STATISTICAL ANALYSIS PLAN
PHASE 4

VERSION:
Amendment No. 1 Final Version
Original Version Date: 22 April 2019

DATE OF PLAN:
09Oct2019

BASED ON:
Protocol Amendment No. 3 Final Version: 16 April 2019
Protocol Amendment No. 2 Final Version: 03 October 2018
Protocol Amendment No. 1 Final Version: 06 September 2018
Original Version Date: 17 May 2018

STUDY DRUG:
NBI-98854

PROTOCOL NUMBER:
NBI-98854-TD4001
NCT03698331

STUDY TITLE:
A Phase 4, Randomized, Double-Blind, Placebo-Controlled Study to Evaluate the Potential for Clinical Dependence and Withdrawal Symptoms Associated with Valbenazine

SPONSOR:

Neurocrine Biosciences, Inc.
12780 El Camino Real
San Diego, CA 92130

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 and/or reviewed by:

[redacted]
Neu ne B osciences, Inc — Signature — Date

[redacted]
Neu cr e B en es, Inc — Signature — Date

This document has been reviewed and accepted by:

[redacted]
Neu cr e B c — Signature — Date

[redacted]
Neu cr e B ci nces, Inc — Signature — Date

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. LIST OF ABBREVIATIONS

Table 1: List of Abbreviations

AE	Adverse event
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
BARS	Barnes Akathisia Rating Scale
BMI	Body mass index
CGI-TD-I	Clinical Global Impression-Tardive Dyskinesia-Improvement
CGI-TD-S	Clinical Global Impression-Tardive Dyskinesia-Severity
C-SSRS	Columbia-Suicide Severity Rating Scale
ECG	Electrocardiogram
eCRF	Electronic case report form
EPS	Extrapyramidal symptoms
ESS	Epworth Sleepiness Scale
ET	Early termination
GGT	Gamma-glutamyl transferase
HAM-A	Hamilton Anxiety Rating Scale
ICF	Informed consent form
ICH	International Conference on Harmonisation
IWRS	Interactive Web Response System
MADRS	Montgomery-Asberg Depression Rating Scale
mCSSA	Modified Cocaine Selective Severity Assessment
MedDRA	Medical Dictionary for Regulatory Activities
NBI	Neurocrine Biosciences, Inc.
PK	Pharmacokinetic
PT	Preferred term
PWC-20	Physician Withdrawal Checklist-20
QTcF	Corrected QT interval using Fridericia's formula
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SAS	Simpson-Angus Scale
SD	Standard deviation
SEM	Standard error of the mean
SOC	System organ class
TD	Tardive dyskinesia
TEAE	Treatment-emergent adverse event
ULN	Upper limit of normal

2. 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 4 study described in Neurocrine Biosciences, Inc. (NBI) Protocol NBI-98854-TD4001.

This SAP was developed in accordance with ICH E9 guidance. All decisions regarding the final analysis, as defined in this SAP document, will be made prior to database lock and unblinding of the study data. Further information related to study design and methodology can be found in the protocol.

3. STUDY OBJECTIVES AND ENDPOINTS

3.1. Study Objectives

3.1.1. Primary Objective

The primary objective of this clinical study is:

- To evaluate the potential for clinical dependence and withdrawal symptoms associated with valbenazine following 4 weeks of once-daily treatment with valbenazine or placebo.

3.1.2. Secondary Objective

The secondary objectives of this study are:

- To evaluate the efficacy of valbenazine administered once daily for up to 4 weeks.
- To evaluate the safety and tolerability of valbenazine administered once daily for up to 4 weeks.

4. STUDY DESIGN

4.1. Summary of Study Design

This is a Phase 4, randomized, double-blind, placebo-controlled study to evaluate the potential for clinical dependence and withdrawal symptoms associated with valbenazine. Approximately 80 medically stable male and female subjects (the proportion of males and females will be consistent with that of the patient population) with neuroleptic-induced tardive dyskinesia (TD) will be enrolled.

Subjects will be screened for eligibility for up to 6 weeks prior to Day -1 (baseline visit).

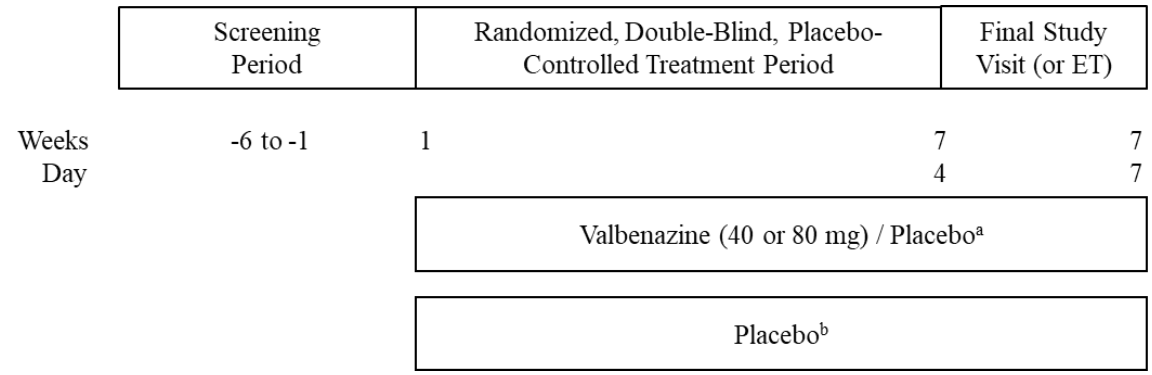
On Day -1, eligible subjects will be randomized (1:1) to 1 of the 2 treatment arms (randomization will be stratified by study site):

- Valbenazine (40 mg for the first week followed by 80 mg for 3 weeks) for the first 4 weeks of the double-blind treatment period followed by placebo for the last 3 weeks of the double-blind treatment period.
- Placebo for the 7 weeks of the double-blind treatment period.

Clinical dependence and withdrawal symptoms, efficacy, pharmacokinetics (PK), safety, and tolerability will be assessed at scheduled times throughout the study.

A schematic of the study design is provided in Figure 1.

Figure 1: Study Design Schematic



ET=early termination.
[a] Subjects will receive valbenazine (40 mg for 1 week increased to 80 mg for 3 weeks) for 4 weeks followed by placebo for 3 weeks.
[b] Subjects will receive placebo for 7 weeks.

4.2. Sample Size Considerations

The sample size for this study is consistent with typical evaluations of dependence and withdrawal.

4.3. Randomization

Subjects will be randomized 1:1 to valbenazine or placebo. Randomization will be stratified by study site.

4.4. Clinical Assessments

Assessments of clinical dependence and withdrawal symptoms include:

- Physician Withdrawal Checklist-20 (PWC-20),
- modified Cocaine Selective Severity Assessment (mCSSA),
- Epworth Sleepiness Scale (ESS), and
- Hamilton Anxiety Rating Scale (HAM-A).

The mCSSA and PWC-20 assessments are collected at baseline (Day -1) and every visit starting after Week 1. The ESS and HAM-A assessments are collected at screening, baseline (Day -1), and every visit starting after Week 1. AEs will also be reviewed for potential withdrawal symptoms.

Efficacy assessments include:

- Clinical Global Impression-Tardive Dyskinesia-Improvement (CGI-TD-I) and
- Clinical Global Impression-Tardive Dyskinesia-Severity (CGI-TD-S).

The CGI-TD-S assessment is collected at baseline (Day -1), Week 4, Week 5 Day 7, Week 6 Day 7, and Week 7 Day 7 or early termination. The CGI-TD-I assessment is collected at Week 4 and Week 7 Day 7 or early termination.

Pharmacokinetic assessments:

Blood samples for plasma drug and metabolite concentration analyses are collected at Week 1 and Week 4. Subjects are asked to record and provide dosing times on the days when these samples are collected.

Safety assessments include:

- AEs,
- Clinical laboratory tests (hematology, clinical chemistry, and urinalysis),
- Vital signs (including orthostatic blood pressure and pulse),
- Physical examinations,
- 12-lead electrocardiogram (ECG),
- Suicidal ideation and behavior, evaluated using the Columbia-Suicide Severity Rating Scale (C-SSRS),
- Drug-induced akathisia, evaluated using the Barnes Akathisia Rating Scale (BARS),
- Drug-induced parkinsonism and extrapyramidal symptoms (EPS), evaluated using the Simpson-Angus Scale (SAS), and
- Changes in the severity of depressive symptoms using the Montgomery-Asberg Depression Rating Scale (MADRS).

See Table 1 of the study protocol for a complete schedule of safety assessments.

5. PLANNED ANALYSES

5.1. Interim Analyses

No interim analysis is planned for this study.

5.2. Final Analyses

Final analyses, as specified in the protocol and in this SAP, will be performed after the study database has been locked and treatment code has been unblinded.

6. GENERAL CONSIDERATIONS FOR DATA ANALYSES AND HANDLING

All analyses described in this plan are considered *a priori* analyses in that they have been defined prior to locking the study database and unblinding the treatment group assignments. Analyses defined subsequent to locking the database and unblinding will be considered *post hoc* analyses and will be applied as exploratory methodology. Any *post hoc* analyses will be clearly identified in the clinical study report.

6.1. General Statistical Procedures

Descriptive statistical methods will be used to summarize the data from this study. The term "descriptive statistics" refers to the number of subjects, mean, median, standard deviation (SD), standard error of the mean (SEM), minimum, and maximum for continuous and ordinal categorical variables; and refers to the number and/or percentage of subjects (or events) for categorical variables.

Unless stated otherwise, "treatment group" refers to the treatment the subject is randomly assigned to receive on Day -1 (ie, valbenazine or placebo).

Select tables will include an "All Subjects" group, which combines both treatment groups and includes all subjects being summarized in the table.

The double-blind (DB) treatment period will be further separated into treatment periods of interest for the analysis of selected variables. The DB treatment period starting after Day -1 through the end of the Week 4 visit will be referred to as the "Valbenazine/Placebo Period." The DB treatment period starting after the Week 4 visit through Week 7 Day 7 (final study visit) will be referred to as the "Withdrawal Period." Subsequent mention of summaries by "treatment period" will be defined as the aforementioned.

Summary statistics will be displayed using the following decimal precision rules: the minimum and maximum will have the same number of decimal places as the data; the mean, median, SD, and SEM will have one more decimal place than the data being summarized. Percentages will be displayed using one decimal place; percentages for 0 counts will be omitted. These rules may be modified if warranted, based on practical considerations.

6.2. Analysis Sets

6.2.1. Definition of Analysis Sets

6.2.1.1. Safety Analysis Set

The safety analysis set will include all subjects who are randomized to a treatment group, take at least one dose of study drug, and have any postbaseline safety data. The safety analysis set will be used for all summaries of safety data (e.g., AEs, clinical laboratory data) and plasma concentration data. For the summaries of data based on the safety analysis set, subjects who are dispensed the incorrect treatment at the time of randomization, and remain on the same incorrect treatment during the study, will be assigned to the treatment actually received in all summary tables and figures. Subjects who are dispensed a combination of correct and incorrect treatments during the study will be assigned to the randomized treatment in all summary tables and figures.

6.2.1.2. Dependence and Withdrawal Analysis Set

The dependence and withdrawal analysis set will include all subjects in the safety analysis set who enter the Withdrawal Period (ie, subject is dispensed a kit at Week 4). The dependence and withdrawal analysis set will be used for summaries and analyses of clinical dependence and withdrawal symptoms data.

6.2.1.3. Efficacy Analysis Set

The efficacy analysis set will include all subjects who are randomized to a treatment group, take at least one dose of study drug, and have a CGI-TD-I or CGI-TD-S assessment at Week 4. The efficacy analysis set will be used for all summaries of efficacy data.

6.2.2. Summary of Analysis Sets

A summary of the number and percentage of subjects included in (and excluded from, as applicable) each analysis set will be provided for each treatment group. The number and percentage of subjects excluded from each analysis set by reason for exclusion will also be provided. An additional "All Subjects" column will be included.

6.2.3. Application of Analysis Sets

Summaries of subject disposition, randomization by study site, analysis set inclusion/exclusion status, and IPDs will include all randomized subjects. All other summaries by analysis set are identified in Table 2.

Table 2: Data Summaries by Analysis Set

Data Summary/Analysis	Analysis Set		
	Safety	Dependence and Withdrawal	Efficacy
Demographics	X	X	
Baseline subject characteristics	X	X	
Medical history	X		
Study drug dose reductions	X		
Study drug compliance	X		
PWC-20		X	
mCSSA		X	
ESS		X	
HAM-A		X	
Plasma concentrations	X		
CGI-TD-S			X
CGI-TD-I			X
Adverse events	X	X	
Clinical laboratory data	X		
Vital signs	X		
Weight	X		
ECG	X		
C-SSRS	X		
BARS	X		
SAS	X		
MADRS	X		
Prior and concomitant medications	X		

6.3. Baseline Definition

The assessments collected at the Day -1 study visit will serve as the baseline value for all assessments. With the exception of C-SSRS, if a Day -1 visit value is not available, then the last measurement collected on or prior to the date of the Day -1 study visit will serve as baseline.

6.4. Derived and Transformed Data

6.4.1. Study Day

Study day is calculated relative to the date of the Day -1 visit. If the date of interest occurs on or after the Day -1 visit, then the study day will be calculated as: date of interest – date of Day -1 visit + 1. If the date of interest occurs prior to the Day -1 visit, then the study day will be calculated as: date of interest – date of Day -1 visit.

6.4.2. Change from Baseline

Change from baseline is calculated as (postbaseline value – baseline value).

If either the baseline or postbaseline value is missing, the change from baseline will also be missing.

6.4.3. Handling of Early Termination Visit Data

An early termination (ET) visit occurs when a subject discontinues from the study prior to completing the scheduled Week 7 Day 7 visit. The data collected at ET visits will be included in summary tables and figures in accordance with the ET visit mapping scheme described in this section.

For the purpose of data summarization, a visit window will be applied to account for ET visits. An ET visit will be mapped to Week 1 if it occurs within 3 days prior to or 3 days after the expected study day of the Week 1 visit. An ET visit which occurs within 3 days prior to or 3 days after the expected study day of the Week 4 visit will be mapped to Week 4. ET visits occurring at a scheduled visit after Week 4 will be mapped to the scheduled visit. If the ET visit occurs after Week 4 and between scheduled visits, the ET visit will be mapped to the next scheduled visit (with the requirement that the scheduled visit prior to the ET visit was completed by the subject).

Early termination visit data which are not mapped to a scheduled visit will not be included in by-visit analyses and summaries. They will be included in any analyses that look across all available assessments during the treatment period, including unscheduled visits. They will also be included in any applicable by-subject data listings.

6.5. Handling of Missing Data

6.5.1. Missing Outcome Measures

6.5.1.1. Clinical Dependence and Withdrawal Assessments

Missing data from clinical dependence and withdrawal assessments (PWC-20, mCSSA, ESS, and HAM-A) during the Withdrawal Period will be handled using the following imputation rules:

- If an individual item or total score from an assessment is intermediate missing (ie, missing between 2 observed visits), the item or total score will be imputed with the higher of the subject's observed value immediately preceding or following the assessment with the missing value. Imputed values for intermediate missing data will be included in by-visit summaries.
- Missing total scores as a result of dropout will be imputed with the highest total score from subjects in the same treatment group with observed total scores at the same visit. This imputation method will only be used to determine the maximum (worst) total score per subject in the Withdrawal Period for dropouts and will not be included in by-visit summaries.

Assessments missed prior to the Withdrawal Period will not be imputed.

6.5.2. Missing Dates

6.5.2.1. Start Dates for Adverse Events and Prior and Concomitant Medications

Missing and incomplete ("partial") dates for AEs and concomitant medications 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 imputation rules for AE start dates are as follows:

- If the date is completely missing, the date will be imputed as the date of the first dose of study drug;
- If only the day is missing, the date will be imputed as the date of the first dose of study drug if the month and year match the month and year of the first dose of study drug; otherwise, the missing day will be imputed as the first day of the month;
- If both the day and month are missing, the date will be imputed as the date of the first dose of study drug if the date is in the same year as the first dose of study drug; otherwise, the missing day and month will be imputed as 01 January;
- If any of the above imputations result in a start date that is later than an existing (not imputed) end date for the event, the start date will be imputed as the end date.

The imputation rules for concomitant medication start dates are as follows:

- If the date is completely missing, the date will be imputed as 01 January in the year of the subject's screening vital signs assessment;
- If only the day is missing, the date will be imputed as the date of the first dose of study drug if the date is in the same month and year as the first dose of study drug; otherwise, the missing day will be imputed as the first day of the month;
- If both the day and month are missing, the date will be imputed as the date of the first dose of study drug if the date is in the same year as the first dose of study drug; otherwise, the missing day and month will be imputed as 01 January;
- If any of the above imputations result in a start date which is later than an existing (not imputed) medication stop date, the start date will be imputed as the stop date.

7. STUDY POPULATION

7.1. Subject Disposition

The summary of subject enrollment and disposition will display the number of subjects who were randomized to each treatment group, who completed through Week 4 of the study period (excluding ET visits mapped to Week 4), who entered the Withdrawal Period (indicated by dispensing of study drug at the Week 4 visit), and who completed the study (ie, up to Week 7 Day 7 excluding ET visits mapped to Week 7 Day 7). The number of subjects who did not complete the study will also be summarized, both overall and according to the reason for early 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.

All of the summaries described in this section will be presented by each treatment group as well as overall.

A listing of randomized subjects will be provided and will include subject ID, informed consent date, randomization date, and randomized treatment group. A listing of lot numbers will also be provided.

7.2. 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 by a committee composed of NBI Clinical Development project team members prior to database lock and unblinding of the randomized treatment assignments. This committee will review a listing of all major protocol deviations reported in the study database and determine which deviations are IPDs. Important protocol deviations include, but are not limited to, the following:

- Failure to obtain informed consent from the subject prior to performing any study procedures.
- Deviations from key inclusion/exclusion criteria.
- Use of prohibited concomitant medications.
- Error in drug dispensing which results in a subject not receiving intended randomized treatment.
- Significant deviation from protocol-specified dosing regimen.

A summary of the number and percentage of subjects with IPDs by deviation category will be presented by treatment group. An additional "All Subjects" column will also be included.

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

7.3. 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. An additional "All Subjects" column will also be included.

Demographics include:

- Age (years)
- Sex
- Ethnicity
- Race

Baseline subject characteristics include:

- Primary clinical diagnosis
- Age at first diagnosis of schizophrenia, schizoaffective disorder, or mood disorder (years)
- Age at TD diagnosis (years)
- Height (measured at screening; cm)
- Weight (presented in both pounds and kilograms)
- Body mass index (BMI; calculated using height collected at the screening visit and weight collected at the Day -1 visit; kg/m^2)
- CYP2D6 genotype

7.4. Medical History and Medical Conditions Present at Entry

Medical history will be summarized in a frequency table (number and percentage of subjects) by Medical Dictionary for Regulatory Activities (MedDRA) System Organ Class (SOC) and Preferred Term (PT) by treatment group, with SOCs and PTs within each SOC sorted alphabetically. An additional "All Subjects" column will also be included.

7.5. Study Drug Dosing, Compliance, and Dose Reductions

7.5.1. Compliance

At the end of Weeks 1, 4, 6, and 7, subjects will return all used and unused study drug and a compliance check will be performed by counting the capsules returned at each study visit. The site will then enter whether the subject's dosing compliance since the previous visit was ≥80% into the eCRF.

The number and percentage of subjects in each treatment group who are dosing compliant (defined as ≥80%) will be presented for each postbaseline visit through Week 7 Day 7.

7.5.2. Dose Reductions

Subjects randomized to the valbenazine arm will receive 40 mg for the first week of treatment and 80 mg until the end of Week 4 (Week 5 Day 1 will be the first dose of placebo for subjects randomized to the valbenazine arm). If a subject is unable to tolerate the 80 mg dose, the daily dosage will be reduced to 40 mg. Subjects unable to tolerate the 40 mg dose (or placebo) will be discontinued from the study.

The number and percentage of subjects with dose reductions will be summarized by treatment group at Week 4. Dose reductions at unscheduled visits prior to Week 4 will be included in the summary.

8. CLINICAL DEPENDENCE AND WITHDRAWAL DATA

8.1. Adverse Events During the Withdrawal Period

A withdrawal-emergent adverse event is an adverse event that begins during the Withdrawal Period.

Summaries of withdrawal-emergent AEs will be presented in frequency tables by treatment group and will include the number and percentage of unique subjects experiencing each event at least once during the Withdrawal Period.

Two versions of the frequency table will be presented:

- Frequency of withdrawal-emergent AEs by SOC and PT, with SOCs and PTs within each SOC sorted by decreasing frequency (in the valbenazine/placebo arm);
- Frequency of withdrawal-emergent AEs by PT, with PT sorted by decreasing frequency (in the valbenazine/placebo arm).

8.2. Physician Withdrawal Checklist-20

The PWC-20 is a validated 20-item physician-rated instrument that assesses potential symptoms of withdrawal on a severity scale in the following areas: gastrointestinal, mood, sleep, motor, somatic, perception, and cognition. Items are rated on a scale from 0 to 3 (0=not present, 1=mild, 2=moderate, and 3=severe), with the total score ranging from 0 to 60.

The PWC-20 total score (observed values and intermediate missing imputed values) will be summarized with descriptive statistics by treatment group and visit (baseline; Week 4; Week 5 Days 1, 3, 5, and 7; Week 6 Days 2, 4, and 7; Week 7 Days 2 and 4; and Week 7 Day 7). The PWC-20 total scores will also be assessed categorically using the following ranges: total scores of 0 to 15, 16 to 30, 31 to 45, and 46 to 60. The number and percentage of subjects who fall into each category will be presented in a table by visit and treatment group.

The mean worst PWC-20 total score per treatment group during the Withdrawal Period will be presented in a table with descriptive statistics by taking the worst total score per subject between Weeks 5 and 7. A subject's worst total score will be based on the maximum of the subject's highest observed total score or, in the case of dropout, the subject's highest imputed total score (as described in Section 6.5.1). The table will also include descriptive statistics for baseline and Week 4 (withdrawal baseline) total scores and the change in the worst total score during the Withdrawal Period from baseline and withdrawal baseline by treatment group.

Additionally, subjects will be assessed to determine if they have worsening symptoms between Weeks 5 and 7 (ie, during the Withdrawal Period) when compared to their Week 4 reported symptoms. A subject's symptoms are considered to be worsening if they meet at least one of the following criteria:

- Subject presents with at least 5 new symptoms at any time during the Withdrawal Period of moderate or severe degree (note: the new symptoms do not have to be reported at the same visit),
- Subject reports a worsening of any of the symptoms (as recorded at the Week 4 visit) by 2 points during at least one visit during the Withdrawal Period.

The number and percentage of subjects with worsening symptoms will be summarized by treatment group in a table.

8.3. Modified Cocaine Selective Severity Assessment

The mCSSA is an 18-item instrument primarily drawn from symptoms commonly reported in the literature as being associated with early cocaine abstinence, including depression, fatigue, anhedonia, anxiety, irritability, sleep disturbance, and inability to concentrate. The instrument includes additional symptoms such as paranoia, carbohydrate craving, bradycardia, and suicidality. The scale has been modified to be specific to study drug (valbenazine or placebo) instead of cocaine; specifically, "cocaine" was replaced with "study drug" in question #4. Items are rated on scales of 0 to 7 or 0 to 8, with separate scale descriptions for each item. The total score ranges from 0 to 127.

The mCSSA total score (observed values and intermediate missing imputed values) will be summarized with descriptive statistics by treatment group and visit (baseline; Week 4; Week 5 Days 1, 3, 5, and 7; Week 6 Days 2, 4, and 7; Week 7 Days 2 and 4; and Week 7 Day 7).

The mean worst mCSSA total score per treatment group during the Withdrawal Period will be presented in a table with descriptive statistics based on the worst total score per subject between Weeks 5 and 7. A subject's worst total score will be based on the maximum of the subject's highest observed total score or, in the case of dropout, the subject's highest imputed total score (as described in Section 6.5.1). The table will also include descriptive statistics for baseline and Week 4 (withdrawal baseline) total scores and the change in the worst total score during the Withdrawal Period from baseline and withdrawal baseline by treatment group.

8.4. Epworth Sleepiness Scale

The ESS is a validated instrument to assess subject's sleepiness. The instrument lists 8 daytime situations in which the subject rates his/her tendency to become sleepy on a scale from 0 (would never doze) to 3 (high chance of dozing). The total score ranges from 0 to 24.

For the ESS total score (observed values and intermediate missing imputed values), descriptive statistics will be presented for baseline and all postbaseline visits by treatment group.

The mean worst ESS total score per treatment group during the Withdrawal Period will be presented in a table with descriptive statistics based on the worst total score per subject between Weeks 5 and 7. A subject's worst total score will be based on the maximum of the subject's highest observed total score or, in the case of dropout, the subject's highest imputed total score (as described in Section 6.5.1). The table will also include descriptive statistics for baseline and Week 4 (withdrawal baseline) total scores and the change in the worst total score during the Withdrawal Period from baseline and withdrawal baseline by treatment group.

8.5. Hamilton Anxiety Rating Scale

The HAM-A is a 14-item scale used to evaluate the severity of anxiety. Each of the items is scored from 0 (not present) to 4 (very severe). The total score ranges from 0 to 56.

For the HAM-A total score (observed values and intermediate missing imputed values), descriptive statistics will be presented for baseline and all postbaseline visits by treatment group.

The mean worst HAM-A total score per treatment group during the Withdrawal Period will be presented in a table with descriptive statistics based on the worst total score per subject between Weeks 5 and 7. A subject's worst total score will be based on the maximum of the subject's highest observed total score or, in the case of dropout, the subject's highest imputed total score (as described in Section 6.5.1). The table will also include descriptive statistics for baseline and Week 4 (withdrawal baseline) total scores and the change in the worst total score during the Withdrawal Period from baseline and withdrawal baseline by treatment group.

9. PLASMA CONCENTRATION DATA

The plasma concentrations of valbenazine (NBI-98854) and its metabolite NBI-98782 will be summarized with descriptive statistics by visit (Week 1 and Week 4) and the valbenazine dose (40 mg or 80 mg) received prior to that visit. Concentrations below the lower limit of quantification will be set equal to zero for all plasma concentration summaries. The lower limits of quantification are as follows: (a) NBI-98854: 1.00 ng/mL and (b) NBI-98782: 0.100 ng/mL.

The following additional descriptive statistics will be included in the plasma concentration summary tables: (a) the number of plasma concentration values greater than or equal to the lower limit of quantification, (b) the geometric mean, and (c) the geometric coefficient of variation (%).

10. EFFICACY

10.1.1. Clinical Global Impression-Tardive Dyskinesia-Improvement

The CGI-TD-I scale is a 7-point scale (range; 1=very much improved to 7=very much worse) used to assess overall improvement in TD symptoms since the initiation of study drug dosing. For the CGI-TD-I, descriptive statistics will be presented for observed values at Week 4 and Week 7 Day 7 visits by treatment group.

The CGI-TD-I categorical responses will be summarized in a table which will display the number and percentage of subjects in each category by treatment group and visit (Week 4 and Week 7 Day 7).

CGI-TD-I responders (number and percentage of subjects) will also be summarized in a frequency table by treatment group and visit. A subject is classified as a responder if their CGI-TD-I score is either a "1" ("very much improved") or a "2" ("much improved").

10.1.2. Clinical Global Impression-Tardive Dyskinesia-Severity

The CGI-TD-S scale is a 7-point scale (range; 1=normal, not at all ill to 7=among the most extremely ill patient) used to assess the overall global severity of TD. For the CGI-TD-S, descriptive statistics will be presented for baseline and postbaseline visits by treatment group. Both observed values and changes from baseline will be summarized. An additional table will be presented with observed values and changes from Week 4 (withdrawal baseline) for assessments collected during the Withdrawal Period. Figures showing the mean (+/-SD) for the observed and change from baseline values vs. study visits will be generated by treatment group.

The CGI-TD-S categorical responses will also be summarized in a table which will display the number and percentage of subjects in each category by treatment group and visit (including baseline).

11. SAFETY

Results will be presented by treatment group using the safety analysis set, as described in Section 6.2.1.1.

11.1. Adverse Events

Adverse events are recorded in the eCRF. Each AE will be coded to a SOC and PT using MedDRA (Version 21.0).

A treatment-emergent adverse event (TEAE) is an AE 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 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.

TEAEs will be summarized in frequency tables by treatment group. The frequency tables will include the number and percentage of unique subjects experiencing each event at least once during the DB treatment period.

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

- Frequency of TEAEs by SOC and PT, with SOCs and PTs within each SOC sorted by decreasing frequency (in the valbenazine/placebo arm);
- Frequency of TEAEs by PT, with PT sorted by decreasing frequency (in the valbenazine/placebo arm).

Overall summary tables will be provided which summarize the number and percentage of unique subjects with any TEAE, any TEAE leading to dose reduction, any TEAE leading to study discontinuation, any serious TEAE, and any TEAE resulting in death in the DB treatment period. The summary tables will also include the frequency distribution of the maximum TEAE intensity (mild, moderate, severe) reported for each subject during the DB treatment period.

11.1.1. Adverse Events Resulting in Premature Discontinuation from Study

Summary tables of TEAEs resulting in early discontinuation from the study will be presented by treatment group. The number and percentage of subjects with a TEAE resulting in study 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 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 discontinuation.

A listing of TEAEs resulting in premature study 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.

11.1.2. Adverse Events Resulting in Study Drug Dose Reductions

Summary tables of TEAEs resulting in study drug dose reductions will be presented by treatment group. The number and percentage of subjects with a TEAE resulting in a dose reduction will be presented by PT within SOC (presented in the same method as the primary TEAE table). More than one AE can contribute to a dose reduction per subject. The first line of the table will display the number and percentage of subjects with at least one TEAE leading to dose reduction.

11.1.3. Deaths and Other Serious Adverse Events

Summary tables of serious adverse events (SAEs) will be presented by treatment group. The number and percentage of subjects with an SAE 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 SAE.

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.

11.2. Clinical Laboratory Data

The hematology and clinical chemistry data will be summarized with descriptive statistics by treatment group at baseline, Week 4, and Week 7 Day 7. Both observed values and changes from baseline will be summarized.

Shift tables 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.

Two shift tables will be presented by treatment period: shifts from baseline to Week 4 (or last available assessment in the Valbenazine/Placebo Period) and shifts from baseline to Week 7 Day 7 (or last available assessment in the Withdrawal Period). Each shift table will have three rows and three columns, with rows reflecting the reference range category at baseline, and columns reflecting the reference range category at the specified postbaseline visit. A "Total" row and "Total" column will also be included. Subjects with a missing baseline value or who do not have postbaseline data will not be included in the tables for that variable. The number and percentage of subjects in each shift category will be displayed in the table; percentages will be based on the number of subjects included in the table.

Shift tables will be presented 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,
- white blood cell count,
- absolute neutrophil count,
- hemoglobin, and
- platelet count.

Summaries of sponsor-defined potentially clinically significant (PCS) values will be presented for the following clinical laboratory variables: ALT, AST, creatine kinase, GGT, total bilirubin, white blood cell count, absolute neutrophil count, creatinine, and BUN. The number and percentage of subjects with PCS values that are reported at any postbaseline visit (scheduled or unscheduled) within each treatment period will be summarized by treatment period and treatment group. The criteria for identifying PCS clinical laboratory values are provided in Table 3.

Table 3: Potentially Clinically Significant Criteria for Clinical Laboratory Variables

Variable	PCS Threshold
ALT	>3 x ULN (upper limit of normal)
AST	>3 x ULN
Creatine kinase	>5 x ULN
GGT	>3 x ULN
Total bilirubin	>1.5 x ULN
White blood cell count	≤2.8 x 1000/µL
Absolute neutrophil count	<1.5 x 1000/µL
Creatinine	>1.5 x baseline value or > 1.5 x ULN
BUN	>30 mg/dL (> 10.71 mmol/L)

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 are: (1) all available lab values will be used in the PCS tables and (2) 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.

11.3. Vital Signs

The vital signs data, including orthostatic blood pressures and heart rate (calculated as standing value minus supine value), will be summarized with descriptive statistics by treatment group at all scheduled visits from baseline through Week 7 Day 7. Observed values, changes from baseline, and changes from Week 4 (to post Week 4 visits) will be summarized.

Summaries of sponsor-defined PCS values will be presented for systolic blood pressure, diastolic blood pressure, and heart rate. The number and percentage of subjects with PCS values that are reported at any postbaseline visit (scheduled or unscheduled) within each treatment period will be summarized by treatment period and treatment group. The criteria for identifying PCS vital signs values are provided in Table 4.

Table 4: Potentially Clinically Significant Criteria for Vital Signs Variables

Variable Name	PCS – Low if:			PCS – High if:		
	Observed Value is:	AND	Decrease from Baseline is:	Observed Value is:	AND	Increase from Baseline is:
Systolic Blood Pressure	<90 mmHg		≥20 mmHg	>180 mmHg		≥20 mmHg
Diastolic Blood Pressure	<50 mmHg		≥10 mmHg	>105 mmHg		≥15 mmHg
Heart Rate	<50 bpm		≥15 bpm	>120 bpm		≥15 bpm

Both supine and standing values of blood pressures and heart rate will be included in the identification and summary of PCS values.

11.4. Body Weight

The body weight data (in units of kilograms) will be summarized with descriptive statistics by treatment group at baseline and each scheduled postbaseline visit through Week 7 Day 7. Both observed values and changes from baseline will be summarized.

11.5. Electrocardiogram

The triplicate values of the quantitative ECG variables (heart rate, PR interval, QRS duration, QT interval, and Fridericia's correction of QT interval [QTcF]) measured at each visit will be averaged (and rounded to one decimal place) 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 three 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, Week 1, Week 4, and Week 7 Day 7. Both observed and change from baseline values will be summarized. Frequency counts and percentages for the investigator ECG interpretation variable categories will be summarized at each visit.

Categorical summaries will be presented for the QT and QTcF interval data for each treatment period. For these summaries, a subject's highest reported postbaseline value (including values reported at unscheduled visits) within each treatment period 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 QT and QTcF intervals (each interval will be summarized separately) for each treatment period. For the first summary, the number and percentage of subjects in each treatment group whose highest reported QT or 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 QT or QTcF increase from their baseline value meets the following thresholds:

- Increase greater than 30 msec
- Increase greater than 60 msec

11.6. Columbia-Suicide Severity Rating Scale

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

- Screening/lifetime assessment by treatment group and for "All Subjects"
- Screening/past 3 months assessment by treatment group and for "All Subjects"
- Baseline (Day -1) assessment by treatment group and for "All Subjects"
- Each treatment period by treatment group.

Each 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 (not reported for the screening/past 3 months assessment)
  - (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 treatment period 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) within each treatment period. Subjects missing either a baseline score or all postbaseline scores will not appear in the table.

11.7. Barnes Akathisia Rating Scale

The BARS is a 4-item scale to assess the presence and severity of drug-induced akathisia. This scale includes both objective items (eg, observed restlessness) and subjective items (eg, subject's awareness of restlessness and related distress) rated on a scale of 0 to 3, together with a global assessment of akathisia. Global assessment is rated on a scale of 0 to 5 (0=absent; 1=questionable; 2=mild akathisia; 3=moderate akathisia; 4=marked akathisia; 5=severe akathisia). The total score is calculated as the sum of the objective and subjective items (items 1-3 of the assessment) and ranges from 0 to 9. If any one of these 3 items is not scored (ie, has a missing value), the associated total score will be set equal to missing.

The total score and the global assessment score at each visit (baseline, Week 4, and Week 7 Day 7) will be summarized with descriptive statistics by treatment group. Changes from baseline at each postbaseline visit will also be summarized.

11.8. Simpson-Angus Scale

The SAS is a 10-item scale to evaluate the presence and severity of drug-induced parkinsonism and other extrapyramidal symptoms. Each item is rated on a 0 to 4 scale of increasing severity with definitions given for each anchor point. The global score is calculated as the average of the 10 items. If any one of these 10 items is not scored (ie, has a missing value), the associated global score will be set equal to missing.

The global score at each visit (baseline, Week 4, and Week 7 Day 7) will be summarized with descriptive statistics by treatment group. Changes from baseline at each postbaseline visit will also be summarized.

11.9. Montgomery-Asberg Depression Rating Scale

The MADRS is a 10-item scale designed to measure changes in severity of depressive symptoms. Each item is scored on a 7-point scale (0 to 6) with increasing number value indicating increasing severity for each item with anchor points provided at 2-point intervals. The total score ranges from 0 to 60. If any one of these 10 items is not scored (ie, has a missing value), the associated total score will be set equal to missing.

The total score at each visit (baseline, Week 4, and Week 7 Day 7) will be summarized with descriptive statistics by treatment group. Changes from baseline at each postbaseline visit will also be summarized.

11.10. Prior and Concomitant Medications

Prior medications and concomitant medications will be summarized by World Health Organization (WHO) Drug Anatomical Therapeutic Chemical Classification (ATC) Level 3 category (or Level 2 if there is not an applicable Level 3 category) and preferred name.

Medications will be assigned to one, two, or three study periods based on the medication start and stop dates relative to study drug dosing and Withdrawal Period:

- Pre-study/screening: medications with a start date prior to study drug dosing
- During the Valbenazine/Placebo Period: medications ongoing at the time of first dose of study drug or with a start date after the first dose of study drug, excluding medications started after the first dose of study drug in the Withdrawal Period (Week 5 Day 1)
- During the Withdrawal Period: medications ongoing at the Week 5 Day 1 visit (first dosing day in the Withdrawal Period) or with a start date on or after the Week 5 Day 1 visit.

A given medication can be assigned to multiple study periods in the tabular summaries, depending on its start and end dates.

The number and percentage of subjects using medications in each WHO Drug ATC category (Level 3/preferred name) will be summarized by treatment group and study period. A subject may take the same medication more than once or multiple medications for a subject may be classified under the same ATC level or preferred name. A subject is counted only once for each level of medication classification within a summary. An “All Subjects” column will be included in these summaries.

12. DEVIATIONS FROM PROTOCOL PLANNED ANALYSIS

Feedback from the FDA on the SAP was received on 31 July 2019 and 18 September 2019 and resulted in the following deviations from the protocol-specified analyses. Adverse events occurring during the Withdrawal Period will not be further classified as "withdrawal-related" based on a prespecified list of preferred terms. Additionally, the difference between treatment groups in number of subjects with withdrawal-related AEs will not be tested for statistical significance. All AEs beginning in the Withdrawal Period will be summarized by treatment group as described in Section 8.1 of this SAP. In addition, key assessments of clinical dependence and withdrawal will not be summarized descriptively by study site as described in the protocol.

13. 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.