NCT #: NCT05165394

STATISTICAL ANALYSIS PLAN
PHASE 2

Version:	1.0
Date:	16 August 2023
Based on Protocol:	Amendment No. 3 Final Version: 21 March 2022 Amendment No. 2 Final Version: 18 November 2022 Amendment No. 1 Final Version: 1 July 2022 Original Final Version: 9 August 2021
Study Treatment:	NBI-1065846
Study Number:	NBI-1065846-MDD2020
Study Title:	A Randomized, Double-Blind, Placebo-Controlled, Proof-of-Concept Study to Evaluate the Efficacy and Safety of Once-Weekly Oral NBI-1065846 in the Treatment of Anhedonia in Major Depressive Disorder (TERPSIS STUDY)
Study Sponsor:	Neurocrine Biosciences, Inc. 12780 El Camino Real San Diego, CA 92130 Telephone: +1-858-617-7600

This study is being conducted in compliance with good clinical practice, including the archiving of essential documents.

CONFIDENTIAL

SIGNATURE PAGE

This document has been prepared by:

[redacted]	eSigned and Dated	
[redacted], Biostatistics Neurocrine Biosciences, Inc.	Signature	Date

This document has been reviewed by:

[redacted]	eSigned and Dated	
[redacted], Biostatistics Neurocrine Biosciences, Inc.	Signature	Date

[redacted]	eSigned and Dated	
[redacted] Neurocrine Biosciences, Inc.	Signature	Date

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Term
AE	Adverse event
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
ANCOVA	Analysis of Covariance
AST	Aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical
BMI	Body mass index
BUN	Blood urea nitrogen
[redacted]	[redacted]
CGI-S	Clinical Global Impression-Severity
COVID-19	Coronavirus Disease 2019
C-SSRS	Columbia-Suicide Severity Rating Scale
CV	Coefficient of variation
DARS	Dimensional Anhedonia Rating Scale
DSM-V	The Diagnostic and Statistical Manual of Mental Disorders, 5th Edition
ECG	Electrocardiogram
eCRF	Electronic case report form
ET	Early termination
FAS	Full analysis set
GGT	Gamma-glutamyl transferase
HAM-D17	Hamilton Depression Rating Scale, 17-item
ICH	International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use
IPD	Important protocol deviation
LS	Least-squares
MADRS	Montgomery Åsberg Depression Rating Scale
MCMC	Markov Chain Monte Carlo
MDD	Major Depressive Disorder

Abbreviation	Term
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple imputation
MNAR	Missing not at random
NBI	Neurocrine Biosciences, Inc.
PCS	Potentially clinically significant
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
PK	Pharmacokinetics
[redacted]	[redacted]
PT	Preferred term
QTcF	Corrected QT interval using Fredericia's formula
SAE	Serious adverse event
SAP	Statistical analysis plan
SD	Standard deviation
SE	Standard error
SOC	System Organ Class
SSRI	Selective Serotonin Reuptake Inhibitor
TEAE	Treatment-emergent adverse event
[redacted]	[redacted]
ULN	Upper limit of normal
US	United States
WHO	World Health Organization

1. INTRODUCTION

The purpose of this statistical analysis plan (SAP) is to describe the planned analyses and data displays that will be prepared to summarize the data from the Phase 2 Study NBI-1065846-MDD2020.

This SAP was developed in accordance with International Council for Harmonization (ICH) E9 guidance. All decisions regarding the final analysis will be made prior to database lock and unblinding of the study data and will be documented in this SAP. Changes to the planned analyses described in this SAP will be statistically justified and described in the study report. Further information related to study design and methodology can be found in the study protocol.

2. STUDY OBJECTIVES

Primary Study Objective:

The primary objective for this study is:

- To evaluate the efficacy of NBI-1065846 compared with placebo on improving symptoms of anhedonia in subjects with MDD.

Secondary Study Objectives:

The secondary objectives for this study are the following:

- To evaluate the efficacy of NBI-1065846 compared with placebo on symptoms of depression in subjects with MDD.
- To evaluate the safety, tolerability, and pharmacokinetics (PK) of NBI-1065846 in subjects with MDD.

3. STUDY DESIGN

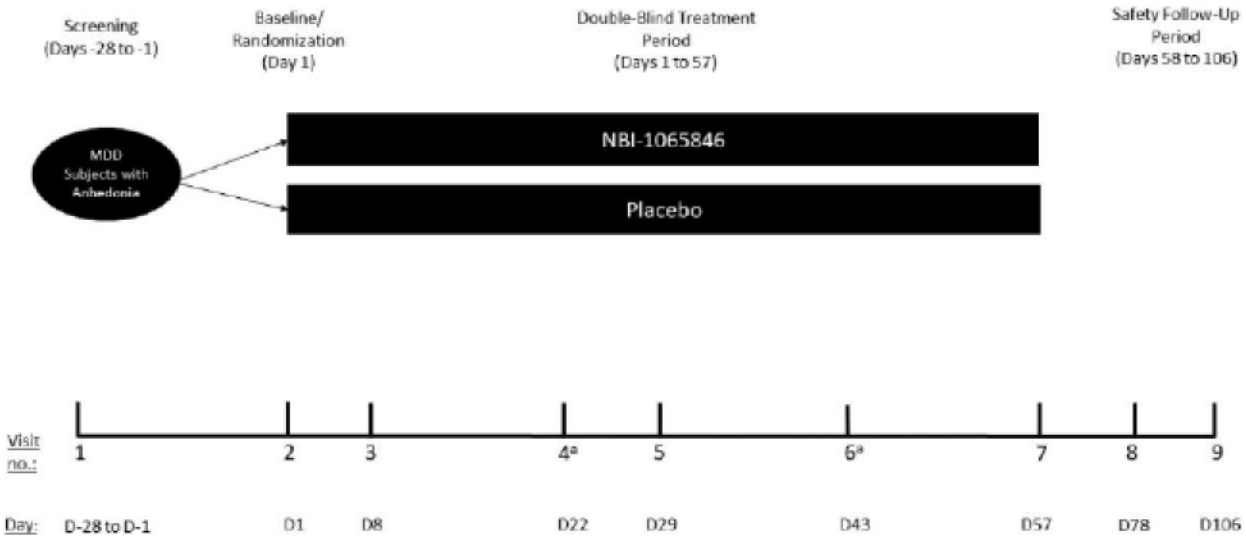
This is a Phase 2, randomized, double-blind, parallel-group, proof-of-concept study comparing use of NBI-1065846 with placebo for improvement of anhedonia in subjects 18 to 65 years of age (inclusive) with MDD who have been treated or are currently being treated with antidepressant medication(s) but continue to have clinically significant anhedonia. Study treatment will be administered as monotherapy or as an adjunctive treatment to the subject's current oral antidepressant medication(s); subjects receiving antidepressant medication(s) at screening are expected to continue using them at the same dose throughout the study.

The study consists of a Screening Period of up to 28 days (Days -28 to -1), an 8-week Treatment Period (Days 1 to 57), and an 8-week Safety Follow-up Period (with visits at 4 and 8 weeks [Days 78 and 106] after last dose of study treatment).

Approximately 88 subjects are planned to be enrolled in the study.

The study design schematic is shown in Figure 1.

Figure 1: Study Design Schematic



D= Day; MDD=major depressive disorder; no.=number

[a] These are virtual visits.

3.1. Randomization

A planned total of approximately 88 subjects will be randomized in a 1:1 ratio in a double-blinded fashion to receive either NBI-1065846 or placebo, administered on a weekly basis, for a period of 8 weeks. The randomization will be stratified based on the Hamilton Depression Rating Scale-17 Item (HAM-D17) score at baseline: remission (HAM-D17 ≤7), mild illness (HAM-D17 from 8 to 18), and moderate to severe illness (HAM-D17 ≥19).

3.2. Blinding

This is a double-blind, placebo-controlled study during which the subjects, investigator, all study center personnel, pharmacists, and the Sponsor, with the exception of the clinical study material supply chain personnel who are not involved in decisions regarding subject's treatment, will be blinded to the subject's treatment assignment.

3.3. Sample Size Considerations

Assuming that 85% of the 88 subjects (randomized in a 1:1 ratio to NBI-1065846 and placebo groups) complete the Day 57 assessment, the study has approximately 80% power to detect an effect size of 0.5 for the primary endpoint at 1-sided significance level of 0.1.

4. ENDPOINTS

The efficacy, safety, pharmacokinetics (PK) and other endpoints are briefly described below. Detailed definitions and planned methods of analysis are provided in the following sections on efficacy (Section 7), pharmacokinetics (Section 8), and safety (Section 9).

4.1. Primary Efficacy

- Change in anhedonia severity, as measured by change in Dimensional Anhedonia Rating Scale (DARS) score from baseline to Day 57.

4.2. Secondary Efficacy

- Change in total Montgomery Åsberg Depression Rating Scale (MADRS) score from baseline to Day 57 in subjects with moderate or higher severity depression.
- Change in Clinical Global Impression-Severity (CGI-S) score from baseline to Day 57.

4.3. Other Endpoints

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

4.4. Safety Endpoints

- Occurrence of TEAEs
- Values and changes from baseline for clinical laboratory tests (hematology and clinical chemistry)
- Values and changes from baseline for vital sign parameters (including orthostatic blood pressure and pulse rate)
- Values and changes from baseline for 12-lead electrocardiogram (ECG) parameters
- Scores from the Columbia-Suicide Severity Rating Scale (C-SSRS)

4.5. Pharmacokinetics Endpoint

The PK endpoint is plasma concentrations for NBI-1065846 after a single dose and at steady state.

5. GENERAL CONSIDERATIONS FOR DATA ANALYSIS

5.1. General Statistical Procedures

Descriptive and inferential statistical methods will be used to evaluate and summarize the data from this study. The term "descriptive statistics" refers to the number of subjects (n), mean, median, standard deviation (SD) or standard error (SE), minimum, and maximum for continuous variables. Ordinal categorical data will be summarized using median, minimum and maximum values. Number and percentage of subjects will be summarized for categorical variables. The term "inferential statistics" refers to hypothesis tests which will be performed to assess differences between the NBI-1065846 treatment group and the placebo group for selected efficacy variables. The overall level of significance (Type I error) for this study is 0.1.

Summary statistics will be presented using the following decimal precision (ie, number of digits to the right of the decimal point): the minimum and maximum will have the same number of decimal places as the data; the mean, median, SD and standard error (SE) will have one more decimal place than the data being summarized; the sample size (N) will be reported as an integer; percentages will be reported to one decimal place (percentages for zero counts are omitted and 100% will be reported as an integer); and p-values will be displayed using four decimal places. P-values less than 0.0001 will be displayed as < 0.0001 and p-values greater than 0.9999 displayed as > 0.9999. Confidence intervals (CI) for means will be reported to the same number of decimal places as mean values; and confidence intervals for percentages will be reported to one decimal place. These rules may be modified if warranted, based on practical considerations.

All available study data will be included in relevant data displays, including data for subjects with incomplete or missing values. Replacement of missing data values with imputed values will generally not be performed unless specified otherwise in relevant endpoint subsections.

5.2. Analysis Sets

The following analysis sets will be defined for this study:

- The Safety Analysis Set will include all randomized subjects who received at least one dose of study treatment. Subjects will be analyzed according to their randomized treatment group, unless they received the incorrect study treatment for the entire treatment duration.
- The Full Analysis Set (FAS) will include all randomized subjects. Subjects will be analyzed according to their randomized treatment group, regardless of compliance with study treatment administration and availability of postbaseline data.
- The PK analysis set will include all randomized subjects who received at least 1 dose of study treatment and who have any measurable NBI-1065846 plasma concentration data.

Summaries of subject disposition, randomization by study site, analysis set inclusion/exclusion status, and important protocol deviations (IPDs) will be based on all randomized subjects.

5.3. Baseline Definition

For efficacy and safety analysis purpose, unless otherwise specified, the assessments collected at the randomization visit (ie, Day1 [Visit 2]) prior to dosing will serve as the baseline value. If the randomization visit value is not available, then the last measurement collected prior to randomization will serve as baseline.

5.4. Derived and Transformed Data

Change from baseline will be calculated as the post-baseline value minus the baseline value; a negative value will represent a decrease at the post-baseline visit. Percent change from baseline will be calculated as (change from baseline/baseline value × 100). If either the baseline or postbaseline value is missing, the change from baseline and/or percent change from baseline will also be missing. If the baseline value is equal to zero, the percent change from baseline will be missing.

5.5. Study Day

Study day will be calculated relative to the date of the first study treatment (Study Day 1) for dates on or after Day 1:

Study Day = Date of Interest – Date of Study Day 1 + 1

Study day will be calculated relative to Study Day 1 for dates before Day 1:

Study Day = Date of Interest – Date of Study Day 1

5.6. Visit Windows

Study day is calculated relative to the date of the Day 1 visit. For efficacy data analysis, the nominal visit number for each visit, including scheduled and early termination/end of study visits will be re-mapped based on the actual study day according to Table 1.

Table 1: Analysis Visit Windows for Efficacy Analysis

Scheduled Visit	Target Study Day	Analysis Window (Analysis Day Range)
Baseline	1	1
Day 29	29	2-43
Day 57	57	44+

For safety data analysis, the nominal visit number for each visit, including scheduled, unscheduled, repeated and early termination/end of study visits, will be re-mapped to an analysis visit according to Table 2: Analysis Visit Windows for Safety Analysis

.

Table 2: Analysis Visit Windows for Safety Analysis

Scheduled Visit	Target Study Day	Analysis Window (Analysis Day Range)
Baseline	1	1 prior to first dosing
Day 1 Post Baseline	1	1 post first dosing
Day 8	8	2-15
Day 22	22	16-25
Day 29	29	26-36
Day 43	43	37-50
Day 57	57	51-last dose day
Day 78	Last dose day+28	Last dose day+1-last dose day+42
Day 106	Last dose day+56	last dose day+43+

If multiple measurements occur within the same visit window after mapping, the measurement that is closest to the target study day will be used for the summary tables where one observation per visit is needed, unless otherwise specified. Where there are ties between the earlier and later observation within the visit window, the earlier observation will be used.

5.7. Handling of Missing Data

5.7.1. Start Dates for Adverse Events

A treatment emergent adverse event (TEAE) is an adverse event not present prior to the initiation of study drug dosing or is an already present event that worsens either in intensity or frequency following the initiation of study drug dosing.

Investigators will be asked to respond "Yes" or "No" on the CRF as to whether the AE started after the subject took the first dose of study treatment. An AE with a response of "Yes" and with an onset date on or before 10 days after the last dose of study treatment will be classified as a TEAE. If the investigator's response is missing, then the treatment emergent status will be derived based on the AE onset date and time relative to the date and time of the subject's first dose of study treatment, as follows: if the AE onset date and time are unknown, it will be assumed that the AE is a TEAE; if the AE onset time is unknown but the AE onset date is the same date as the first dose of study treatment, it will be assumed that the AE is a TEAE. Adverse events that occur more than 10 days after the last dose will not be considered treatment emergent and will be tabulated by treatment group separately.

Missing and incomplete ("partial") dates for AEs will be imputed for the purpose of estimating the time of the event or medication usage in relationship to study treatment. Any data listings will display the original dates as reported in the database.

The missing and incomplete ("partial") dates for AEs will be imputed using the following algorithm for start dates:

- missing day: if the same month and year as the first dose of randomized study treatment, impute the first dose date of randomized study treatment; otherwise impute the 1st of the month.
- missing day and month: if the same year as the first dose of randomized study treatment, impute the first dose date of randomized study treatment; otherwise impute 1st January.

There will be no imputation for AE stop dates.

If any of the above imputations result in a start date that is later than an existing (not imputed) stop date for the event, the start date will be imputed as the stop date.

5.7.2. Start and Stop Dates for Medications

To handle missing/partial dates, the following algorithm will be employed to estimate the time of medication usage relative to study treatment. For start date:

- missing day: impute the 1st of the month unless the same month and year as the study treatment, otherwise impute the first dose date of study drug;
- missing day and month: impute 1st January, unless the same year as the study treatment, otherwise impute the first dose date of study drug.

For stop date:

- missing day: impute the last day of the month unless the same month and year as the study treatment, otherwise impute the last dose date of study drug;
- missing day and month: impute 31st December, unless the same year as the study treatment, otherwise impute the last dose date of study drug.

If any of the above imputations result in a start date that is later than an existing (not imputed) stop date for the event, the start date will be imputed as the stop date. If any of the above imputations result in a stop date that is earlier than an existing (not imputed) start date for the event, the stop date will be imputed as the start date.

5.8. Coding Dictionaries

Medical history and AEs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). Prior and concomitant medications will be coded using the World Health Organization (WHO) Drug Dictionary.

5.9. Impact of COVID-19

Due to the global Coronavirus Disease 2019 (COVID-19) pandemic and in alignment with guidance put forth by the US Food and Drug Administration (FDA; Conduct of Clinical Trials of

Medical Products During the COVID-19 Public Health Emergency [March 2020, updated January 2021]), Health Canada (Management of Clinical Trials During the COVID-19 Pandemic: Notice to Clinical Trial Sponsors [May 2021]), and European Medicines Agency(EMA; Points to Consider on Implications of Coronavirus Disease [COVID-19] on Methodological Aspects of Ongoing Clinical Trials [March 2020]), the following summaries will be conducted to evaluate the impact of the COVID-19 pandemic, including but not limited to:

- A listing of all subjects affected by the COVID-19 pandemic over the course of study participation will be generated. The listing will identify subjects that experience one or more of the following situations due to the COVID-19 pandemic (additional situations may be included):
  - Discontinued study treatment or withdrew from study
  - Presumed or confirmed diagnosis of COVID-19
  - Had at least one COVID-19 pandemic-related major protocol deviation
  - Missed at least one study visit or assessment
  - Had at least one assessment collected in a non-standard way (eg, remotely)
  - Had at least one study treatment interruption
- A table and/or listing will summarize the COVID-19 pandemic-related reasons for treatment and study discontinuation.
- A table and/or listing will summarize the COVID-19 pandemic-related major protocol deviations.
- A table and/or listing will summarize assessments affected by the COVID-19 pandemic (e.g., missing, partial, collected remotely).

The above listings may be omitted if it is known there was no impact of the COVID-19 pandemic on the study conduct.

Adverse events of diagnosed or presumed COVID-19 infections will be included in the standard summaries of TEAEs and SAEs.

The impact of the COVID-19 pandemic on the primary efficacy endpoint will be evaluated as described in Section 7.4.2.

6. STUDY POPULATION

6.1. Disposition

The summary of subject enrollment and disposition will include:

- The total number of subjects who provided informed consent and were screened
- The following categories will be presented by treatment group and overall. The number of subjects randomized will serve as the denominator to calculate percentages.
  - Randomized subjects
  - Randomized but not treated subjects
  - Completed study treatment
  - Discontinued study treatment prematurely, including reasons for discontinuation.
  - Completed study
  - Discontinued study, including reasons for discontinuation.

A separate summary of randomization by study site will be presented. This summary will display the number of subjects randomized to each treatment group by site.

6.2. Summary of Analysis Sets

A summary of the number and percentage of subjects included in each of the following analysis sets will be provided for each treatment group:

- Full Analysis Set
- Safety Analysis Set
- PK Analysis Set

The number and percentage of subjects excluded from each analysis set by reason for exclusion will also be provided.

6.3. Protocol Deviations

Protocol deviations described in the study-specific Protocol Deviation Plan will be entered into the clinical trial management system. Prior to database lock, all major protocol deviations that have been entered into the clinical trial management system will be exported to a file and integrated into the study data.

Important protocol deviations (IPDs) are protocol deviations that might significantly affect the completeness, accuracy and/or reliability of the study data or that might significantly affect a subject's rights, safety, or well-being. An assessment of IPDs will be performed prior to database lock and unblinding of the randomized treatment assignments. The study team will review a listing of all protocol deviations reported in the study database and determine which deviations are IPDs.

A summary of the number and percentage of subjects with IPDs by deviation category and treatment group will be provided.

All major protocol deviations will be presented in a data listing and any that are classified as IPDs will be flagged in the listing.

The COVID-19 pandemic-related major protocol deviations will be summarized and listed.

6.4. Demographic and Baseline Characteristics

Demographic and baseline characteristics data will be summarized using descriptive statistics for continuous variables, and frequency counts and percentages for categorical variables. Results will be presented by treatment group and overall based on the FAS.

Demographics include:

- Age (years)
- Age groups (<35 years, ≥35 to <50 years, and ≥50 years)
- Sex
- Ethnicity
- Race

Baseline subject characteristics include:

- Height (measured at screening, cm)
- Weight (presented in kilograms)
- Body mass index (BMI; kg/m^2)
- HAM-D17 score continuous and categorical (remission [HAM-D17 ≤7], mild illness (HAM-D17 from 8 to 18), and moderate to severe illness [HAM D17 ≥19])
- SHAPS score
- DARS score
- MADRS score
- CGI-S score
- Use of concomitant antidepressant medication(s) (yes, no)
- Use of selective serotonin reuptake inhibitor (SSRI) (yes, no)
- Use of bupropion (yes, no)

6.5. Medical History

Medical history data will be summarized descriptively for the FAS, by treatment group and overall. Medical history data will be coded using MedDRA. The medical history data will be summarized with frequencies and percentages of subjects with at least one medical history item,

and subject frequencies and percentages according to the System Organ Class (SOC) and Preferred Term (PT) levels. The table will be sorted alphabetically by SOC and then, within a SOC, by PT.

7. EFFICACY

The efficacy endpoints and planned analysis methods are described below. Unless otherwise specified, the Full Analysis Set will be used for the primary analyses of all efficacy endpoints. Descriptive statistics will be presented by treatment group according to the study visit. Continuous secondary efficacy endpoints will be analyzed using a similar ANCOVA model as described for the primary endpoint; categorical and ordinal secondary endpoints will be analyzed using the Cochran-Mantel-Haenszel (CMH) Chi-square test adjusted for illness severity (stratification factor) and use of concomitant antidepressant medication(s) at time of randomization.

7.1. Multiple Comparisons and Multiplicity Adjustment

P-values will not be adjusted for multiplicity and should be considered nominal p-values.

7.2. Interim Analysis

There is no planned interim analysis in this study.

7.3. Statistical Hypothesis and Estimand

7.3.1. Statistical Hypothesis

The null hypothesis for the primary endpoint is that the mean change in DARS from baseline to Day 57 in the placebo group is greater than or equal to that in the NBI-1065846 group. The alternative hypothesis is that the mean change in DARS from baseline to Day 57 in the placebo group is less than that in the NBI-1065846 group.

The null hypothesis to be tested is:

$$H_0: \mu_1 \geq \mu_2$$

The alternative hypothesis is:

$$H_1: \mu_1 < \mu_2$$

where μ_1 is the mean change in DARS from baseline to Day 57 in the placebo group, and μ_2 is the mean change in DARS from baseline to Day 57 in the NBI-1065846 group.

7.3.2. Estimand

Consistent with the International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH) E9 Addendum, the 4 attributes of the estimand in conjunction with the primary endpoint are:

- Population:
  - Subjects with a diagnosis of MDD who have been treated with antidepressant medication(s) and who continue to have anhedonia as defined by the study inclusion/exclusion criteria (Section 5 of the protocol).

- Endpoints:
  - Change from baseline in the DARS score at Day 57.
- Intercurrent events:
  - Subjects who discontinue study treatment prematurely or make changes to their background medications will continue to be followed per protocol, therefore the treatment policy strategy will be used to address intercurrent events.
- Summary measures:
  - Difference of means in change from baseline in the DARS score at Day 57 between NBI-1065846 and placebo.

All available assessments up to Day 57 will be used, regardless of occurrence of intercurrent events.

7.4. Primary Efficacy Endpoint

The DARS is a dynamic scale that measures desire, motivation, effort and consummatory pleasure across hedonic domains. The DARS is a self-report questionnaire that measures anhedonia across 4 domains: hobbies/pastimes, food/drinks, social activities, and sensory experiences. The DARS is rated on a five-point Likert scale from 0 (not at all) to 4 (very much), higher values indicating less anhedonia. All items are summed up to a total score in the range of 0 to 68.

Descriptive statistics will be presented by treatment group for the DARS total score observed values, and changes from baseline at each visit. The primary endpoint will be analyzed using the Full Analysis Set by comparing NBI-1065846 and placebo using an Analysis of Covariance (ANCOVA) model. The model will include treatment group, illness severity (stratification factor), use of concomitant antidepressant medication(s) at time of randomization ([ie, yes vs no], covariate), and baseline DARS score (covariate).

Subjects who are missing DARS score at Day 57 will have their data imputed through a multiple imputation (MI) procedure. Missing data will be imputed for subjects in the NBI 1065846 treatment group using retrieved data (i.e., observed data from subjects who discontinued study drug) in that same treatment group, provided there is sufficient retrieved data. Otherwise, missing data will be imputed using observed data from subjects in the placebo treatment group. For subjects in the placebo treatment group, missing data will be imputed using observed data from subjects in the placebo treatment group. The details of imputation method are described below:

Step 1: For subjects with intermittent missing values, a monotone missingness pattern will be generated by imputing for the intermittent missing values using the Markov Chain Monte Carlo (MCMC) method, which assumes a multivariate normal distribution among all variables included in the imputation model. The MI procedure using SAS® software will be employed for this purpose and this first MI step is planned to be repeated 100 times, creating several different datasets with a monotone missing data structure. The imputation is based on the missing at random (MAR) assumption, i.e., the missing data are assumed to follow the

same model as the other subjects in their respective treatment arm.

The following SAS® software code will be used to generate the monotone missing data pattern:

[redacted]

Step 2: The remaining missing data will be imputed using a method for monotone missingness based on missing not at random assumption (MNAR). The trajectories of the subjects in the NBI-1065846 treatment group are assumed to follow the retrieved subjects from the same group after the discontinuation, provided there is sufficient retrieved data (the ratio of retrieved data points to the missing data points is greater than 9 to 1[Lee, Huber, 2021]). Otherwise, the trajectories of the subjects are assumed to follow the placebo group after the discontinuation (i.e., control group-based assumption). Thus, for each created dataset with a monotone missing data pattern, the MI procedure using SAS® software will be employed to impute missing DARS total score values based on a sequential procedure reflecting the monotone missing data pattern. Subjects with the first missing value occurring at Day 29 will have their missing Day 29 value replaced by an imputed value from a regression model with treatment group, illness severity, use of concomitant antidepressant medication(s) at time of randomization, and baseline DARS total score as explanatory variables. In the next step, subjects with their Day 57 value missing will have their missing Day 57 value replaced by an imputed value from a regression model with treatment group, the baseline value, and the Day 29 value as explanatory variables.

An example of SAS® software code is provided below to perform the imputation with the MNAR assumption:

[redacted]

[redacted]

Step 3: The imputed datasets generated with the approach described above contain only non-missing values and are used as input in the model for the primary analysis after change from baseline values have been calculated at Days 29 and 57. Analysis of covariance model described for the primary analysis will thus be run on each of the 100 generated imputed datasets and the difference between the treatment groups at Day 57 will be estimated. Finally, the MIANALYZE procedure using SAS® software will be applied to combine the results from these analyses to derive an overall estimate of the treatment difference at Day 57. In addition to the estimates, corresponding one sided 90% CIs and p-values will be calculated.

An example of SAS® software code is provided below:

[redacted]

Comparison of NBI-1065846 vs. placebo at each visit will be performed by constructing linear contrasts (or equivalent programming code) for differences between treatment group least-squares (LS) means. The LS means and differences in LS means will be presented in summary tables, along with the associated 90% confidence intervals. The treatment difference for the primary endpoint (Day 57) for NBI-1065846 vs placebo will be considered statistically significant if the one-sided nominal p-value is <0.1. Nominal p-values for comparing treatment groups at the individual postbaseline visits will also be reported in the summary tables.

An example of the SAS® software code for the ANCOVA model is provided below:

[redacted]

[redacted]

To convert the two-sided p-values into one-sided p-value, if the treatment difference for NBI-1065846 vs placebo is in the positive direction, the one-sided p-value equals half of the two-sided p-value; if the treatment difference for NBI-1065846 vs placebo is in the negative direction, the one-sided p-value equals one minus half of the two- sided p-value.

Subgroup analyses for the primary efficacy endpoint will be performed for factors defined in Section 7.6.7 using an ANCOVA model, which will be similar to the model described above for the primary endpoint analysis, including additional fixed factors for the subgroup variable and the interaction between the treatment group and subgroup variable. Nominal two-sided p-values for comparing treatment groups and the associated 95% confidence intervals will be reported in summary tables and a forest plot, and p-values for the interaction term between treatment groups and each subgroup variable will be presented. This study was not designed to detect treatment differences with high statistical power within subgroups. No formal statistical testing will be conducted for subgroup analyses; p-values are nominal and intended to be interpreted in a descriptive manner.

7.4.1. Sensitivity Analysis of Primary Efficacy Variable

The following sensitivity analysis will be performed for the primary endpoint to explore the robustness of inferences from the main estimator:

- Tipping point analysis (MNAR): The tipping point assumption will be used, i.e., after the first visit with missing data, the trajectories of the subjects in the NBI-1065846 group are assumed to be worse by an amount of delta. After the MI, as defined previously, an amount of delta will be added to each imputed value at Day 57 in the NBI-1065846 group. Successively harsher deltas will be imposed on the imputed values in the NBI-1065846 group, starting with a DARS total score decrease (worsening) of -1. The delta is further decreased in the steps of 1 (ie, -1, -2, -3, …) until the statistical significance is lost, ie, until the p-value becomes ≥0.1. For the control group, the MI using MAR assumption will be used.

7.4.2. Supplementary Analysis of Primary Efficacy Variable

The following supplementary analyses will be performed for the primary efficacy endpoint:

- Completers analysis using all subjects in the Efficacy Analysis Set with non-missing data for the primary endpoint at the Day 57 visit
- COVID-19 impact analysis excluding all subjects who discontinued study treatment dosing prematurely due to COVID-19.

7.5. Secondary Efficacy Endpoints

7.5.1. Change in total MADRS score from baseline to Day 57 in subjects with moderate or higher severity depression

The MADRS is a 10-item diagnostic questionnaire that psychiatrists use to measure the severity of depressive episodes in subjects with mood disorders. Higher MADRS scores indicate more severe depression, and each item yields a score of 0 to 6.

The MADRS total score will be calculated as the sum of the 10 individual item scores, which ranges from 0 to 60. If more than two individual items are missing, the MADRS total score will not be calculated and will be left as missing. If less than or equal to two individual item scores are missing, the missing item scores will be imputed by the mean of all other available item scores, or the maximum possible values for the missing responses, whichever is smaller, to calculate the MADRS total score.

Descriptive statistics will be presented by treatment group for observed total MADRS score, and changes from baseline at each visit will be presented by treatment group for subjects with moderate or higher severity depression, which is defined as baseline HAM-D17 score ≥19. The changes from baseline to Day 57 will be analyzed using ANCOVA. The ANCOVA model will be similar to the model described above for the primary endpoint analysis.

7.5.2. Change in CGI-S score from baseline to Day 57

The Clinical Global Impression - Severity (CGI-S) uses a 7-point Likert scale to rate the severity of the subject's illness at the time of assessment, relative to the clinician's past experience with subjects who have the same diagnosis. Considering total clinical experience, a subject is assessed on severity of mental illness at the time of rating as 1=normal, not at all ill; 2=borderline mentally ill; 3=mildly ill; 4=moderately ill; 5=markedly ill; 6=severely ill; and 7= extremely ill.

Number and percentage of subjects by CGI-S categories will be summarized by treatment group and visit. In addition, the shift table for CGI-S categories from baseline to Day 57 will be presented.

7.6. Other Efficacy Endpoints

7.6.1. [redacted]

[redacted]

7.6.2. [redacted]

7.6.3. [redacted]

7.6.4. [redacted]

7.6.5. [redacted]

7.6.6. [redacted]

[redacted]

7.6.7. Examination of subgroups

The following subgroups will be used to examine consistency of effect for the primary efficacy endpoint:

- Demographics (age groups [<35 years, ≥35 to <50 years, and ≥50 years], sex, and race)
- Baseline HAM-D17 score (≤ 7 vs. 8 to 18 vs. ≥19)
- Baseline HAM-D17 score (<22 vs. ≥22)
- Baseline total MADRS score (≤ 23 vs. 24 to 27 vs. ≥28)
- Baseline MADRS Item 8 score (≤3 vs. ≥4)
- Baseline DARS score (<33th vs. 33th-66th vs. >66th centile)
- Baseline SHAPS score (<41 vs. ≥41)
- Use of antidepressants at baseline (Yes vs. No)
- Use of selective serotonin reuptake inhibitor (SSRI) (yes, no)
- Use of bupropion (yes, no)

8. PHARMACOKINETICS

The PK endpoint is plasma concentrations for NBI-1065846 after a single dose and at steady state.

8.1. Concentrations of NBI-1065846

The plasma concentrations of NBI-1065846 and metabolites, if appropriate will be summarized with descriptive statistics, which will include the coefficient of variation (CV [%]), geometric mean, and geometric CV (%) in addition to the descriptive statistics described in Section 5.1, by sampling time point (predose, 0.5-2 hour, and 2-36 hour) and the study visit based on the PK analysis set. Concentrations below the lower limit of quantification will be set equal to zero for all plasma concentration summaries.

9. SAFETY

The safety objective of the study is to characterize the safety profile of NBI-1065846 as measured by TEAEs and SAEs, vital signs, clinical laboratory tests, ECG, medications, and safety scales. All outputs for safety endpoints will be based on the Safety Analysis Set. The analysis of the safety data will be based on descriptive statistics and presented by treatment group and the study visit unless otherwise noted. Safety data will not be subject to any imputation and will be summarized on an observed case basis. No formal hypothesis-testing analysis of safety data will be performed.

9.1. Study Treatment Dosing and Compliance

The duration of exposure to randomized study treatment will be calculated as: last dose date – first dose date +1.

Duration of exposure will be summarized with descriptive statistics by treatment group.

The percentage of study drug compliance will be defined as (number of tablets dispensed – number of tablets returned) / 8 x 100%.

Assessment of compliance will be summarized as number and percentage of subjects receiving <80%, 80%-120%, and >120% of prescribed study treatment.

9.2. Prior and Concomitant Medications

Prior medication is defined as any non-study medication taken before the first dose of study treatment.

Concomitant medication is defined as any non-study medication that is started:

- Before the date of first administration of study treatment and ongoing throughout the study or ends on/after the date of first administration of study treatment.
- On/after the date of fist administration of study treatment and is ongoing or ends during the course of study.

Medications with a start date before the first dose of study treatment and an end date after the first dose of study treatment (or ongoing) will be classified as "both prior and concomitant medication".

The number and percentage of patients taking prior and concomitant medications will be tabulated by WHO Drug Dictionary Anatomical Therapeutic Chemical (ATC) Level 3 category or Level 2 if there is not an applicable Level 3 category, and preferred names using the Safety Analysis Set. Subjects will be counted only once for each ATC or preferred name if they have multiple records of the same ATC or preferred name in the database. The table will be sorted alphabetically by ATC class and then, within an ATC class, by preferred names.

9.3. Adverse Events

Adverse events are recorded in the eCRF. Each AE will be coded to SOC and PT using MedDRA.

A treatment-emergent adverse event (TEAE) is an AE that occurs after the initiation of randomized study treatment dosing. Investigators will be asked to respond "Yes" or "No" on the eCRF as to whether the AE started after the subject took the first dose of study drug. An AE with a response of "Yes" will be classified as a TEAE. If the investigator's response is missing, then the treatment emergent status will be derived based on the AE onset date and time relative to the date and time of the subject's first dose of study drug. If the AE onset date and time are unknown, it will be assumed that the AE is a TEAE. If the AE onset time is unknown but the AE onset date is the same date as the first dose of study drug, it will be assumed that the AE is a TEAE.

The frequency tables will include the number and percentage of unique subjects experiencing each event at least once during the study. Unless otherwise specified, summary tables will include events with a start date on or after the date of the first dose of randomized study treatment and up to the last dose of study treatment + 10 days.

Two versions of the primary TEAE frequency tables will be presented by treatment group:

- Frequency of TEAEs by SOC and PT, with SOCs sorted alphabetically and PTs within each SOC sorted by decreasing frequency (number of unique subjects) in the active treatment group.
- Frequency of TEAEs by PT, with PT sorted by decreasing frequency (number of unique subjects) in the active treatment group.

Summary tables of severe TEAEs will be presented by treatment group. The number and percentage of subjects with a severe TEAE will be presented by PT within SOC (presented in the same method as the primary TEAE table). The first line of the table will display the number and percentage of subjects with at least one severe TEAE.

An AE overview summary table will be provided which summarizes the number and percentage of subjects with any TEAE, any TEAE leading to study treatment discontinuation, any serious TEAE, and any TEAE leading to death. The summary table will also include the maximum TEAE severity (mild, moderate, severe) reported for each subject.

In addition, tables will be provided for non-treatment emergent AEs (AE start dates after last dose date +10 days) separately. These tables will be summarized by treatment group and presented by PT within SOC, with SOCs sorted alphabetically and PTs within each SOC sorted by decreasing frequency (number of unique subjects) in the active group.

9.3.1. Adverse Events Leading to Discontinuation of Study Treatment

Summary tables of TEAEs leading to discontinuation of study treatment will be presented by treatment group. The number and percentage of subjects with a TEAE leading to study treatment discontinuation will be presented by PT within SOC (presented in the same method as the primary TEAE table). More than one AE can contribute to study treatment discontinuation per

subject. The first line of the table will display the number and percentage of subjects with at least one TEAE leading to study treatment discontinuation.

A listing of TEAEs leading to study treatment discontinuation will be provided which includes subject ID, treatment group, last treatment received prior to the onset time of the TEAE(s) leading to discontinuation, study day of the discontinuation, and other relevant information from the AE eCRF. Note that "last treatment received prior to the onset time of the TEAE[s] leading to discontinuation" reflects the actual dose level received prior to the AE.

9.3.2. Deaths and Other Serious Adverse Events

Summary tables of Treatment-emergent SAEs and deaths will be presented by treatment group. The tables will include the frequency of SAEs presented by PT within SOC (presented in the same method as the primary TEAE table).

Separate listings of SAEs and fatal TEAEs will also be provided. Each listing will include subject ID, treatment group, last treatment received prior to the onset time of the SAE or fatal TEAE, study day of the SAE or fatal TEAE, and any additional relevant information from the AE eCRF.

9.4. Laboratory Safety Parameters

The hematology and clinical chemistry will be summarized with descriptive statistics by treatment at baseline and at each scheduled postbaseline visit through follow-up visit (Day 106). Both observed values and changes from baseline will be summarized by treatment group.

Shift tables from baseline to Day 57 will be presented for selected clinical laboratory variables based on the reference range-based categories of "Low," "Normal," or "High." A clinical laboratory variable value will be assigned to one of these three categories according to the reference ranges provided by the central clinical laboratory. The shift table will have rows reflecting the reference range category at baseline, and columns reflecting the reference range category on Day 57. A "Total" row and "Total" column will also be included. The number of subjects in each shift category will be displayed in the table.

The shift table will be presented by treatment group for the following clinical laboratory variables:

- aspartate aminotransferase (AST),
- alanine aminotransferase (ALT),
- alkaline phosphatase (ALP),
- gamma-glutamyl transferase (GGT),
- total bilirubin,
- creatine kinase,
- creatinine,
- blood urea nitrogen (BUN),
- white blood cell count,

- absolute neutrophil count,
- hemoglobin,
- platelet count,

Repeat clinical laboratory samples may be collected at any time during this study due to either missing or abnormal results. The general rule for summarizing these data is to include the original sample results in summary tables. Exceptions to this rule is that if there are missing results from the original lab samples at screening, the results of a repeat screening sample will be substituted for the missing results in summary tables.

9.5. Vital Signs

The vital signs data, including orthostatic systolic and diastolic blood pressure (calculated as standing value minus supine value), orthostatic heart rate, respiratory rate (recorded only supine), and body temperature, will be summarized with descriptive statistics by treatment group at baseline and at each scheduled postbaseline visit through follow-up visit. Both observed values and changes from baseline will be summarized.

9.6. Body Weight

The body weight data (in units of kilograms) will be summarized with descriptive statistics by treatment group at baseline and at each scheduled postbaseline visit through follow-up visit. Observed values and absolute and percent changes from baseline will be summarized.

9.7. Electrocardiogram (ECG)

The triplicate values of the quantitative ECG variables (heart rate, PR interval, QRS duration, QT interval, and corrected QT interval using Fridericia's formula [QTcF]) measured at each visit will be averaged for the purpose of analysis. For the categorical ECG interpretation variable (the investigator's assessment of the ECG as "Normal", "Abnormal, not Clinically Significant", or "Abnormal, Clinically Significant"), which is also reported in triplicate, the value that represents the greatest degree of abnormality will be used in all summary tables. If less than 3 values are recorded at an assessment, then the average/greatest abnormality of the available value(s) will be used.

The quantitative ECG variables will be summarized with descriptive statistics by treatment group at baseline and at each scheduled postbaseline visit through follow-up visit. Both observed values and changes from baseline will be summarized. Frequency counts and percentages for the ECG interpretation variable categories will be summarized at each scheduled visit.

Categorical summaries will be presented for the QTcF interval data. For these summaries, a subject's highest reported postbaseline value (including values reported at unscheduled visits) will be used to determine in which category(s) the subject will be counted. The averaged triplicate values will be used when determining each subject's highest reported values.

Two categorical summaries will be presented for the QTcF intervals. For the first summary, the number and percentage of subjects in each treatment group whose highest reported QTcF postbaseline value meets the following thresholds will be summarized:

- Greater than 450 msec
- Greater than 480 msec
- Greater than 500 msec

The second categorical summary will display the number and percentage of subjects in each treatment group whose largest QTcF increase from their baseline value meets the following thresholds:

- Greater than 30 msec
- Greater than 60 msec

9.8. Other Safety Endpoints

9.8.1. Columbia-Suicide Severity Rating Scale (C-SSRS)

The Columbia-Suicide Severity Rating Scale (C-SSRS) is a measure of suicidal ideation and behavior. There are versions of the questionnaire designed for use at screening (Baseline/Screening version) and at baseline and visits throughout the study (Since Last Visit version). Subject responses of "yes" to one or more screening questions will prompt additional questions that evaluate frequency and intensity of suicidal ideation and/or behavior.

The C-SSRS data will be presented in the following summaries:

- Screening (Baseline/Screening version)
- Baseline (Day 1) and postbaseline assessment (Since Last Visit version)

The C-SSRS summary will display the number and percentage of subjects who report "Yes" to specific C-SSRS items or categories of items (a category is assigned a "Yes" value if a "Yes" is reported for any item in the category). These C-SSRS items and categories are as follows:

- Suicidal Ideation Items
  - (1) Wish to be dead
  - (2) Non-specific active suicidal thoughts
  - (3) Active suicidal ideation with any methods (not plan) without intent to act
  - (4) Active suicidal ideation with some intent to act, without specific plan
  - (5) Active suicidal ideation with specific plan and intent
- Suicidal Ideation Category: Any of items (1) through (5)
- Suicidal Behavior Items
  - (6) Preparatory acts or behavior
  - (7) Aborted attempt
  - (8) Interrupted attempt
  - (9) Non-fatal suicide attempt

(10) Completed suicide

- Suicidal Behavior Category: Any of items (6) through (10)
- Suicidal Ideation or Behavior Category: Any of items (1) through (10)

For the "all postbaseline assessments" summary, each subject's C-SSRS responses for all postbaseline assessments during the study will be evaluated, and a "Yes" response for any assessment will be considered as a "Yes" for the subject.

In addition to the summaries described above, shift tables comparing postbaseline suicidal ideation scores to baseline scores will be presented. The shift table scores are defined as the following:

0 = No suicidal ideation

1 = Wish to be dead

2 = Non-specific active suicidal thoughts

3 = Active suicidal ideation with any methods (not plan) without intent to act

4 = Active suicidal ideation with some intent to act, without specific plan

5 = Active suicidal ideation with specific plan and intent

The shift tables will display the number and percentage of subjects within each cell of a 6 x 6 table for each treatment group, with the rows representing the baseline score and the columns representing the maximum score recorded across all postbaseline assessments (including both scheduled and unscheduled visits). Subjects missing either a baseline score or all postbaseline scores will appear in the additional "missing" column/row in the table.

10. PERFORMANCE QUALIFICATION OF SAS® PROGRAMS

The analysis and summary of data from this study will be performed using SAS® 9.4 (or a later release if available). All SAS® programs used in the production of statistical analyses, tables, listings, and figures described in this SAP will undergo performance qualification (verification that the program produces the intended output) in accordance with department standard operating procedures. The performance qualification may include independent programming and/or peer review of the SAS® log files. In addition, tables, figures, listings, and statistical analysis output will be independently reviewed for completeness and accuracy.

11. REFERENCES

European Medicines Agency. Points to Consider on Implications of Coronavirus Disease (COVID-19) on Methodological Aspects of Ongoing Clinical Trials. March 2020.

Health Canada. Management of Clinical Trials During the COVID-19 Pandemic: Notice to Clinical Trial Sponsors. May 2021.

Lee JH, Huber JC Jr. Evaluation of Multiple Imputation with Large Proportions of Missing Data: How Much Is Too Much? Iran J Public Health. 2021 Jul;50(7):1372-1380.

United States Food and Drug Administration. Conduct of Clinical Trials of Medical Products During the COVID-19 Public Health Emergency. Guidance for Industry, Investigators, and Institutional Review Boards. March 2020, updated January 2021.

Signature Page for VV-TMF-54240 v1.0

Reason for signing: Approved	Name: [redacted] Role: Biostatistics Date of signature: 16-Aug-2023 22:51:37 GMT+0000

Reason for signing: Approved	Name: [redacted] Role: Clinical Development Date of signature: 23-Aug-2023 18:08:42 GMT+0000

Reason for signing: Approved	Name: [redacted] Role: Biostatistics Date of signature: 23-Aug-2023 18:42:08 GMT+0000

Signature Page for VV-TMF-54240 v1.0