# STATISTICAL ANALYSIS PLAN
# PHASE 2

**FINAL**

**DATE OF PLAN:**

*05 May 2017*

**STUDY DRUG:**

NBI-98854

**PROTOCOL NUMBER:**

NBI 98854-1501

**STUDY TITLE:**

**A RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED STUDY TO ASSESS THE SAFETY AND EFFICACY OF NBI-98854 IN PEDIATRIC SUBJECTS WITH TOURETTE SYNDROME**

**SPONSOR:**

*Neurocrine Biosciences, Inc.*
12780 El Camino Real
San Diego, CA 92130
Telephone: (858) 617-7600

**CONFIDENTIAL**

# SIGNATURE PAGE

This document has been prepared by:

[redacted] 05 May 2017

<<Printed Name>>, Project Statistician — Signature — Date
**Neurocrine Biosciences, Inc.**

This document has been reviewed and accepted by:

[redacted] 5/5/17

<<Printed Name>>, <<Title>> — Signature — Date
**Neurocrine Biosciences, Inc.**

[redacted] 05 MAY 2017

<<Printed Name>>, <<Title>> — Signature — Date
**Neurocrine Biosciences, Inc.**

# TABLE OF CONTENTS

# LIST OF ABBREVIATIONS

| Abbreviation | Term |
|---|---|
| ADHD | Attention-Deficit Hyperactivity Disorder |
| AE | Adverse event |
| ALT | Alanine aminotransferase |
| ANCOVA | Analysis of covariance |
| ANOVA | Analysis of variance |
| AST | Aspartate aminotransferase |
| ATC | Anatomical Therapeutic Chemical Classification |
| BLQ | Below the limit of quantification |
| CDRS-R | Children's Depression Rating Scale - Revised |
| CGI-Tics | Clinical Global Impression of Tics |
| CGI-TS | Clinical Global Impression of Change – Tourette Syndrome |
| CMH | Cochran-Mantel-Haenszel |
| C-SSRS | Columbia-Suicide Severity Rating Scale |
| CY-BOCS | Children's Yale-Brown Obsessive-Compulsive Scale |
| CYP2D6 | Cytochrome P450 2D6 |
| ECG | Electrocardiogram |
| eCRF | Electronic case report form |
| ESRS-A | Extrapyramidal Symptom Rating Scale - Abbreviated |
| ET | Early termination |
| GGT | Gamma-glutamyl transferase |
| HBsAg | Hepatitis B surface antigen |
| HCV-Ab | Hepatitis C antibody |
| HIV-Ab | Human immunodeficiency virus antibody |
| IPD | Important protocol deviation |
| ITT | Intent-to-treat |
| LS | Least squares |
| MedDRA | Medical Dictionary for Regulatory Activities |
| MMRM | Mixed-effect model repeated measures |
| n, N | Sample size (number of subjects) |
| NBI | Neurocrine Biosciences, Inc. |
| PK | Pharmacokinetic(s) |

| Abbreviation | Term |
|---|---|
| ADHD | Attention-Deficit Hyperactivity Disorder |
| PP | Per-protocol |
| PT | Preferred term |
| PUTS | Premonitory Urge for Tics Scale |
| QTcF | Fridericia's correction of QT interval |
| RTRS | Rush Video-Based Tic Rating Scale |
| SAE | Serious adverse event |
| SAP | Statistical Analysis Plan |
| SAS® | SAS® statistical software |
| SD | Standard deviation |
| SEM | Standard error of the mean |
| SOC | System organ class |
| TEAE | Treatment-emergent adverse event |
| TS | Tourette Syndrome |
| TTS | Total Tic Score |
| WHO | World Health Organization |
| YGTSS | Yale Global Tic Severity Scale |

# 1. INTRODUCTION

This statistical analysis plan (SAP) provides a detailed description of the analyses, tables, figures, and listings that will be prepared to summarize the data from the Phase 2 study described in Neurocrine Biosciences, Inc. (NBI) Protocol NBI-98854-1501.

# 2. STUDY OBJECTIVES

The objectives of this study are:

- To evaluate the efficacy of 2 active doses of NBI-98854 (10 mg and 20 mg in children; and 20 mg and 40 mg in adolescents) administered once daily in pediatric subjects with Tourette Syndrome (TS).
- To assess the safety and tolerability of repeated daily doses of NBI-98854 in pediatric subjects with TS.
- To evaluate plasma exposure of NBI-98854 and its metabolite, NBI-98782, following repeated daily doses of NBI-98854.

## 2.1. Study Design Overview

This is a Phase 2, multicenter, randomized, double-blind, placebo-controlled, parallel group study to evaluate the efficacy, safety, and tolerability of 2 doses of NBI-98854 (10 mg and 20 mg in children, and 20 mg and 40 mg in adolescents) relative to placebo, administered once daily for 6 weeks in pediatric subjects with TS. Approximately 90 male and female subjects, 6 to 17 years of age, with a diagnosis of TS will participate. The subjects will be divided into 2 age groups of approximately 45 children (6 to 11 years of age) and 45 adolescents (12 to 17 years of age). Eligible subjects within each age group will be randomized in a 1:1:1 ratio to placebo or 1 of 2 NBI-98854 treatment groups as depicted in the following table.

**Table 1: Study Randomization Schematic**

| **Age Group**<br>Treatment Group | **Number of Subjects**<br>(approximate) |
|---|---|
| **Children**, 6 to 11 years old | |
| Placebo | 15 |
| NBI-98854 10 mg | 15 |
| NBI-98854 20 mg | 15 |
| **Adolescents**, 12 to 17 years old | |
| Placebo | 15 |
| NBI-98854 20 mg | 15 |
| NBI-98854 40 mg | 15 |

Subjects will be screened to determine eligibility within 21 days (Days -21 to -1) before the start of study drug dosing on Day 1. On Day -1 (baseline; the day before receiving the first dose of study drug), eligible subjects will return to the study center for collection of baseline safety and efficacy assessments and collection of a blood sample for subsequent determination of their CYP2D6 metabolizer status. Subjects who continue to be eligible for the study will then be

randomized (1:1:1) to placebo or 1 of 2 NBI-98854 doses determined by subject age group. A 2-week supply of study drug will be dispensed.

Beginning on Day 1 (the day after the Day -1 visit), study drug will be administered once daily in the morning at home throughout the 6-week treatment period. The higher NBI-98854 dose in each age group (20 mg in children and 40 mg in adolescents) will be titrated in a blinded fashion (children will receive 10 mg and adolescents will receive 20 mg for the first week followed by their randomized dose of 20 mg or 40 mg for the remainder of the 6-week treatment period).

At any time, if the subject is unable to tolerate their current dose, the investigator may decrease the subject's dose. The investigator is allowed to reduce the subject's dose only 1 time during the study. Subjects who have had a dose reduction and are unable to tolerate their new dose will be discontinued from the study.

To maintain the study blind, subjects receiving placebo or the lower dose in each age group (10 mg in children and 20 mg in adolescents) who have a dose reduction will continue to receive their current dose, and subjects receiving the higher dose in each age group (20 mg in children and 40 mg in adolescents) will have their dose reduced to the lower dose, that is, 10 mg in children and 20 mg in adolescents.

During the treatment period, subjects will return to the study center at 2-week intervals (Weeks 2, 4, and 6) for study assessments and dispensing of study drug (Weeks 2 and 4 only). All subjects who have completed the 6-week treatment period will enter a 2-week follow-up period with a follow-up visit at Week 8 (subjects who terminate early will have Week 8 assessments conducted). Efficacy, safety, and study drug exposure will be assessed at scheduled times throughout the study.

## 2.2. Efficacy, Pharmacokinetic Blood Sampling, and Safety Assessments for All Subjects

Efficacy assessments for TS symptomatology include the Total Tic Score (TTS) from the Yale Global Tic Severity Scale (YGTSS) as the primary efficacy variable, and the YGTSS Impairment and Global Tic Severity scores, the Rush Video-Based Tic Rating Scale (RTRS), the Premonitory Urge for Tics Scale (PUTS), the Clinical Global Impression of Change-Tourette Syndrome (CGI-TS)-Improvement scale, and the Clinical Global Impression of Tics (CGI-Tics)-Severity scale as secondary efficacy variables. The YGTSS, RTRS, and PUTS will be administered at screening, on Day -1 (the day before dosing), during the double-blind treatment period (Weeks 2, 4, and 6), and at the follow-up visit (Week 8) or early termination. Note that all efficacy analyses for YGTSS-related variables are based on the certified site rater scores. Computer-generated tandem scores are also collected in the study database and will be presented in data listings only.

The CGI-Tics-Severity will be evaluated at screening and Day -1 (baseline), and then both the CGI-Tics-Severity and CGI-TS-Improvement scales will be administered at Weeks 2, 4, and 6 and at the follow-up visit (Week 8) or early termination.

Blood samples to evaluate plasma concentrations of NBI-98854 and the metabolite NBI-98782 will be collected during the treatment period (Weeks 2, 4, and 6) and at the follow-up visit (Week 8) or early termination.

Safety assessments performed at scheduled times throughout the study include adverse event (AE) monitoring, clinical laboratory tests (hematology, clinical chemistry, and urinalysis), serum prolactin, vital sign measurements, physical examinations, 12-lead ECG, Columbia Suicide Severity Rating Scale (C-SSRS, Children's Version), Extrapyramidal Symptom Rating Scale-Abbreviated (ESRS-A), Children's Depression Rating Scale - Revised (CDRS-R), Children's Yale-Brown Obsessive-Compulsive Scale (CY-BOCS), and Attention-Deficit Hyperactivity Disorder Rating Scale IV: Home Version (ADHD Rating Scale).

# 3. STATISTICAL ANALYSES

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 *post hoc* analyses will be clearly identified in the clinical study report.

## 3.1. General Statistical Procedures

Descriptive and inferential statistical methods will be used to summarize the data from this study. The term "descriptive statistics" refers to the number of subjects (n), mean, median, standard deviation (SD), standard error of the mean (SEM), minimum, and maximum for continuous variables; and refers to the number and/or percentage of subjects (or events) for categorical variables. The term "inferential statistics" refers to hypothesis tests which will be performed to assess differences between the NBI-98854 treatment groups and the placebo group for selected efficacy variables. All hypothesis tests will be tests of the null hypothesis of no difference between the treatment groups being compared versus the two-sided alternative hypothesis that there is a difference.

Descriptive statistics will be presented both by treatment group within each age group and by "pooled" treatment groups (placebo, NBI-98854 low dose, and NBI-98854 high dose) without regard to age group, unless specified otherwise. Only pooled treatment groups will be used in the calculation of inferential statistics. The pooled treatment groups are defined as follows:

- Placebo: Placebo (children and adolescents).
- NBI-98854 low dose: NBI-98854 10 mg (children) and NBI-98854 20 mg (adolescents).
- NBI-98854 high dose: NBI-98854 20 mg (children) and NBI-98854 40 mg (adolescents).

The primary efficacy endpoint in this study is the YGTSS Total Tic Score (TTS) mean change from baseline (Day -1) to Week 6 based on the certified site rater scores (note that all statistical summaries and analyses of the YGTSS data are based on the site rater scores). Inferential statistics will be calculated for this endpoint as well as for secondary efficacy endpoints, which include the YGTSS Impairment score mean change from baseline to Week 6, the YGTSS Global Tic Severity score mean change from baseline to Week 6, the RTRS total score mean change from baseline to Week 6, the PUTS total score mean change from baseline to Week 6, the CGI-TS-Improvement mean score at Week 6 (which is the key secondary efficacy endpoint), and the CGI-Tics-Severity score mean change from baseline to Week 6. A responder analysis will be performed also for the TTS and the CGI-TS-Improvement scale. Inferential analyses of Week 2 and Week 4 values of these endpoints will performed in addition to the Week 6 values (with the exception of the RTRS total score). Nominal (raw) two-sided p-values will be reported

for all hypothesis tests, although a procedure to control for multiplicity will be applied when interpreting the results of the analyses of the primary efficacy endpoint and the key secondary efficacy endpoint.

A comprehensive set of data listings including all randomized subjects will be provided. These listings will include both measured and derived values. Observations in data listings will be sorted by age group, subject, and timepoint (if applicable).

The derived variable "study day" is used in a number of calculations for data summaries and listings. This variable is calculated as the number of days after a subject's Day -1 ("baseline") visit. Study Day 1, then, is the day after the Day -1 visit.

Summary statistics will be presented using a significant figure rule: the median, minimum, and maximum will have the same number of significant figures as the data; the mean will have one more significant figure than the data being summarized; the SD and SEM will have the same number of significant figures as the mean; and the sample size (N) will be reported as an integer. This rule may be modified if warranted, based on practical considerations.

All analyses, tables, figures, and listings will be generated using SAS® software (version 9.3 or later), unless stated otherwise.

## 3.2. Sample Size Calculation

The protocol-specified sample size of 30 subjects per pooled treatment group is based on a power calculation for the TTS change from baseline using a two-sample t-test with a two-sided Type I error of 0.05. This sample size provides approximately 80% power to detect an effect size of 0.75 and approximately 90% power to detect an effect size of 0.85. Note that the effect size is defined as the mean difference between an NBI-98854 treatment group and the placebo group divided by the common SD (eg, a mean difference of 8 divided by an SD of 10 yields an effect size of 0.8).

Standard deviations reported in the literature for TS studies evaluating changes in the TTS (with or without placebo controls) have generally been in the range of 7.5 to 9.5, and in placebo-controlled studies, mean differences between active and placebo arms have been in the range of 5 to 9 (Jankovic et al., 2010; Yoo et al., 2013; ClinicalTrials.gov NCT01727700). The effect sizes of 0.75 and 0.85 mentioned in the above paragraph are representative of effect sizes seen in these published reports.

## 3.3. Pooling of Sites

Approximately 30 study centers are expected to participate in the study. With the exception of the summary of subject randomization by site, study sites will be pooled in all analyses, tables, and figures, since the majority of sites in this study are of a relatively small size, with fewer than five subjects expected to be enrolled. In addition, study site will not be included as a factor in any statistical models.

## 3.4. Handling of Early Termination Visit Data

An early termination (ET) visit occurs when a subject withdraws from the study prior to completing the scheduled Week 8 visit. An ET visit will be mapped to the next scheduled study visit if it occurs within 7 days prior to and 6 days after the expected study day of the next

scheduled visit (with the requirement that the scheduled visit prior to the ET visit was actually completed by the subject). An ET visit at Day 49 or later will be mapped to the Week 8 visit. Early termination visit data which are not mapped to a scheduled visit will be displayed in applicable by-subject data listings but not included in by-visit analyses and summaries.

Table 2 displays the allowable study day range for each scheduled visit for ET visit mapping purposes.

**Table 2: Allowable Study Day Range for Early Termination Visit Mapping**

| Scheduled Visit | Target Study Day | Time Interval (Study Day Range) |
|---|---|---|
| Week 2 | 14 | 7-20 |
| Week 4 | 28 | 21-34 |
| Week 6 | 42 | 35-48 |
| Week 8 | 56 | 49+ |

## 3.5. Handling of Missing Data

Missing values for outcome measures will not be replaced with imputed values except as noted in Table 2 for the ET visit mapping to a scheduled visit for summary and analysis purposes and as described for sensitivity analyses of efficacy endpoints which involve multiple or single imputation.

In the event a Day -1 (baseline) visit value is missing, the screening visit value may be used for the calculation of change (or shift) from baseline or in a statistical model.

Derived scale total scores (eg, the TTS), which are calculated as the sum of the scores of the individual scale items, will be set equal to missing if any of the individual scale item scores are missing.

Special rules for handling missing and incomplete dates are described below.

### 3.5.1. Missing and Incomplete Dates

Missing and incomplete ("partial") dates for AEs and concomitant medications will be imputed only for the purpose of estimating the time of the event or medication usage in relationship to study treatment; however, all 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; otherwise, the 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; otherwise, the 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 complete (not imputed) end date for the event, the start date will be set equal to 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 randomized study drug; otherwise, the 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; 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 complete (not imputed) medication stop date, the start date will be set equal to the stop date.

## 3.6. Coding Dictionary

Medical history and AEs will be coded using the Medical Dictionary for Regulatory Activities (MedDRA, version 12.0). Prior and concomitant medications will be coded using the World Health Organization (WHO) Drug Dictionary.

## 3.7. Analysis Sets

### 3.7.1. Definition of Analysis Sets

Three analysis sets will be defined for this study.

The safety analysis set will include all subjects who are randomized to a treatment group and dispensed study drug, with the following two exclusions: (a) subjects who withdraw from the study and return all previously dispensed study drug with all doses present, and (b) subjects who have no postbaseline safety data collected. The safety analysis set will be used for all summaries of safety data (eg, AEs and clinical laboratory data) and pharmacokinetic 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.

The primary efficacy analysis set is the intent-to-treat (ITT) analysis set, which will include all subjects in the safety analysis set who have a baseline and at least one post-randomization TTS value reported at a scheduled or mapped ET visit during the 6-week treatment period. The ITT analysis set will be used for summaries and analyses of efficacy data. Treatment assignment for all summaries and analyses using the ITT analysis set will be based on the randomization schedule.

The per-protocol (PP) analysis set will include all subjects in the ITT analysis set who have a TTS value for the Week 6 visit (including an ET visit mapped to Week 6), have no efficacy-

related important protocol deviations (IPDs; described in Section 3.9) and, for subjects in the NBI-98854 treatment groups, have a detectable plasma concentration of NBI-98854 at the Week 6 visit, which is a measure of treatment dosing compliance. The PP analysis set will be used for supportive summaries and analyses of the efficacy data. Treatment group assignment for the PP analysis set will follow the rules described above for the safety analysis set.

### 3.7.2. 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 within age group, for the pooled treatment groups, and overall (for all randomized subjects). The number and percentage of subjects excluded from each analysis set by reason for exclusion will also be provided.

### 3.7.3. 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 3.

**Table 3: Data Summaries by Analysis Set**

| Data Summary/Analysis | Analysis Set | | |
|---|---|---|---|
| | Safety | ITT | PP |
| Demographics | X | X | X |
| Baseline subject characteristics | X | X | X |
| Medical history | X | | |
| Study drug dosing (including dose reductions) | X | X | |
| Pharmacokinetic plasma concentration data | X | | |
| YGTSS | | X | X |
| RTRS | | X | X |
| PUTS | | X | X |
| CGI-Tics-Severity | | X | X |
| CGI-TS-Improvement | | X | X |
| Adverse events | X | | |
| C-SSRS | X | | |
| CY-BOCS | X | | |
| CDRS-R | X | | |
| ADHD rating scale IV | X | | |
| ESRS-A | X | | |
| Vital signs | X | | |
| Physical examination | X | | |
| Weight | X | | |
| ECG | X | | |
| Clinical laboratory data | X | | |
| Prior and Concomitant medications | X | | |

## 3.8. Subject Randomization and Disposition

The summary of subject randomization and disposition will display the number of subjects who were randomized into each treatment group, completed the 6-week treatment period (defined as having completed a scheduled Week 6 visit [note that this does not include ET visits mapped to Week 6]), and completed the study (defined as having a scheduled Week 8 visit [note that this does not include ET visits mapped to Week 8]). The number of subjects who did not complete the study will be displayed both overall and by reason 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.

An additional table will display the number of subjects in each treatment group who completed each study visit (screening through Week 8; including mapped ET visits). This table will be based on the safety analysis set.

These summaries will be presented separately for each age group and for the pooled treatment groups. Each summary table will include an "all subjects" column in addition to the columns for each treatment group.

## 3.9. Protocol Deviations

Significant protocol deviations, which are described in the study-specific Protocol Deviation Plan, will be entered into the study database and used to identify important protocol deviations (IPDs). Important protocol deviations 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. This committee will review a listing of all significant protocol deviations reported in the study database and determine which deviations are IPDs. The committee will also indicate which IPDs are efficacy-related deviations that exclude a subject from the PP analysis set. Important protocol deviations may 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.
- Significant deviations in YGTSS administration.

A summary of the number and percentage of subjects with IPDs by deviation category, treatment group within age group, and by pooled treatment groups, will be provided for all randomized subjects. All significant protocol deviations will be presented in a data listing.

## 3.10. Demographic Data and Baseline Characteristics

Demographic data (age, gender, race, and ethnicity) and baseline characteristics (age at TS diagnosis, baseline value of TTS, height, weight (in units of both pounds and kilograms), body mass index [BMI], and CYP2D6 genotype status) will be summarized using descriptive statistics and frequency counts. These data will be summarized by treatment group within age group and by pooled treatment groups. Each summary table will include an "all subjects" column.

## 3.11. Medical History

Medical history will be summarized in frequency tables (number and percentage of subjects) by MedDRA System Organ Class (SOC) and Preferred Term (PT) by treatment group within age group and by pooled treatment groups.

## 3.12. Study Drug Dosing and Compliance

Study drug dosing compliance is calculated as the ratio of the estimated number of doses taken by a subject to the expected number of doses that should have been taken (given the length of time the subject participated in the study treatment period) and is expressed as a percentage. A subject is classified as being compliant with study drug administration if their calculated value is 80% or greater. The number and percentage of subjects in each treatment group (both by treatment group within age group and by pooled treatment groups) who are compliant will be presented for Weeks 2, 4, and 6, and for the full treatment period.

The number of doses taken by each subject between two consecutive visits will be estimated by calculating the difference between the number of capsules dispensed at the first visit and the number of capsules returned at the subsequent visit (note that the number of doses taken is equal to the number of capsules taken, as each dose is 1 capsule of study drug per study protocol). If an ET visit occurs between scheduled visits, the ET visit data will be used in the calculation. If a subject fails to return the study drug kit dispensed at the previous visit, it will be assumed that the number of doses taken by the subject is equal to the number of days between visits (inclusive of the day of the second visit; not to exceed the number of doses dispensed), as subjects are expected to take one dose per day. If a subject is lost to follow-up between scheduled visits, their dosing data will be limited to the last visit at which they were present.

The estimated number of doses taken by a subject across all visits during the treatment period is calculated as the sum of the number of doses taken between each consecutive pair of visits during the treatment period.

The expected number of doses that should have been taken by a subject between or across visits is based on the expected number of days the subject should have dosed with study drug during the interval of interest, which in turn is based on the duration of the subject's study participation during the interval (which may be shorter, for example, if the subject withdraws from the study).

The expected number of days of dosing (and hence, expected number of doses that should have been taken) is calculated as the difference between the visit date of the second of the two visits and the study drug kit dispense date at the first of the two visits. Note that the expected number of days of dosing cannot exceed the total number of doses actually dispensed (each study drug kit in this study contains 17 doses of study drug). If an ET visit occurs between scheduled visits, the ET visit date will be used in the calculation. If a subject is lost to follow-up between scheduled visits, they will be excluded from the compliance calculations for all visits occurring after the last visit at which they were present. If a subject has a missed visit but has previously been dispensed an extra study drug kit that would have been dispensed at the missed visit, they will be excluded from the compliance calculations for all visits beginning with the missed visit, but will be included in the overall treatment period calculations described in the next paragraph.

The dosing compliance calculation across all visits during the study treatment period will follow the logic described above, with the expected number of days of dosing being equal to the difference between the Week 6 (or ET) visit date and the study drug kit dispense date for the Day -1 visit. The estimated number of doses taken by a subject across all visits during the study is calculated as the sum of the number of doses taken between each consecutive pair of visits during the treatment period. If a subject is lost to follow-up prior to Week 6, their dosing compliance will be based on the last study visit at which they were present.

## 3.13. Dose Reduction Summary

The dose reduction summaries will present the number and percentage of subjects with a dose reduction for each treatment group within each age group and for each of the pooled treatment groups at Weeks 2 and 4. Dose reductions at unscheduled visits prior to Week 4 will be included in the summary for the next scheduled visit (ie, Week 2 or Week 4). Dose reductions at unscheduled visits after Week 4 will be summarized as "After Week 4" in the table. The table will include the total number and percentage of subjects with a dose reduction at any time during the treatment period. Note that these summaries include all reports of dose reductions, including

those for the placebo and NBI-98854 low dose treatment groups (which are not true dose reductions).

## 3.14. Pharmacokinetic Data

The plasma concentrations of NBI-98854 and its active metabolite NBI-98782 will be summarized with descriptive statistics by visit (Weeks 2, 4, 6, and 8) and the most recent NBI-98854 dose received by a subject prior to that visit (this is due to the possibility of a subject having a dose reduction from 20 mg to 10 mg or from 40 mg to 20 mg). Note that the dose level at the Week 8 visit will reflect the last dose the subject received during the study treatment period (ie, the dose level at the subject's Week 6 visit). These summary tables will be presented separately for each age group only. There will not be a pooled treatment groups summary.

These summaries will also be generated separately for CYP2D6 poor metabolizers vs. non-poor metabolizers within each age group.

Concentrations below the lower limit of quantification (BLQ) 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 (%).

Plasma concentrations of each PK analyte will be summarized with box plots by NBI-98854 dose level at each visit.

## 3.15. Efficacy Data

### 3.15.1. Efficacy Endpoints

The efficacy endpoints for this study are described in detail in the following sections of this SAP. An overall summary of the efficacy endpoints is presented below in Table 4.

**Table 4: Efficacy Endpoints and Classification**

| Efficacy Endpoint | Endpoint Classification |
|---|---|
| YGTSS Total Tic Score (TTS) mean change from baseline to Week 6 | Primary |
| CGI-TS-Improvement mean score at Week 6 | Key secondary |
| Percentage of subjects classified as YGTSS TTS responders at Week 6 | Secondary |
| YGTSS Impairment score mean change from baseline to Week 6 | Secondary |
| YGTSS Global Tic Severity score mean change from baseline to Week 6 | Secondary |
| Rush Video-based Tic Rating Scale (RTRS) total score mean change from baseline to Week 6 | Secondary |
| Premonitory Urge for Tics Scale (PUTS) total score mean change from baseline to Week 6 | Secondary |
| CGI-Tics-Severity score mean change from baseline to Week 6 | Secondary |
| Percentage of subjects classified as CGI-TS-Improvement responders at Week 6 | Secondary |
| YGTSS Total Tic Score (TTS) mean change from baseline to mean of Week 4 and Week 6 values | Exploratory |
| YGTSS Total Tic Score (TTS) mean change from baseline to Weeks 2 and 4 | Exploratory |
| CGI-TS-Improvement mean score at Weeks 2 and 4 | Exploratory |
| Percentage of subjects classified as YGTSS TTS responders at Weeks 2 and 4 | Exploratory |
| YGTSS Impairment score mean change from baseline to Weeks 2 and 4 | Exploratory |
| YGTSS Global Tic Severity score mean change from baseline to Weeks 2 and 4 | Exploratory |
| Premonitory Urge for Tics Scale (PUTS) total score mean change from baseline to Weeks 2 and 4 | Exploratory |
| CGI-Tics-Severity score mean change from baseline to Weeks 2 and 4 | Exploratory |
| Percentage of subjects classified as CGI-TS-Improvement responders at Weeks 2 and 4 | Exploratory |

### 3.15.2. Primary Efficacy Endpoint

The primary efficacy endpoint for this study is the mean change from baseline (Day -1) to Week 6 in the YGTSS TTS (Total Tic Score) based on the certified site rater scores. Mean changes from baseline to Weeks 2 and 4 are exploratory efficacy endpoints.

The TTS is defined as the sum of the YGTSS motor tic severity score and phonic (vocal) tic severity score. The motor tic severity score is calculated as the sum of the scores for the 5 motor tic items (number, frequency, intensity, complexity, and interference). The score for each item can range from 0 to 5, for a maximum total score of 25. The vocal (phonic) tic severity score is calculated similarly. If any one of the 5 items for the motor or vocal tic severity score is not scored (ie, has a missing value), the associated severity score will be set equal to missing. If any of these items has a missing value at a given subject visit, the TTS value for the subject visit will also be set equal to missing. The TTS value can range from 0 to 50.

Descriptive statistics will be presented by treatment group for the TTS observed values (ie, the raw data) and changes from baseline at each visit from screening through Week 8 (note that changes from baseline will be presented only for postbaseline visits in all such summaries described in this SAP). Descriptive statistics will be presented for both the ITT and the PP analysis sets. These summaries will be presented by treatment group within age group and by pooled treatment groups.

The primary analysis of the TTS will be a mixed-effect model repeated measures (MMRM) analysis, which includes the changes from baseline to Weeks 2, 4, and 6, and is based on the ITT analysis set. The model will include the baseline TTS as a covariate, and age group, pooled treatment groups, visit (Week 2, 4, or 6), pooled treatment groups by visit interaction, and baseline covariate by visit interaction as fixed effects. Subject will be included as a random effect. Study site will not be included in the model, as there is a large number of sites (approximately 30), with most sites enrolling a small number of subjects.

Treatment group comparisons for each NBI-98854 treatment group vs. placebo at each visit (ie, Weeks 2, 4, and 6) will be performed by constructing linear contrasts (or equivalent programming code) for differences between treatment group least-squares (LS) means. Nominal two-sided p-values for testing the significance of these differences and associated 95% confidence intervals will be reported in summary tables; however, interpretation of the p-values for the two active pooled treatment group comparisons to placebo at Week 6 will be based on a procedure which controls for multiplicity that is described below in Section 3.15.4.

A supportive analysis using the MMRM model will be performed with the PP analysis set.

The MMRM analysis will be implemented with the PROC MIXED procedure of SAS®, using the restricted maximum likelihood method, an unstructured within-subject covariance matrix, and denominator degrees of freedom from the Kenward-Roger method. In the event that convergence is not obtained with the unstructured covariance matrix, a Toeplitz covariance structure will be used.

An additional supportive analysis of the TTS change from baseline to Week 6 will be performed using an analysis of covariance (ANCOVA) model. The ANCOVA model will include the baseline TTS as a covariate and age group and pooled treatment groups as fixed effects. This analysis will be performed using both the ITT and PP analysis sets.

An exploratory ANCOVA will be performed for the TTS change from baseline to the mean of the Week 4 and Week 6 TTS values for each subject. The ANCOVA model will be identical to the model described in the previous paragraph with the exception of the response variable. Both ITT and PP analyses will be performed.

Mean (±SEM) values of the TTS at each visit (Day -1 through Week 8) will be summarized in line graphs by treatment group. Similar graphs will be presented for the changes from baseline. These graphs will be based on the ITT analysis set and will be presented by treatment group within age group and by pooled treatment groups.

The LS means (±SEM) from the MMRM analysis will be summarized in line graphs by pooled treatment groups for both the ITT and PP analysis sets.

An additional graph will be presented for the TTS changes from baseline to Week 6. This will be a display of the empirical distribution function for each treatment group (both within age group and pooled), and will be based on the ITT analysis set.

### 3.15.3. Key Secondary Efficacy Endpoint

The key secondary efficacy endpoint for this study is the CGI-TS-Improvement mean score at Week 6. Mean scores at Weeks 2 and 4 are exploratory efficacy endpoints.

Each of the CGI-TS-Improvement response categories will be assigned a numerical score as follows:

Very much improved = 1
Much improved = 2
Minimally improved = 3
Not changed = 4
Minimally worse = 5
Much worse = 6
Very much worse = 7

Descriptive statistics (including both frequency counts [using response categories] and continuous variable statistics [using numerical scores]) will be presented by treatment group for the CGI-TS-Improvement data at each visit from Week 2 through Week 8. Descriptive statistics will be presented for both the ITT and the PP analysis sets. These summaries will be presented by treatment group within age group and by pooled treatment groups.

The primary analysis of the CGI-TS-Improvement numerical scores will be an MMRM analysis which includes the scores at Weeks 2, 4, and 6 and is based on the ITT analysis set. The MMRM model will be similar to the model described above for the TTS analysis, with the exception that the covariate in the model will be the baseline value of the CGI-Tics-Severity numerical score. As indicated above for the TTS, nominal two-sided p-values for testing the significance of pooled treatment group differences and associated 95% confidence intervals will be reported in summary tables; however, interpretation of the p-values for the two active pooled treatment groups comparisons to placebo at Week 6 will be based on a procedure which controls for multiplicity that is described below in Section 3.15.4.

A supportive analysis of the CGI-TS-Improvement numerical scores will be performed using the MMRM analysis based on the PP analysis set.

An additional supportive analysis of the CGI-TS-Improvement numerical scores at Week 6 will be performed using an ANCOVA model. The ANCOVA model will include age group and pooled treatment groups as fixed effects and the baseline value of the CGI-Tics-Severity numerical score as a covariate. This analysis will be performed using both the ITT and PP analysis sets.

Mean (±SEM) values of the CGI-TS-Improvement numerical scores at each visit will be summarized in line graphs by treatment group. These graphs will be based on the ITT analysis set and will be presented by treatment group within age group and by pooled treatment groups.

The LS means (±SEM) from the MMRM analysis of the CGI-TS-Improvement numerical scores at Weeks 2 through 6 will be summarized in line graphs by pooled treatment groups for both the ITT and PP analysis sets.

### 3.15.4. Procedure to Control for Multiplicity

A fixed-sequence testing procedure will be followed for the primary and key secondary efficacy endpoint analyses in order to control for multiple comparisons (ie, comparing each of the two NBI-98854 pooled treatment groups to placebo for each of the two endpoints). The fixed-sequence testing procedure will consist of performing the hypothesis tests in the following pre-specified order:

- Week 6 TTS mean change from baseline: NBI-98854 high dose treatment group vs. placebo treatment group
- Week 6 TTS mean change from baseline: NBI-98854 low dose treatment group vs. placebo treatment group
- Week 6 CGI-TS-Improvement mean score: NBI-98854 high dose treatment group vs. placebo treatment group
- Week 6 CGI-TS-Improvement mean score: NBI-98854 low dose treatment group vs. placebo treatment group

In order for a test result in the above list to be considered statistically significant, all of the test results higher in the list must be significant at the 0.05 level of significance.

### 3.15.5. Sensitivity Analyses for Primary Efficacy Endpoint

An underlying assumption of the MMRM analysis is that missing data are considered to be missing at random (MAR). In addition to assessing the number of subjects who prematurely discontinue from the study prior to Week 6 along with the reasons for premature study discontinuation, three sensitivity analyses of the TTS changes from baseline will be performed to assess the impact of deviations from the assumption that the missing TTS data are MAR. These analyses, described below, are based on the ITT analysis set.

Tipping Point - Sensitivity Analysis #1

*Step 1 of Sensitivity Analysis #1 (Imputation of Missing Data)*

The initial step will be to impute missing TTS values at Weeks 2, 4 and 6 using the SAS® procedure PROC MI. The data will first be examined to determine if the pattern of missing data is monotone or non-monotone across visits from Day -1 (baseline) through Week 6. If the pattern of missing data is monotone (which will be the case if the missingness is purely due to

dropouts) the regression method of PROC MI will be used to impute the missing data. If the pattern of missing data is non-monotone, the Markov chain Monte Carlo (MCMC) method of PROC MI will be used to create a monotone data set, using default system values for the parameters CHAIN, NBITER and NITER (along with IMPUTE=MONOTONE). This data set can then be analyzed using the regression method of PROC MI. In either case, 100 imputed data sets will be generated using a random number generator seed value of 122706. The SAS® statements (or equivalent) provided below will be used for the imputation of missing data. Note that "Treat" represents pooled treatment groups (placebo, NBI-98854 low dose, or NBI-98854 high dose), "Age" is age group, "Baseline" is the Day -1 TTS value, and Week2, Week4, and Week6 are the TTS values at Weeks 2, 4, and 6, respectively.

```
PROC MI NIMPUTE=100 SEED=122706 MINIMUM=0 MAXIMUM=50
OUT=Imputed;
CLASS Treat Age;
MONOTONE REGRESSION;
VAR Treat Age Baseline Week2 Week4 Week6;
```

*Step 2 of Sensitivity Analysis #1 (Creation of Additional Data Sets with NBI-98854 Treatment Group Imputed TTS Scores Made Worse in a Step-Wise Fashion)*

The sensitivity analysis itself will involve the creation of additional data sets based on the fully imputed data sets created in Step 1, with a step-wise "worsening-score" algorithm applied to the NBI-98854 treatment groups, but not to the placebo treatment group. Since the primary endpoint is the TTS mean change from baseline to Week 6, the scores at Weeks 2 and 4, which were used for missing data imputation purposes, will not be used in any subsequent calculations or analyses. Therefore, the worsening-score algorithm will be applied only to the NBI-98854 treatment group imputed Week 6 TTS values (also note that the method of analysis for the imputed data sets described below in Step 3 is an ANCOVA of the Week 6 TTS changes from baseline).

The worsening-score algorithm will involve increasing (or "worsening") the imputed Week 6 TTS values in the two NBI-98854 treatment groups by a percentage of the imputed value in a pre-specified, step-wise fashion. Imputed Week 6 TTS values in the placebo treatment group will not be made worse in this algorithm and will remain at the initially imputed values.

The initial percentage value to be applied is 10%. This will be increased in fixed increments of 10% until the p-value for the comparison of the NBI-98854 high dose treatment group to the placebo treatment group (obtained as described below in Step 3 from PROC MIANALYZE applied to the results of 100 ANCOVAs for each imputed data set) exceeds the 0.05 level of significance. Note that the maximum possible value for the TTS is 50; therefore an imputed Week 6 TTS value adjusted according to this algorithm cannot exceed the value of 50.

This sensitivity analysis represents a tipping point analysis based on "delta adjustments," which is a commonly used approach to assess the impact of missing data in clinical trials (O'Kelly and Ratitch, 2014).

*Step 3 of Sensitivity Analysis #1*

The TTS changes from baseline to Week 6 will be calculated for each of the imputed data sets from the original PROC MI output and for the additional sets of data created according to the

worsened-score algorithm described above in Step 2. The imputed data sets will be analyzed using the ANCOVA model specified above in Section 3.15.2, using either PROC MIXED or PROC GLM.

PROC MIANALYZE will then be used to combine results from the 100 ANCOVAs for the imputed data sets to provide the following statistics:

- LS mean for each treatment group
- LS mean difference between each NBI-98854 treatment group and the placebo treatment group, along with the corresponding two-sided 95% confidence intervals and p-values

Jump to Reference (J2R) - Sensitivity Analysis #2

The J2R method is based on the concept that missing values for subjects in the NBI-98854 treatment groups who drop out prior to the Week 6 visit will tend to be similar to values for subjects in the reference (placebo) group who have similar baseline characteristics. This is plausible under the assumption that treatment with NBI-98854 offers symptomatic, and not disease-modifying treatment, and as such, subjects who stop taking active study drug will no longer benefit from its therapeutic effect, but will still be subject to any placebo (or study participation related) effects.

The J2R analysis will be implemented using the SAS® procedures PROC MI and PROC MIANALYZE. If the pattern of missing data across study visits during the treatment period is non-monotone, a monotone data set will be created using PROC MI with the MCMC method, a seed value of 122706, and system default values for the parameters CHAIN, NBITER, and NITER (along with IMPUTE=MONOTONE). The number of imputed data sets created in this step will be 100.

The actual J2R imputation will follow the sequential model approach described in Chapter 7.4.3 of O'Kelly and Ratitch (2014). With this approach, the placebo treatment group data are used to impute missing values for the NBI-98854 treatment groups, one visit at a time in a sequential fashion. At each visit imputation step, PROC MI with the same seed specified in the preceding paragraph (and 100 imputed data sets) will be implemented, using the monotone regression method with age group and the baseline TTS as explanatory variables. Note that missing data values in the placebo treatment group are imputed under the MAR assumption (using PROC MI with the monotone regression method), prior to imputing missing data values in the NBI-98854 treatment groups.

The final step of this analysis will consist of performing an ANCOVA analysis of the TTS changes from baseline to Week 6 using the imputed data sets and then combining the results of these analyses using PROC MIANALYZE.

Baseline Observation Carried Forward (BOCF) – Sensitivity Analysis #3

The use of a BOCF analysis as a sensitivity analysis is described in a publication authored by Dr. Lisa LaVange of the FDA/CDER Office of Biostatistics (LaVange, 2014). This analysis will be implemented by imputing missing Week 6 TTS values in all three pooled treatment groups with the baseline values (ie, a subject's missing Week 6 value will be set equal to their baseline value). The TTS changes from baseline to Week 6, including BOCF-imputed missing values for Week 6, will then be analyzed using the ANCOVA model, using either PROC MIXED or PROC

GLM. As this is a single-imputation method, the use of PROC MI and PROC MIANALYZE is not required.

### 3.15.6. Sensitivity Analysis for Key Secondary Efficacy Endpoint

Sensitivity analyses will also be performed for the key secondary efficacy endpoint, the CGI-TS-Improvement mean score at Week 6, using the ITT analysis set. Tipping point and J2R analyses similar to those described above for the TTS will be performed, using a random number generator seed value of 339614 for the PROC MI part of the analyses.

A third sensitivity analysis will be performed using a method similar in principle to the BOCF analysis described for the TTS. For this sensitivity analysis, subjects whose CGI-TS-Improvement score at Week 6 is missing will have their missing score imputed with a value of 4, which reflects the CGI-TS-Improvement response category of "not changed". This imputed data set will then be analyzed using an ANCOVA model.

### 3.15.7. Secondary Efficacy Endpoints

#### 3.15.7.1. YGTSS TTS Responder Analysis

A TTS responder is defined, on a per-visit basis, as a subject whose TTS value is reduced by at least 30% from baseline at the specified postbaseline visit. The percentage of subjects classified as TTS responders will be summarized for each treatment group within age group and for the pooled treatment groups at Weeks 2, 4, 6, and 8 for both the ITT and PP analysis sets.

The percentage of subjects classified as TTS responders at Week 6 is a secondary efficacy endpoint. An analysis comparing each pooled NBI-98854 treatment group to the pooled placebo treatment group will be performed for this endpoint with the ITT analysis set using the Cochran-Mantel-Haenszel (CMH) procedure, with age group as a stratification variable. A separate CMH analysis will be performed for each active treatment group comparison to placebo (NBI-98854 high dose vs. placebo and NBI-98854 low dose vs. placebo).

A supportive analysis of the percentage of subjects classified as TTS responders at Week 6 will be performed using the PP analysis set.

The percentages of subjects classified as TTS responders at Weeks 2 and 4 are exploratory endpoints. These endpoints will be analyzed using the CMH procedure at each of these visits with both the ITT and PP analysis sets.

Bar graphs displaying the percentage of subjects classified as TTS responders will be presented by treatment group for each study visit from Week 2 through Week 8 for the ITT and PP analysis sets.

#### 3.15.7.2. YGTSS Impairment Score

The YGTSS Impairment score can range in value from 0 to 50. The YGTSS Impairment score mean change from baseline (Day -1) to Week 6 is a secondary efficacy endpoint. Mean changes from baseline to Weeks 2 and 4 are exploratory efficacy endpoints.

Descriptive statistics will be presented by treatment group within age group and by pooled treatment groups for the YGTSS Impairment score observed values (ie, the raw data) and

changes from baseline at each visit from screening through Week 8. Descriptive statistics will be presented for both the ITT and the PP analysis sets.

The analysis of the YGTSS Impairment score changes from baseline at each visit from Week 2 through Week 6 will be performed using a MMRM model similar to the model described in Section 3.15.2 for the TTS change from baseline, based on the ITT analysis set. The covariate for this MMRM model will be the YGTSS Impairment score at baseline. A supportive analysis of the YGTSS Impairment score changes from baseline will be performed using the PP analysis.

Mean (±SEM) values of the YGTSS Impairment score at each visit (Day -1 through Week 8) will be summarized in line graphs by treatment group within age group and by pooled treatment groups. Similar graphs will be presented for the changes from baseline. These graphs will be based on the ITT analysis set.

The LS means (±SEM) from the MMRM analysis of the YGTSS Impairment score changes from baseline will be summarized in line graphs by pooled treatment group for both the ITT and PP analysis sets.

#### 3.15.7.3. YGTSS Global Tic Severity Score

The YGTSS Global Tic Severity score is the sum of the TTS and the YGTSS Impairment score. The YGTSS Global Tic Severity score at a given subject visit will be set equal to missing if either of the TTS or Impairment scores are missing. The YGTSS Global Tic Severity score value can range from 0 to 100.

The YGTSS Global Tic Severity score mean change from baseline (Day -1) to Week 6 is a secondary efficacy endpoint. Mean changes from baseline to Weeks 2 and 4 are exploratory efficacy endpoints.

Descriptive statistics will be presented by treatment group within age group and by pooled treatment groups for the YGTSS Global Tic Severity score observed values (ie, the raw data) and changes from baseline at each visit from screening through Week 8. Descriptive statistics will be presented for both the ITT and the PP analysis sets.

An analysis of the YGTSS Global Tic Severity score changes from baseline to Weeks 2, 4 and 6 will be performed using a MMRM model similar to the model described in Section 3.15.2 for the TTS change from baseline, based on the ITT analysis set. The covariate for this MMRM model will be the YGTSS Global Tic Severity score at baseline. A supportive analysis of the YGTSS Global Tic Severity scores will be performed using the PP analysis set.

Mean (±SEM) values of the YGTSS Global Tic Severity score at each visit (Day -1 through Week 8) will be summarized in line graphs by treatment group within age group and by pooled treatment groups. Similar graphs will be presented for the changes from baseline. These graphs will be based on the ITT analysis set.

The LS means (±SEM) from the MMRM analysis of the YGTSS Global Tic Severity score changes from baseline will be summarized in line graphs by pooled treatment group for both the ITT and PP analysis sets.

### 3.15.7.4. Rush Video-based Tic Rating Scale (RTRS)

The modified RTRS used in this study includes short video recordings to measure 5 tic variables: number of body areas affected, frequency of motor and phonic tics, and severity of motor and phonic tics. The RTRS videos for Day -1 and Week 6 (or final on-treatment visit for subjects who discontinue from the study prior to their scheduled Week 6 visit) will be reviewed and scored by a pair of blinded central raters using a consensus scoring process. The raters will be blinded to both treatment and visit sequence. The summaries and analyses of the RTRS data, then, will be limited to the scores recorded for these two visits for each subject.

The RTRS total score is calculated as the sum of the 5 domain scores (number of body areas affected, motor tic frequency, phonic tic frequency, severity of motor tics, and severity of phonic tics). The score for each domain can range from 0 to 4, for a maximum possible total score of 20. If any one of the 5 domains is not scored (ie, has a missing value), the total score will be set equal to missing.

The RTRS total score mean change from baseline (Day -1) to Week 6 (or final on-treatment visit if prior to Week 6) is a secondary efficacy endpoint.

Descriptive statistics will be presented by treatment group within age group and by pooled treatment groups for the RTRS total score observed values (ie, the raw data) at each visit and changes from baseline. Descriptive statistics will be presented for both the ITT and the PP analysis sets.

An analysis of the RTRS total score changes from baseline to Week 6 (or final on-treatment visit if prior to Week 6) will be performed using an ANCOVA model similar to the model described in Section 3.15.2 for the TTS change from baseline, based on the ITT analysis set. The covariate for this analysis is the RTRS total score at baseline. A supportive analysis of the RTRS total scores will be performed using the PP analysis set.

Mean (±SEM) values of the RTRS total score for each treatment group within age group and for pooled treatment groups at each visit will be summarized in bar graphs. A similar graph will be presented for the changes from baseline. These graphs will be based on the ITT analysis set.

The LS means (±SEM) from the ANCOVA analysis of the RTRS total score changes from baseline to Week 6 (or final on-treatment visit if prior to Week 6) will be summarized in a bar graph by pooled treatment group for both the ITT and PP analysis sets.

### 3.15.7.5. Premonitory Urge for Tics Scale (PUTS)

The PUTS consists of 9 items, each of which is scored on a 4-point scale (1=not at all true, 2=a little true, 3=pretty much true, 4=very much true). The PUTS total score is calculated as the sum of the scores for the 9 items. The maximum possible total score is 36. If any one of the 9 items is not scored (ie, has a missing value), the PUTS total score will be set equal to missing.

The PUTS total score mean change from baseline (Day -1) to Week 6 is a secondary efficacy endpoint. Mean changes from baseline to Weeks 2 and 4 are exploratory efficacy endpoints.

Descriptive statistics will be presented by treatment group within age group and by pooled treatment groups for the PUTS total score observed values (ie, the raw data) and changes from baseline at each visit from screening through Week 8. Descriptive statistics will be presented for both the ITT and the PP analysis sets.

An analysis of the PUTS total score changes from baseline to Weeks 2, 4, and 6 will be performed using a MMRM model similar to the model described in Section 3.15.2 for the TTS change from baseline, based on the ITT analysis set. The covariate for this MMRM analysis is the PUTS total score at baseline. A supportive analysis of the PUTS total scores will be performed using the PP analysis set.

Mean (±SEM) values of the PUTS total score at each visit (Day -1 through Week 8) will be summarized in line graphs by treatment group within age group and by pooled treatment groups. Similar graphs will be presented for the changes from baseline. These graphs will be based on the ITT analysis set.

The LS means (±SEM) from the MMRM analysis of the PUTS total score changes from baseline to Weeks 2, 4 and 6 will be summarized in line graphs by pooled treatment group for both the ITT and PP analysis sets.

#### 3.15.7.6. CGI-Tics-Severity

Each of the CGI-Tics-Severity response categories will be assigned a numerical score as follows:

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

The CGI-Tics-Severity numerical score mean change from baseline (Day -1) to Week 6 is a secondary efficacy endpoint. Mean changes from baseline to Weeks 2 and 4 are exploratory efficacy endpoints.

Descriptive statistics (including both frequency counts [using response categories] and continuous variable statistics [using numerical scores]) will be presented by treatment group within age group and by pooled treatment groups for the CGI-Tics-Severity data at each visit from screening through Week 8. Changes from baseline for the numerical scores will be summarized also. Descriptive statistics will be presented for both the ITT and the PP analysis sets.

An analysis of the CGI-Tics-Severity numerical score changes from baseline to Weeks 2, 4, and 6 will be performed using a MMRM model similar to the model described in Section 3.15.2 for the TTS changes from baseline, based on the ITT analysis set. The covariate for this MMRM analysis is the CGI-Tics-Severity numerical score at baseline. A supportive analysis of the CGI-Tics-Severity numerical scores will be performed using the PP analysis set.

Mean (±SEM) values of the CGI-Tics-Severity numerical scores at each visit (Day -1 through Week 8) will be summarized in line graphs by treatment group within age group and by pooled treatment groups. Similar graphs will be presented for the changes from baseline. These graphs will be based on the ITT analysis set.

The LS means (±SEM) from the MMRM analysis of the CGI-Tic-Severity numerical score changes from baseline to Weeks 2, 4, and 6 will be summarized in line graphs by pooled treatment group for both the ITT and PP analysis sets.

#### 3.15.7.7. CGI-TS-Improvement Responder Analysis

A subject is classified as a CGI-TS-Improvement responder at a given visit if their CGI-TS-Improvement score is either a "1" ("very much improved") or a "2" ("much improved") at the visit.

The number and percentage of CGI-TS-Improvement responders will be summarized by treatment group within age group and by pooled treatment groups at Weeks 2, 4, 6, and 8. The CGI-TS-Improvement responder summary statistics will be presented for both the ITT and the PP analysis sets.

The percentage of subjects classified as CGI-TS-Improvement responders at Week 6 is a secondary efficacy endpoint. An analysis comparing each pooled NBI-98854 treatment group to the pooled placebo treatment group will be performed for this endpoint with the ITT analysis set using the CMH procedure, with age group as a stratification variable. A separate CMH analysis will be performed for each active treatment group comparison to placebo (NBI-98854 high dose vs. placebo and NBI-98854 low dose vs. placebo).

A supportive analysis of the percentage of subjects classified as CGI-TS-Improvement responders at Week 6 will be performed using the PP analysis set.

The percentages of subjects classified as CGI-TS-Improvement responders at Weeks 2 and 4 are exploratory endpoints. These endpoints will be analyzed using the CMH procedure at each of these visits with both the ITT and PP analysis sets.

Bar graphs displaying the percentage of subjects classified as CGI-TS-Improvement responders at Weeks 2 through 8 will be presented by treatment group within age group and by pooled treatment groups for the ITT and the PP analysis sets.

An additional analysis of the CGI-TS-Improvement responder rates at Week 6 will be performed which incorporates an imputed value of "non-responder" for subjects in the ITT analysis set with missing values at the Week 6 timepoint. This analysis can be considered as a sensitivity analysis with regard to the potential issue of missing data in interpreting the Week 6 analysis results for the CGI-TS-Improvement data.

## 3.16. Safety Data

### 3.16.1. Adverse Events

A treatment-emergent adverse event (TEAE) is any reaction, side effect, or other untoward event, regardless of relationship to study drug, that is newly reported or considered a worsening or change in nature, severity, or frequency of conditions present at the start of the study which occurs any time after first dose of study drug.

TEAEs, categorized by MedDRA (Version 12.0) system organ class (SOC) and preferred term (PT), will be summarized by treatment group within age group and by pooled treatment groups in frequency tables. Each table will also include a separate column for all NBI-98854 treated subjects (ie, excluding placebo subjects). Unless stated otherwise, the frequency tables will

include the number of events reported, and the number and percentage of unique subjects experiencing each event one or more times during the study. SOCs will be displayed in order of decreasing frequency (number of subjects with TEAEs in the SOC based on the "all NBI-98854 treated subjects" column), and within each SOC, PTs will be displayed in a similar fashion in order of decreasing frequency of subjects with TEAEs in the PT based on the "all NBI-98854 treated subjects" column.

Two versions of the primary TEAE frequency table will be presented: (a) the first version as described above which includes both SOC and PT, with each sorted in decreasing frequency based on the "all NBI-98854 treated subjects" column, and (b) a second version which does not include SOCs but does include PTs sorted in decreasing frequency based on the "all NBI-98854 treated subjects" column.

Frequency tables will also be presented including only TEAEs considered to be possibly or definitely related to study drug and, in addition, categorizing TEAEs according to the maximum intensity reported for a given subject. These tables will include both SOC and PT.

AEs with an onset date during screening (prior to study drug dosing) will be presented only in a data listing.

#### 3.16.1.1. Adverse Event Overall Summary

An overall summary table will be provided for each treatment group summary (by treatment group within age group and by pooled treatment groups) which summarizes the number and percentage of subjects with any TEAE, any treatment-related TEAE (ie, possibly or definitely related), any severe TEAE, any TEAE leading to dose reduction, any TEAE leading to study discontinuation, any serious TEAE, and any TEAE resulting in death. These tables will include an "all NBI-98854 treated subjects" column.

#### 3.16.1.2. Adverse Events Resulting in Premature Study Discontinuation

Summary tables of TEAEs resulting in premature study discontinuation will be presented by treatment group within age group and by pooled treatment groups. These summary tables will display the PTs for the TEAEs resulting in study discontinuation, with the PTs sorted in order of decreasing frequency in the "all NBI-98854 treated subjects" column. The first line of each table will display the overall frequency for each treatment group column.

A listing of TEAEs resulting in premature study discontinuation will be provided which includes age group, subject, treatment group, last treatment received prior to the onset time of the TEAE(s) leading to discontinuation, study day of the discontinuation, and other information from the AE eCRF. Note that "last treatment received prior to the onset time of the TEAE[s] leading to discontinuation" reflects the reduced dose for subjects with a dose reduction prior to study discontinuation.

#### 3.16.1.3. Adverse Events Resulting in Study Drug Dose Reductions

Summary tables of TEAEs resulting in study drug dose reductions will be presented by treatment group within age group and by pooled treatment groups. Note that subjects randomized to the placebo and NBI-98854 low dose treatment groups who have a reported dose reduction actually remain on the randomized dose, while subjects randomized to the NBI-98854 high dose treatment groups will have an actual dose reduction.

These summary tables will display the PTs for the TEAEs resulting in a dose reduction, with the PTs sorted in order of decreasing frequency in the "all NBI-98854 treated subjects" column. The first line of each table will display the overall frequency for each treatment group column.

A listing of TEAEs resulting in study drug dose reductions will be provided. This listing will include age group, subject, treatment group, study day of the dose reduction, and additional information from the AE eCRF.

#### 3.16.1.4. Deaths and Other Serious Adverse Events

The frequency of serious TEAEs will be summarized in tables using the table format described above for the summary of TEAEs leading to study discontinuation. Deaths will be presented in a listing only.

Listings of all SAEs and deaths will be provided which include age group, subject, treatment group, last treatment received prior to the onset time of the SAE or TEAE leading to death, study day of the SAE or death, and additional information from the AE eCRF.

### 3.16.2. Clinical Laboratory Data

The hematology, clinical chemistry, and prolactin data will be summarized with descriptive statistics by treatment group within age group and by pooled treatment groups at each scheduled visit from screening (from Day -1 for prolactin) through Week 8. Both observed values and changes from baseline (Day -1) will be summarized.

The prolactin data at each scheduled visit will be summarized by treatment group within age group and by pooled treatment groups for each gender separately, in addition to the summaries described above.

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. Shift tables will be presented by treatment group within age group and by pooled treatment groups.

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 Week 6. Subjects with missing data for a clinical laboratory variable at either baseline or at Week 6 will not be included in tables for that variable. The number and percentage of subjects in each shift category will be displayed in the table.

Shift tables will be presented for the following clinical laboratory variables: aspartate transaminase (AST), alanine transaminase (ALT), gamma-glutamyl transferase (GGT), total bilirubin, creatine kinase, creatinine, blood urea nitrogen, prolactin, 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, creatinine 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) will be summarized by treatment group within age group and by pooled treatment groups. The criteria for identifying PCS clinical laboratory values are provided in Table 5.

**Table 5: Potential 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 Day -1 value or > 1.5 x ULN |
| BUN | >30 mg/dL (> 10.71 mmol/L) |

A listing of subjects with PCS values at any postbaseline visits for any of the variables listed in the table above will be presented. The listing will include age group, subject, treatment group, visit, study day, and, for each subject, all laboratory results (including all scheduled and unscheduled visits) for the variables with a PCS value for that subject. Values that meet the PCS criteria will be flagged with an asterisk in the listing.

Scatter plots of selected variables will be created which display Week 6 values vs. baseline values. Each plot will include a 45 degree ("y=x") reference line. The plots will be generated for ALT, AST, creatine kinase, GGT, total bilirubin, and prolactin by treatment group within age group and by pooled treatment groups.

Additional graphs will be presented for the prolactin data. These graphs will display mean (±SEM) prolactin values by treatment group within age group and by pooled treatment groups at each scheduled visit (Day -1 through Week 8) in a line graph layout. A separate graph will be presented for each gender.

The clinical laboratory data listings will include associated reference ranges if available. In addition, values outside the reference range will be flagged as "L" if below the lower limit of normal and as "H" if above the upper limit of normal. There will also be a flag for clinical significance based on the investigator's assessment of out-of-range values. The urinalysis data will be presented in data listings only.

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. One exception to this rule is when there are missing results from the original samples at screening – in this situation, the results of a

repeat screening sample will be substituted for the missing results in summary tables and graphs. All sample results (original and repeat) will be included in data listings.

### 3.16.3. Physical Examination and Weight

Clinically significant physical examination findings will be presented by subject and visit in a listing. The listing will include age group, subject, treatment group, visit at which the finding was reported, study day, and the clinically significant finding.

Body weight, which is measured during the physical examination, will be summarized in units of kilograms with descriptive statistics (both observed values and changes from baseline [Day -1]) by treatment group within age group and by pooled treatment groups at each scheduled visit from screening through Week 8.

### 3.16.4. 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 within age group and by pooled treatment groups at each scheduled visit from screening through Week 8. Both observed values and changes from baseline (Day -1) will be summarized.

Sponsor-defined PCS values for systolic blood pressure, diastolic blood pressure, and heart rate will be summarized by treatment group within age group and by pooled treatment groups. The number and percentage of subjects with PCS values at any postbaseline visit (scheduled or unscheduled) will be presented in the summary tables. The criteria for identifying PCS vital signs values for each age group are provided in Table 6.

**Table 6: Potentially Clinically Significant Criteria for Vital Signs**

**Children**

| 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 | >130 mmHg | | ≥20 mmHg |
| Diastolic Blood Pressure | N/A | | ≥10 mmHg | >85 mmHg | | ≥10 mmHg |
| Heart Rate | N/A | | ≥15 bpm | >130 bpm | | ≥10 bpm |

**Adolescents**

| 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 |

| Variable Name | PCS – Low if: | | | PCS – High if: | | |
|---|---|---|---|---|---|---|
| | Observed Value is: | AND | Decrease from Baseline is: | Observed Value is: | AND | Increase from Baseline is: |
| Diastolic Blood Pressure | N/A | | ≥10 mmHg | >90 mmHg | | ≥10 mmHg |
| Heart Rate | N/A | | ≥15 bpm | >110 bpm | | ≥10 bpm |

Note that both supine and standing values of blood pressures and heart rate at all postbaseline visits will be included in the identification and summary of PCS values.

A listing of vital signs data at all visits (scheduled and unscheduled) for subjects with PCS values will be presented. The listing will include age group, subject, treatment group, visit, study day, systolic blood pressure (supine and standing), diastolic blood pressure (supine and standing), and heart rate (supine and standing). Values that meet the PCS criteria will be flagged with an asterisk in the listing.

### 3.16.5. 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 for the purpose of analysis. For the overall assessment categorical 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.

The ECG variables will be summarized with descriptive statistics (frequency tables for the overall assessment categorical variable) by treatment group within age group and by pooled treatment groups at each scheduled visit from screening through Week 8. Both observed values and changes from baseline (Day -1) will be summarized (for the overall categorical assessment, only observed values will be summarized).

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.

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 within age group and in each of the pooled treatment groups whose highest reported QT/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 within age group and in each of the pooled treatment groups whose largest QT/QTcF increase from their baseline value meets the following thresholds:

- Greater than 30 msec
- Greater than 60 msec

### 3.16.6. Columbia-Suicide Severity Rating Scale Children's Versions (C-SSRS)

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

- Screening/lifetime assessment by treatment group within age group, by pooled treatment groups, and for all subjects combined.
- Screening/past 1 year assessment by treatment group within age group, by pooled treatment groups, and for all subjects combined.
- Baseline (Day -1) assessment by treatment group within age group, by pooled treatment groups, and for all subjects combined.
- All postbaseline assessments (Day 1 through Week 8, including unscheduled visit assessments) by treatment group within age group and by pooled treatment groups.

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 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 within age group and for each of the pooled treatment groups, 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.

A summary listing of individual subject data will be presented for the C-SSRS which lists data at all visits for subjects with a positive response for suicidal ideation, suicidal behavior, or self-injurious behavior without suicidal intent at any postbaseline assessment during the study. The listing will be in the form of a table, with each row representing a subject visit (scheduled and unscheduled). The listing will include age group, subject, treatment group, visit, timepoint (for screening assessments), study day of the visit, and a column for each suicidal ideation item (1 – 5), each suicidal behavior item (6 – 10), and a final column for self-injurious behavior without suicidal intent. The cells of the table will be populated with "Y" or "N," representing either a positive or negative response, respectively, for each item in the table (ie, for each column of the table).

### 3.16.7. Extrapyramidal Symptom Rating Scale-Abbreviated (ESRS-A)

Amendment 2 of the clinical study protocol added the ESRS-A assessment tool to evaluate extrapyramidal symptoms, therefore it is expected that data will be available for all subject visits that take place after the date of the IRB approval at each site.

The ESRS-A assesses 4 types of movement disorders: parkinsonism, dystonia, akathisia, and tardive dyskinesia. The ESRS-A consists of four subscales, one for each type of movement disorder. The ESRS-A contains 10 items to evaluate parkinsonism, 6 items to evaluate dystonia, 6 items to evaluate dyskinesia, and 2 items to evaluate akathisia. Each item score can range from 0 to 5, for a maximum possible parkinsonism score of 50, maximum possible dystonia score of 30, maximum possible dyskinesia score of 30, and maximum possible akathisia score of 10. A Clinical Global Impression of Severity (CGI-S) is also completed for each type of movement disorder, and is also is scored on a 0 to 5 scale. The ESRS-A is administered at baseline (Day -1) and at Weeks 2, 4, 6, and 8.

The subscale scores for each type of movement disorder (parkinsonism, akathisia, dystonia, and tardive dyskinesia) will be calculated as the sum of the scores of the individual items comprising each subscale. The overall total score will be calculated as the sum of each of the subscale scores. If any one of the items is not scored (ie, has a missing value), the associated subscale score and total score will be set equal to missing. The CGI-S scores will be summarized separately from the subscale scores.

Each of the subscale scores for parkinsonism, akathisia, dystonia, and tardive dyskinesia, and the overall total score at each visit will summarized with descriptive statistics by treatment group within age group and by pooled treatment groups. Changes from baseline at each postbaseline visit will also be summarized.

The CGI-S scores for each subscale at each visit will summarized with descriptive statistics by treatment group within age group and by pooled treatment groups. Changes from baseline at each postbaseline visit will also be summarized.

### 3.16.8. Children's Yale-Brown Obsessive Compulsive Scale (CY-BOCS)

The CY-BOCS is a semi-structured interview designed to rate the severity of obsessive and compulsive symptoms in children.

The CY-BOCS obsession subtotal score is calculated as the sum of the scores for items 1 through 5 of the CY-BOCS scale (excluding item 1b), and the CY-BOCS compulsion subtotal score is calculated as the sum of the scores for items 6 through 10 of the CY-BOCS scale (excluding item 6b). The CY-BOCS total score is the sum of the obsession and compulsion subtotal scores. Each item score ranges from 0 to 4, with a maximum possible obsession subtotal score of 20, a maximum possible compulsion subtotal score of 20, and a maximum possible total score of 40. If any one of these 10 items is not scored (ie, has a missing value), the associated subtotal score and total score will be set equal to missing.

The subtotal scores for obsession and compulsion, and the total score at each visit (screening, Day -1, Weeks 2, 4, 6, and Week 8) will summarized with descriptive statistics by treatment group within age group and by pooled treatment groups. Changes from baseline (Day -1) at each postbaseline visit will also be summarized.

### 3.16.9. Children's Depression Rating Scale, Revised (CDRS-R)

The CDRS-R is a 17-item, semi-structured interview to determine the severity of depression in children. The CDRS-R total score is calculated as the sum of the 17 items making up the CDRS-R. Each item score ranges from 1 to 7 with the exception of items 4, 5, and 16, which range from 1 to 5. The maximum possible total score is 113. If any one of the 17 items is not scored (ie, has a missing value), the total score will be set equal to missing.

The total score at each visit (screening, Day -1, and Weeks 2, 4, 6, and 8) will be summarized with descriptive statistics by treatment group within age group and by pooled treatment groups. Changes from baseline (Day -1) at each postbaseline visit will also be summarized.

### 3.16.10. Attention-Deficit Hyperactivity Disorder Rating Scale IV: Home Version

The ADHD Rating Scale-IV: Home Version will be used to determine the frequency and severity of ADHD symptoms and impairments over the 2 weeks prior to each visit. The 18-item scale consists of 2 subscale scores: Inattention (the sum of the item scores for the odd-numbered 9 items) and Hyperactivity-Impulsivity (the sum of the item scores for the even-numbered 9 items), as well as a Total Scale raw score defined as the sum of the Inattention and Hyperactivity-Impulsivity subscale scores. Each item score ranges from 0 to 3, for a maximum possible Inattention score of 27, a maximum possible Hyperactivity-Impulsivity score of 27, and a maximum possible Total Scale raw score of 54. If any one of the 18 items is not scored (ie, has a missing value), the affected subscale score(s) and Total Scale raw score will be set equal to missing.

The Inattention and Hyperactivity-Impulsivity subscale scores, as well as the Total Scale raw score at each visit (screening, Day -1, and Weeks 2, 4, 6, and 8), will be summarized with descriptive statistics by treatment group within age group and by pooled treatment groups. Changes from baseline (Day -1) at each postbaseline visit will also be summarized.

### 3.16.11. Prior and Concomitant Medications

Prior medications and concomitant medications will be summarized by 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 or two study periods (prestudy or during screening vs. during the treatment or posttreatment follow-up periods) based on the medication start and stop dates relative to study drug dosing. For example, medications that were started and stopped prior to the date of the first dose of study drug will be assigned to the prestudy/screening period only, while medications started prior to the first dose of study drug and either stopped during the study or indicated as "ongoing" will be assigned to both the prestudy/screening and treatment period. A given medication can therefore be assigned to one or two 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 as described in the next paragraph. 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.

The summary of medications taken prestudy or during screening will be presented by treatment group within age group and by pooled treatment groups. An "all subjects" column will be included in these summaries. Medications taken during the study after the initiation of dosing (ie, during the treatment or posttreatment follow-up periods) will be summarized by treatment group within age group and by pooled treatment groups.

## 3.17. Additional Data Presentations

### 3.17.1. Inclusion and Exclusion Criteria

Inclusion and exclusion criteria confirmation results will be presented in a data listing by subject.

### 3.17.2. Serology Test Results

Serology assessments (ie, Human immunodeficiency virus antibody [HIV-Ab], Hepatitis B surface antigen [HBsAg], and Hepatitis C antibody [HCV-Ab]) collected at screening will be reviewed at the study site to ensure the entry criteria are met. The HBsAg and HCV-Ab test results will be presented in a listing by subject. The results of the HIV-Ab test will not be listed, but will be kept on file at the study site.

### 3.17.3. Pregnancy Test Results

Serum and urine pregnancy test results will be presented in a listing by subject.

### 3.17.4. Urine Drug Screen and Alcohol Breath Test

Urine drug screen and alcohol breath test results for adolescents only will be presented in a listing by subject.

# 4. DEVIATIONS FROM PROTOCOL PLANNED ANALYSIS

The methods of analysis described in this SAP are generally consistent with the methods described in the clinical study protocol (Amendment 2). One important change from the protocol is specification of the mean CGI-TS-Improvement score at Week 6 as the key secondary efficacy endpoint. In addition, sensitivity analyses have been added to the SAP for the primary and key secondary efficacy endpoints, and a procedure to control for multiplicity has been included with respect to the primary and key secondary efficacy endpoints. The RTRS total score method of analysis has been changed from the MMRM to the ANCOVA, as the RTRS videos will be scored at a single postbaseline visit. The SAP also includes additional endpoints and data summaries not mentioned in the study protocol, including the TTS responder analysis and the PCS value summaries for clinical laboratory data and vital signs.

# 5. PERFORMANCE QUALIFICATION OF SAS® PROGRAMS

The analysis and summary of data from this study will be performed using SAS® 9.3 (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.

# 6. REFERENCES

ClinicalTrials.gov Identifier: NCT01727700. Otsuka Pharmaceutical Development & Commercialization, Inc. Study evaluating the safety and efficacy of fixed-dose once-daily aripiprazole in children and adolescents with Tourette's Disorder https://www.clinicaltrials.gov/ct2/show/NCT01727700?term=NCT01727700&rank=1 [Accessed 05 June 2015].

Jankovic J, Jimenez-Shahed J, Brown LW. A randomised, double-blind, placebo-controlled study of topiramate in the treatment of Tourette syndrome. J Neurol Neurosurg Psychiatry. 2010 Jan;81(1):70-3.

LaVange LM. The role of statistics in regulatory decision making. Therapeutic Innovation and Regulatory Science. 2014; 48(1) 10-19.

O'Kelly M, Ratitch B. Clinical Trials with Missing Data: A Guide for Practitioners. West Sussex, United Kingdom: John Wiley & Sons, Ltd. 2014; 292-297, 309-310.

Yoo HK, Joung YS, Lee JS, Song DH, Lee YS, Kim JW, et al. A multicenter, randomized, double-blind, placebo-controlled study of aripiprazole in children and adolescents with Tourette's disorder. J Clin Psychiatry. 2013;74(8):e772-80.