STATISTICAL ANALYSIS PLAN
PHASE 3

Version: 1.0

Date: 19 November 2021

Based on Protocol: Amendment No. 5 Final Version: 12 April 2021
[redacted]

Study Drug: NBI-98854

Study Number: NBI-98854-HD3005

Study Title: A Phase 3, Randomized, Double-Blind, Placebo-Controlled Study to Assess the Efficacy, Safety, and Tolerability of Valbenazine for the Treatment of Chorea Associated with Huntington Disease

Study Sponsor: Neurocrine Biosciences, Inc.
12780 El Camino Real
San Diego, CA 92130
Telephone: [redacted]

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]	Electronically Signed and Dated	
[redacted]	Signature	Date

Neurocrine Biosciences, Inc.

This document has been reviewed by:

[redacted]	Electronically Signed and Dated	
[redacted]	Signature	Date

Neurocrine Biosciences, Inc.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Term
AE	Adverse event
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
ANC	Absolute neutrophil count
AS	Anosognosia Scale
ATC	Anatomical Therapeutic Chemical
BARS	Barnes Akathisia Rating Scale
BMI	Body mass index
CAG	Cytosine Adenine Guanine
CGI	Clinical Global Impression
CGI-C	Clinical Global Impression-Change
CGI-S	Clinical Global Impression-Severity
CI	Confidence interval
COVID-19	Coronavirus Disease 2019
CRS-PSP	Clinical Rating Scale for Progressive Supranuclear Palsy
C-SSRS	Columbia-Suicide Severity Rating Scale
CV	Coefficient of variation
ECG	Electrocardiogram
eCRF	Electronic case report form
EQ-5D-5L	EuroQoL 5 Dimensions 5 Levels
ET	Early termination
FAS	Full analysis set
GGT	Gamma-glutamyl transferase
HADS	Hospital Anxiety and Depression Scale
HD	Huntington Disease
HD-HI	Huntington Disease Health Index
IPD	Important protocol deviation
IWRS	Interactive web response system
LS	Least-squares

Abbreviation	Term
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple imputation
MMRM	Mixed-effect model for repeated measures
MoCA	Montreal Cognitive Assessment
NBI	Neurocrine Biosciences, Inc.
Neuro-QoL	Quality of Life in Neurological Disorders
PCS	Potentially clinically significant
PGI-C	Patient Global Impression-Change
PGI-S	Patient Global Impression-Severity
PK	Pharmacokinetics
PT	Preferred term
QTcF	Corrected QT interval using Fredericia's formula
SAE	Serious adverse event
SAP	Statistical analysis plan
SD	Standard deviation
SDQ	Swallowing Disturbance Questionnaire
SE	Standard error
SF-36	Short Form 36 Health Survey
SOC	System Organ Class
TEAE	Treatment-emergent adverse event
TFC	Total functional capacity
TMC	Total Maximal Chorea
TMS	Total Motor Score
UHDRS	Unified Huntington's Disease Rating Scale
ULN	Upper limit of normal
VAS	Visual analogue scale
WBC	White blood cell
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 3 Study NBI-98854-HD3005 (HD3005 hereafter).

This SAP was developed in accordance with International Council for Harmonization 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:

Evaluate the efficacy of valbenazine to reduce chorea associated with Huntington Disease (HD) using the Unified Huntington's Disease Rating Scale (UHDRS) Total Maximal Chorea (TMC) score.

Secondary Study Objectives:

- Evaluate the efficacy of valbenazine to reduce chorea associated with HD using the Clinical Global Impression of Change (CGI-C), the Patient Global Impression of Change (PGI-C), and the Quality of Life in Neurological Disorders (Neuro-QoL) Upper Extremity Function Short Form and the Neuro-QoL Lower Extremity Function Short Form.
- Evaluate the safety and tolerability of valbenazine in subjects with HD.

[redacted]

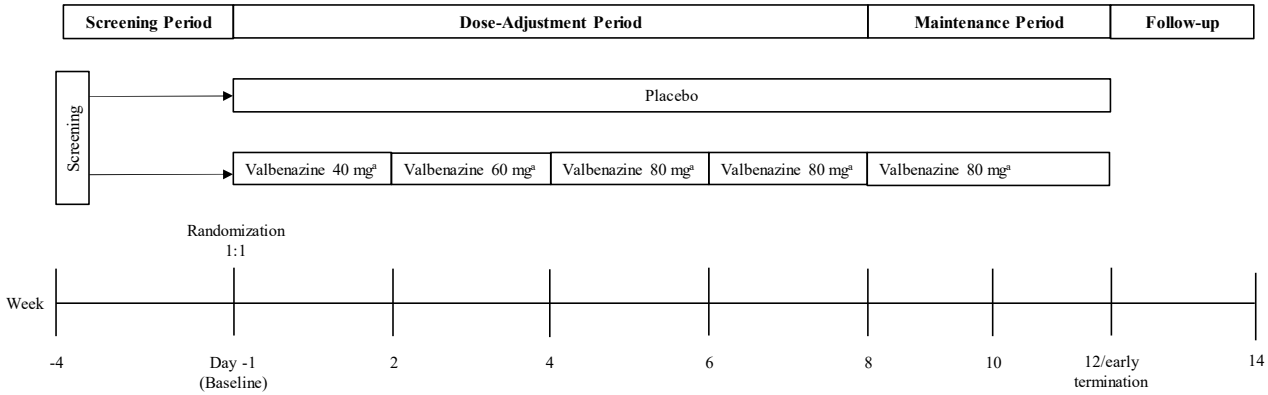
[redacted]

3. STUDY DESIGN

This is a Phase 3, randomized, double-blind, placebo-controlled study to evaluate the efficacy, safety, and tolerability of once-daily valbenazine in subjects with HD. The study includes a 4-week screening period, an 8-week dose-adjustment period, a 4-week maintenance period, and a final study visit 2 weeks following the final dose of study drug.

The study design schematic is shown in Figure 1.

Figure 1: Study Design Schematic



[a] Doses represent maximum daily doses during each 2-week period.
Early termination can happen at any time; the early termination assessments will be the same assessments as those performed at Week 12.

3.1. Randomization

Approximately 120 adult male and female subjects with motor manifest HD and a genetic diagnosis of HD will be enrolled. Eligible subjects will be randomized 1:1 to either valbenazine or placebo on Day -1 (baseline) using an interactive web response system (IWRS).

3.2. Blinding

This study includes a 12-week double-blind placebo-controlled treatment period during which the subject, investigator, all study center personnel, and the sponsor (or designee) will be blinded to the subject's treatment until the end of study. During the treatment period, all subjects will receive valbenazine or placebo as two capsules (identical in appearance) to be self-administered once daily.

3.3. Sample Size Considerations

A total sample size of 120 subjects (60 per treatment group) will provide >95% power to detect a difference between the valbenazine and placebo treatment groups in the change from the screening period baseline (the average of the screening and Day -1 assessments) to maintenance (the average of the Week 10 and Week 12 assessments) in the TMC based on site assessments, using a two-sample t-test with a two-sided Type I error of 0.05. This assumes a difference in means of 2.4 with a

common SD of 3.3. The specified treatment effect is attenuated to account for up to 4 (6.7%) subjects per treatment group who prematurely discontinue study drug dosing.

The above sample size and power estimates were obtained from [redacted].

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

The primary efficacy endpoint is the change from the screening period baseline (the average of the screening and Day -1 assessments) to maintenance (the average of the Week 10 and Week 12 assessments) in the TMC based on on-site assessments.

4.2. Secondary Efficacy

The secondary efficacy endpoints are:

- CGI-C response status at Week 12
- PGI-C response status at Week 12
- The change from baseline (Day -1) to Week 12 in the Neuro-QoL Upper Extremity Function T-score
- The change from baseline (Day -1) to Week 12 in the Neuro-QoL Lower Extremity Function T-score

4.3. Exploratory Efficacy

- Changes from the screening period baseline to each postbaseline study visit (Weeks 2 through 12) in the TMC based on site assessments
- Changes from the screening period baseline to maintenance (the average of the Week 10 and Week 12 assessments) in the TMC based on video recording central rater assessments
- CGI-C response statuses at Weeks 2 through 10
- PGI-C response statuses at Weeks 2 through 10
- The change from baseline to Weeks 4, 8, and 10 in the Neuro-QoL Upper Extremity Function T-score
- The change from baseline to Weeks 4, 8, and 10 in the Neuro-QoL Lower Extremity Function T-score
- Clinical Global Impression of Severity (CGI-S) at Weeks 2 through 12
- Patient Global Impression of Severity (PGI-S) at Weeks 2 through 12
- Short Form 36 Health Survey (SF-36) at Week 12
- Huntington Disease Health Index (HD-HI) at Week 10 and Week 12
- EuroQol 5 Dimensions 5 Levels (EQ-5D-5L) at Week 10 and Week 12

- UHDRS scores for motor, behavior, and functional capacity assessment at Weeks 2 through 12
- Other UHDRS scores including functional assessment and independence scale at Week 12
- Anosognosia Scale (AS) at Week 12
- [redacted]

4.4. Safety Endpoints

- Treatment-emergent adverse events (TEAEs), serious adverse events (SAEs), severe TEAEs, AEs leading to study drug dose reduction, and AEs leading to discontinuation of study drug dosing
- Clinical laboratory tests (including hematology and clinical chemistry)
- Prolactin
- Vital signs (including orthostatic blood pressure and pulse rate, respiratory rate, and oral body temperature)
- 12-lead Electrocardiogram (ECG)
- Columbia-Suicide Severity Rating Scale (C-SSRS)
- UHDRS motor assessment for parkinsonism (items rating retropulsion pull test, finger taps, pronate/supinate hands, rigidity-arms, and bradykinesia-body)
- Barnes Akathisia Rating Scale (BARS)
- Hospital Anxiety and Depression Scale (HADS)

4.5. Pharmacokinetics Endpoint

The PK endpoint is plasma concentrations of valbenazine and the metabolite (NBI-98782).

5. GENERAL CONSIDERATIONS FOR DATA ANALYSIS

All analyses described in this plan are considered a priori analyses in that they have been defined prior to locking the study database. Analyses defined subsequent to locking the database will be considered *post hoc* analyses and will be applied as exploratory methodology. Any pertinent *post hoc* analyses will be statistically justified and described in the clinical study report. Statistical analysis will be conducted, and all tables, figures, and listings generated using SAS® software [redacted], unless stated otherwise.

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 and ordinal categorical variables. 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 valbenazine treatment group and the placebo treatment group for selected efficacy variables. All hypothesis tests will be tests of the null hypothesis of no difference between the groups being compared versus the two-sided alternative hypothesis that there is a difference. The level of significance (type I error) for declaring statistical significance will be 0.05.

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 SE will have one more decimal place than the data being summarized; and the sample size (N) will be reported as an integer. This rule 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 subjects who are randomized to a treatment group, take at least one dose of study drug, and have any post baseline safety data.
- The full analysis set (FAS) will include all subjects who are randomized to a treatment group and who have at least one evaluable TMC change from baseline score (based on on-site assessments) during the 12-week double-blind treatment period.
- The PK analysis set will include all randomized subjects who receive a dose of study drug and have at least one post-dose PK sample qualified. The PK analysis set will be used for summary of plasma concentration.

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

Efficacy summaries and demographics/baseline characteristics are based on FAS. The FAS consists of all randomized subjects analyzed according to the treatment group to which they were randomly assigned.

All other summaries are based on safety analysis set. The safety analysis set consists of all randomized and treated subjects analyzed according to the treatment group to which they were randomly assigned unless they received the wrong treatment for the entire treatment period.

5.3. Baseline and Maintenance Definition

For efficacy analysis purpose, unless otherwise specified, the assessments collected at randomization visit 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. For TMC analysis, the screening period baseline is defined as the average of the screening and Day -1 assessments. If one of the assessments is not available, the available assessment will serve as screening period baseline. The maintenance value is defined as the average of the Week 10 and Week 12 assessments. If one of the assessments is not available, the available assessment will serve as the maintenance value.

For other analyses, the assessments collected at Day -1 prior to study drug will serve as the baseline value for all assessments. If a Day -1 visit value is not available, then the last measurement collected prior to study drug will serve as baseline.

5.4. Derived and Transformed Data

Change from baseline is 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 is 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. The percent change from baseline will also be missing if the baseline value is equal to zero.

5.5. Study Day

Study day is calculated relative to the first dose date. The visit number for each visit, including scheduled, unscheduled, repeat, and early termination/end of study visits, will be re-mapped based on the actual study day according to Table 1. 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.

Table 1: Analysis Visit Windows

Scheduled Visit	Target Study Day	Analysis Window (Study Day Range)
Week 2	14	1-21
Week 4	28	22-35

Scheduled Visit	Target Study Day	Analysis Window (Study Day Range)
Week 6	42	36-49
Week 8	56	50-63
Week 10	70	64-77
Week 12	84	78-91
Week 14	98	92+

5.6. Handling of Missing Data

5.6.1. Missing Assessments

Missing values for assessments will generally not be replaced with imputed values except as noted in Section 5.5 for the ET visit data mapped to scheduled visits for data summary purposes.

Any imputation methods used for the efficacy endpoints are further discussed in the section of efficacy (Section 7).

5.6.2. Missing Dates

5.6.2.1. First and Last Dose Dates

Missing and incomplete ("partial") dates for first and last dose dates will be imputed for the purpose of estimating exposure and defining treatment periods. Missing dates will not be imputed for subjects when the subject is known to have not taken at least one dose of study drug, as documented by the site in the dosing electronic case report form (eCRF).

The imputation rules for first dose date are as follows:

- If the date is completely missing or if both the day and month are missing, the date will be imputed as the randomization date
- If only the day is missing, the date will be imputed as the randomization date if the month and year match the month and year of the randomization date; if the month or year occur after the randomization date, the missing day will be imputed as the first day of the month.

If the date of the last dose of study drug is missing, then the last dose date will be imputed as the earliest of:

- The Week 12 visit date,
- The end of treatment date for subjects who discontinue on or before Week 12,
- The last visit (up to Week 10) prior to discontinuation or the date of an unscheduled dose reduction, whichever occurs later.

5.6.2.2. Start Date for Adverse Events and Start and Stop Dates for 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 drug. Any data listings will display the original dates as reported in the database.

To handle missing/partial start dates for adverse events, 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.

The missing and incomplete ("partial") dates for AEs will be imputed using the following algorithm. For start date: missing day (impute the 1st of the month unless the same month and year as the study drug, otherwise impute the first dose date of study drug); missing day and month (impute 1st January, unless the same year as the study drug, otherwise impute the first dose date of study drug). There will be no imputation for AE stop dates.

To handle missing/partial dates for prior and concomitant medications, the following algorithm will be employed to estimate the time of medication usage relative to study drug. For start date: missing day (impute the 1st of the month, if the same month and year as the study drug, otherwise impute the first dose date of study drug); missing day and month (impute 1st January, if the same year as the study drug, otherwise impute the first dose date of study drug). For stop date: missing day (impute the last day of the month if the same month and year as the study drug, otherwise impute the last dose date of study drug); missing day and month (impute 31st December, if the same year as the study drug, 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.

5.7. 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.

5.8. 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]) and Health Canada (Management of Clinical Trials During the COVID-19 Pandemic: Notice to Clinical Trial Sponsors [May 2021]), 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 will be generated. The listing will identify subjects that experience one of the following situations due to the COVID-19 pandemic (additional situations may be included):

  - Discontinued study drug 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 drug interruption
- A table and/or listing will summarize the COVID-19 pandemic-related reasons for treatment and study discontinuation.
- A table and listing will summarize the COVID-19 pandemic-related major protocol deviations.
- A table will summarize assessments affected by the COVID-19 pandemic (eg, missing, partial, collected remotely).
- 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 (and secondary endpoints, where appropriate) will be performed.

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
  - Full Analysis Set
  - Safety Analysis Set
  - Completed the dose-adjustment period (ie, up to Week 8)
  - Completed the dose maintenance period (ie, up to Week 12)
  - Completed study drug dosing
  - Completed study
  - Discontinued study drug prematurely, including reasons for discontinuation. Discontinuation of study drug due to COVID-19 will also be summarized.
  - 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. 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. This 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.

6.3. Demographic and Other 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 Full Analysis Set.

Demographics include:

- Age (years)
- Age categories (<65 years versus ≥65 years)
- Sex
- Ethnicity
- Race
- Country (US and Canada)

Baseline subject characteristics include, but not limited to:

- Height (measured at screening, cm)
- Weight (presented in kilograms)
- Body mass index (BMI; measured at screening; kg/m^2)
- CAG repeat length
- [redacted] phenotype status
- Montreal Cognitive Assessment (MoCA)
- Swallowing Disturbance Questionnaire (SDQ)
- Baseline TMC based on site assessments
- Baseline CGI-S categories (<4, >=4)
- Baseline PGI-S categories (<3, >=3)

7. EFFICACY

The efficacy endpoints and planned analysis methods are described below. Unless otherwise specified, the Full Analysis Set will be used for all efficacy analyses, and descriptive statistics presented by treatment group (valbenazine vs. placebo) according to the study visit.

7.1. Multiple Comparisons and Multiplicity Adjustment

A fixed-sequence testing procedure will be used to control the family-wise error rate for the primary and secondary efficacy endpoints. Testing of hypotheses at each step of the procedure commences only if all null hypotheses of the prior steps were rejected. The fixed-sequence testing procedure will consist of performing the hypothesis tests in the following prespecified order:

1. Primary endpoint: TMC change from screening period baseline to maintenance (valbenazine versus placebo treatment group).
2. Secondary endpoint: CGI-C response at Week 12 (valbenazine versus placebo treatment group).
3. Secondary endpoint: PGI-C response at Week 12 (valbenazine versus placebo treatment group).
4. Secondary endpoint: Neuro-QoL Upper Extremity Function change from baseline to Week 12 (valbenazine versus placebo treatment group).
5. Secondary endpoint: Neuro-QoL Lower Extremity Function change from baseline to Week 12 (valbenazine versus placebo treatment group).

Each step in the sequential testing procedure uses a local two-sided 0.05 level of significance for the null hypothesis being tested. The null hypothesis at each step of the procedure can only be rejected if all null hypotheses in prior steps were rejected.

7.2. Interim Analysis

No interim analysis is planned for this study.

7.3. Primary Efficacy Endpoint

The UHDRS is a tool developed by the Huntington Study Group (HSG) to assess the clinical features and course of HD. The full UHDRS includes the following assessments: motor, cognitive, behavioral, independence, function, and total functional capacity (TFC).

The TMC is item 12 of the motor assessment and measures the chorea in 7 different body parts including the face, oral-buccal-lingual region, trunk and each limb independently. The maximum score is 28.

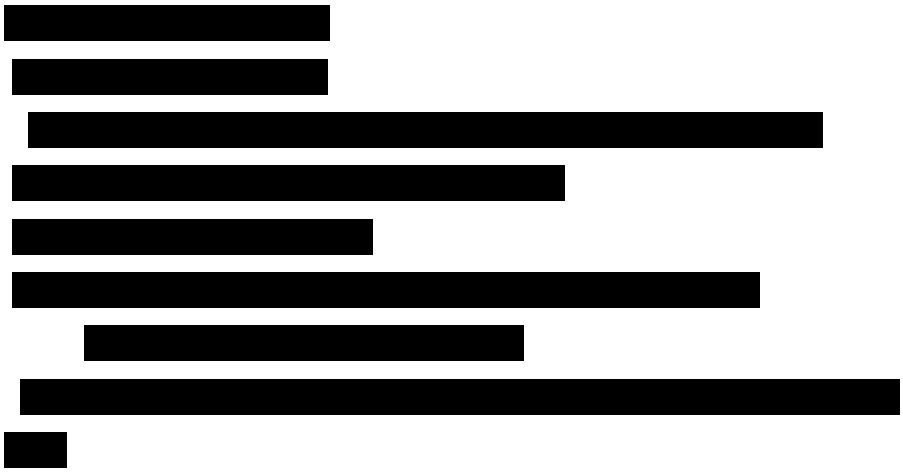
The primary estimator is the mean treatment difference between the valbenazine and placebo treatment groups in the change from the screening period baseline (the average of the screening and Day -1 assessments) in the TMC score based on on-site assessments to maintenance (the average of the Week 10 and Week 12 assessments), in all randomized subjects, regardless of adherence to study drug. The primary analysis will be based on observed TMC scores. For

sensitivity analysis, subjects who are missing TMC scores at treatment visits will have their data imputed using observed data from other subjects through a multiple imputation procedure.

Changes from the screening period baseline to each postbaseline study visit (Weeks 2 through 12) and changes from the screening period baseline to maintenance (the average of the Week 10 and Week 12 assessments) in the TMC based on video recording central rater assessments are exploratory efficacy endpoints.

Descriptive statistics will be presented by treatment group for the TMC observed values, and changes from the screening period baseline at each visit and at maintenance. The changes from the screening period baseline to each postbaseline visit during the treatment period will be analyzed using a mixed-effect model repeated measures (MMRM) analysis. The model will include the screening period baseline TMC as a covariate, and treatment group, visit, treatment group-by-visit interaction, and baseline-by-visit interaction as fixed effects. Subject will be included as a random effect. Treatment group comparisons of the valbenazine treatment group vs. placebo at maintenance and 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 95% confidence intervals. The treatment difference for the primary endpoint (maintenance) will be considered statistically significant if the two-sided p value is <0.05. Nominal (raw) two-sided 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 is provided below:



7.3.1. Sensitivity Analysis of Primary Efficacy Variable

The following sensitivity analyses will be performed for the primary endpoint with differing assumptions to explore the robustness of inferences from the main estimator:

- Multiple imputation (MI) analysis assuming data are missing at random (MAR): This methodology will structure data based on missing data patterns. The method will be based on a missingness pattern having a monotone structure, ie, if among the observations over time one data value is missing, all other values after this missing value will also be treated as missing. For subjects with intermittent missing values, before performing MI, 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 1000 times, creating several different datasets with a monotone missing data structure. The imputation is based on the MAR assumption, ie, 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:

After this step, the remaining missing data will be imputed using a method for monotone missingness, also based on the MAR assumption. Thus, for each created dataset with a monotone missing data pattern, the MI procedure using SAS® software will be employed to impute missing values based on a sequential procedure reflecting the monotone missing data pattern. Subjects with the first missing value occurring at Week 2 will have their missing Week 2 value replaced by an imputed value from a regression model with treatment group and screening period baseline TMC score as explanatory variables. In the next step, subjects with their Week 4 value missing will have their missing Week 4 value replaced by an imputed value from a regression model with treatment group, the screening period baseline value and the Week 2 value as explanatory variables. A similar procedure will be used to replace the missing values at Week 6, Week 8, Week 10, and Week 12.

The following SAS® software code will be used to perform the imputation with the MAR assumption:

The imputed datasets generated with the approach described above contain only non-missing values and are used as input in the model for the sensitivity analysis of the primary endpoint. MMRM models similar to that described for the primary analysis will thus be run on each of the 1000 generated imputed datasets and the difference between the treatment groups at maintenance visit 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 the maintenance visit. In addition to the estimates, corresponding 95% CIs and p-values will be calculated.

The following SAS® software code will be used to complete the three-step process:

- MI analysis assuming data are not missing at random (MNAR): The trajectories of the subjects are assumed to follow the placebo group after the discontinuation (ie, control group-based assumption). The procedure aforementioned will be used to generate monotone missing data. After this, the missing data will be imputed sequentially for each visit (first Week 2, followed by Week 4, followed by Week 6, followed by Week 8, followed by Week 10, and followed by Week 12). Only data from the control group will be used for the imputation. The MNAR statement of the following type will be used to generate the imputation:

- Tipping point analysis (MNAR): The tipping point assumption will be used, ie, after the first visit with missing data, the trajectories of the subjects in the valbenazine group are assumed to be worse by an amount of delta. After the MI using the MAR assumption, as defined previously, an amount of delta will be added to each imputed value in the valbenazine group. Successively harsher deltas will be imposed on the imputed values in the valbenazine group, starting with a TMC score increase (worsening) of 1. The delta is further increased in the steps of 1 (ie, +1, +2, +3, …) until the statistical significance is lost, ie, until the p-value becomes ≥0.05. For the control group, the MI using MAR assumption will be used.

7.3.2. 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 Full Analysis Set with non-missing data for the primary endpoint at all scheduled post-baseline visits up to the Week 12 visit.
- COVID-19 impact analysis excluding all subjects who discontinued study drug dosing prematurely due to COVID-19.

7.4. Secondary Efficacy Endpoints

The CGI-C requires a clinician to rate how much the subject's HD has improved or worsened relative to baseline. A subject's illness improvement status is rated as:

- 1=Very much improved
- 2=Much improved

- 3=Minimally improved
- 4=No change
- 5=Minimally worse
- 6=Much worse
- 7=Very much worse

Subjects whose CGI-C score is either a 1 ("very much improved") or a 2 ("much improved") will be classified as responders. Response status at Week 12 is a secondary endpoint; Weeks 2 through 10 are exploratory efficacy endpoints.

The PGI-C response status based on patient rating is defined similarly.

For analysis of CGI-C and PCI-C response status at Week 12, the descriptive statistics will be presented by treatment group for the number and percentage of subjects classified as responders at each postbaseline visit. Analyses comparing the valbenazine treatment group to placebo will be performed using the Fisher's exact test. Nominal two-sided p-values for comparing treatment groups will be reported in summary tables; however, determining statistical significance of the p-value for the treatment group comparison at Week 12 will be based on a procedure which controls for multiple comparisons.

An example of the SAS® software code for Fisher's exact test is provided below:

A sensitivity analysis will be performed by imputing non-responder to subjects with missing data at Week 12.

The change from baseline (Day -1) to Week 12 in the Neuro-QoL Upper and Lower Extremity Function T-score are secondary efficacy endpoints; the change from baseline to Weeks 4, 8, and 10 are exploratory efficacy endpoints.

For analysis of the change from baseline (Day -1) to Week 12 in the Neuro-QoL Upper and Lower Extremity Function T-score, the descriptive statistics will be presented by treatment group for the observed values and changes from baseline at each visit. The changes from baseline to each postbaseline visit during the treatment period will be analyzed using MMRM. The MMRM model will be similar to the model described above for the TMC analysis, with the exception that the covariate in the model will be the corresponding baseline value (Day -1) of the Neuro-QoL Upper and Lower Extremity Function T-score respectively, and only visits where Neuro-QoL is scheduled to be collected will be included. Nominal two-sided p-values for comparing treatment groups and the associated 95% confidence intervals will be reported in summary tables; however, determining statistical significance of the p-value for the treatment group comparison at Week 12 will be based on a procedure which controls for multiple comparisons.

Similar supplementary analyses, described in Section 7.3.2, will be performed for the secondary efficacy endpoints.

7.5. Other Efficacy Endpoints

The other efficacy endpoints will be summarized descriptively by treatment group and visit.

7.5.1. Clinical Global Impression-Severity (CGI-S)

The CGI scales are commonly used measures of symptom severity and treatment response.

The CGI-S requires a clinician to rate the severity of the subject's HD at the time of the assessment, relative to the clinician's past experience with HD patients. The subject should be assessed on the severity of his or her HD at the time of rating as:

- 1=Normal, not at all ill
- 2=Borderline ill
- 3=Mildly ill
- 4=Moderately ill
- 5=Markedly ill
- 6=Severely ill
- 7=Among the most extremely ill patients

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 Week 12 will be presented.

7.5.2. Patient Global Impression-Severity (PGI-S)

Subjects will rate their improvement in their HD symptoms at the time of the assessment. The subject should be assessed on the severity of his or her HD at the time of rating as:

- 1=None
- 2=Mild
- 3=Moderate
- 4=Severe
- 5=Very Severe

PGI-S categories will be analyzed similarly as that previously described for CGI-S.

7.5.3. Short Form 36 Health Survey (SF-36)

The SF-36 is a 36-item, self-administered questionnaire. It measures health on 8 dimensions: vitality, physical functioning, pain, general health perception, physical role limitations, emotional role functioning, social functioning, and mental health.

Observed values and change from baseline to Week 12 will be summarized descriptively for all dimension scores by treatment group.

7.5.4. Huntington Disease Health Index (HD-HI)

The Huntington Disease Health Index (HD-HI) is a disease-specific patient reported outcome measure designed to evaluate patient disease burden in therapeutic trials. The HD-HI comprises 13 subscales that measure 13 individual areas of HD patient health. Together, each of the subscales can be utilized to generate an estimate of a HD patient's overall disease burden. Each question in the instrument was selected based on its high relevance to the HD population, its ability to be consistently understood by patients and clinicians, its content validity, its face validity, and its potential responsiveness to measure therapeutic benefit of disease progression during clinical trials.

Upon completion of the HD-HI, 14 scores are generated: a subject receives a score for each of the 13 subscales and a total instrument score. The total instrument score is a composite of the 13 subscales. The score for each subscale and the total instrument ranges from 0 to 100 with 100 representing the highest disease burden and a score of 0 representing no disease burden.

Observed values and change from baseline will be summarized descriptively for all subscales and the total instrument score by treatment group and visit.

7.5.5. EuroQol 5 Dimensions 5 Levels (EQ-5D-5L)

The EuroQol 5 Dimensions 5 Levels (EQ-5D-5L) is a general, single index measure for describing and valuing health. It defines health in terms of 5 dimensions: Mobility, Self-Care, Usual Activities, Pain/Discomfort, and Anxiety/Depression. Each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems, and extreme problems. The subject indicates his/her health state by checking the box next to the most appropriate statement. The scores for the 5 dimensions can be combined into a 5-digit number that describes the patient's health state. Number and percentage of subjects in each of the 5 levels for each dimension at baseline and Weeks 10 and 12 will be summarized by treatment group.

Subjects also rate their overall health on a 0 to 100 hash-marked, vertical visual analogue scale (EQ-VAS). The endpoints are labeled 'The best health you can imagine is at the 100' and 'The worst health you can imagine is at the 0'. EQ VAS score at each time point and change from baseline to Week 10 and Week 12 will be summarized by treatment group.

Health states in the EQ-5D-5L can be converted into a single index value, where index values are presented in the country specific value sets. EQ-5D-5L value sets can be used to obtain the EQ5D-5L index values based on the crosswalk for the respective countries. Observed values and change from baseline in EQ-5D index score will be summarized by treatment group and visit.

7.5.6. UHDRS Scores for Motor, Cognitive, Behavior, and Functional Capacity Assessment at Weeks 2 through 12

The motor portion of the UHDRS consists of 15 items that measure the severity of the motor symptoms. Similarly, to the analysis of TMC as described in Section 7.3, the total motor score (TMS) will be summarized by treatment group and visit.

The cognitive portion of the UHDRS consists of 3 executive function tests to evaluate the cognitive performance. The score of individual tests will be summarized by treatment group and visit.

The behavior portion of the UHDRS consists of 11 items (under four subscales: mood, behavior, psychosis, and obsessive) that measure the severity and frequency of behavior symptoms. The frequency and severity of each item is ranked on a 0 to 4 scale with higher scores indicating higher frequency or more severe behavioral symptoms, respectively. The sum of behavior frequency scores (11 total items, range 0-44), and the sum of behavior frequency-times-severity item scores (11 total item products, range 0-176) will be summarized. The behavior milestone questions (items 36 to 40) will be summarized by treatment group.

The functional capacity portion of the UHDRS consists of 5 items. Similarly, to the analysis of TMC as described in Section 7.3, the individual items and total functional capacity score (TFC) will be summarized by treatment group and visit.

7.5.7. Other UHDRS Scores including Functional Assessment and Independence Scale at Week 12

The functional assessment portion of the UHDRS consists of 25 functional questions with binary outcomes (Yes = 1, No = 0). Total scores defined as the sum of the 25 questions at baseline and Weeks 12 will be summarized by treatment group.

The independence scale of UHDRS will be summarized similarly.

7.5.8. Anosognosia Scale (AS) at Week 12

The Anosognosia Scale (AS) is an instrument to screen for anosognosia in daily practice and is specific for HD. This scale requires subjects and investigators to rate the subject's ability to perform tasks relative to individuals of the subject's age and education on a five-point scale, rating 8 items.

Observed values and change from baseline to Week 12 will be summarized for AS total score by treatment group for subjects and investigators ratings, respectively.

[redacted]

7.5.10. Examination of Subgroups

The following subgroups have been pre-planned and will be used to examine consistency of effect for the primary efficacy endpoint:

- Demographics (age <65 years versus ≥65 years, sex, race)
- Baseline CGI-S categories (<4 versus ≥4)
- Baseline PGI-S categories (<3 versus ≥3)

If the categories are sparse in number of subjects (eg, <15), similar categories maybe combined for analysis. The MMRM model will be used similar to the model described above for the primary analysis of the TMC scores, with additional interaction between treatment groups and subgroup variables included in the model. Nominal two-sided p-values for comparing treatment groups and the associated 95% confidence intervals will be reported in summary tables and p-values for the interaction term between treatment groups and subgroup variables will be presented.

8. PHARMACOKINETICS

The plasma concentrations of valbenazine and the metabolite NBI-98782 will be summarized with descriptive statistics, which will include the coefficient of variation (CV[%]), geometric mean, and geometric CV(%) by treatment group, dose (last dose received prior to blood sample being drawn), and visit (Day -1 and Weeks 2, 4, 6, 8, 10, 12, and 14) based on PK analysis set. Concentrations below the lower limit of quantification (valbenazine: 1.00 ng/mL and NBI-98782: 0.10 ng/mL) 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 valbenazine as measured by TEAEs and SAEs, vital signs, clinical laboratory tests, prolactin, ECG 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 Drug Dosing and Compliance

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

Duration of exposure will be summarized with descriptive statistics by treatment group. The number of subjects at each dose level in maintenance period will be summarized.

The number and percentage of subjects who are dose compliant (site evaluation of at least 80% of expected number of doses taken) will be summarized with descriptive statistics by visit (Weeks 2, 4, 6, 8, 10, and 12).

The number and percentage of subjects with dose adjustments will be summarized by visit and treatment group.

9.2. Adverse Events

Adverse events are recorded in the eCRF. Each AE will be coded to SOC and PT using MedDRA (Version 24.0).

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.

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 study drug and up to the last dose of study drug + 14 days.

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

- 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 valbenazine treatment group;
- Frequency of TEAEs by PT, with PT sorted by decreasing frequency (number of unique subjects) in valbenazine 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 dose reduction, any TEAE leading to study drug 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 post-treatment emergent events (AE start dates after last dose of study drug to the date of the last dose of study drug + 14 days) and non-treatment emergent AEs (AE start dates after last dose date + 14 days) separately. The events will be summarized by treatment group and presented by PT within SOC.

9.2.1. Adverse Events Leading to Premature Discontinuation of Study Drug

Summary tables of TEAEs leading to early discontinuation of study drug will be presented by treatment group. The number and percentage of subjects with a TEAE leading to study drug 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 drug 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 drug discontinuation.

A listing of TEAEs leading to premature study drug 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.

9.2.2. Adverse Events Leading to Study Drug Dose Reductions

Summary tables of TEAEs leading to study drug dose reductions will be presented by treatment group. The number and percentage of subjects with a TEAE leading to 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.

9.2.3. Deaths and Other Serious Adverse Events

Summary tables of serious adverse events (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.

9.3. Laboratory Safety Parameters

The following clinical safety laboratory parameters will be analyzed:

Hematology: complete blood count including WBC count with differential, red blood cell (RBC) count, hemoglobin, hematocrit, and platelet count, and absolute neutrophil count (ANC).

Clinical Chemistry: sodium, potassium, calcium, magnesium, phosphorus, chloride, blood urea nitrogen, bicarbonate, creatinine, uric acid, albumin, alkaline phosphatase, lactate dehydrogenase, AST, ALT, GGT, creatine kinase, total bilirubin, total cholesterol, triglycerides, total protein, and glucose.

Prolactin: serum prolactin will be summarized by visit and treatment group for each sex separately and combined.

The hematology, clinical chemistry, and prolactin data will be summarized with descriptive statistics by treatment group at baseline and at each scheduled postbaseline visit through Week 14. 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. The shift tables will present the shift from baseline to Week 12 (using observed values). 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 postbaseline. A "Total" row and "Total" column will also be included. Subjects with a missing baseline value or who do not have postbaseline data will be excluded from the shift 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),
- gamma-glutamyl transferase (GGT),
- total bilirubin,
- creatine kinase,
- creatinine,
- prolactin,
- blood urea nitrogen (BUN),
- white blood cell count (WBC),
- absolute neutrophil count (ANC),
- 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, ANC, creatinine, and BUN. The number and percentage of subjects with PCS values that are reported at any postbaseline visit (scheduled or unscheduled) will be summarized by treatment. The criteria for identifying PCS clinical laboratory values are provided in Table 2.

Table 2: Potentially Clinically Significant Criteria for Clinical Laboratory Variables

Variable	PCS Threshold
ALT	>3 x ULN
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)

ALT=alanine aminotransferase; AST=aspartate aminotransferase; BUN=blood urea nitrogen; GGT=gamma-glutamyl transferase; PCS=potentially clinically significant; ULN=upper limit of normal

Boxplots will be presented for the prolactin data by treatment group at each visit (baseline through Week 14). Separate plots will also be presented for each sex.

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 and graphs. Exceptions to this rule are: (1) all available lab values will be used in the PCS summaries and (2) if there are missing results from the original samples at screening, the results of a repeat screening sample will be substituted for the missing results in summary tables and graphs.

9.4. Vital Signs

The vital signs data, including orthostatic systolic and diastolic blood pressure (calculated as standing value minus supine value), orthostatic pulse rate, respiratory rate (recorded only supine), and oral body temperature, will be summarized with descriptive statistics by treatment group at baseline and at each scheduled postbaseline visit through Week 14. Both observed values and changes from baseline will be summarized.

Summaries of sponsor-defined PCS values will be presented for systolic blood pressure, diastolic blood pressure, heart rate, orthostatic systolic blood pressure, orthostatic diastolic blood pressure, and orthostatic heart rate. The number and percentage of subjects with PCS values that are reported at any postbaseline visit (scheduled or unscheduled) will be summarized by treatment group. The criteria for identifying PCS vital signs values are provided in Table 3.

Table 3: 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	N/A	≥20 mmHg	>145 mmHg	≥20 mmHg
Diastolic Blood Pressure	N/A	≥10 mmHg	>90 mmHg	≥10 mmHg
Heart Rate	<50 bpm	≥15 bpm	>120 bpm	≥15 bpm
Orthostatic Systolic Blood Pressure	Decrease of ≥20 mmHg		Increase of ≥20 mmHg	
Orthostatic Diastolic Blood Pressure	Decrease of ≥10 mmHg		Increase of ≥10 mmHg	
Orthostatic Heart Rate	Decrease of ≥15 bpm		Increase of ≥15 bpm	

PCS=potentially clinically significant

Both supine and standing values of blood pressures and heart rate will be included in the identification and all vital sign values will be used for the summary of PCS values.

9.5. 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 Week 14. Both observed values and changes from baseline will be summarized.

9.6. 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 Week 14. 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 QT and 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 QT and QTcF intervals (each interval will be summarized separately). 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:

- Greater than 30 msec
- Greater than 60 msec

9.7. Other Safety Endpoints

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

The Columbia-Suicide Severity Rating Scale (C-SSRS) is a measure of suicidal ideation and behavior. 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 (not reported for the Screening/past 1 year 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 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 not appear in the table.

9.7.2. UHDRS Motor Score (Items for Parkinsonism)

A subset of the UHDRS Motor Assessment (items rating retropulsion pull test, finger taps, pronate/supinate hands, rigidity-arms, and bradykinesia-body) will be used to assess for parkinsonism at each visit.

Observed values and change from baseline in parkinsonism total motor scores will be summarized descriptively by treatment group and visit.

9.7.3. Barnes Akathisia Rating Scale (BARS)

The Barnes Akathisia Rating Scale (BARS) is a validated four-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, subjects' awareness of restlessness and related distress), together with a global assessment of akathisia. Objective akathisia, subjective awareness of restlessness and subjective distress related to restlessness are rated on a 4-point scale from 0 – 3 and are summed yielding a total score ranging from 0 to 9. Global assessment is made on a 6-point scale of 0 to 5 (0=absent; 1=questionable; 2=mild akathisia; 3=moderate akathisia; 4=marked akathisia; 5=severe akathisia).

The BARS total score will be summarized by treatment group and visit. The BARS global assessment of akathisia score summary will be presented for each treatment group.

9.7.4. Hospital Anxiety and Depression Scale (HADS)

The HADS is a commonly used instrument to determine the levels of anxiety and depression that a person is experiencing. The HADS is a 14-item scale; 7 of the items relate to anxiety and 7 relate to depression. Each item is answered on a 4-point (0 to 3) response categories so the possible scores range from 0 to 21 for anxiety and 0 to 21 for depression. The HADS has been validated as a measure of depression and anxiety.

Observed values and change from baseline in HADS scores for anxiety and depression will be summarized descriptively by treatment group and visit.

10. DEVIATIONS FROM PROTOCOL PLANNED ANALYSIS

The analyses described in this SAP are consistent with the protocol planned analyses. The following assessments will be reviewed by study personnel but will not be included in tables, listings, or figures: medical history, prior and concomitant medications, serology, urinalysis, and pregnancy test results.

11. PERFORMANCE QUALIFICATION OF SAS® PROGRAMS

The analysis and summary of data from this study will be performed using SAS® [redacted]. 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.

12. REFERENCES

None.