STATISTICAL ANALYSIS PLAN
PHASE 2

VERSION: FINAL

DATE OF PLAN:

25-SEP-2018

BASED ON:

Protocol Amendment No. 2 Final Version: 31 July 2018

Amendment No. 1 Final Version Date: 19 October 2017

Original Final Version Date: 26 July 2017

STUDY DRUG:

NBI-98854

PROTOCOL NUMBER:

NBI-98854-TS2003

STUDY TITLE:

A Phase 2b, Randomized, Double-Blind, Placebo-Controlled, Dose Optimization Study to Assess the Safety, Tolerability, and Efficacy 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.

CONFIDENTIAL

SIGNATURE PAGE

This document has been prepared and/or reviewed by:

[redacted]

Project Statistician
Neurocrine Biosciences, Inc

Signature

26-Sep-2018
Date

[redacted]

Director – Biostatistics
Neurocrine Biosciences, Inc

Signature

26 Sep 2018
Date

This document has been reviewed and accepted by:

[redacted]

Chief Medical Officer
Neurocrine Biosciences, Inc

Signature

26-Sep-2018
Date

[redacted]

Chief Regulatory Officer
Neurocrine Biosciences, Inc

Signature

9/27/18
Date

TABLE OF CONTENTS

LIST OF TABLES

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
ANCOVA	Analysis of Covariance
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
CMH	Cochran-Mantel-Haenszel
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
FAS	Full Analysis Set
GGT	Gamma-Glutamyl Transferase
IPD	Important Protocol Deviation
LOCF	Last Observation Carried Forward
LS	Least-squares
MAR	Missing at Random

MCMC	Markov Chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities Terminology
MI	Multiple Imputation
MMRM	Mixed-effect Model Repeated Measures
NRI	Non-responder Imputation
PCS	Potentially Clinically Significant
PP	Per-Protocol
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-TS2003.

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

3. STUDY OBJECTIVES AND ENDPOINTS

3.1. Study Objectives

3.1.1. Primary Objective

The primary objective of this clinical study is:

- To evaluate the efficacy of NBI-98854, titrated to the subject's optimal dose in the range of 20 mg to 60 mg for subjects <50 kg and 40 mg to 80 mg for subjects ≥50 kg, administered once daily for the treatment of Tourette syndrome (TS).

3.1.2. Secondary Objective

The secondary objectives of this study are:

- To evaluate the safety and tolerability of NBI-98854, titrated to the subject's optimal dose in the range of 20 mg to 60 mg for subjects <50 kg and 40 mg to 80 mg for subjects ≥50 kg, administered once daily.
- To evaluate plasma exposure measures of NBI-98854 and its active metabolite, NBI-98782, following titrated doses of NBI-98854 from 20 mg to 60 mg for subjects <50 kg and 40 mg to 80 mg for subjects ≥50 kg.

4. STUDY DESIGN

4.1. Summary of Study Design

This is a Phase 2b, randomized, double-blind, placebo-controlled, dose-optimization study to evaluate the efficacy, safety and tolerability of NBI-98854 titrated to the subject's optimal dose in the range of 20 mg to 60 mg for subjects <50 kg and 40 mg to 80 mg for subjects ≥50 kg compared to placebo, administered once daily (qd) for a total of 12 weeks of treatment in pediatric subjects with TS. Approximately 120 male and female pediatric subjects, 6 to 17 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. Eligible subjects will be randomized in a 1:1 ratio to NBI-98854 or placebo treatment. Randomization will be stratified based on subject's weight group at baseline (<50 kg, ≥50 kg).

For subjects randomized to active treatment, 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, which 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. To maintain the study blind, subjects randomized to placebo in each weight group will be subjected to the same dose escalation requirements, but will receive only placebo during the treatment period. Dose escalations will occur at the end of Weeks 2 and 4 based on the following 2 criteria: 1) the subject's tics are not sufficiently controlled per physician investigator assessment; and 2) an evaluation by the physician investigator indicates that the subject is tolerating the study drug at the current dose and would likely be able to tolerate the next dose level. During the first 6 weeks of the 12-week double-blind treatment period, the physician investigator may escalate a subject's dose to the next dose level, continue with the subject's current dose, or reduce to the subject's prior tolerated dose (in subjects who have had a dose escalation). After Week 6, subjects will continue to receive their optimized dose of NBI-98854 or placebo for an additional 6-week dose maintenance period. If a subject's optimal dose has already been established at Week 2 (or at Week 4), no further dose escalation will be allowed during the dose optimization period and the subject will continue at that dose until the end of the 12-week treatment 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 12-week treatment period. The investigator may reduce the subject's dose only one time. 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 14 (2 weeks after the last dose of the study drug).

4.2. Sample Size Considerations

The protocol-specified sample size of 60 subjects per treatment group is based on a power calculation for the Yale Global Tic Severity Scale (YGTSS) Total Tic Score (TTS) change from baseline using a two-sample t-test with a two-sided Type I error of 0.05. Power calculations based on this sample size under three dropout rate scenarios for two hypothesized effect sizes are summarized in Table 2. Note that the effect size is defined as the mean difference between the

NBI-98854 treatment group and the placebo group divided by the common standard deviation (SD) (e.g., a mean difference of 8 divided by an SD of 10 yields an effect size of 0.8).

Table 2: Power Calculations

Dropout Rate	Number of Subjects per Treatment Group	Effect Size	Power
0%	60	0.75	98%
		0.85	99%
10%	54	0.75	97%
		0.85	99%
15%	51	0.75	96%
		0.85	98%

Standard deviations reported in the literature for TS studies evaluating changes in the YGTSS 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 above are representative of effect sizes seen in these published reports.

4.3. Randomization

Eligible subjects will be randomized in a 1:1 ratio to receive either NBI-98854 or placebo treatment. Randomization will be stratified based on subject's weight group at baseline (<50 kg vs. ≥50 kg).

4.4. Clinical Assessments

Efficacy 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).

Efficacy assessments are collected at every visit with the exception of CGI-TS-Improvement, which is only collected during postbaseline visits.

Safety assessments include:

- adverse event (AE) monitoring,
- clinical laboratory tests (hematology, clinical chemistry, and urinalysis),

- serum prolactin
- 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).

Safety assessments are collected at every visit with the exception of serum prolactin and the ADHD Rating Scale-5, which are not collected at screening, and physical examinations, which are not conducted at Weeks 4, 8, and 10.

Blood samples for plasma drug and metabolite concentration analyses are collected at Day 1 (baseline) and at Weeks 2, 4, 6, 8, 10, 12, and 14 (final study visit or at early termination). Subjects/caregivers will be asked to record and provide dosing times from the evening before the treatment period visits when these samples are collected.

5. PLANNED ANALYSES

5.1. Interim Analyses

No interim analysis is planned for this study.

5.2. Final Analyses

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

6. GENERAL CONSIDERATIONS FOR DATA ANALYSES AND HANDLING

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

6.1. General Statistical Procedures

Descriptive and inferential statistical methods will be used to summarize the data from this study. The term "descriptive statistics" refers to the number of subjects, mean, median, SD, standard error of the mean (SEM), minimum, and maximum for numerical variables; and refers to the number and/or percentage of subjects (or events) for categorical variables. Additional descriptive statistics may be presented for selected variables. The term "inferential statistics" refers to hypothesis tests which will be performed to assess differences between the NBI-98854 treatment group 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. The level of significance (Type I error) for declaring statistical significance will be 0.05.

Descriptive statistics will generally be presented separately for each baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups. The treatment groups for all 3 weight group categories are NBI-98854 and Placebo. The NBI-98854 treatment group includes all possible study drug doses (ranging from 20 mg to 80 mg). Select tables will include an "All Subjects" treatment group, which combines all treatment groups and subjects included in the table. Select tables will separate the NBI-98854 treatment group by dispensed dose.

Separate summaries may be combined in the same table for reporting purposes.

Inferential statistics will only be presented for the pooled weight groups.

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. P-values will be displayed using 4 decimal places. These rules may be modified if warranted, based on practical considerations.

6.2. Analysis Sets

6.2.1. Definition of Analysis Sets

6.2.1.1. Safety Analysis Set

The safety analysis set will include all subjects who are randomized to a treatment group, take at least one dose of study drug, and have any postbaseline safety data. The safety analysis set will be used for all summaries of safety data (e.g., AEs and clinical laboratory data) and plasma concentration data. For the summaries of data based on the safety analysis set, subjects who are

dispensed the incorrect treatment at the time of randomization, and remain on the same incorrect treatment during the study, will be assigned to the treatment actually received in all summary tables and figures. Subjects who are dispensed a combination of correct and incorrect treatments during the study will be assigned to the randomized treatment in all summary tables and figures.

6.2.1.2. Full Analysis Set

The full analysis set (FAS) will include all subjects who are randomized to a treatment group and have at least one evaluable TTS change from baseline value reported at a visit (either scheduled or mapped early termination [ET] visit) during the 12-week double-blind treatment period. The FAS will be used for summaries and analyses of efficacy data. Treatment assignment for all summaries and analyses using the FAS will be based on the randomization schedule.

6.2.1.3. Per-Protocol Analysis Set

The per-protocol (PP) analysis set will include all subjects in the FAS who do not have any efficacy-related important protocol deviations (IPDs; Section 7.2) and have a detectable plasma concentration of NBI-98854 at Week 12 (unless randomized to the placebo treatment group). Treatment assignment for all summaries and analyses using the PP analysis set will be based on the randomization schedule.

6.2.2. Summary of Analysis Sets

A summary of the number and percentage of subjects included in (and excluded from, as applicable) each analysis set will be provided for each treatment group. The number and percentage of subjects excluded from each analysis set by reason for exclusion will also be provided. An additional "All Subjects" column will be included. Summaries will be presented for each baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups using all randomized subjects.

6.2.3. Application of Analysis Sets

Summaries of subject disposition, randomization by study site, analysis set inclusion/exclusion status, and IPDs will include all randomized subjects. All other summaries by analysis set are identified in Table 3.

Table 3: Data Summaries by Analysis Set

Data Summary/Analysis	Analysis Set		
	Safety	FAS	PP*
Demographics		X	
Baseline subject characteristics		X	
Medical history	X		
Study drug exposure	X		
Study drug adjustments and dosing		X	
Study drug compliance		X	
Plasma concentrations	X		
YGTSS		X	X
CGI-Tics-Severity		X	X
CGI-TS-Improvement		X	
PUTS		X	
C&A-GTS-QOL		X	
Adverse events	X		
Clinical laboratory data	X		
Serum prolactin	X		
Vital signs	X		
Weight	X		
ECG	X		
C-SSRS	X		
CY-BOCS	X		
CDRS-R	X		
ESRS-A	X		
ADHD rating scale-5	X		
Prior and concomitant medications	X		

* - Pooled weight group only; select tables.

6.3. Baseline Definition

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

6.4. Derived and Transformed Data

6.4.1. Study Day

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

6.4.2. Change from Baseline

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

Percent change from baseline is calculated as (change from baseline/baseline value * 100).

If either the baseline or postbaseline value is missing, the change from baseline and/or percent change from baseline will also be missing. The percent change from baseline will also be missing if the baseline value is equal to zero.

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 14 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). An ET visit at Day 90 or later will be mapped to the Week 14 visit.

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

Table 4: 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 6	42	35-48
Week 8	56	49-62
Week 10	70	63-76
Week 12	84	77-90
Week 14	98	>90

6.5. Handling of Missing Data

6.5.1. Missing Outcome Measures

Missing values for outcome measures will generally not be replaced with imputed values except as noted in Table 4 for the ET visit data mapped to scheduled visits for data summary purposes.

Any imputation methods used for the efficacy endpoints are discussed in the Efficacy Section (Section 9).

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 electronic case report form (eCRF).

The imputation rules for first dose date are as follows:

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

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

- the Week 12 visit date,
- study discontinuation date for subjects who discontinue on or before Week 12,
- the last visit (up to Week 10) prior to discontinuation + 17 or the date of an unscheduled dose reduction +17, whichever occurs later.

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

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

The imputation rules for AE start dates are as follows:

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

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

- If the date is completely missing, the date will be imputed as 01 January in the year of the subject's screening vital signs assessment;

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

7. STUDY POPULATION

7.1. Subject Disposition

The summary of subject enrollment and disposition will display the number of subjects who were randomized to each treatment group, who received at least one dose of study drug, who completed the dose optimization period (i.e., up to Week 6, excluding ET visits mapped to Week 6), who completed the dose maintenance period (i.e., up to Week 12, excluding ET visits mapped to Week 12), and who completed the study (i.e., up to Week 14, excluding ET visits mapped to Week 14). The number of subjects who did not complete the study will also be summarized, both overall and according to the reason for early discontinuation.

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

All of the summaries described in this section will be presented for each baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups using all randomized subjects. Each summary table will include an "All Subjects" column in addition to the columns for each treatment group.

A listing of randomized subjects will also be provided and will include subject ID, informed consent/assent date, randomization date, baseline weight group, and randomized treatment group.

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. 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 and unblinding of the randomized treatment assignments. This committee will review a listing of all 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 include, but are not limited to, the following:

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

A summary of the number and percentage of subjects with IPDs by deviation category and treatment group will be provided. Separate summaries will be presented for each baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups using all randomized subjects. Each summary table will include an "All Subjects" column in addition to the columns for each treatment group.

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. Results will be presented by treatment group. An additional "All Subjects" column will also be included. Separate summaries will be presented for each baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups using the FAS.

Demographics include:

- Age (years)
- Age category (child [ages 6-11], adolescent [ages 12-17])
- Sex
- Ethnicity
- Race

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

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) by treatment group, with SOCs and PTs within each SOC sorted alphabetically. An additional "All Subjects" column will also be included. Separate summaries will be presented for each baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups using the safety analysis set.

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 by treatment group. The frequency and percentage of subjects with the following exposure categories will also be presented:

- >0 to <2 weeks
- ≥2 to <4 weeks
- ≥4 to <6 weeks
- ≥6 to <8 weeks
- ≥8 to <10 weeks
- ≥10 to <12 weeks
- ≥12 weeks

Separate summaries will be presented for each baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups using the safety analysis set.

7.5.2. Dose Adjustments and Dosages

Dose adjustments will be summarized for each postbaseline visit prior to Week 12. The number and percentage of subjects whose dose is escalated (Weeks 2 and 4 only), maintained, or reduced will be presented by treatment group. Dose reductions at unscheduled visits will be included in the summary for the next scheduled visit. Dose reductions at unscheduled visits after Week 10 will be summarized as "After Week 10" in the table. Subjects who prematurely discontinue study at a visit will not be included in the visit summary unless they had a dose reduction at an unscheduled visit after the previous scheduled visit. The table will include the total number and percentage of subjects with a dose reduction at any time during the treatment period. Note that changes in dosage will be presented for the placebo group, even though no actual changes in study drug are occurring.

A summary of study drug dosages in subjects in the NBI-98854 treatment group will be presented for each postbaseline visit through Week 12. Descriptive statistics of the last dose assigned prior to the visit (including at unscheduled visits) will be presented. The number and percentage of subjects receiving each dose will also be presented for each postbaseline visit.

All summaries described in this section will be presented by baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups using the FAS.

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 in each treatment group who are dosing compliant will be presented for each postbaseline visit though Week 12. Separate summaries will be presented by baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups using the FAS.

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 14 visit will reflect the last dose the subject received during the study treatment period (i.e., the dose level at the subject's Week 12 visit). These summary tables will be presented for each baseline weight group (<50 kg vs. ≥50 kg) only. There will not be a pooled weight groups summary.

The summary tables will also be generated for CYP2D6 poor metabolizers vs. non-poor metabolizers within each baseline weight 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 analyte will be summarized with box plots by NBI-98854 dose level at each visit. These will be presented for each baseline weight group (<50 kg vs. ≥50 kg).

The safety analysis set will be used for all plasma concentration summaries.

9. EFFICACY

The efficacy endpoints and analyses 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 5.

Table 5: Efficacy Endpoints and Classification

Efficacy Endpoint	Endpoint Classification
YGTSS TTS change from baseline to Week 12	Primary
CGI-Tics-Severity change from baseline to Week 12	Secondary
YGTSS TTS response at Week 12	Secondary
CGI-TS-Improvement response at Week 12	Secondary
YGTSS TTS change from baseline to Weeks 2-10	Exploratory
CGI-Tics-Severity change from baseline to Weeks 2-10	Exploratory
YGTSS TTS response at Weeks 2-10	Exploratory
CGI-TS-Improvement response at Weeks 2-10	Exploratory
YGTSS Impairment score change from baseline to Weeks 2-12	Exploratory
YGTSS Global Tic Severity score change from baseline to Weeks 2-12	Exploratory
CGI-TS-Improvement score at Weeks 2-12	Exploratory
PUTS total score change from baseline to Weeks 2-12	Exploratory
C&A-GTS-QOL total score change from baseline to Weeks 2-12	Exploratory

9.1. General Considerations

Unless otherwise specified, the FAS will be used for all efficacy analyses. The PP analysis set will also be used for select analyses.

9.2. Statistical Models

9.2.1. Mixed-Effect Model Repeated Measures (MMRM)

The primary analysis of numerical efficacy variables will be a mixed-effect model repeated measures (MMRM) analysis, which includes the changes from baseline (or other dependent variable) to each postbaseline visit (Weeks 2 through 12). The model will include the baseline value as a covariate, and baseline weight group (<50 kg, ≥50 kg), treatment group (NBI-98854, Placebo), visit, treatment group-by-visit interaction, and baseline value-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 55), with most sites anticipated to enroll a small number of subjects.

Treatment group comparisons of the NBI-98854 treatment group vs. Placebo at each visit will be performed by constructing linear contrasts (or equivalent programming code) for differences between treatment group least-squares (LS) means. Nominal (raw) two-sided p-values for testing the significance of these differences and associated 95% confidence intervals will be reported in summary tables.

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. If convergence is not obtained with the unstructured covariance matrix, a Toeplitz covariance structure, followed by the heterogeneous autoregressive covariance structure, followed by the compound symmetry will be used.

9.2.2. Analysis of Covariance

An analysis of covariance (ANCOVA) model will be used for select continuous endpoints. The model will include the baseline value as a covariate, and baseline weight group (<50 kg, ≥50 kg) and treatment group (NBI-98854, Placebo) as fixed effects.

9.3. Controlling for Multiple Comparisons

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

1. Primary endpoint: YGTSS TTS change from baseline to Week 12 (NBI-98854 treatment group vs. Placebo) using the FAS
2. Secondary endpoint: CGI-Tics-Severity change from baseline to Week 12 (NBI-98854 treatment group vs. Placebo) using the FAS
3. Secondary endpoint: TTS response at Week 12 (NBI-98854 treatment group vs. Placebo) using the FAS
4. Secondary endpoint: CGI-TS-Improvement response at Week 12 (NBI-98854 treatment group vs. Placebo) using the FAS

Each step in the sequential testing procedure uses a local 2-sided 0.05 level of significance for the null hypothesis being tested. Testing of hypotheses at each step of the procedure commences only if all null hypotheses in prior steps were rejected.

9.4. Analysis of the Primary Efficacy Endpoint

9.4.1. Primary Efficacy Analysis

The primary efficacy endpoint is the change from baseline to Week 12 in the YGTSS TTS as generated by the certified site rater using the RaterStation™. Changes from baseline to Weeks 2 through 10 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. The TTS value can range from 0 to 50, with higher scores representing greater severity. 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 visit, the TTS value for the visit will also be set equal to missing.

Descriptive statistics will be presented by treatment group for the TTS observed and change from baseline values at baseline and at each postbaseline visit. These summaries will be presented by baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups. The pooled weight groups summary will be repeated for the PP analysis set.

Descriptive statistics will also be presented by visit and treatment group where the NBI-98854 group is separated by the last dose level received prior to the visit (including dose changes that occur at unscheduled visits). These summary tables will be presented for each baseline weight group (<50 kg vs. ≥50 kg) only.

The primary analysis of the change from baseline in TTS will be an MMRM analysis as described in Section 9.2.1. Statistical significance for the treatment comparison at the primary endpoint (Week 12) will be determined using the multiple comparison procedure described in Section 9.3. The primary MMRM analysis will be repeated for the PP analysis set.

As a supportive analysis, the MMRM analyses will be repeated with the following modification: Weeks 2 and 4 will be excluded, as subjects may not yet have had the opportunity to reach their optimal dose.

An additional supportive analysis of the TTS change from baseline to Week 12 will be performed for the pooled weight groups using an ANCOVA model, as described in Section 9.2.2. The analysis will be performed using observed values and repeated using last observation carried forward (LOCF) imputation for missing Week 12 values using the following imputation rules:

- If a subject has a Week 12 visit, including mapped ET visits, then the Week 12 value will be used. If these subjects are missing the Week 12 value, then the last postbaseline observation prior to Week 12 will be used;
- If a subject discontinues study prior to Week 12, the last postbaseline observation on or prior to discontinuation will be used.

Mean (±SEM) values of the TTS at each visit (baseline through Week 14) will be summarized in line graphs by treatment group. Similar graphs will be presented for the changes from baseline. These graphs will be presented by baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups.

The LS means (±SEM) from the primary MMRM analysis results will be summarized in similar line graphs for the pooled weight groups.

An additional graph will be presented for the TTS changes from baseline to Week 12. This will be a display of the empirical cumulative distribution function by treatment group by baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups.

9.4.2. Sensitivity Analyses of the Primary Efficacy Results

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 12 along with the reasons for premature study

discontinuation, 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 FAS.

Tipping Point – Sensitivity Analysis #1

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). To assess the robustness of the MAR assumption, any imputed Week 12 values (from the multiple imputation [MI] procedure described below) for subjects in the NBI-98854 treatment group will be gradually worsened until the treatment difference at Week 12 is no longer statistically significant.

The implementation procedure for the tipping point analysis is described Appendix 14.1.

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 group who drop out prior to the Week 12 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 implementation procedure for the J2R analysis is described Appendix 14.2.

9.5. Analysis of the Secondary Efficacy Endpoints

9.5.1. CGI-Tics-Severity

Change from baseline to Week 12 in the CGI-Tics-Severity score is a secondary endpoint. Changes from baseline to Weeks 2 through 10 are exploratory efficacy endpoints.

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

Analysis of the numerical scores will be identical to those presented for YGTSS TTS (Section 9.4.1), and includes the following:

- Descriptive statistics at baseline and at each postbaseline visit for observed and change from baseline values by treatment group, presented by baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups; the pooled weight groups summary will be repeated using the PP analysis set;

- Descriptive statistics at each postbaseline visit for observed and change from baseline values by treatment group where the NBI-98854 group is separated by the last dose assigned prior to the visit, presented by baseline weight group (<50 kg vs. ≥50 kg) only;
- Primary MMRM analysis for change from baseline through Week 12 for the pooled weight groups; repeated using the PP analysis set;
- MMRM analysis for change from baseline through Week 12, excluding Weeks 2 and 4, for the pooled weight groups;
- ANCOVA analysis for change from baseline to Week 12 on observed data for the pooled weight groups;
- ANCOVA analysis for change from baseline to Week 12 with LOCF imputation for pooled weight groups.

Statistical significance for the treatment comparison for the secondary endpoint (Week 12) will be determined using the multiple comparison procedure described in Section 9.3.

Figures for the numerical scores will be identical to those presented for YGTSS TTS (Section 9.4.1), and includes the following:

- Line graphs of mean (±SEM) values by visit by treatment group, presented by baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups;
- Line graphs of mean (±SEM) change from baseline values by visit by treatment group, presented by baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups;
- Line graphs of LS means (±SEM) from the primary MMRM analysis by visit by treatment group, for the pooled weight group;

In addition to the above analyses, frequency counts using the response categories will be presented by treatment group and visit. This summary will be presented by baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups.

9.5.2. 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. Response status at Week 12 is a secondary endpoint; Weeks 2 through 10 are exploratory efficacy endpoints.

Descriptive statistics will be presented by treatment group for the number and percentage of subjects classified as TTS responders at each postbaseline visit, using observed data. These summaries will be presented by baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups.

For the pooled weight groups, the primary analysis comparing the NBI-98854 treatment group to the Placebo treatment group at each postbaseline visit (excluding Week 14) will be performed using the Cochran-Mantel-Haenszel (CMH) procedure, with baseline weight group as a stratification variable. Any missing response values through Week 12 will be imputed using non-responder imputation (NRI), where any subjects with missing response values for a visit are imputed as non-responders for that visit. Statistical significance for the treatment comparison for

the secondary endpoint (Week 12) will be determined using the multiple comparison procedure described in Section 9.3.

As a supportive analysis, the CMH procedure will also be performed using observed data.

The percentage of subjects classified as TTS responders at each postbaseline visit will be summarized in line graphs by treatment group, using observed values. These graphs will be presented by baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups. The pooled weight groups graph will be repeated (excluding Week 14) using NRI for missing values.

9.5.3. 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 (see Section 9.6.2) is either a "1" ("very much improved") or a "2" ("much improved") at the visit. Response status at Week 12 is a secondary endpoint; Weeks 2 through 10 are exploratory efficacy endpoints.

Descriptive statistics will be presented by treatment group for the number and percentage of subjects classified as CGI-TS-Improvement responders at each postbaseline visit, using observed data. These summaries will be presented by baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups.

For the pooled weight groups, the analysis comparing the NBI-98854 treatment group to the Placebo treatment group at each postbaseline visit (excluding Week 14) will be performed using the CMH procedure, with baseline weight group as a stratification variable. Any missing response values through Week 12 will be imputed using NRI, where any subjects with missing response values for a visit are imputed as non-responders for that visit. Statistical significance for the treatment comparison for the secondary endpoint (Week 12) will be determined using the multiple comparison procedure described in Section 9.3.

As a supportive analysis, the CMH procedure will also be performed using observed data.

The percentage of subjects classified as CGI-TS-Improvement responders at each postbaseline visit will be summarized in line graphs by treatment group. These graphs will be presented by baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups. The pooled weight groups graph will be repeated (excluding Week 14) using NRI for missing values.

9.6. Analysis of the Exploratory Efficacy Endpoints

9.6.1. YGTSS

9.6.1.1. YGTSS Impairment Score

The YGTSS Impairment score can range in value from 0 to 50, with higher scores representing more severe impairment.

Descriptive statistics will be presented by treatment group for the observed and change from baseline values at baseline and at each postbaseline visit. These summaries will be presented by baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups.

The change from baseline at each postbaseline visit through Week 12 will be analyzed using an MMRM analysis for the pooled weight groups, as described in Section 9.2.1.

Mean (±SEM) values of the observed values at each visit (baseline through Week 14) will be summarized in line graphs by treatment group. Similar graphs will be presented for the changes from baseline at each postbaseline visit. These graphs will be presented for the pooled weight groups.

The LS means (±SEM) from the MMRM analysis results will be summarized in a similar line graph for the pooled weight groups.

9.6.1.2. 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.

Descriptive statistics will be presented by treatment group for the observed and change from baseline values at baseline and at each postbaseline visit. These summaries will be presented by baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups.

The change from baseline at each postbaseline visit through Week 12 will be analyzed using an MMRM analysis for the pooled weight groups, as described in Section 9.2.1.

Mean (±SEM) values of the observed values at each visit (baseline through Week 14) will be summarized in line graphs by treatment group. Similar graphs will be presented for the changes from baseline. These graphs will be presented for the pooled weight groups.

The LS means (±SEM) from the MMRM analysis results will be summarized in a similar line graph for the pooled weight groups.

9.6.2. 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

Descriptive statistics will be presented by treatment group for the observed numerical score values at each postbaseline visit. These summaries will be presented by baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups.

The observed values at each postbaseline visit through Week 12 will be analyzed using an MMRM analysis for the pooled weight groups. The model will be similar to that described in Section 9.2.1 with the following changes:

1. Observed values are used instead of changes from baseline;

2. Instead of using baseline score, which is not applicable for this outcome variable, the covariate in the model will be the baseline value of the CGI-Tics-Severity numerical score.

Mean (±SEM) values of the observed values at each visit (Week 2 through Week 14) will be summarized in line graphs by treatment group. These graphs will be presented for the pooled weight groups.

The LS means (±SEM) from the MMRM analysis results will be summarized in a similar line graph for the pooled weight groups.

In addition to the above analyses, frequency counts using the response categories will be presented by treatment group and visit. This summary will be presented by baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups.

9.6.3. 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 by treatment group for the observed and change from baseline values at baseline and at each postbaseline visit. These summaries will be presented by baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups.

The change from baseline at each postbaseline visit through Week 12 will be analyzed using an MMRM analysis for the pooled weight groups, as described in Section 9.2.1.

Mean (±SEM) values of the observed values at each visit (baseline through Week 14) will be summarized in line graphs by treatment group. Similar graphs will be presented for the changes from baseline. These graphs will be presented for the pooled weight groups.

The LS means (±SEM) from the MMRM analysis results will be summarized in a similar line graph for the pooled weight groups.

9.6.4. 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 by treatment group 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. These summaries will be presented by baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups.

For the normalized total scores, the change from baseline at each postbaseline visit through Week 12 will be analyzed using an MMRM analysis for the pooled weight groups, as described in Section 9.2.1.

Mean (±SEM) values of the observed normalized total scores values at each visit (baseline through Week 14) will be summarized in line graphs by treatment group. Similar graphs will be presented for the changes from baseline. These graphs will be presented for the pooled weight groups.

The LS means (±SEM) from the MMRM analysis results will be summarized in a similar line graph for the pooled weight groups.

10. SAFETY

Unless otherwise specified, summaries for each table described in this section will be presented separately for each baseline weight group (<50 kg vs. ≥50 kg) and for the pooled weight groups. Results will be presented by treatment group (NBI-98854 vs. Placebo) using the safety analysis set, as described in Section 6.2.1.1.

Select assessments will include a "Treatment Endpoint" summary. For subjects who complete through Week 12, the Week 12 data will be used. For subjects who discontinue the study prior to Week 12 or have missing data, the last value at any scheduled postbaseline visit or ET visit prior to Week 12 or discontinuation (whichever occurs first) will be used.

10.1. Adverse Events

Adverse events are recorded in the eCRF. Each AE will be coded to SOC and PT using MedDRA (Version 12.0).

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

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) in the NBI-98854 treatment group (or combined NBI-98854 groups, if applicable);
- Frequency of TEAEs by PT, with PT sorted by decreasing frequency (number of unique subjects) in the NBI-98854 treatment group (or combined NBI-98854 groups, if applicable).

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

An AE overview summary table will be provided which summarizes the number and percentage of subjects with any TEAE, any TEAE leading to dose reduction, any TEAE leading to study 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 by treatment group. The number and percentage of subjects with a TEAE resulting in study discontinuation will be presented by PT within SOC (presented in the same method as the primary TEAE table). More than one AE can contribute to study discontinuation per subject. The first line of the table will display the number and percentage of subjects with at least one TEAE leading to study discontinuation.

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

10.1.2. Adverse Events Resulting in Study Drug Dose Reductions

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

10.1.3. Deaths and Other Serious Adverse Events

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

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

10.1.4. Summaries by NBI-98854 Dose

To help understand any potential dose-related safety signals, select summaries will be presented where subjects are analyzed by the NBI-98854 dose dispensed at the Week 4 visit (regardless of any subsequent dose reductions):

- Placebo
- NBI-98854 20 mg
- NBI-98854 40 mg
- NBI-98854 60 mg
- NBI-98854 80 mg
- NBI-98854 All Doses

These summaries will include all subjects who did not discontinue study on or prior to Week 4, and will be presented for the following previously described tables:

- AE overview table
- TEAEs by SOC and PT
- SAEs by SOC and PT

Any TEAEs with a start date on or after the Week 4 visit date will be included.

10.2. Clinical Laboratory Data

The hematology, clinical chemistry, and prolactin data will be summarized with descriptive statistics by treatment group at baseline and at each scheduled postbaseline visit through Week 14. A summary at "Treatment Endpoint" will also be included. Both observed values and changes from baseline will be summarized.

In addition to the summaries described above, the prolactin data will also be summarized by visit and treatment group for each sex separately.

Shift tables will be presented for selected clinical laboratory variables based on the reference range-based categories of "Low," "Normal," or "High." A clinical laboratory variable value will be assigned to one of these three categories according to the reference ranges provided by the central clinical laboratory.

Two shift tables will be presented: shift from baseline to Week 12 (using observed values) and shift from baseline to "Treatment Endpoint". Each shift table will have three rows and three columns, with rows reflecting the reference range category at baseline, and columns reflecting the reference range category at postbaseline. A "Total" row and "Total" column will also be included. Subjects with a missing baseline value or who do not have postbaseline data will 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.

Summaries of sponsor-defined potentially clinically significant (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) will be summarized by treatment. The criteria for identifying PCS clinical laboratory values are provided in Table 6.

Table 6: 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)

Scatter plots of selected variables will be created which display baseline vs. Week 12 values and baseline vs. “Treatment Endpoint” values. Subjects with a baseline and postbaseline value will be included. 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.

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

Repeat clinical laboratory samples may be collected at any time during this study due to either missing or abnormal results. The general rule for summarizing these data is to include the original sample results in summary tables and graphs. Exceptions to this rule are: (1) all available lab values will be used in the PCS tables and (2) if there are missing results from the original samples at screening, the results of a repeat screening sample will be substituted for the missing results in summary tables and graphs.

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 by treatment group at

baseline and at each scheduled postbaseline visit through Week 14. A summary at "Treatment Endpoint" will also be included. 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 by treatment group. The criteria for identifying PCS vital signs values for children (6 to 11 years of age at baseline) and adolescents (12 to 17 years of age at baseline) are provided in Table 7 and Table 8, respectively.

Table 7: Potentially Clinically Significant Criteria for Vital Signs Variables in 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

Table 8: Potentially Clinically Significant Criteria for Vital Signs Variables in 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
Diastolic Blood Pressure	N/A		≥10 mmHg	>90 mmHg		≥10 mmHg
Heart Rate	N/A		≥15 bpm	>110 bpm		≥10 bpm

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

10.4. Body Weight

The body weight data (in units of kilograms) will be summarized with descriptive statistics by treatment group at baseline and at each scheduled postbaseline visit through Week 14. A summary at "Treatment Endpoint" will also be included. 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 in triplicate, the value that represents the greatest degree of abnormality will be used in all summary tables. If less than 3 values are recorded at an assessment, then the average/greatest abnormality of the available value(s) will be used.

The quantitative ECG variables will be summarized with descriptive statistics by treatment group at baseline and at each scheduled postbaseline visit through Week 14. A summary at "Treatment Endpoint" will also be included. Both observed values and changes from baseline will be summarized. Frequency counts and percentages for the ECG interpretation variable categories will be summarized at each scheduled visit. A summary at "Treatment Endpoint" will also be included.

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

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

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

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

- Greater than 30 msec
- Greater than 60 msec

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

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

- Screening/lifetime assessment by treatment group and for "All Subjects"
- Screening/past 1 year assessment by treatment group and for "All Subjects"
- Baseline (Day 1) assessment by treatment group and for "All Subjects"
- All postbaseline assessments (through Week 14, including unscheduled visit assessments) by treatment group.

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

- Suicidal Ideation Items

  (1) Wish to be dead
  (2) Non-specific active suicidal thoughts
  (3) Active suicidal ideation with any methods (not plan) without intent to act
  (4) Active suicidal ideation with some intent to act, without specific plan
  (5) Active suicidal ideation with specific plan and intent

- Suicidal Ideation Category: Any of items (1) through (5)
- Suicidal Behavior Items (not reported for the Screening/past 1 year assessment)

  (6) Preparatory acts or behavior
  (7) Aborted attempt
  (8) Interrupted attempt
  (9) Non-fatal suicide attempt
  (10) Completed suicide

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

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

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

0 = No suicidal ideation
1 = Wish to be dead
2 = Non-specific active suicidal thoughts
3 = Active suicidal ideation with any methods (not plan) without intent to act
4 = Active suicidal ideation with some intent to act, without specific plan
5 = Active suicidal ideation with specific plan and intent

The shift tables will display the number and percentage of subjects within each cell of a 6 x 6 table for each treatment group, with the rows representing the baseline score and the columns representing the maximum score recorded across all postbaseline assessments (including both scheduled and unscheduled visits). Subjects missing either a baseline score or all postbaseline scores will not appear in the table.

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 by treatment group 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 by treatment group 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 by treatment group 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 by treatment group 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 by treatment group 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 treatment group and study period. A subject may take the same medication more than once or multiple medications for a subject may be classified under the same ATC level or preferred name. A subject is counted only once for each level of medication classification within a summary. An “All Subjects” column will be included in these summaries.

11. DEVIATIONS FROM PROTOCOL PLANNED ANALYSIS

The methods of analysis described in this SAP are consistent with the methods described in the clinical study protocol (Amendment 2) with the following modification:

- The TEAE by maximum intensity summary was replaced with a severe TEAE summary; TEAE by maximum intensity will be included in the TEAE overview table.

12. PERFORMANCE QUALIFICATION OF SAS® PROGRAMS

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

13. 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 28 June 2017].

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.

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 Aug;74(8):e772-80.

14. APPENDICES

The following appendices provide additional details regarding the sensitivity analyses for the primary endpoint. The final SAS procedures and code may vary from the provided examples.

14.1. Implementation of Tipping Point Sensitivity Analysis

Part 1: Imputation of Missing Data

1) Create a dataset with one row per subject that includes the following variables: subject, treatment group, baseline weight group, and separate variables for each observed TTS value (baseline, Week 2 through Week 12). Create an additional flag that indicates if the Week 12 value is missing.
2) If the pattern of missing data across study visits during the treatment period is non-monotone, use the Markov chain Monte Carlo (MCMC) method of PROC MI to impute any interim missing TTS values and create a monotone missing data pattern, using the default system values for the parameters CHAIN, NBITER, and NITER (along with IMPUTE=MONOTONE). Create 100 imputed datasets (NIMPUTE=100). Use random number generator seed value of 86753.
3) Use the regression method of PROC MI (random number generator seed = 86753) to impute missing values. Example SAS® statements are provided below using the following variables: "Treatment" represents treatment group (NBI-98854 vs. Placebo), "Weight" represents baseline weight group (<50 kg vs. ≥50 kg), "Baseline" is the baseline TTS value, and Week2-Week12 represent the Week 2 – Week 12 TTS values.

   ```
   PROC MI NIMPUTE=1 SEED=86753 MINIMUM=0 MAXIMUM=50 OUT=Imputed;
   BY _imputation_;
   CLASS Treatment Weight;
   MONOTONE REGRESSION;
   VAR Treatment Weight Baseline Week2 Week4 Week6 Week8 Week10 Week12;
   ```

   Note: if previous step to create 100 monotone datasets was not needed, use NIMPUTE=100 and omit the BY statement.

Part 2: Analysis and delta-adjustment of imputed values

1) For the 100 imputed datasets created in Part 1, if a subject in the NBI-98854 treatment group was originally missing the Week 12 value, add *delta* (see rules below) to the imputed Week 12 value. Note that the maximum delta-adjusted imputed value cannot exceed 50 (the maximum TTS value). Calculate change from baseline to Week 12 for all subjects:
   - If Week 12 is non-missing: change from baseline = Week 12 value – baseline value;
   - If Week 12 is missing and subject is in the Placebo treatment group: change from baseline = imputed Week 12 value (from Part 1, Step 3) – baseline value;

- If Week 12 is missing and subject is in the NBI-98854 treatment group: change from baseline = delta-adjusted imputed Week 12 value (from Part 2, Step 1) – baseline value.

2) Analyze the change from baseline to Week 12 in each of the 100 imputed datasets using ANCOVA (as specified Section 9.2.2) using PROC MIXED. Use PROC MIANALYZE to combine the results and obtain the LS mean statistics (within treatment, and treatment differences).

3) If the p-value in the previous step was not significant (i.e., >0.05), then no additional analyses are needed. If the p-value was significant, repeat the previous Part 2 steps using the next value of *delta*. Continue to do this until significance is lost, or the maximum *delta* value is reached.

The following rules will be used for defining *delta*:

- The initial delta will be = 0 (i.e., an analysis of the original, multiply imputed datasets)
- The second delta will be = 1. If this is significant, continue to increase delta by 1 until significance is lost, or until reaching delta = 50 (the maximum value for TTS).

If significance is not lost or is lost slowly (such as after delta = 20), some values of delta may be omitted from the final table.

The tipping point will be defined as the value of delta that results in loss of significance.

14.2. Implementation of Jump to Reference Sensitivity Analysis

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 86753, 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 group, one visit at a time in a sequential fashion. At each visit imputation step, PROC MI (using BY _imputation_ to separate by the 100 imputed data sets created in the previous paragraph) will be implemented with the same seed specified in the preceding paragraph, using the monotone regression method with weight 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 12 using the imputed data sets and then combining the results of these analyses using PROC MIANALYZE.