STATISTICAL ANALYSIS PLAN

Nobias Therapeutics, Inc.

NB-001-01

Protocol Title:	A randomized, placebo-controlled crossover trial to assess the safety and efficacy of NB-001 in children and adolescents with 22Q11 Deletion Syndrome
Protocol Version and Date:	Version: 3.1; 06 June 2022
Sponsor:	Nobias Therapeutics, Inc. [redacted]
Prepared By:	[redacted]
Document Version and Date	Version 1.0; 16 June 2023

STATISTICAL ANALYSIS PLAN APPROVAL

Sponsor:	Nobias Therapeutics, Inc.
Clinical Protocol Number:	NB-001-01
Protocol Title:	A randomized, placebo-controlled crossover trial to assess the safety and efficacy of NB-001 in children and adolescents with 22Q11 Deletion Syndrome
Document File and Name:	Nobias_NB-001-01_Statistical Analysis Plan_v1.0_2023-06-16.pdf
Document Version and Effective Date:	Version 1.0; 16 June 2023

Approved By:

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

ABBREVIATIONS

Table 1 List of Abbreviations

Abbreviation	Definition
22q11DS	22q11.2 Deletion Syndrome
ADHD	attention deficit hyperactivity disorder
ADHD-RS-5	Attention Deficit Hyperactivity Disorder-Rating Scale-5
AE	adverse event
ASD	autism spectrum disorder
ATC	Anatomical Therapeutic Chemical
BID	twice daily
BMI	body mass index
CGI	Clinical Global Impression
CGI-I	Clinical Global Impression – Improvement
CGI-S	Clinical Global Impression – Severity
CI	confidence interval
COVID-19	coronavirus disease 2019
C-SSRS	Columbia–Suicide Severity Rating Scale
CSR	clinical study report
CTCAE	Common Terminology Criteria for Adverse Events
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
eCRF	electronic case report form
ECG	electrocardiogram
EOT	end of trial
ENR	Enrolled Analysis Set
ET	early termination
FAS	Full Analysis Set
GAD	Generalized Anxiety Disorder
ICH	International Council for Harmonisation
IP	investigational product
MedDRA	Medical Dictionary for Regulatory Activities
OLS	Ordinary least squares
PARS	Pediatric Anxiety Rating Scale
PK	pharmacokinetic
PPS	Per Protocol Set
Q1	1st quartile (25th percentile)
Q3	3rd quartile (75th percentile)
OLS	ordinary least squares
QTc	corrected QT
SAE	serious adverse event
SAF	Safety Analysis Set

Abbreviation	Definition
SAP	Statistical Analysis Plan
SAD	separation anxiety disorder
SD	standard deviation
SE	standard error
SI	Système International
SIB	suicidal ideation and behavior
SoP	social phobia
SRS™-2	Social Responsiveness Scale, Second Edition
TEAE	treatment-emergent adverse event
TEAE-SI	treatment-emergent adverse event of special interest
WHODDE	World Health Organization Drug Dictionary Enhanced

1. INTRODUCTION

The purpose of this Statistical Analysis Plan (SAP) is to provide comprehensive and detailed descriptions of the methods and presentation of data analyses proposed for Nobias Therapeutics, Inc. Protocol NB-001-01 (A randomized, placebo-controlled crossover trial to assess the safety and efficacy of NB-001 in children and adolescents with 22Q11 Deletion Syndrome [22Q11DS]). Descriptions of planned analyses are provided in order to avoid post hoc decisions that may affect the interpretation of the statistical analysis. The statistical methods applied in the design and planned analyses of this study are consistent with the International Council for Harmonisation (ICH) guidelines *Statistical Principles for Clinical Trials (E9)* (1998) and *Statistical Principals for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials (E9[R1])* (2017).

This SAP will be finalized prior to data analysis and before treatment unblinding and database lock to provide comprehensive details of the tables, figures, and listings to be presented in the Clinical Study Report (CSR). Any changes between the statistical methods provided in the clinical study protocol and this SAP will be explained herein; any changes or deviations from this SAP relative to the final analysis will be fully documented in the CSR. Minor changes or deviations from the templates for tables, figures, and listings need not be documented in the CSR.

2. STUDY OBJECTIVES

2.1 Primary Study Objective

The primary objective of this study is to assess the safety and tolerability of NB-001 during 6 weeks of treatment with NB-001 in subjects with 22q11DS.

2.2 Secondary Study Objectives

The secondary objectives of this study are to assess:

- The efficacy of NB-001 after 6 weeks of treatment with NB-001 using the Clinical Global Impression (CGI) endpoint;
- The efficacy of NB-001 using a scale specific to the following domain/symptom of greatest impairment for each subject based upon baseline clinical assessment:
    - Anxiety (generalized anxiety disorder [GAD], social phobia [SoP], or separation anxiety disorder [SAD]);
    - Attention and executive deficits (i.e., attention deficit hyperactivity disorder [ADHD]);
    - Social interaction and repetitive tendencies (i.e., autism spectrum disorder [ASD]);
- The efficacy of NB-001 using scales specific to the additional symptom domains above.

3. INVESTIGATIONAL PLAN

3.1 Overall Study Design

This is a Phase 2, randomized, placebo-controlled crossover trial to assess the safety and efficacy of NB-001 in children and adolescents with 22q11DS that manifest commonly associated neuropsychiatric conditions.

Approximately 34 subjects will be randomized across 3-5 trial centers located in the United States and Canada. Additional subjects may be enrolled, at the discretion of the Sponsor, to replace subjects who terminate early from the trial.

The trial is designed to allow all visits to be conducted via telemedicine or by a home health nurse. An in-person visit is required at Screening unless site or government mandates restrict this due to coronavirus disease 2019 (COVID-19). Other in-person visits may occur, if indicated, based on the Investigator's clinical judgement.

Subjects will be screened to confirm eligibility and then randomized in a 1:1 ratio to one of two treatment sequences: NB-001 followed by placebo or placebo followed by NB-001.

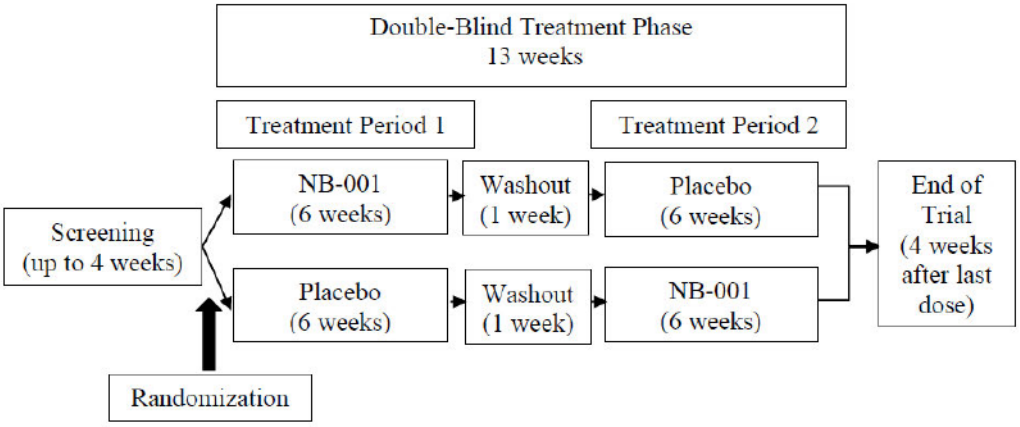
Subjects will receive investigational product (IP) corresponding with the first treatment assignment for 6 weeks (Treatment Period 1). This will be followed by an intervening wash-out period of 1 week, and then subjects will receive their second treatment assignment for the subsequent 6-week period (Treatment Period 2). The wash-out period may be extended beyond one week (and within reason) to accommodate the subject's schedule, after consultation with the Clinical Research Associate or Medical Monitor. During the Double-Blind Treatment Phase, the subject and/or parent/guardian will be contacted at Day 0 to complete baseline symptom scales and will begin dosing with the IP (NB-001 or placebo) on the morning of Day 1.

Subjects or their parent/guardian will administer the IP twice daily (BID) and will be contacted at Days 0, 1, 14, 28, 42, 49, 50, 63, 77 and 91 to evaluate measures of safety and efficacy, including the completion of symptom scales. All symptom scales will be centrally and/or locally administered. In addition, the subject and/or parent/guardian will be contacted at Days 7, 21, 35, 56, 70 and 84 to assess subject safety. Blood samples for pharmacokinetic (PK) analysis, 4β-hydroxycholesterol and plasma proline will be collected at the timepoints noted in the schedule of assessments.

The subject and/or parent/guardian will be contacted for an End of Trial (EOT) Visit to occur 4 weeks following the last dose of IP to assess safety. If the Investigator perceives a subject is experiencing clinical benefit to treatment and the subject has completed the trial, provisions may be made to allow the subject to continue to receive NB-001. Details will be provided in a separate open-label extension protocol.

A schematic of the trial design is provided in the figure below.

Figure 1 Trial Design



3.2 Schedule of Assessments

For the complete schedule of assessments, refer to Table 3 in Section 6.1 of the clinical study protocol.

3.3 Treatments

3.3.1 Treatments Administered

Due to the cross-over study design, all subjects will receive both NB-001 and placebo capsules. Each NB-001 capsule contains 100 mg of the active ingredient. Placebo capsules contain excipients only and match the NB-001 capsules in appearance.

NB-001 and matching placebo will be supplied to the Investigator or designee in blinded plastic bottles, each containing 40 capsules.

NB-001 or placebo capsules will be administered orally to the subject. Subjects will take two capsules BID, two in the morning and two in the evening (i.e., four capsules daily). The subject will begin dosing for Treatment Period 1 on the morning of Day 1, after collection of pre-dose samples. Similarly, the subject will begin dosing for Treatment Period 2 on the morning of Day 50 after collection of pre-dose samples.

3.3.2 Method of Assigning Subjects to Treatment Groups

Upon confirmation of trial eligibility, subjects will be randomized at the conclusion of the Screening Visit. Randomization will occur in a 1:1 ratio to one of two treatment sequences: NB-001 followed by placebo or placebo followed by NB-001.

3.3.3 Blinding Procedures

The randomization scheme will be generated using a computer program and verified for accuracy using strict quality control procedures. Each IP bottle will be labeled with a unique bottle number which will correspond to the randomization code and blinded treatment assignment.

In the event of a medical emergency in which it is important for the treating physician to know the treatment assignment, the blind may be broken by the Investigator. Whenever possible (i.e., if time allows), the Investigator will contact the Medical

Monitor, or designee, prior to accessing the treatment code for the subject. Instructions for emergency unblinding of treatment codes by the Investigator will be provided in a separate data access plan.

3.4 Efficacy and Safety Variables

3.4.1 Efficacy Variables

3.4.1.1 Primary Efficacy Variable

The primary efficacy endpoint is the Clinical Global Impression-Improvement (CGI-I) score per the central rater.

The CGI-I scale will be performed by a central rater and where possible (i.e., optional), a local rater. The CGI-I assessments at Days 14, 28, 42 and 49 will use the Day 0 Clinical Global Impression-Severity (CGI-S) baseline rating as a reference for change. The CGI-I assessments at Days 63, 77, 91 and 119 will use the Day 49 CGI-S baseline rating as reference for change.

The CGI-S scale will be used to establish the degree of disease severity for each subject. The CGI-I scale will be used as an outcome measure for estimating level of functioning in response to treatment. The clinician's global assessments of the severity of the symptoms and changes in symptoms from baseline (Day 0 or Day 49, per above) will be based on reports from the subject and/or parent/guardian, and minimal direct observation (Guy 1976; Busner and Targum 2007).

At the visits noted in the schedule of assessments, the rater(s) will assess the subject's improvement relative to his/her symptoms at baseline (Day 0 or Day 49, per above) using a 7-point scale ranging from 1 (very much improved) to 7 (very much worse).

3.4.1.2 Secondary Efficacy Variables

Secondary efficacy endpoints include the following assessments:

CGI-S

The CGI-S scale will be used to rate the severity of the subject's condition on a 7-point scale ranging from 1 (Normal, not at all ill) to 7 (Among the most extremely ill subjects). It will be administered at Screening and Days 0 and 49 prior to dosing, and at additional visits noted in the schedule of assessments.

Pediatric Anxiety Rating Scale (PARS)

The PARS is a clinician-rated instrument for assessing, over time, the severity of anxiety symptoms that are associated with common Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) (American Psychiatric Association 2013), anxiety disorders (GAD, SAD, and SoP) in subjects aged 6 to 17 years. The PARS includes two sections. The first section is a 50-item symptom checklist, which the clinician rates as present or absent during the past week. The second section is comprised of 7 severity/impairment items that are rated on a 6-point Likert scale. The 7 severity/impairment items reflect the severity/impairment of all symptoms noted in

Section 1 of the PARS (during the previous week) (Research Units on Pediatric Psychopharmacology Anxiety Study Group 2002). The clinician will interview both the subject and the parent/guardian to complete this assessment at the visits described in the schedule of assessments. The PARS severity score is calculated as the sum of items 2, 3, 5, 6, and 7 and will be provided in the external data transfer along with the individual item scores.

Attention Deficit Hyperactivity Disorder-Rating Scale-5 (ADHD-RS-5)

The ADHD-RS-5 will be administered at the visits noted in the schedule of assessments to document the subject's ADHD symptoms over the previous week. The ADHD-RS-5 Home Version is a parent/guardian reported scale that was developed to measure the behaviors of children with ADHD, with separate forms for children (ages 5-10 years) and adolescents (11-17 years). The ADHD-RS-5 consists of 18 items designed to reflect the symptomatology of ADHD based on the DSM-5 criteria. Each item will be scored on a scale ranging from 0 (reflecting no symptoms) to 3 (reflecting severe symptoms). The 18 items will be grouped into two subscales: inattention and hyperactivity/impulsivity. The Inattention subscale raw score is computed by summing the item scores for Items 1 through 9 and the Hyperactivity-Impulsivity subscale raw score is computed by summing the item scores for Items 10 through 18. Both subscale scores will be recorded on the electronic case report form (eCRF).

Additionally, the ADHD-RS-5 incorporates two impairment scales keyed to the inattention and hyperactivity-impulsivity dimensions, which allow the clinician to assess the extent to which ADHD-related problems adversely affect the home and/or school functioning of children and adolescents (DuPaul et al. 2016). Further details are provided in Appendix 4 of the clinical study protocol.

Social Responsiveness Scale, Second Edition (SRS™-2)

The SRS™-2 identifies the presence and severity of social impairment within the autism spectrum and differentiates it from that which occurs in other disorders. It is a 65-item, parent completed questionnaire which incorporates 5 content areas of social deficits: social awareness, social cognition, social communication, social motivation, and autistic mannerisms. The SRS™- 2 will be proctored by the local rater and derived and subscale raw and T-data scores will be reported on the eCRF (Constantino et al. 2003; Constantino and Gruber 2005). Further details are provided in Appendix 5 of the clinical study protocol.

3.4.2 Safety Variables

Safety assessments will include monitoring and recording of adverse events (AEs), clinical laboratory tests, vital sign measurements, physical exam, electrocardiograms (ECGs), and the Columbia–Suicide Severity Rating Scale (C-SSRS).

3.4.2.1 Adverse Events

An AE is defined as any untoward medical occurrence in a clinical investigation subject administered an IP and which does not necessarily have a causal relationship with the product. An AE can therefore be any unfavorable and unintended sign (including an abnormal laboratory finding), symptom, or disease temporally associated with the use

of a pharmaceutical product, whether or not it is considered related to the pharmaceutical product.

Pre-existing events that increase in frequency or severity in nature during or as a consequence of use of an IP will also be considered as AEs. Any medical condition or clinically significant laboratory abnormality with an onset date before the date of informed consent signature is pre-existing and should be documented as medical history. Any AE that occurs between the date of informed consent signature and the date/time of first IP administration will be considered a pre-treatment-emergent AE. Any AE that occurs after the date/time of first IP administration will be considered a treatment-emergent adverse event (TEAE).

A treatment-emergent adverse event of special interest (TEAE-SI) is an AE of scientific and medical concern specific to the Sponsor's IP/device or program, which may warrant ongoing monitoring and/or rapid communication by the investigator to the Sponsor or designee. Such an event might warrant further investigation in order to characterize and understand it. TEAEs-SI in this trial will include: development of psychotic symptoms, suicidal ideation and behavior (SIB), seizure, overdose with untoward medical sequalae, and confirmed cases of COVID-19.

All AEs will be assessed for severity according to the Common Terminology Criteria for Adverse Events (CTCAE) Version 5.0 (effective 27 November 2017).

3.4.2.2 Laboratory Parameters

Clinical laboratory tests will include chemistry, hematology, and urinalysis as listed below in Table 2. Samples will be obtained at the Screening visit, Day 0, Day 42, Day 91/Early Termination (ET), and Day 119/EOT Visits.

Table 2 Clinical Laboratory Tests

Chemistry	**Hematology**	**Urinalysis**
Alkaline phosphatase	Hemoglobin	Glucose
Aspartate aminotransferase	Hematocrit	Protein
Alanine aminotransferase	Mean corpuscular hemoglobin	Specific gravity
Total bilirubin	Mean corpuscular hemoglobin	pH
Total protein	concentration	Nitrite
Albumin	Mean corpuscular volume	Bilirubin
Glucose	Erythrocyte count	Urobilinogen
Carbon dioxide	Leukocyte count	Ketone
Blood urea nitrogen	Neutrophil count and percentage	Blood
Creatinine	Lymphocyte count and	Leukocyte esterase
Sodium	percentage	
Potassium	Monocyte count and percentage	**Other**
Chloride	Eosinophil count and percentage	eGFR (Screening
Calcium	Basophil count and percentage	only, revised
	Platelet count	Schwartz
		equation)

Other laboratory tests will include serum and urine pregnancy tests in females of childbearing potential, and urine drug tests.

Reports containing these laboratory test results will be generated by the laboratory performing the tests. Laboratory values outside of the normal range will be assessed for clinical significance by the Investigator. Abnormal laboratory tests may be repeated at the discretion of the Investigator or Sponsor.

3.4.2.3 Vital Signs

Vital signs will include blood pressure (systolic and diastolic), heart rate, respiratory rate, and temperature and will be performed after the subject has been at rest in the seated position for at least 5 minutes. Vital signs will be assessed at the Screening Visit and as clinically indicated per the schedule of assessments.

3.4.2.4 Physical Examination

A complete physical examination will be performed at the Screening Visit, Day 0, Day 42, Day 91/ET Visit, and Day 119/EOT Visit, and will include an assessment of the following body systems: general appearance; weight; height (at the Screening Visit only) mental status; head, eyes, ears, nose, and throat; dermatologic; cardiovascular; respiratory; gastrointestinal; musculoskeletal; and neurological at Screening. Additional body systems may be examined at the Investigator's discretion. An abbreviated, clinically focused physical examination based on the Investigator's clinical judgement will be administered at all other assessment intervals or as clinically indicated.

3.4.2.5 Electrocardiogram

A 12-lead ECG in triplicate will be obtained at the at the Screening Visit, Day 0, Day 42, Day 91/ET Visit, and Day 119/EOT Visit, with the subject in a supine position following at least a 5-minute rest. The Investigator will review and report the results of each ECG; the reviews on Days 42 and 91 will include a comparison of each ECG to the Day 0 ECG.

3.4.2.6 Columbia-Suicide Severity Rating Scale

Consistent with Food and Drug Administration (FDA) Guidance, *Guidance for Industry Suicidal Ideation and Behavior: Prospective Assessment of Occurrence in Clinical Trials*, (FDA 2012), the C-SSRS is used to evaluate SIB. The scale consists of 10 categories with binary responses (yes/no) to assess the subject's suicidal thoughts. An additional question regarding self-injurious behavior without suicidal intent will be evaluated with a binary response (yes/no). The C-SSRS will be used at the visits noted in the schedule of assessments to evaluate the subject's suicidal tendencies.

3.5 Data Quality Assurance

Report summaries will be generated using validated Base SAS® software, version 9.4 or higher, on a PC or server-based platform. Additional validated software may be used to generate analyses, as needed.

All SAS programs that create outputs or supporting analysis datasets will be validated by a second statistical programmer or biostatistician. At a minimum, validation of programs will consist of a review of the program log, review of output or dataset format and structure, and independent confirmatory programming to verify output results or dataset content. Additionally, all outputs will undergo a review by a senior level team member before finalization.

The content of the source data will be reviewed on an ongoing basis by project statistical programmers and statisticians. Data will be checked for missing values, invalid records, and extreme outliers through defensive programming applications, analysis-based edit checks, and other programmatic testing procedures. All findings will be forwarded to the project data manager for appropriate action and resolution.

4. STATISTICAL METHODS

4.1 General Methodology

Data will be analyzed by Emanate biostatistics personnel. Statistical analyses will be reported with tables, figures, and listings, presented in rich text format, and using recommended ICH numbering. Output specifications for all tables, figures, and listings will be in conformance with guidelines specified by the ICH in Appendix 7 of the Electronic Common Technical Document Specification (Apr 2003).

4.1.1 Reporting Conventions

Tables and figures will be summarized by either treatment sequence or treatment group, as further described in Section 4.4 and Section 4.5 of the SAP. In general, all data collected and any derived data will be presented in subject data listings, for all enrolled subjects. Listings will be ordered by site, subject number, treatment sequence, treatment group, and assessment or event date. The treatment sequence presented in listings will be based on the planned assignment, unless otherwise noted.

In general, continuous variables will be summarized to indicate the study population sample size (N), number of subjects with available data (n), mean, standard deviation (SD), median, first (Q1) and third (Q3) quartiles, minimum, and maximum values. Categorical variables will be summarized by the population size (N), number of subjects with available data (n), number of subjects in each category, and the percentage of subjects in each category. Unless otherwise noted, the denominator to determine the percentage of subjects in each category will be based on the number of subjects with available data. Select ordinal data may be summarized using both descriptive statistics and counts and percentages of subjects in each category, as appropriate.

Non-zero percentages will be rounded to one decimal place. Rounding conventions for presentation of summary statistics will be based on the precision of the variable of summarization, as it is collected in its rawest form (i.e., on the eCRF or as provided within an external file) and are outlined as follows:

- The mean and median will be rounded to one more decimal place than the precision of the variable of summarization;

- Measures of variability (e.g., SD, standard error [SE]) will be rounded to two more decimal places than the precision of the variable of summarization; and

- Minimum and maximum values will be presented using the same precision as the variable of summarization.

Other statistics (e.g., confidence intervals [CIs]) will be presented using the same general rules outlined above, or assessed for the most appropriate presentation based on the underlying data.

Unless otherwise specified, statistical significance testing will be two-sided and performed using α=0.05. P-values will be reported for all statistical tests, rounded to four decimal places. P-values less than 0.0001 will be displayed as "<0.0001." P-values greater than 0.9999 will be displayed as ">0.9999." Tests of interaction terms, if applicable, will be two-sided and performed using α=0.10.

4.1.2 Summarization by Visit

Data summarized by study visit will be based on the nominal, scheduled visit label as reported on the eCRF for each treatment group. Each treatment period consists of 42 days and study visits in Treatment Period 2 will be labeled according to the respective day within that treatment period. For example, the Day 63 visit occurs during Treatment Period 2 and will be labeled as Day 14 within that period. Data collected for the last subject treatment visit completed within each treatment period will be summarized separately for:

- The scheduled visit for those subjects who complete the scheduled end of treatment visit in each treatment period per protocol. This will be the scheduled Day 42 for subjects in Treatment Period 1 and Day 91 for subjects in Treatment Period 2. This summary will be labeled as "Day 42" in the analysis; and

- The last treatment visit completed on-study, combining data collected for subjects who complete the scheduled end of treatment visit, as well as the visit data for those subjects who discontinue the study early. This summary will be labeled as "Day 42/ET" in the analysis.

Data collected at unscheduled visits within the double-blind treatment period will not be included in by-visit summaries, but will be considered when endpoint derivations potentially include multiple visits (e.g., determination of baseline value, determination of worst post-baseline value, etc.). All data will be included in subject listings.

Any summarization of data collected during the follow-up period will be presented separately and will be based on the nominal, scheduled visit label as reported on the eCRF.

4.1.3 Data Handling Rules

Unless otherwise noted, values reported as greater than or less than some quantifiable limit (e.g., "< 1.0") will be summarized with the sign suppressed in summary tables and figures, using the numeric value reported. Data will display on subject listings to include the sign.

4.1.4 Standard Calculations

Where appropriate, the calculated study day of each assessment or event will be presented with the assessment or event date on subject data listings, where study day will be determined as:

- The assessment/event date minus the date of first dose of IP, if the assessment/event date is prior to the date of first dose; and
- The assessment/event date minus the date of first dose of IP, plus one, if the assessment/event date is on or after the date of first dose.

Unless otherwise noted, study day will be calculated relative to the date of first dose of IP.

Other variables requiring calculations will be derived using the following formulas:

- **Days:** A duration between two dates expressed in days will be calculated using the following conventions:
    - Later date – earlier date + 1, if the earlier date is on or after the date of first dose of IP; or
    - Later date – earlier date, if the earlier date is prior to the date of first dose of IP.
- **Months:** A duration expressed in months will be calculated by dividing the duration in days by (365.25 / 12).
- **Years:** A duration expressed in years will be calculated by dividing the duration in days by 365.25.
- **Change from Baseline:** Change from baseline will be calculated as the post baseline value minus the baseline value.
- **Percentage Change from Baseline:** Percentage change from baseline will be calculated as the change from baseline divided by the baseline value, multiplied by 100.

4.2 Analysis Sets

The analysis sets are defined as follows:

- Enrolled Analysis Set (ENR): The ENR consists of all subjects who meet all eligibility criteria and consent to participate in the study. All listings will be based on the ENR unless otherwise noted.
- Full Analysis Set (FAS): The FAS consists of ENR subjects who have at least one valid post-baseline efficacy evaluation within each treatment period. Subjects will be analyzed based on the treatment to which they were randomized. Subjects in the FAS are referred to as evaluable subjects.

- Safety Analysis Set (SAF): The SAF consists of all subjects who receive at least one capsule (100 mg) of IP. Subjects will be analyzed based on the treatment received.

- Per Protocol Analysis Set (PPS): The PPS includes all subjects in the FAS who have no major protocol violations. Subjects will be analyzed based on the treatment received.

Data summaries to be presented on the SAF, FAS, and PPS will only be produced on all three analysis sets if there is a difference in the population groups.

4.3 Study Subjects

4.3.1 Disposition of Subjects

Subject disposition will be summarized for the SAF by treatment sequence and over all subjects combined. Summaries will include the number and percentage of subjects in each analysis set, completing the study, and discontinuing the study early by the primary reason for discontinuation.

4.3.2 Protocol Deviations

Major protocol violations will be summarized by treatment group for the FAS. Major protocol violations are protocol deviations captured on-study that are deemed by the Sponsor to potentially impact the efficacy or safety conclusions of the study.

All major protocol violations will be determined and appropriately categorized prior to database lock and prior to breaking the blind of the treatment group assignments. The number and percentage of subjects with any major protocol violations as well as the number and percentage of subjects with violations within each category will be presented.

4.4 Efficacy Evaluation

4.4.1 Datasets Analyzed

All efficacy summaries will be based on the FAS; select efficacy summaries will also be produced on the PPS as a sensitivity analysis. A data listing of subjects excluded from the FAS or PPS, to include the reason for exclusion, will be presented.

4.4.2 Demographic and Other Baseline Characteristics

Demographic and baseline characteristic variables will be summarized by treatment sequence and over all subjects combined for the SAF, FAS, and PPS.

Demographic variables include age, sex, ethnicity, and race. Age will be reported as collected on the eCRF and summarized using descriptive statistics. Sex, ethnicity, and race will be summarized with the number and percentage of subjects in each parameter category.

Baseline characteristics include height, weight, body mass index (BMI), medical history, WASI-II assessment, and domain/symptom of greatest impairment. Height, weight, BMI at baseline, and WASI-II assessment scores will be summarized using descriptive statistics. The domain/symptom of greatest impairment will be determined using the subject inclusion criteria described in Section 7.1 of the clinical study protocol and is further described later in Section 4.4.7.8. The number of subjects in each category (anxiety symptoms in clinical range, anxiety symptoms in subclinical range, ADHD symptoms in clinical range [majority related to inattention], ADHD symptoms in subclinical range [majority related to inattention], autism symptoms in clinical range, and autism symptoms in subclinical range) will be summarized with frequency counts and percentages.

Medical history conditions will be mapped to preferred terms and system organ classes using the Medical Dictionary for Regulatory Activities (MedDRA, version 24.1). Frequency counts and percentages to summarize subjects reporting abnormal medical history by system organ class and preferred term will be presented.

4.4.3 Measurements of Treatment Compliance

Compliance with the study treatment regimen will be determined as the total number of doses received divided by the expected number of doses received, multiplied by 100. The expected number of doses is defined as the date of last dose of IP minus the date of first dose of IP, plus one, times two. The total number of doses received will be calculated as the number of expected doses minus the number of missed doses as reported on the *Subject Checks* eCRF. Dosing compliance will be summarized using descriptive statistics, by treatment group, based on the SAF. The number and percentages of subjects who are < 80% compliant and ≥ 80% compliant within each treatment group will be summarized.

4.4.4 Primary Efficacy Endpoint Analysis Methods

The primary estimand is the difference in means between treatment groups in the CGI-I scale per central rater for all subjects in the FAS. Data collected at post-baseline evaluation time points will be combined across the two treatment periods depending on the treatment received.

The null hypothesis to be tested is that there is no difference between Placebo and NB-001:

$$H_0: \mu_A = \mu_B;$$

Where μ_A and μ_B represent the mean values for Placebo and NB-001, at the end of each respective treatment period. The alternate hypothesis to be tested is that the treatment group means differ:

$$H_1: \mu_A \neq \mu_B;$$

The treatment effect will be evaluated using a mixed-effects model that includes treatment, visit, treatment-by-visit interaction, period, and sequence as fixed effects, subject nested within sequence as a random effect, and the baseline CGI-S value as a covariate. The within-subject errors will be modeled using an unstructured covariance

matrix. The Kenward-Roger approximation will be used to estimate denominator degrees of freedom. If the model fails to converge, other covariance structures will be evaluated until model convergence is met.

4.4.5 Secondary Endpoint Analysis Methods

The CGI-I per local rater, CGI-S per central rater, CGI-S per local rater, PARS Total Severity Score per central rater, PARS Total Severity Score per local rater, ADHS-RS-5 (Inattention and Hyperactivity Subscale Scores), and the SRS™-2 Total Derived T-Score will be analyzed using the same statistical methodologies as applied to the primary efficacy endpoint, described in Section 4.4.4. For all endpoints where baseline is collected (i.e., excluding the improvement scales), the baseline value will be added as a covariate in the mixed-effects model.

The analyses for the CGI-I per central rater, CGI-S per central rater, PARS per central rater, ADHS-RS-5, and SRS™-2 as described above will also be presented in separate summaries by treatment period. For these analyses, the mixed-effects model used to evaluate the treatment effect will include treatment, visit, and treatment-by-visit interaction as fixed effects and the baseline value as a covariate.

[redacted]

For all efficacy measures, mean scores over time will be plotted and presented by treatment group as well as by treatment sequence. Spaghetti plots will also be provided to show individual scores over time by treatment sequence.

For the CGI-I per central rater, scores collected during the washout and follow-up periods will be summarized by treatment sequence using descriptive statistics.

4.4.7 Statistical/Analytical Issues

4.4.7.1 Adjustments for Covariates

For all endpoints where baseline is collected (i.e., all except the improvement scales), the mixed-effects model to compare treatment groups will include a covariate adjustment for the baseline value.

4.4.7.2 Handling of Dropouts or Missing Data

No explicit imputations will be performed on missing data; all analyses will be based on observed data only.

4.4.7.3 Interim Analyses and Data Monitoring

There are no interim analyses planned, nor is there a plan to establish a data monitoring committee for this study.

4.4.7.4 Multicenter Studies

Efficacy data collected from all study sites will be pooled for data analysis. The effect of study site on the efficacy analysis results may be explored post hoc, as needed.

4.4.7.5 Multiple Comparisons/Multiplicity

An overall treatment effect evaluation will be performed using the global test for multiple endpoints (O'Brien 1984; Dmitrienko and Tamhane 2009). This test relies on pooling the evidence of effectiveness across several endpoints. The global alternative hypothesis to be tested is that NB-001 is uniformly better than placebo in the mean changes from baseline in the PARS Total Severity Score, ADHD-RS-5 Inattention Subscale Score, and SRS™-2 Total Derived T-Score scores when evaluated as a single instrument.

The procedure will be based on the ordinary least squares (OLS) estimate of the common effect size. In this method, the OLS test statistic will be computed using the following formula:

$$t_{OLS} = \frac{t_1 + t_2 + \cdots + t_m}{\sqrt{1 + R_{11} + \cdots + R_{1m} + 1}}$$

where $m = 3$ is the number of endpoints, $t_1, t_2, \ldots, t_m$ correspond to the t-statistics computed for each individual endpoint score, and $R_{11}, \ldots, R_{1m}$ correspond to the Pearson correlation coefficients as displayed in the correlation matrix below:

$$\hat{R} = \begin{bmatrix} 1 & \cdots & R_{1m} \\ \vdots & \ddots & \vdots \\ R_{1m} & \cdots & 1 \end{bmatrix}$$

The global one-sided p-value corresponding to this test statistic will be calculated using the Logan-Tamhane formula (Logan and Tamhane 2004), and will be considered significant at the 0.025 level.

The data will be checked for normality and a nonparametric global testing approach will be applied if major deviations from normality are detected.

4.4.7.6 Use of an "Efficacy Subset" of Subjects

The primary efficacy analysis will be performed on the FAS; the PPS will be utilized as a sensitivity analysis. The PPS will exclude subjects with major protocol violations.

A supplemental sensitivity analysis may also be performed for subjects with efficacy assessments performed out of window for the Day 42 and/or Day 91 study visits.

4.4.7.7 Active-Control Studies Intended to Show Equivalence

This study does not include an active-control product and is not intended to demonstrate equivalence between any two drug products.

4.4.7.8 Examination of Subgroups

The efficacy analyses for PARS Total Severity Score per central rater, PARS Total Severity Score per local rater, ADHS-RS-5 (Inattention and Hyperactivity Subscale Scores), and SRS™-2 Total Derived T-Score will be first evaluated within the subset of subjects with the corresponding significant domain/symptom of greatest impairment (e.g., the treatment effect on the PARS will be assessed within the subset where the symptom of greatest impairment is anxiety). The domain/symptom of greatest impairment will be determined at baseline using the psychiatric symptoms described below (inclusion criteria # 3 in Section 7.1 of the clinical study protocol):

a) Psychiatric symptoms in the **clinical** range for at least one of three disorders (anxiety, attention/executive deficits, and social interaction/repetitive tendencies) as demonstrated by scores at or above the following numbers on at least 1 of 3 scales:

   - PARS 5-Item Severity Score ≥ 12
   - ADHD-RS-5 Scores of 2 or 3 on at least 6 questions, with the majority of symptoms related to inattention rather than hyperactivity
   - SRS™-2 Derived T-Score ≥ 60; or

b) Psychiatric symptoms in the **subclinical** range for at least two of three scales:

   - PARS 5-Item Severity Score of 10 or 11
   - ADHD-RS-5 Scores of 2 or 3 on 4 or 5 questions, with the majority of symptoms related to inattention rather than hyperactivity
   - SRS™-2 Derived T-Score of 55-59.

Subjects may be counted in up to three subgroups, corresponding to any clinical or subclinical symptoms they report for the PARS, ADHD-RS-5, and SRS™-2 scales. The six possible subgroups in which a subject may be counted are listed below:

- Subjects with Anxiety Symptoms in the Clinical Range at Enrollment
- Subjects with Anxiety Symptoms in the Subclinical Range at Enrollment
- Subjects with ADHD Symptoms in the Clinical Range (Majority Related to Inattention) at Enrollment
- Subjects with ADHD Symptoms in the Subclinical Range (Majority Related to Inattention) at Enrollment
- Subjects with Autism Symptoms in the Clinical Range at Enrollment
- Subjects with Autism Symptoms in the Subclinical Range at Enrollment

For example, if a subject reports symptoms in the clinical range for anxiety and symptoms in the subclinical range for ADHD, then this subject would be included in both the Subjects with Anxiety Symptoms in the Clinical Range at Enrollment and the Subjects with ADHD Symptoms in the Subclinical Range at Enrollment subgroups.

In addition, the analyses for the CGI-I per Central Rater and the CGI-S per Central Rater will be repeated for three subgroups of the Full Analysis Set:

- Subjects with Anxiety Symptoms at Enrollment
- Subjects with ADHD Symptoms at Enrollment
- Subjects with Autism Symptoms at Enrollment

For each of these subgroups, all subjects with any symptoms in the clinical or subclinical range at enrollment as described above in inclusion criteria # 3 in Section 7.1 of the clinical study protocol will be included in the analysis.

Additional subgroup analyses may be performed post hoc, as appropriate.

4.4.8 Plasma Concentrations

Plasma concentration measurements will be summarized in a separate report and are outside the scope of this SAP.

4.4.9 Pharmacokinetic Analysis

Pharmacokinetic analysis will be summarized in a separate report and is outside the scope of this SAP.

4.5 Safety Evaluation

Safety analysis will be carried out for the SAF, to include all subjects who receive at least one capsule of IP. Subjects who do not complete both treatment periods, for whatever reason, will have all available data up until the time of termination included in the analysis. For safety analysis presented by study visit, unless otherwise stated, the

baseline value will be the last measurement reported prior to the first dose of any IP at Day 1 during Treatment Period 1 or Day 50 during Treatment Period 2.

4.5.1 Extent of Exposure

Extent of exposure to study treatment will be summarized for the SAF by treatment group. The duration of exposure will be presented in days and calculated as the date of last dose of IP minus the date of first dose of IP, plus one. Duration of exposure and total dose received (mg) will be summarized using descriptive statistics.

4.5.2 Adverse Events

Treatment-emergent AEs are defined as those AEs with onset after the first dose of IP or existing events that worsened after the first dose during the study.

Treatment-emergent AEs will be summarized by treatment group through the end of Treatment Period 2. Select TEAE summaries through the end of Treatment Period 2 will also be presented by treatment period. Any AEs reported during the post treatment Follow-up period will be summarized separately by treatment sequence.

Each TEAE will be assigned to a single treatment using the date and time of onset. A TEAE that occurs after the first dose of IP in Treatment Period 1 but before the first dose of the IP in Treatment Period 2 will be assigned to the treatment associated with Treatment Period 1. This includes TEAEs that may occur in the washout interval between treatment periods.

A TEAE that continues over both treatment periods will be counted only once in the period during which the event started unless the TEAE worsens in severity, in which case the TEAE will be attributed to the treatment period during which it progressed to its most severe intensity. Events reported with a partial onset date (e.g., month and year are reported but the day is missing) will be considered to be treatment-emergent if it cannot be confirmed that the event onset was prior to the first dose of IP based on the available date entries.

Verbatim terms on case report forms will be mapped to preferred terms and system organ classes using the Medical Dictionary for Regulatory Activities (MedDRA, version 24.1).

Summaries that are displayed by system organ class and preferred terms will be ordered by descending incidence of system organ class and preferred term within each system organ class. Summaries displayed by preferred term only will be ordered by descending incidence of preferred term. Summaries of the following types will be presented:

- Overall summary of number of unique TEAEs and treatment-emergent serious adverse events (SAEs) and subject incidence of TEAEs meeting various criteria;
- Subject incidence of TEAEs by MedDRA system organ class and preferred term;
- Subject incidence of the most frequently-occurring TEAEs (i.e., TEAEs occurring in ≥ 10% of the SAF) by MedDRA preferred term;

- Subject incidence of TEAEs by CTCAE grade, MedDRA system organ class, and preferred term;
- Subject incidence of TEAEs by relationship to IP, MedDRA system organ class, and preferred term;
- Subject incidence of the most frequently-occurring TEAEs related to IP (i.e., related TEAEs occurring in ≥ 10% of the SAF) by MedDRA preferred term;
- Subject incidence of ≥ Grade 3 TEAEs related to IP by MedDRA system organ class and preferred term;
- Subject incidence of SAEs by MedDRA system organ class and preferred term;
- Subject incidence of serious TEAEs by CTCAE grade, MedDRA system organ class, and preferred term; and
- Subject incidence of serious TEAEs by relationship to IP, MedDRA system organ class, and preferred term.

In addition, subject incidence of TEAEs by system organ class and preferred term will be summarized by treatment sequence for the washout and follow-up periods.

At each level of summarization (e.g., any AE, system organ class, and preferred term), subjects experiencing more than one TEAE will be counted only once. In the summary of TEAEs by CTCAE grade, subjects will be counted once at the highest grade reported at each level of summarization; in the summary of TEAEs by relationship, subjects will be counted once at the closest relationship to IP. Adverse event data will be presented in data listings by subject, treatment sequence, treatment group, and event.

4.5.3 Deaths, Other Serious Adverse Events, and Other Significant Adverse Events

All deaths during the study, including the post treatment follow-up period, will be listed by subject, to include the primary cause of death. Serious AEs and other significant AEs, including those that led to withdrawal, interruption, or dose reduction of the IP, will be provided in separate subject data listings.

Treatment-emergent AE-SIs in this trial will include the development of psychotic symptoms, SIB, seizure, overdose with untoward medical sequalae, and confirmed cases of COVID-19, and will be provided in a subject data listing.

4.5.4 Clinical Laboratory Evaluation

All descriptive summaries of laboratory results will be based on data analyzed by the central laboratory and presented in Système International (SI) units, as suggested by the Center for Biologics Evaluation and Research and the Center for Drug Evaluation and Research *Position on Use of SI Units for Lab Tests* (Oct 2013). Summaries will also be presented in conventional units. All data will be included in by-subject data listings. Laboratory measurements identified as abnormal (i.e., outside the normal range) will also be listed separately by subject, laboratory test, and unit. In addition, normal ranges provided by the central laboratory will be presented in a separate listing.

Clinical laboratory measurements, including serum chemistry and hematology, will be summarized by treatment group. Descriptive statistics will be presented for observed values and changes from baseline at each visit where parameters were scheduled to be collected per the clinical study protocol.

Where applicable, laboratory results will be classified as "low," "normal," or "high" with respect to the parameter-specific reference ranges (i.e., below the lower limit of the normal range, within the normal range, or above the upper limit of the normal range). Three-by-three contingency tables will be presented for each laboratory parameter to summarize the shift from the baseline category to the worst post-baseline measurement, defined as the value numerically farthest outside of the normal range across all post-baseline visits through the end of the study.

Summary results will include the count and percentage of subjects within each shift category and treatment group.

4.5.5 Vital Signs, Physical Findings, and Other Observations Related to Safety

4.5.5.1 Vital Signs

Vital sign parameter measurements will be summarized by treatment group. Descriptive statistics will be presented for results and change from baseline at each visit where parameters were scheduled to be collected.

4.5.5.2 12-Lead Electrocardiogram

Select twelve-Lead ECG interval parameters to include heart rate, PR interval, QRS duration, and corrected QT (QTc) interval will be summarized by treatment group. Summaries of ECGs performed in triplicate will be based on the mean of the triplicate measurements at each time point. For visits where the interpretation is provided in triplicate, the worst interpretation will be used for analysis. Descriptive statistics will be presented for observed values and changes from baseline at each visit where parameters were scheduled to be collected.

Twelve-lead ECGs will be classified by the investigator as "normal," "abnormal, not clinically significant," or "abnormal, clinically significant." Three-by-three contingency tables will be presented to summarize the shift from the baseline category to the worst post-baseline value. Summary results will include the count and percentage of subjects within each shift category and treatment group.

4.5.5.3 Physical Examination

Results of the physical examination will be presented in subject data listings by subject, study visit, and body system.

4.5.5.4 Prior and Concomitant Medications

Medications will be coded using the World Health Organization Drug Dictionary Enhanced (WHODDE), version September 1, 2021. Medications entered on the eCRF will be mapped to Anatomic Therapeutic Chemical (ATC) drug class (level 4) and drug name.

Prior and concomitant medications will be summarized separately and the study phase of each medication will be determined programmatically based on medication start and end dates. A prior medication is defined as any medication administered prior to the date of the first dose of IP in Treatment Period 1. A concomitant medication is defined as any medication administered on or after the date of the first dose of IP in Treatment Period 1. A medication may be defined as both prior and concomitant. If it cannot be determined whether a medication was received prior to the start of IP dosing due to partial or missing medication start and/or end dates, it will be considered a prior medication. Likewise, if it cannot be determined whether a medication was received after the start of IP dosing, it will be considered concomitant. Any medications administered during the post treatment follow-up period will be listed only.

For both prior and concomitant medications summaries, the number and percentage of subjects receiving any medication will be summarized by treatment group, as will the number and percentage receiving any medication by ATC drug class and generic drug name. Prior medications will also be summarized over all subjects combined. Subjects reporting use of more than one medication at each level of summarization (any medication received, ATC class, and generic drug name) will be counted only once. ATC class terms will be displayed by descending order of incidence, as will generic drug names within each ATC class. The study phase during which each medication was received (e.g., prior, concomitant, or both) will be presented on the listing of prior and concomitant medications.

4.6 Determination of Sample Size

A total of 30 evaluable subjects, with 15 subjects in each treatment sequence, are planned for enrollment in this trial. The sample size is chosen to guarantee approximately 80% power for demonstrating a statistically significant treatment effect at a 2-sided alpha of 0.05 if the true mean difference (NB-001 versus placebo) for the CGI-I score at the end of each treatment period is 1 and the standard deviation is 1.25. This sample size calculation is based on the paired t-test.

4.7 Changes in the Conduct of the Study or Planned Analyses

The protocol defined the ENR as all subjects who meet all eligibility criteria and receive at least 1 complete daily dose (400 mg) of IP. This definition was restricted in the SAP to remove the “at least 1 complete daily dose” requirement. All subjects who meet eligibility criteria and consent to study participation will be include in this analysis set.

Section 15.4 of the protocol states that response rates will be compared between treatment groups using Fisher’s exact test. This test is not appropriate for this cross-over study design, and the McNemar’s test will be used instead.

5. REFERENCE LIST

American Psychiatric Association. Diagnostic and statistical manual of mental disorders (5th ed.). Washington, DC: Publisher. 2013.

Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. Psychiatry (Edgmont). 2007; 4(7):28-37.

Constantino JN, Davis SA, Todd RD, et al. Validation of a brief quantitative measure of autistic traits: Comparison of the social responsiveness scale with the autism diagnostic interview revised. J. Autism Dev. Disord. 2003; 33(4):427-433.

Constantino JN, Gruber CP. Social responsiveness scale: Manual. Los Angeles, CA: Western Psychological Services; 2005.

Dmitrienko A, Tamhane AC. Analysis of multiple endpoints in clinical trials. In Multiple Testing Problems in Pharmaceutical Statistics, Dmitrienko A, Tamhane AC, Bretz F (editors). Chapman and Hall/CRC Press: New York, 2009; 131-163.

DuPaul GJ, Power TJ, Anastopoulos, AD, et al. ADHD Rating Scale-5 for children and adolescents: Checklists, norms, and clinical interpretation. Guilford Press. 2016.

FDA Guidance for Industry Suicidal Ideation and Behavior: Prospective Assessment of Occurrence in Clinical Trials. 2012.

Guy W. The Clinical Global Impression Scale. In: ECDEU Assessment Manual for Psychopharmacology-Revised. Rockville, MD: US Dept. of Health, Education & Welfare, ADAMHA, NIMH Psychopharmacology Research Branch. 1976, pp 218-222.

Logan BR, Tamhane AC. On O'Brien's OLS and GLS tests for multiple endpoints. Recent Developments in Multiple Comparison Procedures. IMS Lecture Notes and Monograph Series. Benjamini Y, Bretz F, Sarkar S. (editors). 2004, 76-88.

O'Brien PC. Procedures for comparing samples with multiple endpoints. Biometrics. 1984; 40:1079-1087.

Research Units on Pediatric Psychopharmacology Anxiety Study Group. The Pediatric Anxiety Rating Scale (PARS): Development and psychometric properties. Journal of the American Academy of Child & Adolescent Psychiatry. 2002; 41(9):1061–1069.