QT_CF interval, and heart rate (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 the QT_C correction method by treatment group. Fridericia's correction for QT_C will be applied as follows:

Fridericia's Correction (QT_CF)

$$QTc_f = \frac{QT_{msec}}{\sqrt[3]{RR}}$$

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

9.5. Physical Examinations

The number and percentage of participants with normal and abnormal findings in the physical examination at screening, baseline and each post-baseline visit will be displayed for each treatment group. Full physical examination will be done at Screening and at Week 12/Early Termination and abbreviated physical exams will be completed at Baseline, Week 4, Week 8, Week 15 (Follow-Up). 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, and other body systems as required. Height will be measured at Screening and weight will be measured at Screening and Week 12/Early Termination.

Abbreviated physical examinations will include assessments of the eyes, ears, nose, throat, cardiovascular system, respiratory system, abdomen/gastrointestinal system, and other body systems 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, Week 4, Week 8, Week 12/Early Termination, and Follow-Up Week 15. Participants enrolled in the UK will additionally have C-SSRS administered at Week 1 and Week 2. Results will be summarized by treatment group and study visit. C-SSRS data will also be listed.

[redacted] [redacted]

9.7. Concomitant Medication

Prior and concomitant treatments will be summarized descriptively by treatment group using the number and proportion of participants by PT and ATC Class Level 4. Prior treatments will be flagged separately from concomitant treatments. Treatments that started before the first dose of study medication will be considered prior treatments, whether or not they were stopped before the first dose of study medication. Any treatments continuing or starting after the first dose of study medication will be considered to be concomitant. If a treatment starts before the first dose of study medication and continues after the first dose of study medication, it will be considered both prior and concomitant.

10. Pharmacokinetic Analysis

Systemic concentrations of BNC210 will be determined at Weeks 4, 8, and 12. Plasma concentration data will be presented nominally. A data dependent population PK analysis will be outlined under a separate analytical plan.

11. Changes from Planned Analysis

ANCOVA has been added in this version to analyze the difference in the mean CGI-I and mean PGI-I scores between BNC210 and placebo at Week 12.

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 Harmonised 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. 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.
6. Molenberghs G., Kenward M.G. 2007. *Missing Data in Clinical Studies*. New York: Wiley.
7. 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.

8. 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.
9. 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.
10. European Medicines Agency. 2010. Guideline on Missing Data in Confirmatory Clinical Trials.
11. 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.
12. 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.
13. Little R., Rubin D. 1987. *Statistical Analysis with Missing Data*. New York: John Wiley.
14. Verbeke G., Molenberghs, G. 2000. *Linear Mixed Models for Longitudinal Data*. New York: Springer.
15. Lu, K., Mehrotra, D.V. 2010. Specification of covariance structure in longitudinal data analysis for randomized clinical trials. *Statistics in Medicine,* 29, 474-488.