Statistical Analysis Plan for Interventional Studies

Sponsor Name: Emalex Biosciences, Inc.

Protocol Number: EBS-101-CL-001

Protocol Title:
A Multicenter, Placebo-Controlled, Double-Blind, Randomized, Parallel-Group, Phase 2b Study to Evaluate the Efficacy and Safety of Ecopipam Tablets in Children and Adolescent Subjects with Tourette's Syndrome

Protocol Version and Date: (DD-Mmm-YYYY):
1.0, 24-Dec-2018
2.0, Amendment 1, 07-Feb-2019
3.0, Amendment 2, 28-Feb-2019
4.0, Amendment 3, 28-May-2019
5.0, Amendment 4, 25-Oct-2019
6.0, Amendment 5, 13-Apr-2020

Syneos Health Project Code: 7001736

Author: PPD

Revision History

Version #	Date (DD-Mmm-YYYY)	Document Owner	Revision Summary
1.0	26-May-2020	PPD	Final Version 1.0
1.1	3-May-2021	PPD	Updates based on FDA 01 September 2020 comments and sponsor requests. Major changes include: clearly defined the estimand of primary interest, providing details about sensitivity analyses and adding programming codes to the appendix.
1.2	28-May-2021	PPD	Updates based on sponsor comments and additional requests. Major changes include: adding subgroups for efficacy summaries, providing tipping point analysis and programming code, adding details about AEs of special interest and adding AE by subgroup and time interval summaries.
2.0	Jun-2021	PPD	Clean up and final version.
3.0	Sep-2021	PPD	Updates based on FDA response. Updates include edits to the primary estimand and intercurrent events.

I confirm that I have reviewed this document and agree with the content.

Approvals

Syneos Health Approval

PPD Lead Biostatistician	PPD Signature	PPD Date (DD-Mmm-YYYY)

Emalex Biosciences, Inc. Approval

PPD PPD	PPD Signature	PPD Date (DD-Mmm-YYYY)

Table of Contents

1. Glossary of Abbreviations

Abbreviation	Description
ADR	Adverse Drug Reaction
AE	Adverse Event
AESI	Adverse Event of Special Interest
AIMS	Abnormal Involuntary Movement Scale
ATC	Anatomic Therapeutic Classification
BARS	Barnes Akathisia Rating Scale
C&A-GTS-QOL	Child and Adolescent Gilles De La Tourette Syndrome-Quality of Life Scale
CaGI-C	Caregiver Global Impression of Change
CDRS-R	Children's Depression Rating Scale-Revised
CGI	Clinical Global Impression
CGI-TS-I	Clinical Global Impression Tourette's Syndrome of Improvement
CGI-TS-S	Clinical Global Impression Tourette's Syndrome of Severity
CHW	Cui-Hung-Wang
CSR	Clinical Study Report
CTC	Common Toxicity Criteria
CTCAE	Common Toxicity Criteria for Adverse Events
CY-BOCS	Child Yale-Brown Obsessive Compulsive Scale
DSMB	Data Safety Monitoring Board
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
GCP	Good Clinical Practices
GS	Global score
HCl	Hydrochloride
HbA1c	Hemoglobin A1c
ICH	International Conference on Harmonization
MAR	Missing at Random
MNAR	Missing Not at Random
MedDRA	Medical Dictionary for Regulatory Activities
mITT	Modified Intent to Treat

Abbreviation	Description
NCI	National Cancer Institute
PARS	Pediatric Anxiety Rating Scale
PK	Pharmacokinetics
PO	Oral (per os)
PP	Per-Protocol
SAE	serious adverse event
SAP	Statistical Analysis Plan
SD	standard deviation
SNAP-IV	Swanson, Nolan, and Pelham questionnaire
SS	Safety Set
TEAE	Treatment emergent adverse event
TS	Tourette's Syndrome
TTS	Total Tic Score
YGTSS-TTS	Yale Global Tic Severity Scale-Total Tic Score

2. Purpose

The purpose of this statistical analysis plan (SAP) is to ensure that the data listings, summary tables, and figures which will be produced, and the statistical methodologies that will be used, are complete and appropriate to allow valid conclusions regarding the study objectives.

2.1. Responsibilities

Syneos Health will perform the statistical analyses and are responsible for the production and quality control of all tables, figures, and listings.

2.2. Timings of Analyses

The primary analysis of safety and efficacy and/or pharmacokinetics is planned after all subjects complete the final study visit or terminate early from the study.

Interim analyses will be completed in coordination with the Data Safety Monitoring Board (DSMB) meetings to primarily assess safety and preliminary efficacy. An interim analysis will be conducted after approximately 50% of subjects have enrolled and completed treatment or early termination procedures are completed. See Section 11 for complete details.

3. Study Objectives

3.1. Primary Objective

The primary objective of this study is to evaluate the efficacy of ecopipam tablets in pediatric subjects (aged ≥ 6 to <18 years) with Tourette's Syndrome (TS).

3.2. Secondary Objective(s)

The secondary objectives of this study are to evaluate the safety of ecopipam tablets in pediatric subjects (aged ≥6 to <18 years) with TS and characterize the pharmacokinetics (PK) of ecopipam.

3.3. Brief Description

This is a multicenter, placebo-controlled, double-blind, randomized, parallel-group, Phase 2b study in pediatric subjects (aged ≥6 to <18 years) with TS. Following a Screening period up to 28 days and Baseline visit, eligible subjects will be randomized 1:1 to receive either a target steady-state dose of 2 mg/kg/day ecopipam HCl or matching placebo for a 4-week Titration period followed by an 8-week Treatment period. Following titration, subjects who weigh ≥18 kg to ≤23 kg will receive a 37.5-mg active dose (or placebo), subjects who weigh >23 kg to ≤34 kg will receive a 50-mg active dose (or placebo), subjects who weigh >34 kg to ≤44 kg will receive a 75-mg active dose (or placebo), subjects who weigh >44 kg to ≤68 kg will receive a 100-mg active dose (or placebo), subjects who weigh >68 kg to ≤83 kg will receive a 150-mg active dose (or placebo), and subjects who weigh >83 kg will receive a 200-mg active dose (or placebo).

After the Screening period, subjects will return to the clinic at Baseline and at 4, 6, 8, and 12 weeks after Randomization. Subjects will have telephone visits on Weeks 2 and 10 to assess adverse events and other safety parameters. Efficacy assessments will be conducted at Weeks 4, 6, 8, and 12 and safety assessments will be conducted at all visits as well. A Follow Up visit will be conducted in the clinic 7 and 14 days after the last dose of study medication. Subjects will also be contacted 30 days after the last dose of study medication to determine any adverse events. At study completion or early termination, subjects will be tapered off study drug as outlined in the protocol.

3.4. Subject Selection

3.4.1. Inclusion Criteria

Subjects eligible for inclusion in this study must fulfill all of the criteria listed in section 8.1 of the protocol.

3.4.2. Exclusion Criteria

Subjects will not be eligible if they meet any of the criteria listed in section 8.2 of the protocol.

3.5. Determination of Sample Size

A sample size of 75 subjects per group will provide >80% power to detect difference between groups of 3 points on change from baseline the Yale Global Tic Severity Scale-Total Tic Score (YGTSS-TTS), with standard deviation = 6.1 and alpha = 0.05, assuming 10% drop out.

3.6. Treatment Assignment & Blinding

Subjects will be randomized 1:1 to either ecopipam or matching placebo. Randomization will be stratified by weight band (≥18 to ≤23 kg, >23 to ≤34 kg, >34 to ≤44 kg, >44 to ≤68 kg, >68 to ≤83 kg, >83 kg).

Throughout the study, subjects and all personnel involved with the conduct and interpretation of the study, including the subject, parents/guardians, investigators, site personnel, and sponsor staff will be blinded to the treatment codes. Randomization data will be kept strictly confidential, filed securely by an appropriate group with the Sponsor or Syneos and accessible only to authorized persons (e.g., Safety) until the time of unblinding.

A master list of all treatments and the subject numbers associated with them will be maintained electronically by the unblinded clinical supply vendor and the unblinded personnel at Syneos. The process to request unblinding and the randomization code will be outlined in a separate document. The site will be trained in this process and it should be used only in an emergency. These codes should only be broken if knowledge of the subject's randomization code will affect his/her medical treatment. If possible, before breaking the blind, the Investigator should consult with the Sponsor to ascertain the necessity of breaking the code. The Investigator is to record the date and time of requesting the code and the reason for breaking the code.

3.7. Administration of Study Medication

During the 4-week titration period, the following ecopipam HCl doses or matching placebo will be administered oral (per os) (PO) for each of the weight bands:

- Those who weigh ≥18-≤23 kg will titrate from 12.5 mg daily to 37.5 mg daily: 12.5 mg during Week 1, 25 mg during Week 2, and 37.5 mg during Weeks 3 and 4.
- Those who weigh >23-≤34 kg will titrate from 12.5 mg daily to 50 mg daily: 12.5 mg during Week 1, 25 mg during Week 2, 37.5 mg during Week 3 and 50 mg during Week 4.
- Those who weigh >34-≤44 kg will titrate from 12.5 mg daily to 75 mg daily: 12.5 mg during week 1, 25 mg during week 2, 50 mg during Week 3 and 75 mg during Week 4.
- Those who weigh >44-≤68 kg will titrate from 25 mg daily to 100 mg daily: 25 mg during Week 1, 50 mg during Week 2, 75 mg during Week 3 and 100 mg during Week 4.
- Those who weigh >68-≤83 kg will titrate from 25 mg daily to 150 mg daily: 25 mg during Week 1, 50 mg during Week 2, 100 mg during Week 3 and 150 mg during Week 4.
- Those who weigh >83 kg will titrate from 25 mg daily to 200 mg daily: 25 mg during Week 1, 50 mg during Week 2, 100 mg during Week 3, and 200 mg during Week 4.

Subjects who do not tolerate the dose titration up to the full designated dose for their weight stratum will be discontinued from the study. These subjects will also be tapered off their current dose of study drug according to their weight stratum.

During the 8-week treatment phase, the following ecopipam HCl doses or matching placebo will be administered for each of the weight bands:

- Those who weigh ≥18-≤23 kg will receive 37.5 mg daily.
- Those who weigh >23-≤34 kg will receive 50 mg daily.
- Those who weigh >34-≤44 kg will receive 75 mg daily.
- Those who weigh >44-≤68 kg will receive 100 mg daily.
- Those who weigh >68-≤83 kg will receive 150 mg daily.

- Those who weigh >83 kg will receive 200 mg daily.

All doses will be administered PO once daily in the evening. Subjects who have changes in weight during the study will not have their doses adjusted for the duration of the study.

At the end of the 8-week treatment phase, subjects will titrate off therapy and receive ecopipam HCl doses or matching placebo that will be reduced by 25 mg/day until off of study drug. Signs of symptoms of withdrawal will be monitored.

3.8. Study Procedures and Flowchart

After the screening period, subjects will return to the clinic at Baseline and 4, 6, 8 and 12 weeks after Randomization. Subjects will have telephone visits on Weeks 2 and 10 to assess adverse events and other safety parameters. Efficacy assessments will be conducted at weeks 4, 6, 8 and 12. Safety assessments will be conducted at these visits as well. A Follow Up visit will be conducted in the clinic 7 and 14 days after the last dose of study medication. Subjects will also be contacted 30 days after the last dose of study medication to determine any adverse events. At study completion or early termination, subjects will be tapered off study drug as outlined in the protocol. See Appendix section 14.1 for the Study of Assessments Flowchart.

4. Endpoints

4.1. Primary Efficacy Endpoint

The primary efficacy endpoint is the change in the Yale Global Tic Severity Scale – Total Tic Score (YGTSS-TTS, i.e., sum of the motor and phonic tic scores) from baseline (YGTSS-TTS score from Baseline visit) to end of therapy (YGTSS-TTS score from Week 12).

4.1.1. Primary Efficacy Estimand

The mean change from baseline in the Yale Global Tic Severity Scale – Total Tic Score to end of therapy are compared between ecopipam and placebo for patients aged 6 to less than 18 years eligible for the trial.

The primary estimand is defined as the following:

Treatment: 4-week titration period followed by 8-week treatment phase with ecopipam HCl doses or matching placebo.

Target Population: all children and adolescent patients with Tourette's syndrome who were randomized, received at least one dose of study drug, and had at least one post Baseline scoring of the YGTSS (mITT population).

Variable: the change in the Yale Global Tic Severity Scale (YGTSS) from baseline to Week 12.

Intercurrent events: For the intercurrent events, receiving prohibited concomitant medication, change in dosing of allowed concomitant medication, and no dosing adjustment with BMI change, the treatment policy strategy will be used for the primary analysis, meaning the observed data at Week 12 will be used for the change from baseline in YGTSS. For the intercurrent event of study treatment discontinuation due to lack of efficacy and treatment related adverse events, for subjects with no YGTSS at Week 12, the hypothetical strategy will be used, meaning the non-collected data of YGTSS at Week 12 will be multiple imputed using similar subjects (relevant demographic/baseline characteristics) from the placebo arm. The intercurrent event 'death prior to Week 12' will be handled using the 'while alive and on-treatment strategy' using the last observed YGTSS assessment for the analysis of Week 12.

Population-level summary: treatment difference between ecopipam and placebo treatment groups using MMRM model in mITT population after having applied the hypothetical strategy and the while-on-treatment strategy, respectively, to the non-collected / non-observed Week 12 YGTSS assessments specified above.

The primary efficacy analysis approach is in line with the primary estimand. Potential confounders will be identified and sensitivity and/or supplementary analysis may be used to deal with missing data and assess the robustness of conclusions.

4.2. Secondary Efficacy Endpoints

Key Secondary Efficacy Endpoint:

- Change in Clinical Global Impression of Tourette Syndrome Severity (CGI-TS-S) from Baseline to Week 12

Other Secondary efficacy endpoints include:

- Change in Clinician Global Impression Tourette Syndrome of Improvement (CGI-TS-I) from Week 4 to Week 12
- Change in YGTSS-Global Score (GS) from Baseline to Week 12
- Change in Caregiver Global Impression of Change (CaGI-C) from Week 4 to Week 12
- Change in Gilles de la Tourette Syndrome–Quality of Life Scale for Children and Adolescents (C&A-GTS-QOL) from Baseline to Week 12
- Percentage of subjects with a 25%, 50%, and 75% improvement on the YGTSS-TTS
- Percentage of subjects with complete remission of tics on the YGTSS-TTS (100% improvement)
- Change in YGTSS-TTS from Baseline to Weeks 4, 6, and 8
- Change in CGI-TS-S from Baseline to Weeks 4, 6, and 8
- Change in CGI-TS-I from Baseline to Weeks 4, 6, and 8
- Change in YGTSS-GS from Baseline to Weeks 4, 6, and 8
- Change in CaGI-C from Baseline to Weeks 4, 6, and 8
- Change in the C&A-GTS-QOL from Baseline to Weeks 4, 6, and 8.

4.3. Exploratory Endpoints

To evaluate the PK/pharmacodynamic relationship between plasma levels of ecopipam and treatment effect if possible.

5. Analysis Sets

5.1. Screened Set

The Screened Set will include all subjects screened. Unless specified otherwise, this set will be used for subject listings and for summaries of subject disposition.

5.2. Safety Set

The Safety Set (SS) will include all subjects who received at least one dose of study drug. The SS will be used for the analysis of the safety endpoints.

5.3. Modified Intent-to-Treat Set

The Modified Intention-to-Treat (mITT) set will include all randomized subjects who received at least one dose of study drug and have at least one post Baseline scoring of the YGTSS. The mITT set will be used for the analysis of primary and secondary efficacy endpoints. Subjects will be analyzed according to randomized treatment.

5.4. Per Protocol Set

The Per-Protocol (PP) set will include subjects from mITT set who have no major protocol deviations. Before data are released for statistical analysis, a blinded review of all data will be performed by the Sponsor's clinical team to identify protocol deviations that may potentially affect the results. At this time, it will be determined if subjects and/or data should be excluded from the PP set. The list of subjects or observations to be excluded from the PP set, along with the reason for exclusion, will be finalized prior to database unblinding.

5.5. Pharmacokinetic Set

The Pharmacokinetics (PK) set will include all subjects from safety set who have a valid concentration measurement.

5.6. Protocol Deviations

Protocol deviations will be collected and categorized as shown in the protocol deviation and non-compliance management plan.

6. General Aspects for Statistical Analysis

6.1. General Methods

All analyses and outputs will be produced using SAS® version 9.4 or later. Unless otherwise specified, efficacy and safety summaries will be presented for each treatment. All other summaries will be presented for each treatment and overall. Continuous variables will be summarized using the number of observations (n), mean, standard deviation (SD), median, minimum, and maximum. Categorical variables will be summarized using number of observations (n), frequency, and percentages of subjects.

All relevant subject data will be included in listings. All randomized subjects entered into the database will be included in subject data listings. All by visit summaries will use the nominal visit. Unscheduled visits will not be summarized but will be included in the listings.

6.2. Key Definitions

Treatment emergent adverse events (TEAEs) are defined as any adverse events that start or increase in intensity on or after the first dose of Double-blind treatment on Day 1.

Timing specific adverse events are defined as any adverse events that occur or increase in intensity during any specifically defined study milestones: treatment phase (TPAE), taper down (TDAE) and off study drug (ODAE).

Duration of exposure is defined as (Date of Last Dose – Date of First Dose) +1. If date of last dose is unknown then the date of last clinical visit will be used to impute the date of last dose.

Baseline is defined as the last measurement taken before the first dose of double-blind treatment on Day 1.

6.3. Missing Data

Partial dates of medications will be imputed solely for the purpose of defining prior/concomitant status for medications. Dates will be defined using the hierarchy of derivations below.

- For missing start day where month and year are present, the start day will be set to the 1st of the month, unless the month and year are the same as the first dose month and year and the 1st of the month is before the first dose date, in which case, the start date will be set to the first dose date.

- For missing start day and month where year is present, the start day and month will be set to January 1st, unless the year is the same as the first dose year and January 1st is before the first dose date, in which case, the start date will be set to the first dose date.

- For missing end day where month and year are present, the end day will be set to the last day of the month, unless the month and year are the same as the trial termination month and year, in which case, the end date will be set to the trial termination date.

- For missing end day and month, where year is present, the end date will be set to the trial termination date if the years are the same. If the trial termination year is greater than the end year, the end day and month will be set to December 31st.

Partial onset dates for adverse events will be imputed in the same manner described above for partial start dates for the purpose of defining treatment emergent status.

6.4. Visit Windows

Visit window will not be used in the summarization or analysis of data.

6.5. Pooling of Centers

Analyses will not include study site as a factor in the model. Region (North America and Europe) will be included, so sites will not be pooled based on the number of subjects, but by region.

6.6. Subgroups

Subgroups to be analyzed for YGTSS, CGI-S and CGI-I, CaGI-C and C&A-GTS-QOL as below:

- Baseline Age Group (6-11 years, 12-17 years)
- Region: North America and Europe
- Gender: Male and Female
- Prior TS Treatment: Yes or No
- YGTSS TTS Baseline Score Group: > = mean, < mean
- YGTSS TTS Baseline Score Group: > 2 SD from mean, within 2 SD of mean
- Body Weight Group: > 100Kg.

7. Demographic, Other Baseline Characteristics and Medication

7.1. Subject Disposition and Withdrawals

Subject disposition will be presented for all subjects, which include the following:

- Number of subjects enrolled
- Number of screened and reason for screen failure
- Number (%) of subjects randomized

Among the randomized subjects, the following will be summarized

- Number (%) of subjects in the Safety Set
- Number (%) of subjects in the Modified Intent-to-Treat Set
- Number (%) of subjects in the Per-Protocol Set
- Number (%) of subjects who completed the study
- Number (%) of subjects who discontinued study prematurely and their reason

A separate by-subject listing of subject disposition and withdrawal will also be provided. Subjects who screen failed will be listed along with the date and reason for the screen failure.

Randomized subjects not included in an analysis set and their reason for exclusion will be summarized and listed.

7.2. Demographic and Other Baseline Characteristics

Demographics and other baseline characteristics will be summarized for the Modified Intent-to-Treat Set by randomized treatment and overall. Summary statistics and by-subject listings will be provided.

Demographics and baseline characteristics will include age, sex, ethnicity, race, weight, height, body mass index (BMI).

Age at Study day 1 = (Study day 1 visit date - date of birth + 1) / 365.25 and truncated to complete years.

Height (in cm) = height (in inches) * 2.54

Weight (in kg) = weight (in lbs) * 0.4536

BMI (kg/m^2) = Weight(kg)/[Height(m)^2]

7.3. Medical History and Concomitant Diseases

A summary table of the number and percentage of subjects by medical history, system organ class (SOC) and preferred term will be produced from the Safety Set. Medical history will be sorted alphabetically by SOC and in descending order of subjects per preferred term within each SOC.

Medical history will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) version 22.0 or higher.

A separate by-subject listing of medical history will also be provided.

7.4. Medication

All prior and concomitant medications will be summarized based on classification using the Anatomical Therapeutic Chemical (ATC) classification and preferred drug name from the World Health Organization Drug Dictionary, version Mar 2019, or later.

A separate by-subject listing of medications will also be provided.

7.4.1. Prior Medication

Prior medications are defined as medications entered in the eCRF that either start, or end before the first dose of study medication. Prior medications will be summarized by ATC level 2 and preferred drug name for the Safety Set.

Prior medications which continue after first dose of study medication will also be classified as a concomitant medication.

7.4.2. Concomitant Medication

Concomitant medications are defined as medication on the eCRF that are taken on or after or are ongoing at the start date of dosing. Concomitant medications will be summarized by ATC level 2 and preferred drug name for the Safety Set.

8. Efficacy

The mITT set will be used for the analysis of the efficacy endpoints.

8.1. Primary Efficacy Endpoint and Analysis

8.1.1. Primary Efficacy Analysis

The Yale Global Tic Severity Scale (YGTSS) is a clinician-completed rating scale used to quantify overall tic severity as well as specific subdomains of tic number, frequency, duration, intensity, and complexity. Each of these subdomains is scored, on a 5-point scale, separately for motor and vocal tics and then summed across both motor and vocal tics to yield a total tic score (TTS) ranging from 0 to 50. The YGTSS also provides for an overall impairment rating (0 = "none" to 50 = "severe"). The YGTSS has demonstrated acceptable internal consistency, good interrater reliability, and acceptable convergent and divergent validity.

YGTSS will be assessed at Screening, Baseline and at Weeks 4, 6, 8 and 12.

The primary efficacy endpoint for this trial will be the change from Baseline (score of YGTSS-TTS {i.e., sum of the motor and phonic tic scores} at Baseline) to end of therapy (score of YGTSS-TTS at Week 12). The analysis of primary efficacy endpoint will use the MITT set. Results for the PP set will also be provided.

Statistical analysis:

The primary endpoint will be analyzed with a Mixed Model for Repeated Measures (MMRM) model using the SAS MIXED procedure with REPEATED statement. The model will include fixed effects for treatment, visit, interaction between treatment and visit, region, age group (0 = children 6-11, 1 = adolescents 12-17) and covariate for baseline YGTSS-TTS.

If there is no sample size increase at the interim analysis, then a single MMRM based on all the data will be performed. However, if the sample size is increased from the planned (150) then two MMRM models will be performed and combined using the CHW (Cui-Hung-Wang) method, which controls the type 1 error, one MMRM model on the data from Stage 1 (the interim analysis data) and the second on data from Stage 2 (post-interim data). Information on treatment effect from the two stages will be combined in the weighted test-statistic:

$$T_{CHW} = \sqrt{\frac{n_1}{N}}T_1 + \sqrt{\frac{n_2}{N}}T_2$$

T_{CHW} will be compared to a standard Gaussian distribution to determine the level of statistical significance. Here n_1 is the **actual** number of subjects in Stage 1 (planned to be 75 subjects), n_2 is the **planned** number of Stage 2 subjects before any sample size increase (i.e., 150 minus the actual number in Stage 1), N is the sum of n_1 and n_2 (i.e., 150), and T_1 and T_2 are the t-statistics for treatment difference from MMRM models on data from, respectively, Stage 1 and Stage 2 data. At a given stage, the mean difference Δ_{e-p} (treatment effect of ecopipam versus placebo) can be estimated by the difference $\hat{\delta} = LSM_e - LSM_p$ in least squares (LS) means of change from baseline at Week 12 between the ecopipam and placebo treatment groups, respectively. The LSMEANS statement with COV option will report the estimated covariance of LS means, which can be used to derive the variance $\widehat{var}(\hat{\delta})$ of the difference $\hat{\delta}$. The t-statistic T_1 or T_2 can then be calculated as:

$$T = \frac{\hat{\delta}}{\sqrt{\widehat{var}(\hat{\delta})}}$$

If the planned sample size is adjusted and the CHW method used to analyze the primary endpoint, the MMRM model and sensitivity analyses of the complete dataset as if the sample size had been planned from the beginning will be performed to provide complementary interpretation to the CHW analysis.

8.1.2. Sensitivity Analysis

The primary efficacy analysis on change in the YGTSS-TTS at week 12 is to use MMRM model using all observed data. This model is appropriate when data is missing at random (MAR). To assess the robustness of the primary efficacy analysis, the following additional sensitivity analyses for missing data will be included for the mITT set:

1. Missing at Random (MAR) predictive mean matching multiple imputation method
2. Missing Not at Random (MNAR) jump to reference multiple imputation method
3. Tipping point under MNAR assumption

Details of sensitivity analysis for missing data and the statistical model are explained below.

8.1.2.1. Sensitivity analyses for missing data

Sensitivity analyses for missing data includes MAR predictive mean matching multiple imputation method, MNAR jump to reference multiple imputation method, and MNAR tipping point analysis. Details for the sensitivity analysis for missing data are given below:

1) MAR predictive mean matching multiple imputation method

All missing data will be imputed for patients missing data at week 12 using the predictive mean matching multiple imputation method (Heitjan and Little 1991, Schenker and Taylor 1996) for a MAR data assumption using fully conditional specification (FCS) method. The imputation model will include age group at baseline, region, treatment arm, baseline YGTSS-TTS, and YGTSS-TTS at all visits up to week 12 where the YGTSS is scheduled to be collected. This imputation will be performed using the SAS MI procedure with fcs specification. The resulting complete, imputed datasets will each be analyzed using the same model as the primary analysis model, and the resulting statistics combined using methodology provided by Rubin (1987) and Little and Rubin (2002). The SAS code for the multiple imputation methodology is detailed in the Appendix (Section 14.3).

2) MNAR jump to reference multiple imputation method

Data will be imputed for patients missing data at week 12 using similar predictive mean matching multiple imputation method. Under a MNAR data assumption, patients will have their missing data at week 12 imputed as if they behaved like placebo treated patients after dropout. Only placebo patients with complete data will be included in the imputation model. The imputation models will include age group at baseline, region, baseline YGTSS-TTS score, and YGTSS-TTS scores at all visits up to week 12 where the YGTSS is scheduled to be collected. After obtaining complete data sets, the complete data sets will be used in the analysis. The resulting complete, imputed datasets will each be analyzed using the same model as the primary analysis model, and the resulting statistics combined using methodology provided by Rubin (1987) and Little and Rubin (2002). The SAS code for the multiple imputation methodology is detailed in the Appendix (Section 14.4).

3) Tipping point analysis:

In the event that statistical significance in favor of ecopipam is determined from the primary analysis, subjects in the ecopipam group with a missing YGTSS-TTS score (change from baseline) will be assigned successively more extreme values, whilst subjects in the placebo group with a missing YGTSS-TSS score (change from baseline) will be assigned successively less extreme values, to find the point at which statistical significance is lost (i.e. the 2-sided P-value becomes greater than 0.05)

Details for tipping point analysis are described below:

a. For patients that completed the study, intermittent missing data are considered as MAR and will be imputed by predictive mean matching multiple imputation method (Heitjan and Little 1991, Schenker and Taylor 1996). The imputation model will include age group at baseline, region, treatment arm, baseline TTS score, and TTS scores at all visits up to the imputed visit where the TTS is scheduled to be collected.

b. For patients who discontinued due to any reason, all missing data will be imputed using the same predictive mean matching multiple imputation method. The imputed values will be adjusted under MNAR assumption:

   - Patients randomized to ecopipam: the first missing value after patients dropout will be analyzed assuming the unobserved outcome is on average worsened by δ1 (where δ1 = 1 to 10 or estimated treatment effect, whichever is higher, in step of 1) compared to the patients who have no missing value;

   - Patients treated by placebo: the first missing value after patients dropout will be analyzed assuming the unobserved outcome is on average better by δ2 (where δ2 = 1 to 10 or estimated treatment effect, whichever is higher, in step of 1) compared to the patients who have no missing value;

   The penalty or reward will not be applied to the values from later time points as these values are penalized or rewarded through condition on previous time points. For each combination of (δ1, δ2), 100 imputed datasets will be obtained.

c. The resulting complete, imputed datasets will each be analyzed using the same model as the primary analysis model, and the resulting statistics combined using methodology provided by Rubin (1987) and Little and Rubin (2002).

The complete multiple imputation methodology, along with SAS code, is detailed in Appendix (Section 14.5).

8.2. Secondary Efficacy Endpoint(s) and Analyses

A hierarchical (fixed-sequence) testing approach will be used for the analysis of the primary and secondary endpoints to maintain the experiment-wise type I error rate of 5% (2.5% for one-sided tests). If an endpoint is not statistically significant, confirmatory hypothesis testing will not be carried out on the remaining hypotheses, and remaining hypotheses testing will be considered exploratory rather than confirmatory. The order of testing following the primary efficacy endpoint is outlined below:

- Change in Clinical Global Impression of Tourette Syndrome Severity (CGI-TS-S) from Week 4 to Week 12

- Change in Clinical Global Impression of Tourette Syndrome of Improvement (CGI-TS-I) from Baseline to Week 12
- Change in the YGTSS-global Score from Baseline to Week 12
- Caregiver Global Impression of Change (CaGI-C) at Week 12
- Change in Gilles de la Tourette Syndrome-Quality of Life Scale for Children and Adolescents (C&A-GTS-QOL) from Baseline to Week 12
- Percentage of subjects with a 25%, 50% and 75% improvement on the YGTSS-TTS
- Percentage of subjects with complete remission of tics on the YGTSS-TTS (100% improvement)
- Change in the YGTSS-TTS from Baseline to Weeks 4, 6, 8
- Change in the CGI-TS-S from Baseline to Weeks 4, 6, 8
- Change in the CGI-TS-I from Baseline to Weeks 4, 6, 8
- Change in the YGTSS-GS from Baseline to Weeks 4, 6, 8
- Change in the C&A-GTS-QOL from Baseline to Weeks 4, 6, 8

Each endpoint and analyses are defined in the following sections (8.2.1-8.2.13). Continuous endpoints using MMRM analyses will use the CHW method if the sample size is adjusted, as defined in section 8.1.

8.2.1. Change in Clinical Global Impression of Tourette Syndrome Severity (CGI-TS-S) from Baseline to Week 12 (Key Secondary Efficacy Endpoint)

The CGI consists of two reliable and valid 7-item Likert scales used to assess severity and change in clinical symptoms. The improvement scale will be used at every visit after the Screening and Baseline Visits. The scale ranges from 1 = "very much improved" to 7 = very much worse." The CGI severity scale will be used at each study site visit and ranges from 1 = "not ill at all" to 7 = "among the most extremely ill."

A secondary analysis will be performed on the change in CGI-TS-S from Baseline to Week 12.

Statistical analysis:

The analysis method will be the same as the primary efficacy analysis as described in **Section 8.1.1.**

Summary statistics will be presented along with a full data listing by treatment group.

8.2.2. Change in Clinician Global Impression Tourette Syndrome of Improvement (CGI-TS-I) from Baseline to Week 12

The CGI consists of two reliable and valid 7-item Likert scales used to assess severity and change in clinical symptoms. The improvement scale (CGI-TS-I) will be used at every visit after the Screening and Baseline Visits. The scale ranges from 1 = "very much improved" to 7 = very much worse." The CGI severity scale (CGI-TS-S) will be used at each study site visit and ranges from 1 = "not ill at all" to 7 = "among the most extremely ill."

A secondary analysis will be performed on the CGI-TS-I at Week 12.

Statistical analysis:

The analysis method will be the same as the primary efficacy analysis as described in **Section 8.1.1**, with the exception of no baseline score will be used as a covariate.

Summary statistics will be presented along with a full data listing by treatment group.

8.2.3. Change in YGTSS-Global Score (GS) from Baseline to Week 12

A secondary analysis will be performed on the change in YGTSS-Global Score (YGTSS-GS) from Baseline to Week 12.

Statistical analysis:

The analysis method will be the same as the primary efficacy analysis as described in **Section 8.1.1.**.

Summary statistics will be presented along with a full data listing by treatment group.

8.2.4. Change in Caregiver Global Impression of Change (CaGI-C) from Baseline to Week 12

The CaGI-C scale is a 7 point Likert scale that asks the caregiver the following question: overall, how have the patient's symptoms changed (if at all) since the beginning of the study (before starting treatment)? The CaGI-C is rated as follows: 1 (very much improved), 2 (much improved), 3 (minimally improved), 4 (no change), 5 (minimally worse, 6 (much worse), and 7 (very much worse).

A secondary analysis will be performed on the change in CaGI-C from Baseline to Week 12.

Statistical analysis:

The analysis method will be the same as the primary efficacy analysis as described in **Section 8.1.1**, with the exception of no baseline score will be used as a covariate.

Summary statistics will be presented along with a full data listing by treatment group.

8.2.5. Change in Gilles de la Tourette Syndrome – Quality of Life Scale for Children and Adolescents (C&A-GTS-QOL) from Baseline to Week 12

Gilles de la Tourette Syndrome–Quality of Life Scale for Children and Adolescents (C&A-GTS-QOL) is a patient-reported health related quality of life measure developed for children and adolescents. The C&A-GTS-QOL is a 27-item questionnaire specific to TS patients that asks the patient to assess the extent to which their quality of life is impacted by their symptoms. The C&A-GTS-QOL contains 6 subscales (cognitive, coprophenomena, psychological, physical, obsessive-compulsive, and ADL) and uses a 5 point Likert scale ranging from no problem to extreme problem. Patients will also be asked how satisfied they feel overall with their life at that moment by using a VAS scale between 0 and 100 (Su et al 2017). Following are the questions assessed in each C&A GTS QOL subscale:

- Cognitive (questions 11, 12, 13, 14, 18, 20, 21, 23) (range: 0- 32)
- Psychological (questions 15, 16, 17, 19, 25, 27) (range: 0 – 24)
- Obsessive-compulsive (questions 7, 8, 9, 10) (range: 0 – 16)
- Physical (questions 1, 3, 4) (range: 0 – 12)
- Coprophenomena (questions 5, 6, 22) (range: 0 -12)
- ADL (questions 2, 24, 26) (range: 0-12)

Scores for the six subscales are generated by summing items and, for ease of interpretation, transformation to a range of 0 to 100 (100x [(observed score - min possible score)/ (max possible score - min possible score)]). The total score, resulting from the sum of the subscale scores, is also normalized to a 0–100 range. For C&A GTS-QOL, if a response to 1 question is missing within the subscale, the missing response will be replaced with the average of the remaining responses within the subscale; if

responses to 2 or more questions within a subscale are missing, the missing responses will not be replaced and the subscale score will be set to missing; if at least 1 subscale is missing then the total score will be set to missing. For all analysis and summary, transformed scores will be used.

A secondary analysis will be performed on the change in C&A-GTS-QOL total score and VAS from Baseline to Week 12.

Statistical analysis:

The overall VAS score, QOL subscale scores for Cognitive, Psychological, Obsessive-compulsive, Physical, Coprophenomena, and ADL and the QOL total score will be analyzed using the same method as the primary efficacy analysis as described in 8.1 above.

Furthermore, summary statistics for the overall VAS score, QOL subscale scores for Cognitive, Psychological, Obsessive-compulsive, Physical, Coprophenomena, and ADL and QOL total scorewill be presented along with a full data listing by treatment group.

8.2.6. Percentage of subjects with a 25%, 50% or 75% improvement on the YGTSS-TTS

The percent of subjects with a 25% improvement on the YGTSS-TTS at any time between the Baseline visit and the Week 12 visit will be considered as a Response. Any subject who has at least 1 Response will be considered as a Responder; otherwise the subject will be considered a Non-Responder.

Statistical analysis:

The percentage of subjects with a 25% improvement on the YGTSS-TTS will be analyzed through a logistic regression model to obtain an estimate of the population odds ratio and associated confidence intervals between ecopipam and matching placebo. The model will include fixed effects for treatment, center, and covariate for baseline YGTSS-TTS score.

Summary statistics will be presented along with a full data listing.

Similar analyses will be performed for 50% and 75% improvement.

8.2.7. Subjects with complete remission of tics on the YGTSS-TTS

Subjects with complete remission of tics on the YGTSS-TTS (i.e. YGTSS-TTS=0) at any time between the Baseline visit and the Week 12 visit will be considered as Responders. Any subject who has at least 1 Response will be considered as a Responder; otherwise the subject will be considered a Non-Responder.

Statistical analysis:

The percentage of subjects with complete remission of tics on the YGTSS-TTS will be analyzed through a logistic regression model to obtain an estimate of the population odds ratio and associated confidence intervals between ecopipam and matching placebo. The model will include fixed effects for treatment, center, and covariate for baseline YGTSS-TTS score.

Summary statistics will be presented along with a full data listing.

8.2.8. Change in YGTSS-TTS from Baseline to Weeks 4, 6, and 8

A secondary analysis will be performed on the change in YGTSS-TTS from Baseline to Weeks 4, 6 and 8.

Statistical analysis:

The analysis method will be the same as the primary efficacy analysis as described in **Section 8.1.1.**, with the treatment difference at weeks 4, 6, and 8 estimated based on MMRM.

Summary statistics will be presented along with a full data listing by treatment group.

8.2.9. Change in CGI-TS-S from Baseline to Weeks 4, 6, and 8

A secondary analysis will be performed on the change in CGI-TS-S from Baseline to Weeks 4, 6 and 8.

Statistical analysis:

The analysis method will be the same as the primary efficacy analysis as described in **Section 8.1.1.**, with the treatment difference at weeks 4, 6, and 8 estimated based on MMRM.

Summary statistics will be presented along with a full data listing by treatment group.

8.2.10. CGI-TS-I at Week 4, 6, and 8

A secondary analysis will be performed on the CGI-TS-I at Weeks 4, 6 and 8.

Statistical analysis:

The analysis method will be the same as the primary efficacy analysis as described in **Section 8.1.1.**, with the exception of no baseline score will be used as a covariate. The treatment difference at weeks 4, 6, and 8 will be estimated based on MMRM.

Summary statistics will be presented along with a full data listing by treatment group.

8.2.11. Change in YGTSS-GS from Baseline to Weeks 4, 6, and 8

A secondary analysis will be performed on the change in YGTSS-GS from Baseline to Weeks 4, 6 and 8.

Statistical analysis:

The analysis method will be the same as the primary efficacy analysis as described in **Section 8.1.1.**, with the treatment difference at weeks 4, 6, and 8 estimated based on MMRM.

Summary statistics will be presented along with a full data listing by treatment group.

8.2.12. Change in CaGI-C from Baseline to Weeks 4, 6 and 8

See section 8.2.4 for a description of the Caregiver Global Impression of Change (CaGI-C).

A secondary analysis will be performed on the CaGI-C at Weeks 4, 6 and 8.

Statistical analysis:

The analysis method will be the same as the primary efficacy analysis as described in **Section 8.1.1**, with the exception of no baseline score will be used as a covariate. The treatment difference at weeks 4, 6, and 8 will be estimated based on MMRM.

Summary statistics will be presented along with a full data listing by treatment group.

8.2.13. Change in the C&A-GTS-QOL from Baseline to Weeks 4, 6, and 8

See section 8.2.5 for a description of the Gilles de la Tourette Syndrome–Quality of Life Scale for Children and Adolescents (C&A-GTS-QOL). A secondary analysis will be performed on the change in C&A-GTS-QOL VAS and total score from Baseline to Weeks 4, 6 and 8.

Statistical analysis:

The analysis method for the VAS and total score will be the same as the primary efficacy analysis as described in **Section 8.1.1.**, with the treatment difference at weeks 4, 6, and 8 estimated based on MMRM.

Summary statistics for the VAS, subscales, and total score will be presented along with a full data listing by treatment group.

8.3 COVID-19 Analyses

Due to COVID-19 pandemic, additional analyses will be conducted:

- The amount of missing data by treatment group and overall by endpoint for data missing due to COVID-19 will be summarized in a table.
- Efficacy endpoint analyses will have additional analysis modeling using the terms *Remote* (Yes/No) and *Type* (Audio/Video), to account for any bias that these data collection methods may have introduced.

9. Analysis of Pharmacokinetics

Blood samples will be collected to measure concentrations of ecopipam and/or its major (active) metabolites at Weeks 4 and 12.

For the Week 4 visit, a morning appointment is suggested. Parents/caregivers will be instructed not to administer the study medication to the subject on the evening prior to the Week 4 visit and to record the time of administration of the last dose of study medication taken. The study drug administration will occur on the day of the Week 4 visit at the site under the supervision of the study investigator. An intravenous catheter will be placed, and the subject will have samples collected at the following time windows: one sample at predose (34 to 44 hours since the last dose), one sample between 0.5 and 1.5 hour after administration of study medication, and one sample between 2 and 4 hours after study drug administration. Any samples should be collected at least 30 min apart.

For the Week 12 visit, parents/caregivers will be asked to record the time of study drug administration on the evening prior to the visit. An appointment is suggested to be made for the morning of the Week 12 visit. A blood sample will be collected at the end of the visit. The time of sample collection will be recorded. Blood samples will be processed as outlined in the protocol and serum/plasma will be frozen, shipped and analyzed for ecopipam and SCH40853.

Plasma concentration-time data will be summarized in the Clinical Study Report (CSR). Population pharmacokinetic analysis will be conducted and summarized separately, using data from this study along with data from adults using the same tablet formulation (if available). The methodology for these pharmacokinetic analyses will be reported in a separate analysis plan.

10. Safety

The population used for safety analyses will be the Safety Set (SS). Safety will be assessed by monitoring and recording all adverse events (AEs) and Serious Adverse Events (SAEs) (all Visits), regular monitoring of hematology, blood chemistry, and urine values (Screening, Baseline, Week 12 and 7-day and 14-day Follow Up visits). HbA1c will also be measured during these visits except on the 14 day Follow Up visit. Regular measurement of vital signs and the performance of a physical examination and an ECG will occur at Screening, Baseline and Weeks 4, 6, 8 and 12 and 7 day and 14 day Follow Up visits. An additional assessment will include the Columbia-Suicide Severity Rating Scale (C-SSRS) (all Visits except 30 day Follow up visit). Additional safety outcomes (Baseline and Weeks 4, 6, 8 and 12) will include the Abnormal Involuntary Movement Scale (AIMS), the Barnes Akathisia Rating Scale (BARS), the Swanson, Nolan and Pelham (SNAP-IV) questionnaire, the Children's Yale-Brown Obsessive Compulsive Scale (CY-BOCS), the Children's Depression Rating Scale-Revised (CDRS-R), and the Pediatric Anxiety Rating Scale (PARS).

All safety data will be summarized in Safety Set according to the actual treatment.

Treatment-emergent adverse events (TEAEs) are defined as adverse events that are newly occurring or worsening after the first dose of study medication. The incidence of TEAEs will be summarized by treatment group, by severity and by relationship to study medication. Serious TEAEs and TEAEs leading to the study termination will be summarized by treatment group. Laboratory data, vital sign, physical examination and ECG will be summarized by treatment and visit.

Timing specific adverse events are defined as any adverse events that occur or increase in intensity during any specifically defined study milestones: treatment phase (TPAE), taper down (TDAE) and off study drug (ODAE). Treatment phase adverse events (TPAEs) will be adverse events that start or increase in intensity between the first day of dosing and end of the 8-week treatment phase (day before taper down). Taper down adverse events (TDAEs) will be adverse events that start or increase in intensity between the first day of dose tapering (after 8-week treatment phase) and last dosing date. Off study drug adverse events (ODAE) will be adverse events that start or increase in intensity between the first day off study drug (day after last dosing date) and end of study. The incidence of TPAEs, TDAEs and ODAEs will be summarized by treatment group, by severity and by relationship to study medication and will be displayed such that comparisons can be made. The C-SSRS, AIMS, SNAP-IV questionnaire, CY-BOCS, CDRS-R and PARS will also be summarized descriptively.

10.1. Extent of Exposure

Duration of exposure will be summarized using descriptive statistics by treatment group. The duration of exposure will also be categorized (<7, 7-<14, 14-<28, 28-<42, 42-<56, 56-<70, 70-<84, ≥84), and tabulated by treatment group. Duration of exposure (weeks) will be calculated as (Date of Last Dose – Date of First Dose) +1. If date of last dose is unknown then the date of last clinical visit will be used.

A listing including study drug administration information from the eCRF will be presented.

10.2. Treatment Compliance

Subjects are expected to take 1 tablet of study medication once per day.

Treatment compliance, as a percentage, will be calculated as compliance (%) = (Sum of Study drug taken= Yes) / (number of days in study) x 100. The number of actual doses taken will be recorded on the eCRF.

Subjects will be considered compliant overall for study medication if the compliance is ≥80%. Descriptive statistics (number of subjects, mean, SD, minimum, median, and maximum) for number of actual tablets taken, number of expected tablets taken, and treatment compliance will be summarized by treatment group.

10.3. Adverse Events / Adverse Drug Reactions

An adverse event (AE) is any untoward medical condition that occurs in a patient while participating in this clinical study. Adverse events will be coded according to the Medical Dictionary for Regulatory Activities (MedDRA 22.0) terminology and the severity of the toxicities will be graded according to the National Cancer Institute (NCI) Common Toxicity Criteria for Adverse Events (CTCAE), v4.03, where applicable. An adverse drug reaction (ADR) means any AE caused by a drug or biologic.

Treatment-Emergent Adverse Events (TEAEs) are those adverse events/adverse drug reactions that are recorded during or following the initiation of study treatment administration, and do not necessarily have a causal relationship to the use of the study medication. Treatment-Emergent Adverse Events (simply referred to as adverse events in summary tables) will be summarized.

The following adverse event summary tables will be summarized by treatment group:

1) An overall summary with the number and percentage of patients reporting AEs, serious AEs, grade 3 or higher AEs, treatment-related AEs, serious TEAEs, AEs leading to study treatment discontinuation and AEs with outcome of deaths.

2) AEs overall and by system organ class and preferred term.

   In this summary, a patient is counted once at the system organ class and once at each preferred term within the system organ class.

3) AEs overall and by system organ class, preferred term and initial daily dose level.

4) AEs overall and by system organ class, preferred term and region.

5) AEs overall and by system organ class, preferred term and weight band.

6) AEs overall and by system organ class, preferred term in subjects over 100Kg in weight.

7) AEs overall and by system organ class, preferred term and week within the titration phase.

8) AEs overall and by system organ class, preferred term and visit within the maintenance phase.

9) AEs overall and by system organ class, preferred term and visit within the down titration phase and follow up.

10) AEs overall and by system organ class, preferred term and highest CTCAE grade.

    In this summary, a patient is counted once at the highest grade for which the event occurred in the system organ class and the highest grade for each unique preferred term within that system organ class. Therefore, patients may only contribute once to each preferred term and once to each system organ class. The missing severity grade will be reported in a separate category.

11) Study-treatment-related AEs overall and by system organ class and preferred term.

All those AEs with relationship to treatment marked as "Possibly Related" or "Probably Related" or missing will be reported in the table.

12) Study-treatment related AEs by system organ class, preferred term and Highest CTCAE grade

13) AEs leading to study treatment termination by system organ class and preferred term.

14) AEs leading to study treatment termination by system organ class, preferred term and highest CTCAE grade

15) CTCAE Grade 3 or higher AEs, overall and by system organ class and preferred term

16) AEs by maximum severity, maximum relatedness and milestones (titration phase, maintenance phase, taper down, and off study drug).

A Serious Adverse Event (SAE) is an AE which falls into one or more of the following categories:

a. Results in death
b. Is life-threatening
c. Requires inpatient hospitalization or prolongation of existing hospitalization
d. Results in persistent or significant disability or incapacity
e. Is a congenital anomaly/birth defect
f. Is any other important medical event

Serious adverse events will be summarized by treatment group and combined over both treatment groups. A summary by age group (children 6-11 and adolescents 12-17) will also be provided.

17) SAEs overall and by system organ class and preferred term.

In this summary, a patient is counted once at the system organ class and once at each preferred term within the system organ class.

Adverse Events of Special Interest (AESI) for ecopipam include Convulsions, Depression and Suicide/Self Injury, Drug Abuse, Dependence and Withdrawal, Dyslipedaemia, Hepatic Disorders, Hyperglycaemia/New Onset Diabetes Mellitus, NMS, and Psychosis and Psychotic Disorders.

AESI for Parkinsonism include Ataxia, Confusional state, Coordination abnormal, Somnolence, Disturbance in attention, Dyskinesia, Hyperkinesia, Gait disturbance, Memory impairment, Tremor, Vertigo, Agitation, Anxiety, Disturbance in attention, Dystonia, Bradykinesia, Musculoskeletal stiffness, Dysstasia, Gait disturbance, Posture abnormal, Balance disorder, Movement disorder, Movement disorder, Dyskinesia, Muscle rigidity, Coordination abnormal, Muscle contracture, Bradykinesia, Gait disturbance, Terminal insomnia, Nightmare, Poor quality sleep, Sleep disorder, Fatigue, Dizziness, Balance disorder, Restlessness, Amnesia, Confusional state, Dementia, Mental impairment, Speech disorder, Dysphonia, Vocal cord dysfunction, Parosmia, Anosmia, Urinary incontinence, Anxiety, Apathy, Joint stiffness, Reduced facial expression, Confusional state, Flat affect, Constipation, Depression, Dysphagia, Drooling, Fall, Fear of falling, Loss of visual contrast sensitivity, Muscle tightness, Dysgraphia, Dyskinesia, and Weight decreased.

AESI summary tables are listed below:

1) AESI for ecopipam overall and by system organ class and preferred term

2) AESI for Parkinsonism overall and by system organ class and preferred term.

In this summary, a patient is counted once at the system organ class and once at each preferred term within the system organ class.

10.4. Laboratory Evaluations

Blood samples for hematology, biochemistry, and urine samples for urinalysis are to be collected at screening, Baseline, Weeks 12 and 7 day and 14 day Follow Up visits. All results will be provided using International System of Units (SI). HbA1c will be measured at the baseline and completion/early termination visits.

Descriptive statistics for hematology, biochemistry and HbA1c will be provided for each test parameter and for change from baseline treatment by visit and for each treatment group (specific emphasis will be placed on mean changes between ecopipam and placebo in fasting glucose levels and blood lipid parameters). Shift tables (i.e., low-normal-high at Baseline versus low-normal-high at each post-baseline visit in a 3-by-3 contingency table) will be provided for hematology, biochemistry, HbA1c and urine to assess changes from Baseline in laboratory values by visit for each treatment group.

Separate listings will be provided for all laboratory evaluations (hematology, biochemistry, HbA1c and urinalysis.

10.5. Vital Signs

Vital sign measurements including sitting systolic blood pressure, sitting diastolic blood pressure, sitting pulse rate, supine orthostatic systolic blood pressure, supine orthostatic diastolic blood pressure, supine pulse rate, standing orthostatic systolic blood pressure, standing orthostatic diastolic blood pressure, and standing pulse rate at each scheduled time and changes from baseline in vital sign measurements will be summarized with descriptive statistics at each scheduled time point by treatment group.

Body height will be measured at Screening only and will be summarized by treatment group. Weight will be measured at each schedule visit and, along with changes from baseline, will be summarized with descriptive statistics at each scheduled time point by treatment group. Weight gain ≥ 7% of body weight will be tabulated. Weight will be displayed in kilograms; height will be displayed in centimeters.

Body mass index (BMI) will be derived from the weight at each visit and from height (at screening), along with changes from baseline, will be summarized with descriptive statistics at each schedule timepoint by treatment group. BMI will be displayed in kilograms per meter squared.

All vital signs data will be presented in a listing.

10.6. ECG

The following ECG data will be collected at the Screening, Baseline, Week 4, Week 6, Week 8, Week 12, 7 Day Follow Up and 14 day Follow Up visits:

- PR interval (msec)
- QRS interval (msec)
- QT interval

- QTc interval (msec)
- Overall ECG result (Normal; Abnormal, Not Clinically Significant; Abnormal, Clinically Significant).

An ECG is optional at unscheduled visits, per the investigator's discretion.

The actual ECG results along with change from baseline will be summarized at each visit by treatment group. All ECG data will be presented in a listing.

10.7. Physical Examination

A Physical examination will be performed at the Screening, Baseline, Week 4, Week 6, Week 8, Week 12, 7 Day Follow Up and 14 day Follow Up visits. This will include physical examination of the following body areas and systems: Head, Eyes, Ears, Nose, Mouth, Throat, Neck (including Thyroid); Thorax; Abdomen; Urogenital; Extremities; Neurological; Skin and Mucosae; and Other.

A physical exam is optional at unscheduled visits, per the investigator's discretion.

The Physical examination results will be summarized at each visit by treatment group. All physical examination data will be presented in a listing.

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

The C-SSRS is a low burden (approximately 5 minutes for completion) instrument to assess both suicidal behavior and ideation. The scale is appropriate for subjects from age 6 through to an elderly population.

The number and percentage of subjects with suicidal behavior and ideation will be summarized by treatment group at the Baseline, Week 4, Week 6, Week 8 and Week 12 visits and all questionnaire data will be displayed in a listing.

10.9. Other Safety

10.9.1. Abnormal Involuntary Movement Scale (AIMS)

Abnormal Involuntary Movement Scale (AIMS) records the occurrences of tardive dyskinesia (TD) in subjects receiving neuroleptic medications. The test is used to detect TD and to follow the severity over time. It consists of rating the presence and severity of movement disorders involving the face, mouth, extremities, and trunk as well as three items of global judgment from a scale of 0 (none) to 4 (severe).

The AIMS total score will by calculated by summing the severity of the 10 items that are rated on a scale of 0 to 4. The total score will be summarized by treatment group at the Baseline, Week 4, Week 6, Week 8 and Week 12 visits and all questionnaire data will be displayed in a listing.

10.9.2. Barnes Akathisia Rating Scale (BARS)

The Barnes Akathisia Rating Scale (BARS) is used to assess the severity of drug-induced akathisia. Objective and subjective items in the scale measure the level of subject's restlessness, ranging from zero (normal) to 3 (most severe). The BARS also includes a global assessment of akathisia, ranging from zero (absent) to 5 (severe).

The objective, subjective, and global scores will be summarized by treatment group at the Baseline, Week 4, Week 6, Week 8 and Week 12 visits and all questionnaire data will be displayed in a listing.

10.9.3. Swanson, Nolan and Pelham (SNAP-IV) questionnaire

Swanson, Nolan and Pelham (SNAP-IV) questionnaire measure designed assess ADHD and oppositional defiant disorder (ODD) symptoms in children and adolescents. The SNAP-IV is based on a 0 to 3 rating scale: Not at All = 0, Just A Little = 1, Quite A Bit = 2, and Very Much = 3. Subscale scores on the SNAP-IV are calculated by summing the scores on the items in the subset and dividing by the number of items in the subset. The score for any subset is expressed as the Average Rating-Per-Item.

The items from the DSM-IV (1994) criteria for Attention-Deficit/Hyperactivity Disorder (ADHD) are included for the two subsets of symptoms: inattention (items #1-#9) and hyperactivity/impulsivity (items #11-#19). Also, items are included from the DSM-IV criteria for Oppositional Defiant Disorder (ODD, items #21-#28) since it often is present in children with ADHD.

In addition to the DSM-IV items for ADHD and ODD, the SNAP-IV contains items from the Conners Index Questionnaire (Conners, 1968) and the IOWA Conners Questionnaire (Loney and Milich, 1985). The IOWA was developed using divergent validity to separate items which measure inattention/overactivity (I/O — items #4, #8, #11, #31, #32) from those items which measure aggression/defiance (A/D — items #21, #23, #29, #34, #35). The Conners Index (items #4, #8, #11, #21, #32, #33, #36, #37, #38, #39) was developed by selecting the items which loaded highest on the multiple factors of the Conners Questionnaire, and thus represents a general index of childhood problems.

The SNAP-IV also includes the 10 items of the Swanson, Kotkin, Agler, MyInn, and Pelham (SKAMP) Rating Scale. These items are classroom manifestations of inattention, hyperactivity, and impulsivity (i.e., getting started, staying on task, interactions with others, completing work, and shifting activities). The subscales of the SKAMP include academic (items #81-86) and deportment (items #87-90).

The subscale scores will be summarized by treatment group at the Baseline, Week 4, Week 6, Week 8 and Week 12 visits and all questionnaire data will be displayed in a listing.

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

Children's Yale-Brown Obsessive Compulsive Scale (CY-BOCS) is a reliable and valid scale to both determine severity of OCD and to monitor improvement during treatment. The scale is a clinician-rated, 10-item scale that includes questions about the amount of time spent on obsessions/compulsions, level of impairment or distress, and how much resistance and control subjects have over these thoughts.

Severity of compulsions and obsessions are rated on a five-point scale from 0 to 4. CY BOCS total score is computed as the sum of the 10 items, ranging from 0 to 40, with higher scores indicating more severe compulsions and obsessions.

The total score will be summarized by treatment group at the Baseline, Week 4, Week 6, Week 8 and Week 12 visits and all questionnaire data will be displayed in a listing.

10.9.5. Children's Depression Rating Scale-Revised (CDRS-R)

Children's Depression Rating Scale-Revised (CDRS-R) is a clinically validated rating scale designed to assess psychiatric signs and symptoms of depressions. Fourteen signs and symptoms are rated from 1 (normal) to 7 (most severe), and 3 signs and symptoms are rated from 1 (normal) to 5 (most severe). The raw summary score is the sum of all 17 items, ranging from 17 to 113.

The raw summary score will be summarized by treatment group at the Baseline, Week 4, Week 6, Week 8 and Week 12 visits and all questionnaire data will be displayed in a listing.

10.9.6. Pediatric Anxiety Rating Scale (PARS)

Pediatric Anxiety Rating Scale (PARS) is a clinician-rated instrument for assessing the severity of anxiety symptoms associated with common anxiety disorders and generalized anxiety in children and adolescents. The PARS has two sections: the symptom checklist and the severity items. The symptom checklist is used to determine the child's repertoire of symptoms during the past week. The 7-severity item is used to determine severity of symptoms and the PARS total score. Each severity item is coded from zero (none) to 5 (most extreme). Not applicable is coded to 8, and does not know is coded to 9. The total score for the PARS is total of the 7 severity items. The total score ranges from 0 to 35. Codes "8" and "9" are not included in the summation.

The total score will be summarized by treatment group at the Baseline, Week 4, Week 6, Week 8 and Week 12 visits and all questionnaire data will be displayed in a listing.

10.10. **COVID-19 Analyses**

Due to COVID-19 pandemic, additional analyses will be conducted:

- Listings of COVID-19 related study disruptions by subject and by site, including descriptions of the disruption.
- The amount of missing data by treatment group and overall by endpoint for data missing due to COVID-19 will be summarized in a table.
- Protocol deviations due to COVID-19 will be summarized by type and treatment in a table.

11. Interim Analyses

An unblinded interim analysis will be conducted when approximately 50% of the 150 planned subjects have been randomized (1:1 ratio between ecopipam and placebo) and completed treatment or are terminated early. The interim analysis will be performed by an independent statistician and governed by an external Data Safety Monitoring Board (DSMB). The conditional power will be calculated to help the DSMB assess the probability of a successful study based on observed data.

The primary efficacy endpoint of the study is the YGTSS-TTS. A smaller YGTSS-TTS score indicates a better condition in subjects with Tourette's syndrome, and we may expect that the ecopipam group has a larger reduction than the placebo group. Let μ_e be the mean change YGTSS-TTS score from Baseline to Week 12 for the ecopipam group, and μ_p be the mean change YGTSS-TTS score from Baseline to Week 12 for the placebo group. Then, given H_0: $\mu_e \geq \mu_p \equiv \Delta_{e-p} \geq 0$ vs. H_1: $\mu_e < \mu_p \equiv \Delta_{e-p} < 0$. Since the original sample size calculation is based on two-sided test at a significance level of α=5%, we will use α/2=2.5% to match the one-sided test.

The primary endpoint will be analyzed with a Mixed Model for Repeated Measures (MMRM) model using the SAS MIXED procedure with REPEATED statement. The model will include fixed effects for treatment, visit, visit interaction between treatment and visit, region, age group (0 = children 6-11, 1 = adolescents 12-17) and covariate for baseline YGTSS-TTS.

To determine if the total sample size should be increased from N (to N^*), the conditional power based on the MMRM will be calculated at the interim analysis using the observed effects (where n=75) and the planned sample size (N=150).

If it is determined that the sample size should be increased, the CHW (Cui-Hung-Wang) statistic, which controls the type I error, will be used. The CHW statistic is a weighted statistic defined as:

$$T_{CHW} = \sqrt{\frac{n_1}{N}} T_1 + \sqrt{\frac{n_2}{N}} T_2^*$$

The validity of the CHW test statistic requires that pre-interim (Stage 1) and post-interim (Stage 2) statistics T_1 and T_2^* be stochastically independent of each other. Under the various scenarios described below, the T_1 statistic will be based the least-square means from the Stage 1 MMRM model. Specifically, at Stage 1 the mean difference Δ_{e-p} (treatment effect of ecopipam versus placebo) is the difference in least-square means of change from baseline at Week 12 from the MMRM. The LSMEANS statement with COV option will report the covariance of LS means and the variance $var(\delta_1)$ of the difference can be calculated. Thereby, the t-statistic T_1 at the interim analysis can be calculated as:

$$T_1 = \frac{\delta_1}{\sqrt{var(\delta_1)}}, \text{ where } \delta_1 = LSM_e - LSM_p$$

The T_2^* statistic will be defined in two ways to produce three conditional power calculations, as follows:

Firstly, T_2^* will be based on Stage 2 sample means assuming all Stage 2 subjects complete the Week 12 assessment. Then the Stage 2 (post-interim analysis) t-statistic T_2^* at the interim analysis can be calculated as:

$$T_2^* = \frac{\hat{\delta}_2}{\sigma\sqrt{2/k_2^*}}$$

where $\widehat{\delta}_2$ is the observed treatment mean difference in the change of YGTSS-TTS from baseline to Week 12 when the k_2^* subjects in Stage 2 complete Week 12 of study for each group.

Given the information at Stage 1,

$$E(T_{CHW}|T_1) = \sqrt{\frac{n_1}{N}}T_1 + \sqrt{\frac{n_2}{N}}\frac{\delta}{\sigma\sqrt{2/k_2^*}} = \sqrt{\frac{n_1}{N}}T_1 + \sqrt{\frac{n_2 k_2^*}{2N}}\frac{\delta}{\sigma}$$

$$Var(T_{CHW}|T_1) = \frac{n_2}{N}var(T_2^*) = \frac{n_2}{N}$$

Thus, the conditional power is:

$$CP = P(T_{CHW} < -Z_{\alpha/2}|T_1) \approx 1 - \Phi\left(\sqrt{\frac{N}{n_2}}Z_{\alpha/2} + \sqrt{\frac{n_1}{n_2}}T_1 + \sqrt{\frac{k_2^*}{2}}\frac{\delta}{\sigma}\right)$$

The condition power will then be calculated under the following scenarios for the treatment difference δ and standard deviation σ:

1) Using the values hypothesized in the protocol. Treatment difference $\delta = 3$ and standard deviation $\sigma = 6.1$
2) Using the estimates from the Stage 1 data. Treatment difference δ is the Stage 1 observed sample mean difference in change from baseline between the two treatment groups, and standard deviation σ is the pooled standard deviation of the change from baseline, both based on the completers in the respective treatment groups.

If the drop-out rate based on the first stage is *f*, the corresponding sample size at 2nd stage should be $n_2^* = k_2^*/(1-f)$, and the new total sample size is $N^* = n_1 + n_2^*$.

Secondly, T_2^* will be based on Stage 2 least-squares mean difference assuming all Stage 2 subjects will have the same trend in treatment difference and variability, as does Stage 1. The treatment difference δ will be estimated by the least-squares means difference between treatments observed in Stage 1. As above, define $T_2^* = \frac{\hat{\delta}_2}{\sqrt{var(\hat{\delta}_2)}}$, where $\hat{\delta}_2$ is the future estimate of treatment effect from Stage 2. The inequality $T_{CHW} < -Z_{\alpha/2}$ can be written as:

$$\frac{\hat{\delta}_2 - \delta}{\sqrt{var(\hat{\delta}_2)}} < \sqrt{\frac{N}{n_2}}\left(-Z_{\alpha/2} - \sqrt{\frac{n_1}{N}}T_1\right) - \frac{\delta}{\sqrt{var(\hat{\delta}_2)}}$$

As above, n_1, n_2, N, and T_1 are known at the time of the interim analysis, and the treatment effect, δ, is hypothesized as that estimated in Stage 1. The remaining term, $var(\hat{\delta}_2)$, is estimated as:

$$var(\hat{\delta}_2) = n_1 Var(\widehat{\delta_1})/k_2^*$$

To justify this equation, note that the variance of an MMRM estimate of treatment effect depends on sample size, other aspects of study design incorporated in the model (e.g., covariate, stratification factors, etc.), and underlying population variability. Define D to be the component of $var(\hat{\delta}_1)$ that incorporates all design and covariate features but sample size, i.e., $var(\hat{\delta}_1) = D/n_1$, so that D = $n_1 var(\hat{\delta}_1)$. Assuming that the model components and underlying population variability in Stage 2 are the same as in Stage 1, we have $var(\hat{\delta}_2) = \frac{D}{k_2^*}$.

It follows that $var(\hat{\delta}_2) = n_1 Var(\hat{\delta}_1)/k_2^*$. Therefore, the conditional power is:

$$CP = \Phi\left(\sqrt{\frac{N}{n_2}}\left[-Z_{\alpha/2} - \sqrt{\frac{n_1}{N}}T_1\right] - \frac{\delta}{\sqrt{n_1 Var(\hat{\delta}_1)/k_2^*}}\right)$$

The condition power will then be calculated under the following for the treatment difference δ and standard error $Var(\hat{\delta}_1)$:

3) Using the estimates from the stage 1 data. Treatment difference δ is the Stage 1 MMRM treatment effect $(\hat{\delta}_1)$ and $Var(\hat{\delta}_1)$ is the variance of this estimate of treatment difference based on the covariance matrix of the LSmeans.

Using these results, the DSMB may make one of the following recommendations:

1) If the conditional power at the planned sample size (150) is less than 30% then the DSMB may recommend that the study be terminated for futility.
2) If the conditional power at the planned sample size (150) is greater than 30% and less than 80%, then the planned sample size will be increased up to 200 to try to achieve a conditional power of at least 80%. If the conditional power at the maximum sample size (200) is still less than 80%, the DSMB may recommend continuing the study at the maximum sample size (200).
3) If the conditional power is at least 80%, then the DSMB may recommend continuing this study with the planned sample size (150).

Full details of the DSMB including primary responsibilities, its relationship with other study components, its membership, and the purpose and timings of its meetings are documented in a DSMB charter. These details will also include procedures to ensure confidentiality and proper communication, the guidelines to be implemented by the DSMB and an outline of the content of the closed reports (unblinded) and open reports (blinded) that will be provided to the DSMB.

12. Changes from Analysis Planned in Protocol

There are no changes from the analysis planned in protocol.

13. Reference List

1. Robertson, M.M., Banerjee, S., Kurlan, R., Cohen, D.J., Leckman, J.F., McMahon, W., Pauls, D.L., Sandor, P., van de Wetering, B.J.M. The Tourette Syndrome Diagnostic Confidence Index: Development and Clinical Associations. Neurology Dec 1999, 53 (9) 2108-2112.
2. Leckman JF, Riddle MA, Hardin MT, Ort SI, Swartz KL, Stevenson J, et al. The Yale Global Tic Severity Scale: initial testing of a clinician-rated scale of tic severity. J Am Acad Child Adolesc Psychiatry 1989;28(4):566-73
3. Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, Currier GW, Melvin GA, Greenhill L, Shen S, Mann JJ. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry. 2011 Dec;168(12):1266-77
4. First MB, Williams JBW, Karg RS, Spitzer RL: Structured Clinical Interview for DSM-5 Disorders, Clinical Trials Version (SCID-5-CT). Arlington, VA, American Psychiatric Association, 2015
5. Leckman JF, Towbin KE, Ort SI, Cohen DJ. Clinical assessment of tic disorders severity. In: Cohen DJ, Bruun RD, Leckman JF, editors. Tourette's Syndrome and Tic Disorders: Clinical Understanding and Treatment. New York: John Wiley & Sons; 1988
6. Guy, William (1976). "Clinical Global Impressions". ECDEU Assessment Manual for Psychopharmacology—Revised. Rockville, MD: U.S. Department of Health, Education, and Welfare; Public Health Service, Alcohol; Drug Abuse, and Mental Health Administration; National Institute of Mental Health; Psychopharmacology Research Branch; Division of Extramural Research Programs
7. Abnormal involuntary movement scale [AIMS] (1976). National Institute of Mental Health Psychopharmacology Research Branch. Guy, W. (1976). ECDEU assessment manual for psychopharmacology. Rev. Rockville, MD: U.S. National Institute of Health, Psychopharmacology Research Branch. Pg. 534-537
8. Barnes T.R. A rating scale for drug-induced akathisia. British Journal of Psychiatry. 1989;154:672-676.
9. Swanson JM, Kraemer HC, Hinshaw SP, Arnold LE, Conners CK, Abikoff HB, Clevenger W, Davies M, Elliott GR, Greenhill LL, Hechtman L, Hoza B, Jensen PS, March JS, Newcorn JH, Owens EB, Pelham WE, Schiller E, Severe JB, Simpson S, Vitiello B, Wells K, Wigal T, Wu M. Clinical relevance of the primary findings of the MTA: success rates based on severity of ADHD and ODD symptoms at the end of treatment. J Am Acad Child Adolesc Psychiatry. 2001;40(2):168-79
10. Scahill L, Riddle MA, McSwiggin-Hardin M, Ort SI, King RA, Goodman WK, Cicchetti D, Leckman JF. Children's Yale-Brown Obsessive Compulsive Scale: reliability and validity. J Am Acad Child Adolesc Psychiatry. 1997 Jun;36(6):844-52
11. Poznanski EO, Freeman LN, Mokros HB. Children's Depression Rating Scale Revised (September 1984). Psychopharmacol Bull 1985; 21:979-989.

12. Su M, McFarlane F, Cavanna A, et al. The English Version of the Gilles de la Tourette Syndrome–Quality of Life Scale for Children and Adolescents (C&A-GTS-QOL). J Child Neurol 2017;32(1):76-83.

13. Research Units on Pediatric Psychopharmacology Anxiety Study Group. The Pediatric Anxiety Rating Scale (PARS): development and psychometric properties. J Am Acad Child Adolesc Psychiatry 2002;41;1061-9.

14. Heitjan, Daniel F., and Roderick J. A. Little. "Multiple Imputation for the Fatal Accident Reporting System." *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 40, no. 1, 1991, pp. 13–29. *JSTOR*, www.jstor.org/stable/2347902.

15. Schenker, Nathaniel and Taylor, Jeremy M. G., (1996), Partially parametric techniques for multiple imputation, Computational Statistics & Data Analysis, 22, issue 4, p. 425-446.

16. Rubin, Donald B. (1987) Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons.

17. Little, R.J., & Rubin, D.B. (2002). Statistical Analysis with Missing Data, 2nd. ed. UNC 2011 SSIL Wiley

14. Appendices

14.1. Schedule of Assessments

Assessment	Screening[a]	Baseline	Titration		Treatment			Completion/ Early Termination	7 Day Follow-up	14 Day Follow-up	30 day Follow Up*
Study Week	Up to -4	0	2*	4[g]	6[g]	8[g]	10*	12[g]	13/14[g]	14/15[g]	16/17
Study Days	Up to -28	0	14±3	28±3	42±3	56±3	70±3	84±3	7 Days Post Last Dose +3	14 Days Post Last Dose +3	30 Days Post Last Dose +3
Study Visit Number	1	2	3	4	5	6	7	8	9	10	11
Informed Consent	X										
Inclusion/ Exclusion	X	X									
Medical/Psychiatric/medication History	X	X									
Randomization		X									
Physical Exam/Vital Signs[b]	X	X		X	X	X		X	X	X	
ECG	X	X		X	X	X		X	X	X	
Laboratory tests (Hematology and Chemistry)[c]	X	X						X	X	X[c]	
Urine Drug Screen	X	X						X			
Urine Pregnancy Test	X	X						X		X	
DSM-5 Criteria for TS	X										

Assessment	Screening[a]	Baseline	Titration		Treatment			Completion/ Early Termination	7 Day Follow-up	14 Day Follow-up	30 day Follow Up*
Study Week	Up to -4	0	2*	4[g]	6[g]	8[g]	10*	12[g]	13/14[g]	14/15[g]	16/17
Study Days	Up to -28	0	14±3	28±3	42±3	56±3	70±3	84±3	7 Days Post Last Dose +3	14 Days Post Last Dose +3	30 Days Post Last Dose +3
Study Visit Number	1	2	3	4	5	6	7	8	9	10	11
Yale Global Tic Severity Scale	X	X		X	X	X		X			
Clinical Global Impression - Tourette's Syndrome of Severity		X		X	X	X		X			
Clinical Global Impression - Tourette's Syndrome of Improvement				X	X	X		X			
Caregiver Global Impression of Change				X	X	X		X			
Columbia Suicide Severity Rating Scale	X	X	X	X	X	X	X	X	X	X	
Abnormal Involuntary Movement[d]		X		X	X	X		X			
Barnes Akathisia Rating Scale		X		X	X	X		X			
Swanson, Nolan, and Pelham Questionnaire		X		X	X	X		X			
Children's Yale-Brown Obsessive Compulsive Scale		X		X	X	X		X			

Assessment	Screening[a]	Baseline	Titration		Treatment			Completion/ Early Termination	7 Day Follow-up	14 Day Follow-up	30 day Follow Up*
Study Week	Up to -4	0	2*	4[g]	6[g]	8[g]	10*	12[g]	13/14[g]	14/15[g]	16/17
Study Days	Up to -28	0	14±3	28±3	42±3	56±3	70±3	84±3	7 Days Post Last Dose +3	14 Days Post Last Dose +3	30 Days Post Last Dose +3
Study Visit Number	1	2	3	4	5	6	7	8	9	10	11
Children's Depression Rating Scale - Revised		X		X	X	X		X			
Gilles de la Tourette Syndrome Quality of Life Scale		X		X	X	X		X			
Pediatric Anxiety Scale		X		X	X	X		X			
PK Blood Draws[e,h]				X				X			
Adverse Events	X	X	X	X	X	X	X	X	X	X	X*
Concomitant Medications	X	X		X	X	X		X	X	X	
Dispense Study Drug		X		X	X	X		X[f]			
Collect Unused Study Drug/Assess drug compliance				X	X	X		X	X		

Abbreviations: AE = adverse event; BP = blood pressure; ECG = electrocardiogram; HR = heart rate; PK = pharmacokinetic(s)

[a] Informed consent must be performed prior to any screening procedures. All screening procedures are to occur after washout of any applicable medications. Rescreening is allowed after approval by the Medical Advisor.

[b] Vital signs will include, pulse, BP, orthostatic BP (done 5 minutes after being supine and then 3 minutes after standing), height and weight.

[c] Subjects should be in a fasting state (8 hours) for laboratory tests. HbA1c will be measured at the baseline and completion/early termination visits.

[d] Every attempt should be made to perform this assessment at the same time at each visit.

[e] At Week 4 Study dose should not be taken until after pre-dose blood draw at the site. PK draws should be pre-dose, between 0.5 and 1.5 hours after study drug administration, and between 2 and 4 hours after study drug administration. Both Week 4 and Week 12, visits should be scheduled for the morning. At Week 12, blood sample will be collected in the afternoon at the end of the visit.

[f] Study drug will be dispensed for dose down titration.

* Weeks 2 and 10 will be telephone visits. If there are any abnormal findings, the subject will be brought to the site for full assessment. A telephone call will be made to the subject 30 days after the last visit to assess any adverse events that may have occurred.

[g] Assessments for these visits may be completed in locations other than study clinic and/or safety and efficacy assessments required at these visits will be made available for remote administration due to restrictions as a consequence of the COVID-19 pandemic or other qualifying event.
[h] If visits at Weeks 4 and 12 are completed in locations other than study clinic due to restrictions as a consequence of the COVID-19 pandemic or other qualifying event, then collection of labs for PK assessments are optional

14.2. Instruments

Exact copies of instruments and scoring algorithms are provided in additional documentation, literature references are stated in Section 13.

14.3. MAR multiple imputation methodology

The FCS method will be used to impute missing values for TTS of the YGTSS change values and will use age group at baseline, region, treatment arm, baseline YGTSS-TTS score and YGTSS-TTS scores at all visits up to week 12 to determine the scores (100 imputations will be created).

The SAS code for the imputation is as follows:

```
proc mi data=final out=mi1 seed=123 nimpute=100;
 class agegrp region arm ;
 var agegrp region arm base _4 _6 _8 _12;
 fcs regpmm; run;
```

14.4. MNAR jump to reference multiple imputation methodology

The monotone method will be used to impute missing values for TTS of the YGTSS change values and will use age group at baseline, region, and baseline TTS score to determine the scores (100 imputations will be created). Only placebo patients with complete data will be included in the imputation model.

The SAS code for the imputation is as follows:

```
proc mi data=final out=mi2 seed=123 nimpute=100;
class agegrp region arm;
  monotone reg;
  mnar model( _4 /modelobs=(arm='Placebo'));
  mnar model( _6 /modelobs=( arm='Placebo'));
  mnar model( _8 /modelobs=( arm='Placebo'));
```

```
  mnar model( _12 /modelobs=( arm='Placebo'));
  var agegrp region base _4 _6 _8 _12;
run;
```

14.5. Tipping Point Analysis

Let delta1 and delta2 be the penalty to ecopipam group and reward for placebo group, respectively. Delta1 and delta2 = 1 to 10 (or estimated treatment effect from primary analysis if it is larger than 10) in step of 1. General steps for this analysis includes:

Step 1:

For each combination of (delta1, delta2), 100 imputations will be created.

In order to induce monotonicity, an imputation model will first be used to impute missing values for TTS of the YGTSS week 4 change values and will use age group at baseline, treatment arm, and baseline TTS score to determine the scores.

The SAS code for the imputation is as follows:

```
proc mi data=final out=mi seed=666 nimpute=100;
 class agegrp region1 arm;
 var agegrp region1 arm base _4;
 monotone regpmm; run;

* Create delta1, delta2 in the dataset;

data mitv1;
 set mi;
 do i=1 to 10;
  do j=1 to 10;
    delta1=i; delta2=j; output;
  end;
 end;
run;

data mitv1;
 set mitv1;
 if firstdrp=4 and missdm in ("MNAR') and index(arm, 'ecopipam') then _&cv=_&cv+delta1;
else if firstdrp=4 and missdm in ("MNAR') and arm='Placebo' then _&cv=_&cv-delta2;
run;
```

For each visit after week 4:

```
%macro tippimp (cv, pv, pvs, seed);

proc mi data=mitv&pv out=mitv&currentv seed=&seed nimpute=1;
 class agegrp region1 arm;
 var agegrp region1 arm base _&cv &pvs;
 monotone regpmm;
 by _Imputation_ &delta1 &delta2; run;

data mitv&cv;
if firstdrp=&cv and missdm in ("MNAR') and index(arm, 'ecopipam') then _&cv=_&cv+delta1;
else if firstdrp=&cv and missdm in ("MNAR') and arm='Placebo' then _&cv=_&cv-delta2;
run;
%mend;


% tippimp (6, 4, _4 , 2626);
% tippimp (8, 6, 4, %str(_4 _6), 2828);
% tippimp (12, 8, 6, 4 %str(_4 _6 _8), 2929);
```

Step 2

The resultant complete data sets will be analyzed using the same model as the primary analysis. The SAS code is as follows:

```
proc sort data=mitv12; by delta1 delta2 _imputation_ descending arm avisitn; run;

proc mixed data=all order=data; by delta1 delta2 _imputation_;
  class usubjid arm avisitn agegrp region1;
  model chg=arm avisitn arm*avisitn agegrp region1 base / ddfm=sat;
  repeated avisitn / type=un subject=usubjid;
  lsmeans arm*avisitn / diff cl;
  ods output diffs=diff(where=(avisitn=_avisitn) lsmeans=lsm; run;
```

Step 3

The results from the imputed data analysis will be combined using PROC MIANALYZE which will provide the LS mean and standard error of the LS mean for each treatment group at each visit as well as the LS mean difference, 95% CI and resultant p-value between the ecopipam and placebo groups at each visit. The SAS code is as follows:

```
proc sort data=diff; by delta1 delta2 avisitn arm; run;
proc sort data=lsm; by delta1 delta2 avisitn arm; run;

proc mianalyze data=diff;  by delta1 delta2 avisitn arm;
 modeleffects estimate;
 stderr stderr;
 ods output ParameterEstimates=estdiff; run;

proc mianalyze data=lsm; by delta1 delta2 avisitn arm;
 modeleffects estimate;
 stderr stderr;
 ods output ParameterEstimates=estlsm; run;
```