STATISTICAL ANALYSIS PLAN
PHASE 2B

VERSION: Final

DATE OF PLAN:

19-FEB-2020

BASED ON:

Protocol Amendment No. 1 Final Version Date: 06 November 2017

Original Version Date: 05 September 2017

STUDY DRUG:

NBI-98854

PROTOCOL NUMBER:

NBI-98854-TS2004

STUDY TITLE:

Open-Label Safety and Tolerability Study of Optimized Doses of NBI-98854 for the Treatment of Pediatric Subjects with Tourette Syndrome

SPONSOR:

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

858-617-7600

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

SIGNATURE PAGE

This document has been prepared and/or reviewed by:

[redacted]
Neurocrine Biosciences, Inc

[redacted]
Signature

[redacted]
Date

[redacted]
Neurocrine Biosciences, Inc

[redacted]
Signature

[redacted]
Date

This document has been reviewed and accepted by:

[redacted]
Neurocrine Biosciences, Inc

[redacted]
Date

[redacted]
Neurocrine Biosciences, Inc

[redacted]
Date

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. LIST OF ABBREVIATIONS

Table 1: List of Abbreviations

Abbreviation	Term
ADHD	Attention-Deficit Hyperactivity Disorder
ADHD Rating Scale-5	Attention-Deficit Hyperactivity Disorder Rating Scale-5: Home Version
AE	Adverse Event
ALP	Alkaline Phosphatase
ALT	Alanine Aminotransferase
AST	Aspartate Transaminase
ATC	Anatomical Therapeutic Chemical Classification
BLQ	Below the Lower Limit of Quantification
BMI	Body Mass Index
C&A-GTS-QOL	Gilles de la Tourette Syndrome – Quality of Life Scale for Children and Adolescents
CDRS-R	Children's Depression Rating Scale - Revised
CGI-S	Clinical Global Impression of Movement Severity
CGI-Tics-Severity	Clinical Global Impression of Tics-Severity scale
CGI-TS-Improvement	Clinical Global Impression of Tourette Syndrome-Improvement scale
C-SSRS	Columbia-Suicide Severity Rating Scale
CY-BOCS	Children's Yale-Brown Obsessive Compulsive Scale
DSM-IV or -V	Diagnostic and Statistical Manual of Mental Disorders, 4th or 5th Editions
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
ESRS-A	Extrapyramidal Symptom Rating Scale-Abbreviated
ET	Early Termination
GGT	Gamma-Glutamyl Transferase
IPD	Important Protocol Deviation
MedDRA	Medical Dictionary for Regulatory Activities Terminology
NBI	Neurocrine Biosciences, Inc.
PCS	Potentially Clinically Significant
PD	Pharmacodynamic

PT	Preferred Term
PUTS	Premonitory Urge for Tics Scale
QTcF	Fridericia's Correction of QT Interval
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
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
ULN	Upper Limit of Normal
VAS	Visual Analogue Scale
WHO	World Health Organization
YGTSS	Yale Global Tic Severity Scale

2. INTRODUCTION

The purpose of this statistical analysis plan (SAP) is to describe the planned analyses and data displays that will be prepared to summarize the data from the Phase 2b study described in Neurocrine Biosciences, Inc. (NBI) Protocol NBI-98854-TS2004.

This SAP was developed in accordance with ICH E9 guidance. Further information related to study design and methodology can be found in the protocol.

3. STUDY OBJECTIVES AND ENDPOINTS

The objectives of this clinical study are as follows:

- To determine the long-term safety and tolerability of up to 24 weeks of treatment with NBI-98854 in pediatric subjects with Tourette Syndrome (TS).
- To evaluate the long-term pharmacodynamic (PD) effects of NBI-98854 administered once daily in pediatric subjects with TS.
- To evaluate plasma exposure of NBI-98854 and its metabolite, NBI-98782, following repeated daily doses of NBI-98854.

4. STUDY DESIGN

4.1. Summary of Study Design

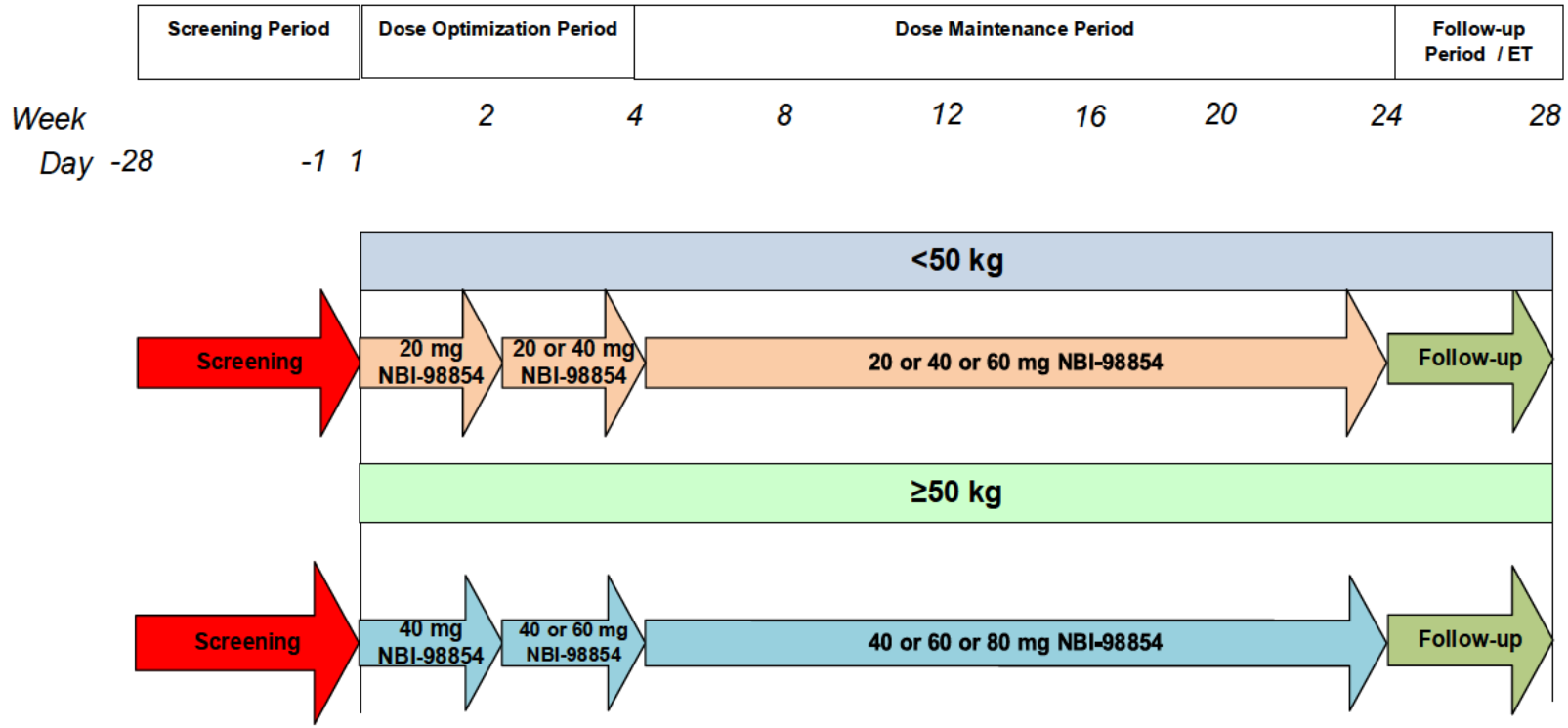
This is a Phase 2b, multicenter, open-label, dose-optimization study of the safety, tolerability, PD, and plasma exposure of NBI-98854 doses in the range of 20 mg to 60 mg for subjects <50 kg and 40 mg to 80 mg for subjects ≥50 kg, administered each night at bedtime for a total of 24 weeks of treatment in pediatric subjects with TS. This study will only include subjects who participated in and completed the previous Phase 2b clinical study NBI-98854-TS2003.

Up to 120 male and female pediatric subjects, 6 to 18 years of age, with a Diagnostic and Statistical Manual of Mental Disorders, 4th or 5th Editions (DSM-IV or -V) diagnosis of TS will participate.

The starting dose will be NBI-98854 20 mg for subjects <50 kg at baseline and NBI-98854 40 mg for subjects ≥50 kg at baseline. The dose may be escalated in increments of 20 mg every 2 weeks to a maximum of 60 mg for subjects <50 kg and 80 mg for subjects ≥50 kg to achieve an optimal dose of NBI-98854 for each subject. Dose escalations may occur at the end of Weeks 2 and 4. After Week 4, subjects will continue to receive their optimized dose of NBI-98854 for an additional 20-week dose maintenance period. At any time after Week 2, the physician investigator may decrease the dose to the previous dose for any subject who had a dose escalation and who is unable to tolerate a given dose increase. The subject will continue at that dose until the end of the treatment period. Subjects who are unable to tolerate the starting dose or resumption of the previous dose will be discontinued from the study. Follow-up assessments will be conducted at the end of Week 28 (4 weeks after the last dose of the study drug). Full details of the dose optimization and maintenance are provided in the Protocol.

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

Figure 1: Study Design Schematic



4.2. Sample Size Considerations

The sample size for this open-label safety study is based on practical considerations and not on a statistical power calculation.

4.3. Randomization

This is an open-label study. Subjects are not randomized to treatment in this study.

4.4. Clinical Assessments

Safety assessments include:

- adverse event (AE) monitoring,
- clinical laboratory tests (hematology, clinical chemistry, and urinalysis),
- serum prolactin
- hemoglobin A1c
- vital sign measurements,
- physical examinations, including weight,
- 12-lead electrocardiogram (ECG),
- Columbia-Suicide Severity Rating Scale (C-SSRS), Children's Version,
- Children's Depression Rating Scale - Revised (CDRS-R),
- Children's Yale-Brown Obsessive Compulsive Scale (CY-BOCS),
- Attention-Deficit Hyperactivity Disorder Rating Scale-5: Home Version (ADHD Rating Scale-5), and
- Extrapyramidal Symptom Rating Scale-Abbreviated (ESRS-A).

PD assessments for TS include:

- Yale Global Tic Severity Scale (YGTSS),
- Clinical Global Impression of Tics-Severity scale (CGI-Tics-Severity),
- Clinical Global Impression of Tourette Syndrome-Improvement scale (CGI-TS-Improvement),
- Premonitory Urge for Tics Scale (PUTS), and
- Gilles de la Tourette Syndrome – Quality of Life Scale for Children and Adolescents (C&A-GTS-QOL).

Blood samples for plasma drug and metabolite concentration analyses are also collected. Subjects/caregivers will be asked to record and provide dosing times from the evening before the treatment period visits when these samples are collected.

The schedule of assessments can be found in the protocol.

5. PLANNED ANALYSES

5.1. Interim Analyses

An interim analysis is not planned for this study.

5.2. Final Analyses

A single, final analysis will be performed after the study database has been locked.

6. GENERAL CONSIDERATIONS FOR DATA ANALYSES AND HANDLING

6.1. General Statistical Procedures

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

A single treatment group (NBI-98854) will be used for all analyses. Descriptive statistics will generally be presented separately for each baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups. Each table will therefore have 3 columns:

- NBI-98854 (<50 kg baseline weight group)
- NBI-98854 (≥50 kg baseline weight group)
- NBI-98854 (All Subjects)

Select tables will display results by NBI-98854 dose level.

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

6.2. Analysis Populations

For purposes of defining analysis sets, "enrolled subjects" refers to subjects enrolled into the study at the Day 1 visit (i.e., are not screen failures) as specified by the subject enrollment electronic case report form (eCRF).

Unless otherwise specified, the safety analysis set will be used for all analyses in this study. The safety analysis set will include all enrolled subjects who take at least one dose of study drug and have any data collected after the initial dose of study drug. If it is unknown whether or not the subject took at least one dose of study drug but the subject was dispensed study drug and has postbaseline data in the study database, the subject will be included in the safety analysis set.

Summaries of subject disposition, analysis set inclusion/exclusion status, and important protocol deviations (IPDs) will include all enrolled subjects.

6.3. Baseline Definition

The assessments collected at the Day 1 study visit will serve as the baseline value for all assessments, excluding any assessments that occur after the first dose of study drug. If a Day 1 visit value is not available, then the last measurement collected on or prior to the date of the first dose of study drug will serve as baseline.

6.4. Derived and Transformed Data

6.4.1. Study Day

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

6.4.2. Change from Baseline

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

6.4.3. Handling of Early Termination Visit Data

An early termination (ET) visit occurs when a subject discontinues from the study prior to completing the scheduled Week 28 visit. The data collected at ET visits will be included in summary tables and figures in accordance with the ET visit mapping scheme described in this section.

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) for visits that are expected to occur every 2 weeks and within 14 days prior to and 13 days after for visits that are expected to occur every 4 weeks. ET visits occurring after the Week 24 visit will be mapped to the Week 28 follow-up visit if it occurs after day 181.

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

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	Visit Window (Study Day Range)
Week 2	14	7-20
Week 4	28	21-34
Week 8	56	42-69
Week 12	84	70-97
Week 16	112	98-125
Week 20	140	126-153
Week 24	168	154-181
Week 28	196	>181

6.5. Handling of Missing Data

6.5.1. Missing Outcome Data

Missing values for outcome measures will generally not be replaced with imputed values except as noted in Section 6.4.3 for the ET visit data mapped to scheduled visits for data summary purposes.

6.5.2. Missing Dates

6.5.2.1. First and Last Dose Dates

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

The imputation rules for first dose date are as follows:

- If the date is completely missing or if both the day and month are missing, the date will be imputed as the date when the first study drug kit was assigned;
- If only the day is missing, the date will be imputed as the date when the first study drug kit was assigned if the month and year match the kit assignment date; if the month and year occur after the kit assignment date, the missing day will be imputed as the first day of the month.

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

- the Week 24 visit date,
- study discontinuation date for subjects who discontinue before Week 24,
- the date when the last study drug kit(s) was assigned + the number of doses dispensed with the kit(s).

6.5.2.2. Missing Start Dates for Adverse Events and Prior and Concomitant Medications

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

The imputation rules for AE start dates are as follows:

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

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

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

7. STUDY POPULATION

7.1. Subject Disposition

The summary of subject disposition will display the number of subjects who were enrolled, who received at least one dose of study drug, who completed the dose optimization period (i.e., up to Week 4, excluding ET visits mapped to Week 4), who completed the dose maintenance period (i.e., up to Week 24, excluding ET visits mapped to Week 24), and who completed the study. The number of subjects who did not complete the study will also be summarized, both overall and according to the reason for early discontinuation.

7.2. Protocol Deviations

Protocol deviations described in the study-specific Protocol Deviation Plan will be entered into the study database and used to identify IPDs. IPDs are protocol deviations that might significantly affect the completeness, accuracy and/or reliability of the study data or that might significantly affect a subject's rights, safety, or well-being. An assessment of IPDs will be performed by a committee composed of NBI Clinical Development project team members prior to database lock. This committee will review a listing of all protocol deviations reported in the study database and determine which deviations are IPDs.

A summary of the number and percentage of subjects with IPDs by deviation category will be provided using all enrolled subjects.

All protocol deviations entered into the study database will be presented in a data listing. Any IPDs will be flagged in the listing.

7.3. Demographic and Baseline Characteristics

Demographic and baseline characteristics data will be summarized using descriptive statistics for continuous variables, and frequency counts and percentages for categorical variables.

Demographics include:

- Age (years)
- Age category (child [ages 6-11], adolescent [ages 12-18])
- Sex
- Ethnicity
- Race

Study baseline subject characteristics include:

- Age at TS diagnosis (years)
- Baseline value of TTS
- Height (measured at screening; cm)
- Weight (presented in both pounds and kilograms)
- Body mass index (BMI; measured at screening; kg/m^2)
- CYP2D6 genotype status (measured at baseline in clinical study NBI-98854-TS2003)

7.4. Medical History and Medical Conditions Present at Entry

Medical history will be summarized in frequency tables (number and percentage of subjects) by Medical Dictionary for Regulatory Activities (MedDRA) System Organ Class (SOC) and Preferred Term (PT), with SOCs and PTs within each SOC sorted alphabetically

7.5. Study Drug Dosing and Compliance

7.5.1. Exposure to Study Drug

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

Duration of exposure will be summarized with descriptive statistics. The number and percentage of subjects with the following exposure categories will also be presented:

- >0 to <2 weeks
- ≥2 to <4 weeks
- ≥4 to <8 weeks
- ≥8 to <12 weeks
- ≥12 to <16 weeks
- ≥16 to <20 weeks
- ≥20 to <24 weeks
- ≥24 weeks

7.5.2. Dose Adjustments and Dosages

The number and percentage of subjects with a dose reduction at any time will be presented. Percentages will be based on the number of subjects with at least one dose escalation.

Descriptive statistics of study drug dosages for the last dose received in the study will also be presented.

7.5.3. Compliance

Subjects will bring all unused study drug and empty study drug packaging material to the center at each study visit for drug accountability and reconciliation by study center personnel. A compliance check will be performed by counting the capsules returned at each study visit. The site will then enter whether the subject's dosing compliance since the previous visit was ≥80% into the eCRF.

The number and percentage of subjects who are dosing compliant at each visit during the treatment period will be provided.

8. PLASMA CONCENTRATION DATA

The plasma concentrations of NBI-98854 and its active metabolite NBI-98782 will be summarized with descriptive statistics by visit and by the last dose level received prior to the blood sample being drawn (20, 40, 60, or 80 mg). The dose at the Week 28 visit will reflect the last dose the subject received during the study treatment period (i.e., the dose level at the subject's Week 24 visit). These summary tables will be presented separately for each baseline weight group (<50 kg vs. ≥50 kg). There will not be a pooled weight groups summary.

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 (%).

9. PHARMACODYNAMIC ASSESSMENTS

9.1. Yale Global Tic Severity Scale

The following three analysis variables are based on the YGTSS: (a) the Total Tic Score (TTS), (b) the Impairment score, and (c) the Global Tic Severity score.

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 (i.e., 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, with higher scores representing greater severity.

The YGTSS Impairment score is captured directly in the eCRF. This score can range in value from 0 to 50, with higher scores representing more severe impairment.

The YGTSS Global Tic Severity score is the sum of the TTS and the 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.

For each of these variables, descriptive statistics will be presented for the observed and change from baseline values at baseline and at each postbaseline visit.

9.2. 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

For the numerical scores, descriptive statistics will be presented for the observed and change from baseline values at baseline and at each postbaseline visit. Frequency counts using the response categories will also be presented.

9.3. CGI-TS-Improvement

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

For the numerical scores, descriptive statistics will be presented for the observed values at each postbaseline visit. Frequency counts using the response categories will also be presented.

9.4. Premonitory Urge for Tics Scale

The PUTS is an instrument for quantifying the premonitory urge phenomena associated with tics. It 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 PUTS total score value can range from 9 to 36. If any one of the 9 items is not scored (i.e., has a missing value), the PUTS total score will be set equal to missing.

Descriptive statistics will be presented for the observed and change from baseline values at baseline and at each postbaseline visit.

9.5. Gilles de la Tourette Syndrome-Quality of Life Scale for Children and Adolescents

The C&A-GTS-QOL has two parts:

(1) a 27-item scale which includes 4 factors (subscales), with each of the 27 items scored on a 5-point scale (1=never, 2=rarely, 3=sometimes, 4=often, 5=always); and

(2) a visual analog scale (VAS) which ranges in value from 0 to 100, with 0 representing extremely unhappy/dissatisfied with life, and 100 representing extremely happy/satisfied with life.

The C&A-GTS-QOL total score is calculated as the sum of the scores for the 27 items. If any one of the 27 items is not scored (i.e., has a missing value), the total score will be set equal to missing.

Scores will be calculated for each of the 4 C&A-GTS-QOL factors in a similar fashion. These factors and the corresponding item numbers are as follows:

- Psychological (16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 27)
- Physical and activities of daily living (1, 2, 3, 4, 5, 6, 26)
- Obsessive-compulsive (7, 8, 9, 10, 15)
- Cognitive (11, 12, 13, 14)

The total score and each factor will be normalized to a value that can range from 0 to 100 by using the following formula:

$$100 \times \frac{observed\ score - minimum\ possible\ score}{maximum\ possible\ score - minimum\ possible\ score}$$

Two versions of this instrument are used in this trial: 1 version for children aged 6 to 12 years and 1 version for adolescents aged 13 to 18 years. Scoring is identical for both versions and all results will be combined for analyses.

Descriptive statistics will be presented for the observed and change from baseline C&A-GTS-QOL normalized total scores, the 4 normalized factor scores, and the VAS scores at baseline and at each postbaseline visit.

10. SAFETY

10.1. Adverse Events

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

A treatment-emergent adverse event (TEAE) is an AE not present prior to the initiation of study drug dosing, or is an already present event that worsens either in intensity or frequency following the initiation of study drug dosing. The determination of whether an AE is treatment-emergent is based on the AE onset date relative to the date of the subject's first dose of study drug. If the AE onset date and date of the first dose of study drug are the same, or if the AE onset date is unknown, it will be assumed that the AE is a TEAE.

The frequency tables will include the number and percentage of unique subjects experiencing each event at least once during the study. Unless otherwise specified, summary tables will include events with a start date on or after the date of the first dose of study drug and on or before the date of the last dose of study drug + 30 days. Missing or incomplete AE onset dates or study drug dosing dates will be imputed as described in Section 6.5.2.

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

- Frequency of TEAEs by SOC and PT, with SOCs and PTs within each SOC sorted by decreasing frequency (number of unique subjects)

- Frequency of TEAEs by PT, with PT sorted by decreasing frequency (number of unique subjects)

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

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

10.1.1. Adverse Events Resulting in Premature Discontinuation from Study

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

A listing of TEAEs resulting in premature study discontinuation will be provided which includes weight group, subject ID, last study drug dose level received prior to the onset date of the TEAE(s) leading to discontinuation, study day of the discontinuation, and other relevant information from the AE eCRF.

10.1.2. Adverse Events Resulting in Study Drug Dose Reductions

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

10.1.3. Deaths and Other Serious Adverse Events

Summary tables of serious adverse events (SAEs) will be presented. The tables will include the frequency of SAEs presented by PT within SOC (presented in the same method as the primary TEAE table).

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

10.2. Clinical Laboratory Data

By visit summaries

The clinical chemistry, hematology, hemoglobin A1c, and prolactin data will be summarized with descriptive statistics at baseline and at each postbaseline visit. Both observed values and changes from baseline will be summarized.

In addition to the summaries described above, the prolactin data will also be summarized for each sex separately.

Shift tables

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.

Each shift table will have three rows and three columns, with rows reflecting the reference range category at baseline, and columns reflecting the reference range category at the subject's last study visit during the treatment period of the study. A "Total" row and "Total" column will also be included. Subjects with a missing baseline value or who do not have postbaseline data will not be included in the tables for that variable. The number and percentage of subjects in each shift category will be displayed in the table; percentages will be based on the number of subjects included in the table.

Shift tables will be presented for the following clinical laboratory variables: aspartate transaminase (AST), alanine transaminase (ALT), alkaline phosphatase (ALP), gamma-glutamyl transferase (GGT), total bilirubin, creatine kinase, creatinine, blood urea nitrogen, white blood cell count, absolute neutrophil count, hemoglobin, and platelet count.

Potentially clinically significant (PCS) values

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

Table 3: Potentially Clinically Significant Criteria for Clinical Laboratory Variables

Variable	PCS Threshold
ALT	>3 x ULN (upper limit of normal)
AST	>3 x ULN
ALP	>2.5 x ULN
Creatine kinase	>5 x ULN
GGT	>3 x ULN
Total bilirubin	>1.5 x ULN
White blood cell count	≤2.8 x 1000/µL
Absolute neutrophil count	<1.5 x 1000/µL
Creatinine	>1.5 x baseline value or >1.5 x ULN
BUN	>30 mg/dL (>10.71 mmol/L)

10.3. Vital Signs

The vital signs data, including orthostatic blood pressures and heart rate (calculated as standing value minus supine value), will be summarized with descriptive statistics at baseline and at each postbaseline visit. Both observed values and changes from baseline will be summarized.

Summaries of sponsor-defined PCS values will be presented for systolic blood pressure, diastolic blood pressure, and heart rate. The number and percentage of subjects with PCS values that are reported at any postbaseline visit (scheduled or unscheduled) will be summarized. The criteria for identifying PCS vital signs values for children (6 to 11 years of age at baseline) and adolescents (12 to 18 years of age at baseline) are provided in Table 4 and Table 5, respectively. Both supine and standing values will be included in the identification and summary of PCS values.

Table 4: Potentially Clinically Significant Criteria for Vital Signs Variables in Children

	PCS – Low if:	PCS – High if:		
Variable Name	**Decrease from Baseline is:**	**Observed Value is:**	**AND**	**Increase from Baseline is:**
Systolic Blood Pressure	≥20 mmHg	>130 mmHg		≥20 mmHg
Diastolic Blood Pressure	≥10 mmHg	>85 mmHg		≥10 mmHg
Heart Rate	≥15 bpm	>130 bpm		≥10 bpm

Table 5: Potentially Clinically Significant Criteria for Vital Signs Variables in Adolescents

	PCS – Low if:	PCS – High if:		
Variable Name	**Decrease from Baseline is:**	**Observed Value is:**	**AND**	**Increase from Baseline is:**
Systolic Blood Pressure	≥20 mmHg	>145 mmHg		≥20 mmHg
Diastolic Blood Pressure	≥10 mmHg	>90 mmHg		≥10 mmHg
Heart Rate	≥15 bpm	>110 bpm		≥10 bpm

10.4. Body Weight

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

10.5. 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 categorical ECG interpretation variable (the investigator's assessment of the ECG as "Normal", "Abnormal, not Clinically Significant", or "Abnormal, Clinically Significant"), which is also reported for each replicate, the value that represents the greatest degree of abnormality will be used in all summary tables. If less than 3 values are recorded at an assessment, then the average of the available value(s) for the quantitative variables and the greatest degree of abnormality values(s) for the interpretation variable will be used.

The quantitative ECG variables will be summarized with descriptive statistics at baseline and at each scheduled postbaseline visit. Both observed values and changes from baseline will be summarized. Frequency counts and percentages for the ECG interpretation variable categories will also be presented at baseline and at each scheduled postbaseline visit.

Categorical summaries will be presented for the QT and QTcF interval data. For these summaries, a subject's highest reported postbaseline value (including values reported at

unscheduled visits) will be used to determine in which category(s) the subject will be counted. The averaged triplicate values will be used when determining each subject's highest reported values.

Two categorical summaries will be presented for the QT and QTcF intervals (each interval will be summarized separately). For the first summary, the number and percentage of subjects whose highest reported QT or QTcF postbaseline value meets the following thresholds will be summarized:

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

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

- Greater than 30 msec
- Greater than 60 msec

10.6. Columbia-Suicide Severity Rating Scale Children's Versions

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

- Screening/lifetime assessment
- Screening/past 1 year assessment
- Baseline (Day 1) assessment
- All postbaseline assessments (through Week 28, including unscheduled visit assessments)

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 at any postbaseline assessment for a particular item or category will be considered as a "Yes" for that item or category 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, 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.

10.7. Extrapyramidal Symptom Rating Scale-Abbreviated

The ESRS-A assesses 4 types of movement disorders: parkinsonism, dystonia, akathisia, and 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 Movement Severity (CGI-S) is also completed for each type of movement disorder and is also is scored on a 0 to 5 scale.

The subscale scores for each type of movement disorder (parkinsonism, akathisia, dystonia, and 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 (i.e., 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 dyskinesia, and the overall total score will be summarized with descriptive statistics at baseline and at each scheduled postbaseline visit. Both observed values and changes from baseline will be summarized.

The CGI-S scores for each subscale will be summarized with descriptive statistics at baseline and at each scheduled postbaseline visit. Both observed values and changes from baseline will be summarized.

10.8. Children's Yale-Brown Obsessive-Compulsive Scale

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 (i.e., has a missing value), the associated subtotal score and total score will be set equal to missing.

The obsession subtotal, compulsion subtotal and total scores will be summarized with descriptive statistics at baseline and at each scheduled postbaseline visit. Both observed values and changes from baseline will be summarized.

10.9. Children's Depression Rating Scale, Revised

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 (i.e., has a missing value), the total score will be set equal to missing.

The total score will be summarized with descriptive statistics at baseline and at each scheduled postbaseline visit. Both observed values and changes from baseline will be summarized.

10.10. Attention-Deficit Hyperactivity Disorder Rating Scale-5: Home Version

The ADHD Rating Scale-5: 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 scale consists of 2 symptom subscales: Inattention (9 items) and Hyperactivity-Impulsivity (9 items). The Inattention subscale and Hyperactivity-Impulsivity subscale scores are each derived by summing up the 9 relevant item scores. Each item score ranges from 0 to 3, therefore each subscale score has a range of 0 to 27. The Total Scale raw score is defined as the sum of the Inattention and Hyperactivity-Impulsivity subscale scores, and has a range of 0 to 54. If any one of the 18 items is not scored (i.e., has a missing value), the affected subscale score(s) and Total Scale raw score will be set equal to missing.

The scale also assesses 6 domains of impairment that are common among children and adolescents with ADHD: relationships with family members, peer relationships, academic functioning, behavioral functioning, homework performance, and self-esteem. Each domain impairment item ranges from 0-3 and is assessed after completing each of the 2 subscales (i.e., each impairment item is assessed twice). The score for each domain is defined as the higher (worst) of the two scores. A total impairment score is defined as the sum of the 6 impairment

items (using the higher of the two scores) and has a range of 0 to 18. If any one of the 12 individual items is not scored (i.e., has a missing value), the total impairment score will be set equal to missing.

Two versions of this instrument are used in this trial: 1 version is for children aged 5 to 10 years and 1 version is for adolescents aged 11 to 17 years. Scoring is identical for both versions and all results will be combined for analyses.

The Inattention and Hyperactivity-Impulsivity subscale scores, the Total Scale raw score, and the Total Impairment score will be summarized with descriptive statistics at baseline and at each scheduled postbaseline visit. Both observed values and changes from baseline will be summarized.

10.11. Prior and Concomitant Medications

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

Medications will be assigned to one or two study periods (pre-study 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. 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/posttreatment 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 the study periods defined in the previous 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.