Statistical Analysis Plan

1042-TSC-3001

A Phase 3, Double-blind, Randomized, Placebo-controlled Trial of Adjunctive Ganaxolone (GNX) Treatment in Children and Adults with Tuberous Sclerosis Complex (TSC)-related Epilepsy (TrustTSC)

Protocol Numbers	1042-TSC-3001
Investigational Product	Ganaxolone (GNX)
Indication	Tuberous Sclerosis Complex (TSC)
Reference Product	Placebo
Phase	3
Final	25 March 2022, Version 1.0
Amendment 1	14 September 2023, Version 2.0
Amendment 2	02 August 2024, Version 3.0

Statistical Analysis Plan Approval

Author:

[redacted], Ph.D. _____ Date (mm/dd/yyyy)
[redacted] Biostatistics
Marinus Pharmaceuticals, Inc.

Review:

[redacted], M.D. _____ Date (mm/dd/yyyy)
[redacted]
Marinus Pharmaceuticals, Inc.

Approval:

[redacted] _____ Date (mm/dd/yyyy)
[redacted] Biostatistics
Marinus Pharmaceuticals, Inc.

Table of Contents

List of Tables

LIST OF ABBREVIATIONS

AE	Adverse event
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
ADAMS	Anxiety, Depression, and Mood Scale
ANCOVA	Analysis of Covariance
AST	Aspartate aminotransferase
BMI	Body Mass Index
CI	Confidence Interval
CMH	Cochran-Mantel-Haenszel
CRF	Case Report Form
CGI-CSID	Caregiver Global Impression of Change in Seizure Intensity/Duration
CGI-S	Clinical Global Impression - Severity
CGI-I	Clinical Global Impression of Improvement
DB	Double-blind
ECG	Electrocardiogram
eDiary	Electronic Diary
ELDQOL	Epilepsy and Learning Disabilities Quality of Life
EMA	European Medicines Agency
GNX	Ganaxolone
kg	Kilogram
kg/m^2	Kilogram per Meter Squared
MedDRA	Medical Dictionary for Regulatory Activities

mg	Milligram
Peds-QL-FIM	Pediatric Quality of Life – Family Impact Module
ROW	Rest of World
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SF-36	Short Form 36
SOC	System Organ Class
TEAE	Treatment-emergent Adverse Event
TLF	Table, Listing and Figure
ULN	Upper Limit of Normal

1. INTRODUCTION

This is a Phase 3, global, double-blind, randomized, placebo-controlled study of adjunctive Ganaxolone (GNX) treatment in children and adults with TSC-related epilepsy. The study consists of a 4-week prospective baseline phase, defined as the first 28 days following screening, followed by a double-blind phase consisting of a 4-week titration period and a 12-week maintenance period.

Following completion of the treatment period, participants who are compliant with study conduct will have the option to enroll and be treated with GNX in a separate OLE study, 1042-TSC-3002. All participants entering the OLE will have their dose of study medication adjusted in a double-blind cross-titration over 4 weeks such that all are receiving GNX at study completion. Participants who do not continue in the OLE will undergo a 2- to 4-week double-blind down-titration and taper if discontinuing IP.

Sections 2 and 5 of the Protocol provides a detailed description of the investigational product, target patient population, and potential risks and benefits of treatment with GNX. The purpose of this Statistical Analysis Plan (SAP) is to define the methodology for analyzing and summarizing the data collected during the conduct of Study 1042-TSC-3001.

2. STUDY INFORMATION, OBJECTIVES, AND ENDPOINTS

2.1. Protocol and Case Report Form Version

This SAP is based on 1042-TSC-3001 Protocol Amendment 3 v4.0 dated 12SEP2023, Protocol Addendum (Amd 2) dated 18AUG2023, and case report forms (CRFs) approved 20APR2023. Unless superseded by an amendment, this SAP will be effective for all subsequent Protocol amendments and CRF versions.

2.2. Study Objectives

2.2.1. Primary Objective

The primary objective of the study is: to assess the safety and efficacy of GNX compared to placebo (PBO) as adjunctive therapy for seizures associated with TSC in children and adults as assessed by the change from baseline in the frequency of countable major motor and focal seizures (primary endpoint seizures) during the double-blind phase.

2.2.2. Secondary Objectives

The secondary objectives of the study are to evaluate the efficacy of GNX vs. PBO:

- To determine the percentage of change from baseline in 28-day primary endpoint seizure frequency during the maintenance period.
- To assess the change in focal seizure frequency from baseline during the double-blind phase.
- To assess changes in mood, behavior, and quality of life using the following:
  - ADAMS
  - Peds-QL-FIM
  - SF-36
  - ELDQOL
- To assess overall clinical outcome using the CGI-I scores by the clinician and the parent/caregiver.
- To evaluate the changes in seizure intensity and duration using the CGI CSID.

[redacted]

[redacted]

2.2.4 Safety Objectives

To assess the safety and tolerability of GNX compared with placebo as adjunctive therapy for seizures associated with TSC.

2.3. Study Endpoints

2.3.1. Primary Endpoint

The primary efficacy endpoint is the percentage change from baseline in 28-day primary endpoint seizure frequency during the double-blind phase.

2.3.2. Key Secondary Efficacy Endpoints

- Percentage change from baseline in 28-day primary endpoint seizure frequency during the maintenance period.
- Number (%) of participants considered treatment responders during the double-blind phase.
- Number (%) of participants considered treatment responders during the maintenance period.
- CGI-I at the last scheduled visit in the double-blind phase.

2.3.3. Secondary Efficacy Endpoints (Behavior/Neuropsychiatric/Quality of life):

- Change from baseline in ADAMS total score and sub-score.
- Change from baseline in quality-of-life scales: Peds-QL-FIM, SF-36, and ELDQOL.

2.3.4. Secondary Efficacy Endpoints (Seizure Control):

- Change from baseline in the percentage of seizure-free days during the double-blind phase, based on primary endpoint seizure type.
- Change from baseline in the percentage of seizure-free days during the maintenance phase, based on primary endpoint seizure type.
- CGI-CSID at the last scheduled visit in the double-blind phase.
- Participants with a $\geq 25\%$ and $\geq 75\%$ reduction from baseline in primary endpoint seizure frequency during the double-blind phase.

- Participants with a ≥ 25%, ≥ 50%, and ≥ 75% reduction from baseline in primary endpoint seizure frequency during the maintenance period.
- Responder analysis for primary endpoint seizures and all seizures during the double-blind phase using the following response categories: ≤ 0%, > 0% to < 25%, ≥ 25% to < 50%, ≥ 50% to < 75%, and ≥ 75% to 100%.
- Percent change in 28-day frequency of all seizures.
- Change from baseline in the percentage of seizure-free days, based on all seizure types.
- Change from baseline in the longest seizure-free interval, based on primary endpoint seizure types and all seizure types.

2.3.5. Exploratory Endpoints

Except for the endpoints of seizure frequency during the titration and maintenance periods of the study, [redacted]

2.3.6. Safety Endpoints

- Incidence and severity of AEs, SAEs, and withdrawals and dose-reductions due to AEs
- Other measures of safety including physical/neurological/developmental examinations, vital sign measurements (eg, blood pressure, heart rate, respiratory rate, and body temperature), 12-lead ECGs, clinical laboratory tests, and suicidal ideation and behavior risk monitoring.

3. STUDY DESIGN

3.1. Randomization

Subjects will be randomized to receive either GNX or PBO in a 1:1 ratio and randomization will be stratified based on current use of Epidiolex. Block randomization will be used with a block size specified in Rave RTSM (randomization and trial supply management) study randomization list settings document.

3.2. Sample Size Considerations

The planned sample size for this study is 128 randomized and dosed subjects. These subjects will be allocated into 2 treatment groups, GNX vs. PBO, in a 1:1 ratio. Sixty-four subjects per group provides at least 90% power for the primary endpoint assuming an estimate of the treatment effect (the actual analysis will use a Wilcoxon rank-sum test, which has approximately the same power as the ANOVA) difference that is at least 25% and a common standard deviation of 43% or less.

3.3. Multiplicity

All endpoints will be assessed descriptively, by the treatment to which the participants are randomized, with point estimates and 95% confidence intervals. The primary and secondary endpoints will be assessed in an inferential manner. Multiplicity control will be used for the primary and key secondary endpoints. A gate keeping approach will be used to control the study-wide Type-I error rate. There is a single primary efficacy endpoint, and formal hypothesis testing will be performed for this endpoint first. If the null hypothesis is rejected for the primary efficacy endpoint at the 2-sided 5% alpha level, then statistical hypothesis testing will be performed sequentially on the 4 key secondary endpoints as the order listed in Section 2.3.2. The testing will stop if the testing result is non-statistically significant. All other non-key secondary and exploratory endpoints will use a 2-sided 5% alpha with no adjustment for multiplicity.

4. STATISTICAL METHODOLOGY

4.1. General Methodology

Unless otherwise stated, SAS®software (SAS Institute Inc, Cary, NC; SAS Enterprise Guide 8.3 Version or later) will be used for the generation of all tables, graphs, and statistical analyses. Summary statistics including the number of subjects (n), mean, standard deviation, median, minimum and maximum will be presented for continuous variables. Generally, the minimum and maximum values will be presented to the same decimal precision as the raw values, the mean and median values to one more, and the standard deviation, to two more decimal places than the raw values. For categorical variables, per category, the absolute counts (n) and percentages (%) of subjects with data, and if appropriate, the number of subjects with missing data, will be presented. Percentages will be presented to one decimal place.

For AEs, medical history and concomitant medications reported on a per-subject basis, the denominator for the percentage calculation will be the number of subjects in each treatment group. A subject will be considered at risk if the subject is in the analysis set and in the subgroup of interest.

If a p-value is less than 0.0001 it will be displayed as "< 0.0001". All p-values larger than 0.0001 will be rounded to 4 decimal places.

4.2. Analysis Sets

For the purposes of analysis, the following analysis sets are defined:

- Intent-to-treat (ITT) Set – Includes all randomized participants, who dosed and had at least one post-baseline efficacy assessment. Participants will be summarized within the treatment group to which they were randomized. This set will be used for efficacy analyses.
- Per-protocol (PP) Set – Includes all participants in the ITT set who received IP for at least 6 weeks, provided at least 5 weeks of post-baseline seizure data, and without major protocol deviations that affect the efficacy endpoints (defined prior to database lock).
- Safety Analysis Set – Includes all randomized participants who receive at least 1 dose of the IP. Participants will be summarized within the treatment group for which they received treatment. This set will be used for the safety analyses.

The ITT analysis set is used to analyze endpoints related to the efficacy objectives and the safety analysis set is used to analyze the endpoints and assessments related to safety. The primary analysis will be performed within the ITT population; supportive analyses in the PP population will also be conducted.

4.3. Statistical Methods for Efficacy Analyses

4.3.1. Estimand Framework for the Primary Endpoint

Population: The target study population comprises children and adults with TSC-related epilepsy who also meet the inclusion and exclusion criteria as specified in the study protocol. The analysis population is the Intent-to-Treat (ITT) Population defined as all randomized participants.

Variable: The variable is the primary efficacy endpoint, the percentage change from baseline in 28-day primary endpoint seizure frequency during the double-blind phase.

Population-Level Summary: The population-level summary will be the difference of the percentage change on primary efficacy endpoint between GNX and PBO groups and its 2-sided 95% confidence interval.

Accounting for Intercurrent Events (ICEs): Intercurrent events and their handling rules are as follows:

- Discontinuations due to lack of efficacy;
- Discontinuations due to expected (per IB) adverse events.

Both ICEs will be addressed by the treatment policy strategy. The use of rescue medication is allowable at any time during the study. A patient may require rescue medication during the trial but continues with the trial post-rescue medication. The efficacy of the trial treatment will be assessed regardless of the need for rescue medication. Alternative approaches to handling ICEs will be addressed in sensitivity analyses, see Section 4.3.3.

4.3.2. Primary Efficacy Endpoint

Hypothesis testing will be performed for the primary endpoint: the percentage change in 28-day primary endpoint seizure frequency during the DB phase relative to the baseline, based on the primary endpoint seizure types. The hypothesis testing is:

$$H_o: \theta_{PBO} = \theta_{GNX} \qquad H_a: \theta_{PBO} \neq \theta_{GNX}$$

where θ_{GNX} is the percentage change of seizure frequency during the DB phase relative to the baseline for participants who treated with GNX and θ_{PBO} is similarly defined for participants receiving placebo.

Primary endpoint seizure types are defined as the following: focal motor seizures without impairment of consciousness or awareness, focal seizures with impairment of consciousness or awareness with motor features, focal seizures evolving to bilateral, tonic-clonic seizures, and generalized motor seizures including tonic-clonic, bilateral tonic, bilateral clonic, or atonic/drop seizures. Seizures that do not count toward the primary endpoint include: focal aware seizures without motor features (eg, focal nonmotor seizures with or without impaired awareness, absence or myoclonic seizures), infantile or epileptic spasms, and myoclonic seizures.

Post-baseline 28-day seizure frequency will be calculated as the total number of seizures in the double-blind phase divided by the number of days with seizure data in the same period, multiplied by 28. Baseline 28-day seizure frequency will be calculated as the total number of seizures in the baseline phase divided by the number of days with seizure data in the same period, multiplied by 28. The calculation for percent change from baseline in 28-day seizure frequency will be done as follows for each participant:

$$\left(\frac{[(\text{Post-baseline 28-day seizure frequency}) - (\text{Baseline 28-day seizure frequency})]}{(\text{Baseline 28-day seizure frequency})}\right)\times 100\%$$

The baseline, post-baseline, and arithmetic and percent changes from baseline in 28-day seizure frequency will be summarized using descriptive statistics.

The difference between the treatment groups in the percent changes will be tested for statistical significance. The estimate of the median difference and corresponding 95% confidence interval based on the Hodges–Lehmann approach will be provided. Since the percent differences are anticipated to display skewness and/or outliers, the test will be performed using the Wilcoxon Rank-Sum statistic using a 2-sided significance level of 0.05.

4.3.3. Sensitivity Analysis

Five sensitivity analyses of the primary efficacy endpoint of change in 28-day frequency of the primary seizure types will be performed:

1. Intermittent (random/sporadic) missing data during the DB phase and any missing data during the baseline phase will be assumed missing completely at random and the collected data will be used to calculate the 28-day seizure frequencies. For early drug termination prior to the end of the DB phase, caregivers will be instructed to continue to provide daily seizure records until the end of the 16-week DB phase (hence further preventing missingness).

   In the first sensitivity analysis, the following imputation approach will be used for the primary outcome measure when a subject stops recording measurements permanently prior to the end of the DB phase: for those missing days, the corresponding median PBO data will be imputed (irrespective of treatment arm) as follows:

- Partition the subjects in the PBO arm into quartiles, where the quartile equals Floor [Rank*4/(n+1)], where Floor [.] is the integer floor function, Rank is the rank of the 28- day seizure frequency in the PBO arm based on all available PBO measurements during the 4-week baseline phase (tied scores receive tied ranks), and n is the number of subjects in the PBO arm.
- Assign each subject for whom imputation is needed into the quartile that covers the 28-day seizure frequency based on that subject's available measurements during the 4-week baseline phase. If a frequency is lower than the lower bound of the lowest quartile (which could happen only in the GNX arm), then assign that subject to the lowest frequency quartile; likewise, if a frequency is higher than the upper bound of the highest quartile, then assign that subject to the highest frequency quartile. If a frequency falls between two quartiles (which, again, could happen only in the GNX arm), then based on the distances between the frequency and the upper and lower bounds of the lower and higher quartiles, respectively, assign the subject to the closer quartile (in case of a tie, assign to the lower quartile).
- Within each quartile, compute the median 28-day seizure frequency in the PBO arm based on all available PBO measurements during the DB treatment phase. Label that 'X' and define 'A' to be 'X'/28, (i.e., the daily average on placebo during the DB phase).
- For any days (whether they be on the PBO or GNX arm) with missing counts due to a subject stopping recording measurements permanently, impute 'A' from the corresponding quartile on that day.

2. The second sensitivity analysis will explore the possibility that subjects who stop recording their seizure counts tend to have higher seizure counts than the other subjects. The imputation method described in the 3rd and 4th bullets above, except that quartiles are not used, will be modified to use the median of the 5 highest counts (equivalent to the 3rd highest count) among all the PBO subjects with data.

3. The third sensitivity analysis will be performed on subjects who achieved their maintenance dose at the start of the maintenance period. This analysis will exclude those subjects who are still titrated continuously during maintenance period.

4. The fourth sensitivity analysis will be performed for the primary outcome measure by excluding all intercurrent events.

4.3.4. Key Secondary Efficacy Endpoints

Hypothesis testing will be performed for the three key secondary endpoints:

1. Percentage change from baseline in 28-day primary endpoint seizure frequency during the maintenance period.

   The hypothesis testing is:

   $$H_o\text{: } \theta_{PBO} = \theta_{GNX} \qquad H_a\text{: } \theta_{PBO} \neq \theta_{GNX}$$

   where θ_{GNX} is the percentage change of seizure frequency during the maintenance period relative to the baseline for participants who treated with GNX and θ_{PBO} is similarly defined for participants receiving placebo.

   Similar to the primary efficacy endpoint, the test will be performed using the Wilcoxon Rank-Sum statistic using a 2-sided significance level of 0.05.

2. Number (%) of participants considered treatment responders during the DB phase.

   The hypothesis testing is:

   $$H_0\text{: } \gamma_{\mathrm{PBO}} = \gamma_{\mathrm{GNX}} \quad vs. \quad H_a\text{: } \gamma_{\mathrm{PBO}} \neq \gamma_{\mathrm{GNX}}$$

   where γ_{GNX} is the proportion of participants who treated with GNX experiencing ≥ 50% reduction from baseline in 28-day seizure frequency based on the primary seizure types and γ_{PBO} is similarly defined for participants receiving placebo.

   The test comparison will be analyzed using Fisher's Exact test and the corresponding p-value will be provided. The difference in proportions along with the 95% confidence interval for the difference also will be presented. The number and percentage of subjects experiencing a reduction (improvement) of at least 50% from baseline in the 28-day seizure frequency of the primary seizure types will be summarized. In addition, a cumulative responder curve figure will be provided, in which the X-axis represents amount of improvement in increments of 5%, and the Y-axis represents the percentage of subjects improving by at least the amount on the X-axis. The treatment groups will be presented separately within the figure.

3. Number (%) of participants considered treatment responders during the maintenance phase.

   This is similar to #2 above. The only difference will be replacing the DB phase by the maintenance phase.

4. CGI-I at the last scheduled visit in the DB phase.

   The hypothesis testing is:

   $$H_0\text{: } OR_{\mathrm{GNX/PBO}} = 1 \quad vs. \quad H_a\text{: } OR_{\mathrm{GNX/PBO}} \neq 1$$

   where $OR_{\mathrm{GNX/PBO}}$ is the odd ratio between GNX and PBO in regard to CGI-I.

The CGI-I is 7-point scale, and the number and percentage of subjects with each score will be summarized. The scores range from 1=very much improved and 7=very much worse. The CGI-I is completed by both parent/caregiver/LAR and clinician. The score at the last visit in the DB treatment phase will be analyzed using ordinal logistic regression. Proportional odds modelling will be carried out by including treatment group as a factor. The estimated odds ratios (GNX vs. PBO), 95% CI for the odds ratios, and the p-value testing the null hypothesis that the odds ratio is equal to 1, will be presented.

4.3.5. Secondary Efficacy Endpoints

Summary statistics including confidence intervals for the treatment differences will be used to summarize the results for the secondary endpoints. Testing, as outlined below, will be used for the secondary endpoints.

Change from baseline in ADAMS total score and sub-score.

The ADAMS (Anxiety, Depression, and Mood Scale) is a rating scale designed to screen for anxiety and depression in individuals with intellectual disability. The 28-question scale is filled out by the parent/caregiver/LAR and is based on the subject's behavior during the past 6 months at the Baseline visit and since the last visit at the subsequent visits. Each of the questions are scored as 0 = Not a problem, 1 = Mild problem, 2 = Moderate problem, or 3 = Severe problem. The questions are grouped into 5 subscales:

Manic/ Hyperactive Behavior	Depressed Mood	Social Avoidance	General Anxiety	Obsessive/ Compulsive Behavior
3.	**5.**	**2.**	**1.**	**8.**
4.	**9.**	**6.**	**3.**	**16.**
12.	**10.**	**13.**	**7.**	**20.**
17.	**14.**	**19.**	**11.**	
22.	**18.**	**21.**	**15.**	
	23.	**25.**	**24.**	
	28.	**27.**	**26.**	

The subscale totals will be summarized with descriptive statistics of the baseline and post-baseline values and the arithmetic changes from baseline.

Change from baseline in quality-of-life scales: Peds-QL-FIM, SF-36, and ELDQOL.

- **Peds-QL-FIM**

  The Peds-QL-FIM is designed to measure the impact of pediatric chronic health conditions on parents and the family. The Peds-QL-FIM measures parent self-reported physical, emotional, social, and cognitive functioning, communication, and worry. The module also measures parent reported family daily activities and family relationships (see Protocol Section 10.14).

  Total Score is computed by averaging all 36 items. Parent HRQOL Summary Score is computed by averaging 20 items in Physical, Emotional, Social, and Cognitive Functioning. Family Summary Score is computed by averaging 8 items in Daily Activities and Family Relationships.

- **SF-36**

  The SF-36 is a multi-purpose survey designed to capture participant or parent/caregiver perceptions of own health and well-being. The SF-36 has 36 items grouped in 8 dimensions: physical functioning, physical and emotional limitations, vitality, social functioning, bodily pain, general, and mental health (see Protocol Section 10.15).

  Scoring the 36-Item Health Survey is a two-step process. First, precoded numeric values are recoded per the scoring key given in Table 1. Note that all items are scored so that a high score defines a more favorable health state. In addition, each item is scored on a 0 to 100 range so that the lowest and highest possible scores are 0 and 100, respectively. Scores represent the percentage of total possible score achieved. In step 2, items in the same scale are averaged together to create the 8 scale scores. Table 2 lists the items averaged together to create each scale. Items that are left blank (missing data) are not taken into account when calculating the scale scores. Hence, scale scores represent the average for all items in the scale that the respondent answered.

Table 1: Recoding Items for SF-36

Item numbers	Change original response category	To recoded value of:
1, 2, 20, 22, 34, 36	1 →	100
	2 →	75
	3 →	50
	4 →	25
	5 →	0
3, 4, 5, 6, 7, 8, 9, 10, 11, 12	1 →	0
	2 →	50
	3 →	100
13, 14, 15, 16, 17, 18, 19	1 →	0
	2 →	100
21, 23, 26, 27, 30	1 →	100
	2 →	80
	3 →	60
	4 →	40
	5 →	20
	6 →	0
24, 25, 28, 29, 31	1 →	0
	2 →	20
	3 →	40
	4 →	60
	5 →	80
	6 →	100

Item numbers	Change original response category	To recoded value of:
32, 33, 35	1 →	0
	2 →	25
	3 →	50
	4 →	75
	5 →	100

Table 2: Averaging Items to Form Scales for SF-36

Scale	Number of items	After recoding per Table 1, average the following items
Physical functioning	10	3 4 5 6 7 8 9 10 11 12
Role limitations due to physical health	4	13 14 15 16
Role limitations due to emotional problems	3	17 18 19
Energy/fatigue	4	23 27 29 31
Emotional well-being	5	24 25 26 28 30
Social functioning	2	20 32
Pain	2	21 22
General health	5	1 33 34 35 36

- **ELDQOL**

  The ELDQOL is a 70-item parent/caregiver-reported measure that examines seizure severity, seizure related injury, AED side effects, behavior, mood, physical status, cognitive and social functioning, communication, overall health and quality of life, and family concerns. The higher the score, the poorer the participant's quality of life (see Protocol Section 10.16). See the scoring instructions below:

Scoring instructions for 4 subscales of the re-validated version of ELDQOL

Seizure Severity Scale (Qs. 1 – 14)

Qs. 1, 5, 7, 8, 9, 10, 11, 13

Reverse the individual item scoring so that

1 = 4
2 = 3
3 = 2
4 =1
5 - define as a missing value so that it is not included in the calculation of the scale score

Qs. 4, 12

Reverse the individual item scoring so that

1 = 4
2 = 3
3 = 2
4 =1
5 = 0
6 - define as a missing value

Qs. 2, 14

5 - Define 5 as a missing value.
Other scores unchanged.

Qs. 3, 6

5 = 0
Define 6 as a missing value.
Other scores unchanged

Total scale score range = 10-56

Side-effects Profile (Qs. 17 & 18) (Only items in Q.18 used to compute sub-scale.)

Score is the sum of the individual scores for each of the 19 items in Q.18

For each item, score of 5 - define as a missing value.

Any additional problems itemised by respondents should be recorded, but are not added to the total scale score.

Total scale score range = 19-76, where high scores = high level of side effects reported.

Behaviour (Qs. 19 – 27)

Qs.19, 24, 25, 26

Define 5 as a missing value
Other scores unchanged

Qs. 20, 21, 27

Reverse scoring so that:

1 = 4
2 = 3
3 = 2
4 =1
5 - define as a missing value

Qs.22, 23

Recode so that:

5 = 4
6 – define as a missing value
Other scores unchanged

Total scale score range = 9-36

Mood Scale (Q. 28)

Reverse scoring of negative items (aggressive, irritable, tearful, hyperactive, sad, agitated, restless, tantrum-prone, frustrated, withdrawn), so that:

1 = 4
2 = 3
3 = 2
4 =1
5 - define as a missing value

Total Scale score range = 16-64

General points

For each subscale, higher score = poorer QOL / functioning.

Give non-responses a missing value (suggest 9 or -9).

The mean score for each subscale is computed as the sum of the items divided by the number of items answered. To take account of any missing data in a subscale, the mean of the completed items can be imputed, providing at least 50% of the items are completed. If more than 50% of the items in the subscale are missing, then the total subscale score should be classed as a missing value.

All scales and subscales will be summarized separately with descriptive statistics of the baseline and post-baseline values and the arithmetic changes from baseline.

Change from baseline in the percentage of seizure-free days during the DB phase, based on primary endpoint seizure type.

The percentages of seizure-free days will be based on the primary seizure types. Post-baseline percentage of seizure-free days will be calculated as the number of days in the DB treatment phase with zero seizures divided by the number of days with seizure data in the phase, multiplied by 100. The baseline percentage of seizure-free days will be calculated as the number of days in the 4-week baseline phase with zero seizures divided by the number of days with seizure data in the baseline phase, multiplied by 100.

The baseline and post-baseline values and the arithmetic changes from baseline will be summarized using descriptive statistics.

CGI-CSID at the last scheduled visit in the DB phase.

The CGI-CSID is from the 7-point Caregiver Global Impression of Change in Seizure Intensity/Duration instrument. The scores range from 1=very much improved and 7=very much worse. The number and percentage of subjects with each score will be summarized.

Participants with a ≥ 25% and ≥ 75% reduction from baseline in primary endpoint seizure frequency during the DB phase.

Number (%) of subjects considered treatment responders, defined as those with a ≥25% and ≥75% reduction from baseline in primary seizure frequency during the DB phase will be summarized.

Participants with a ≥ 25%, ≥ 50%, and ≥ 75% reduction from baseline in primary endpoint seizure frequency during the maintenance period.

Number (%) of subjects considered treatment responders, defined as those with a ≥25%, ≥ 50%, and ≥75% reduction from baseline in primary seizure frequency during the maintenance period will be summarized.

Responder analysis for primary endpoint seizures and all seizures during the double-blind phase using the following response categories: ≤ 0%, > 0% to < 25%, ≥ 25% to < 50%, ≥ 50% to < 75%, and ≥ 75% to 100%.

Number (%) of subjects considered treatment responders categories, defined as those with a ≤ 0%, > 0% to < 25%, ≥ 25% to < 50%, ≥ 50% to < 75%, and ≥ 75% to 100% reduction from baseline in seizure frequency during the DB phase will be summarized using descriptive statistics (in both a table and a figure). The analysis will be conducted for primary endpoint seuzures and all seizures, respectively.

Percent change in 28-day frequency of all seizures.

This analysis is similar to primary efficacy endpoint during the DB phase, except all seizure types will be applied.

Change from baseline in the percentage of seizure-free days, based on all seizure types.

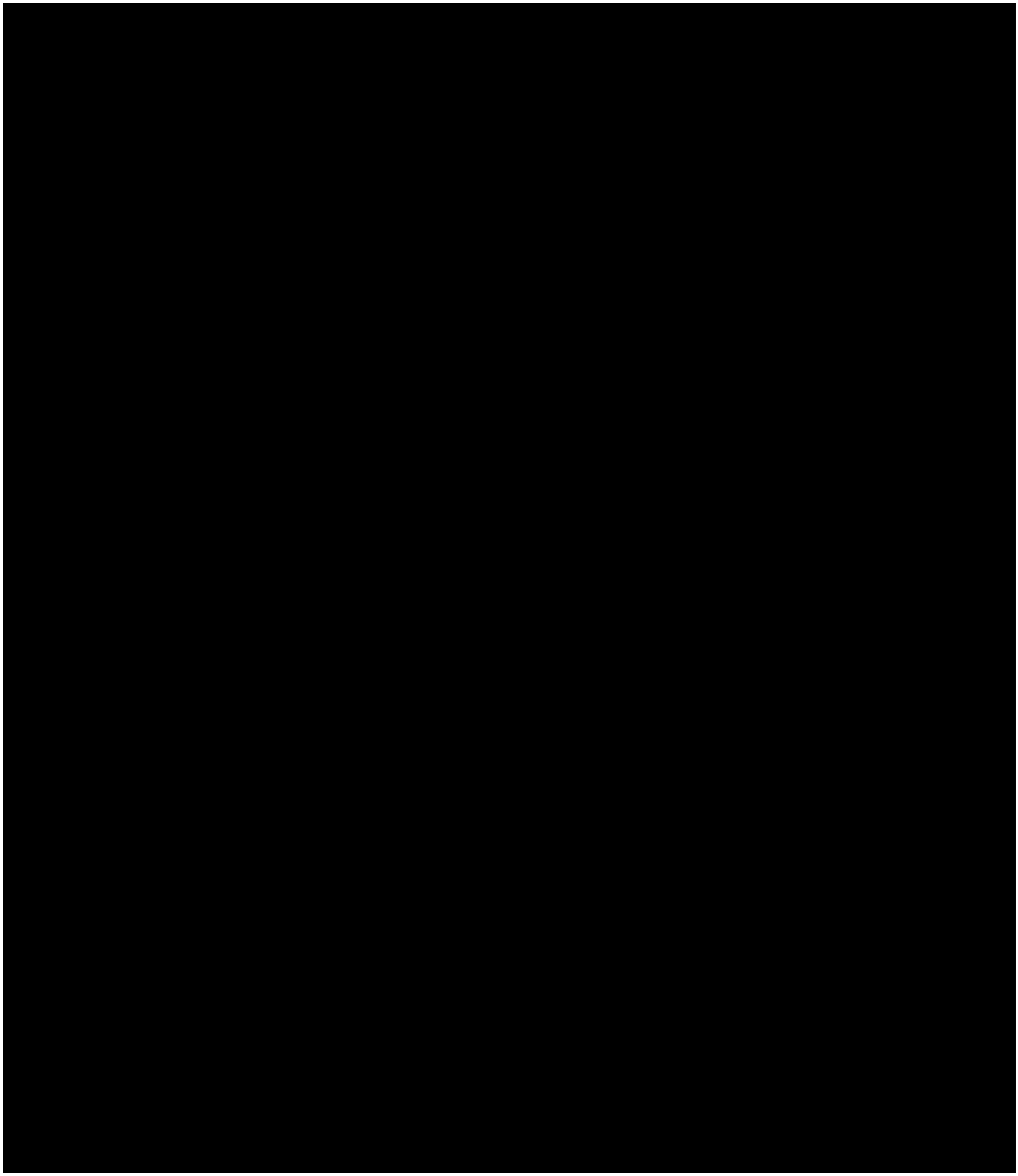
The percentages of seizure-free days will be based on all seizure types. Post-baseline percentage of seizure-free days will be calculated as the number of days in the DB treatment phase with zero seizures divided by the number of days with seizure data in the phase, multiplied by 100. The baseline percentage of seizure-free days will be calculated as the number of days in the 4-week baseline phase with zero seizures divided by the number of days with seizure data in the baseline phase, multiplied by 100.

The baseline and post-baseline values and the arithmetic changes from baseline will be summarized using descriptive statistics.

Change from baseline in the longest seizure-free interval, based on primary endpoint seizure type and all seizure types.

The longest seizure-free intervals (percentage of days) will be based on both the primary seizure types and all seizure types. The post-baseline longest period of time seizure-free is defined as the longest seizure-free period (days) in the DB treatment phase divided by the days with seizure data during the phase, and then multiplied by 100%. The baseline longest period of time seizure-free is defined as the longest individual seizure-free period (days) in the baseline phase divided by the days with seizure data during the baseline phase, and then multiplied by 100%. For both phases, the longest period is based on the number of days without an interruption of a day with a seizure. If there is an interruption of a day without any seizure data, then that day will not count. For example, if there are seizures on January 1st, no seizures on January 2nd through 4th, no seizure data on January 5th, no seizures on January 6th, and there are seizures on January 7th, then the longest period of time seizure free in this interval would be 4 days (January 2nd through 4th and January 6th).

The longest seizure-free interval will be summarized using descriptive statistics for the baseline and post-baseline values, as well as the arithmetic changes from baseline.



4.3.7. Subgroup Analyses for Primary Endpoint

Summaries with descriptive statistics and frequency counts on the primary efficacy endpoint for both groups will be provided for the subgroups indicated below. Forest plots will also be created for subgroup analyses and will display treatment differences. The subgroups will be:

- Gender (female vs. male)
- Age Group (12 months to 23 months; 2 to 11 years; 12 to 17 years; 18 to 64 years; and 65 years and above)
- Country/Region (US, China, EU, ROW)
-
- Concomitantly taking everolimus (yes vs. no)
- Concomitantly taking cannabidiol (yes vs. no)
- mTOR medication (taking sirolimus, everolimus, tacrolimus vs. no)

4.3.8. Handling of Missing Data

Available data from patients who withdraw from the study prematurely will be included in analyses, unless otherwise indicated. Summary statistics will be reported based upon observed data. No data will be imputed or carried forward for missing values or for patients who discontinue early. All missing data and missing or partial dates for AEs and medications will be queried for a value. In the unlikely event that no value can be obtained, substitutions will be made as described below.

- **Missing and Partial Start/Stop Dates - AEs and Concomitant Medications**

  The handling of partial start and stop dates for AEs are described in Table 3 below. Similar algorithms for handling missing and partial dates of concomitant medication usage are described in Table 4. In both cases, if a stop date is complete and an imputed start date is after the stop date, the start date will be set to the stop date.

- **Missing Seizure/Medication Diary Entries**

  The calculation of seizure frequency endpoints will include only days for which diary entries are available. Missing seizure data will not be imputed.

Table 3: Adverse Event Start/Stop Date Imputation

Parameter	Missing	Additional Conditions	Imputation
Start date for AEs	D	M and Y same as M and Y of first dose of IP	Date of first dose of IP
		Y same but M prior to month of first dose of IP	Last day of month
		Y same but M after month of first dose of IP	First day of month
		Y is prior to year of first dose of IP	Last day of month
		Y is after year of first dose of IP	First day of month
	M	Y is same as Y of first dose of IP	Month of first dose of IP
		Y is prior to year of first dose of IP	M = December
		Y is after Y of first dose of IP	M = January
	D and M	Y same as Y of first dose of IP	Date of first dose of treatment
		Y prior to Y of first dose of IP	M and D will be December 31
		Y after Y of first dose of IP	M and D will be January 1
	Y, or M, D, Y	Y and/or start date missing	Date of first dose of treatment
Stop date for AEs	D	M and Y not missing	Use last day of month (i.e. D may take on values of 28, 29, 30, or 31, depending on month)
	M	Y not missing; if D also missing, impute D as described above	M = December
	Y, or M, D, Y	Y and/or stop date missing.	No imputation. Date left missing.

D=day, M=month, Y=year

Note: In all cases, if an estimated start date is after a complete stop date, the start date will be set to the AE stop date. Similarly, if the estimated stop date is before a complete or imputed start date, use the last day of the start day month.

Table 4: Imputation for Prior/Concomitant Medication Missing and Partial Dates

Parameter	Missing	Additional Conditions	Imputation
Start date for con meds	D only	M and Y are not missing or imputed.	Use 1st day of M.
	M only	D and Y are not missing or imputed.	M = January
	M and D	Y is not missing or imputed.	Use Jan 01 of Y
	M, D, and Y	None - date completely missing	No imputation but considered concomitant unless stop date is prior to first dose of IP.
Stop date for con meds	D only	M and Y are not missing or imputed.	Last day of month
	M only	D and Y are not missing or imputed.	M = December
	M and D	Y is not missing or imputed	Use Dec 31 of Y
	M, D, and Y	None - date completely missing and NOT ongoing	No imputation

Note: In all cases, if an estimated start date is after a complete stop date, the start date will be set to the end date of medication.

5. BASELINE, DISPOSITION, AND EXPOSURE

5.1. Study Day and Analysis Window

For reporting purposes, Study Day will be calculated from the date of the Baseline Treatment Visit (Visit 2, Day 1). Similarly, Study Week will be calculated from the date of the Baseline Treatment Visit.

For all assessments performed on or after the first day of treatment, i.e. reference date ≥ date of first dose of IP, Study Day is calculated as:

- Study Day = (date of assessments/events - first day of treatment) + 1

For assessments performed prior to the first day of treatment, i.e. reference date < date of first dose of IP, Study Day is calculated as:

- Study Day = date of assessment/event – first day of treatment.

5.2. Definition of Baseline

The Baseline Visit is defined as Visit 2 (Day 1). Baseline values for non-seizure efficacy and safety assessments will be the last values collected prior to the first administration of IP (i.e., GNX), unless otherwise specified.

Baseline seizure frequency data is collected during the four weeks between the Screening Visit (Visit 1, Week -4) and the Baseline Visit (Visit 2, Day 1). This interval may range from no less than 28 days to no more than 34 days.

5.3. Analysis Visit Windows

Efficacy Analyses

Efficacy analyses will use data only from scheduled visits. Data will be reported by designated visit. Data collected at unscheduled visits will be provided in listings. For analysis of data collected across specific study periods, the following definitions will be applied:

- Baseline period: Visit 1 to Visit 2; Study Week -4 (Day -28) through Study Day 1 (date of first dose of IP)
- Titration period: Visit 2 + 1 day to Visit 7; Study Day 1 through Study Week 7 (Day 28 ± 3 days; date of Visit 7)
- Maintenance period: Visits 7 + 1 day to Visit 12; Study Week 5 (± 3 days) through Study Week 16 (± 3 days)
- 16-week treatment period: titration period + maintenance period (including taper and safety visits if not continuing in OLE)

Safety Analyses

Safety data will be summarized by actual visit. No visit windowing will be applied.

5.4. Demographics and Baseline Characteristics

Demographics and baseline characteristics will be summarized and listed for all patients in the Safety Population. Demographic characteristics, including age, height, weight, and BMI, will be summarized with descriptive statistics (n, mean, median, SD, minimum and maximum). Categorical characteristics such as age group (≤ 17 years, > 17 years), weight group (≤ 28 kg, > 28 kg), sex, race and ethnicity, will be summarized as counts and percentages.

Additionally, baseline clinical characteristics including results of genetic testing for pathogenic or likely pathogenic TSC1 or TSC2 variant, Tanner Staging, number of AEDs taken and stopped prior to screening, neurological exam results, developmental examination results (pediatric patients aged 2 to 17, inclusive), and Seizure Identification and Diagnostic Review Form (SIF/DRF) responses including baseline seizure types reported as primary and non-primary, IQ, intellectual disabilities, and EEG findings will be summarized.

Baseline values of efficacy and safety parameters will be summarized in the respective analysis tables.

5.5. Subject Disposition

The number of patients screened for this study, the number of screen failures and the number of patients who received at least one dose of IP will be summarized. Reasons for screen failure will be provided in the data listing. The final disposition of all patients who receive IP will be summarized. Reasons for discontinuation of study treatment and for discontinuation of study participation will be summarized. The total number and percentage of patients who complete the study will be summarized.

A listing of patient disposition will be provided. In addition, a summary of the number of patients included in each analysis set will be presented.

5.6. Eligibility Criteria and Protocol Deviations

The clinical team will review and identify deviations and the deviations will be recorded into the database. A data listing of subjects who violate any of the inclusion/exclusion characteristics will be provided as well as a data listing of subjects with other protocol violations.

5.7. Prior and Concomitant Medications

All investigator terms for medications recorded on the CRF will be coded using the World Health Organization Drug Dictionary Enhanced (WHODDE) Anatomical Therapeutic Chemical (ATC) drug dictionary. The number (percentage) of subjects who took prior and concomitant medications will be summarized in the Safety population by Anatomic Therapeutic Chemical (ATC) classification level 2 and Preferred Term (PT).

Prior medications are defined as medications/therapies that started and stopped prior to the first dose of study drug. Concomitant medications are defined as medications (other than the study drug) administered on or after the first dose of the study, regardless of when the medications started.

The number and percentage of patients in the Safety population reporting use of prior and concomitant treatments will be summarized and listed separately. In addition, antiepileptic prior and concomitant drugs (AEDs) and rescue medications will be presented separately.

5.8. Medical History

Medical history will be coded using Medical Dictionary for Regulatory Activities (MedDRA), version 25.1. The number and percent of subjects with clinically significant medical history at screening will be summarized by system organ class (SOC) and preferred term (PT) in the Safety population. Medical history will also be listed.

5.9. Study Drug Exposure

Exposure to study medication (GNX) will be summarized as the number of days on treatment, percentage of days dosed, maximum daily dose received, and total dosage received. The denominator for the percentage of days of study drug is the number of days the patient is on treatment in the respective part of the study. If the patient stops taking IP during the study, the last day that drug was taken will be used as the last day of treatment. For patients who are ongoing at the end of a specific part of the study, the last known treatment date will be used for the last day on treatment.

A listing of exposure data for each patient including total daily dose, start and end dates, and reason for any dose change will also be provided.

5.10. Study Drug Compliance

The percent compliance with study drug will be estimated using dosing information as follows:

$$\frac{[total\ doseage\ during\ maintenance\ period]}{[(last\ day\ of\ maintenance\ period) - (first\ day\ of\ maintenance\ period) + 1] \times (target\ daily\ dose)}\ x100$$

Reported dosing by visit and calculated compliance values will be listed.

5.11. Cross-Over Titration Summary

The end date of Double-Blind Maintenance Period and the start date of Cross-over Titration period are supposed to be two different days. However, one overlapped day may exist in some rare scenarios. When this appears in the EDC data, the following rule will be implemented to handle the analysis and data summary:

1. The last visit date in DB maintenance period will not be changed. This date will be set as the end date of the DB phase.
2. All scheduled/unscheduled tests (Lab, Vital Sign, ECG etc.) have been done on the last visit date in DB maintenance period will be summarized in the DB phase.
3. All the TEAEs, CMs and seizures occurred on the last visit date in DB maintenance period will also be summarized in the DB phase.

The following safety endpoints will be summarized descriptively:

- Incidence and severity of AEs, SAEs, and withdrawals and dose-reductions due to AEs
- Physical/neurological/developmental examinations
- Vital sign measurements
- 12-lead ECGs
- Clinical laboratory tests
- Suicidal ideation and behavior risk monitoring

6. ANALYSIS OF SAFETY ENDPOINTS

6.1. Adverse Events

Adverse events (AEs) are collected from the time of informed consent/assent through the defined follow-up period. AEs are coded using MedDRA dictionary, version 25.1 and are categorized by system organ class (SOC) and preferred term (PT). A treatment-emergent AE (TEAE) is defined as an AE that starts or worsens on or after the first day of dosing with IP. Detailed by-patient listings of all AEs reported during the study will be provided and will include verbatim and coded terms for each AE.

For the safety population, the following tabular summaries will be created:

- An overall summary of TEAEs will present the number and percentage of subjects with at least 1 reported TEAE, study drug related TEAE, serious TEAE, TEAE of Grade 3 or higher, study drug infusion interruption, and death
- TEAEs by system organ class and preferred term
- TEAEs by system organ class, preferred term, and maximum severity
- Study drug related TEAEs by system organ class and preferred term
- TEAEs by decreasing frequency of preferred term
- Grade 3 or higher TEAEs by system organ class and preferred term
- Grade 3 or higher study drug related TEAEs by system organ class and preferred term
- TEAEs that led to study drug infusion interruption by system organ class and preferred term
- TEAEs that led to study discontinuation by system organ class and preferred term
- Serious TEAEs by system organ class and preferred term
- Serious study drug related TEAEs by system organ class and preferred term

For incidence reporting, if a subject reported more than 1 AE that was coded to the same preferred term/system organ class, the subject will be counted only once for that specific preferred term/system organ class. For TEAEs presented by severity, the worst severity for each event during the clinical trial will be presented for each subject. If the severity is missing for a TEAE, then a “Serious” will be assigned. The imputed values for severity assessment will be used for incidence summaries, while the actual values will be used in data listings.

For TEAEs presented by relationship to study drug, the closest relationship to study drug for each event during the clinical trial will be presented for each subject. If the relationship to investigational product is missing for a TEAE, a causality of "Related" will be assigned. The imputed values for relationship assessment will be used for incidence summaries, while the actual values will be presented in data listings.

All AEs, SAEs, TEAEs leading to study drug interruption, TEAE leading to study discontinuation, SAEs leading to death and Grade 4 or higher AEs will be displayed in separate listings.

6.2. Clinical Laboratory Evaluations

Laboratory data collected in this study include serum chemistry, hematology values and urinalysis results. The baseline laboratory value is defined as the last value observed prior to the first administration of study drug. Any values collected after the administration of study drug are regarded as post-baseline. Change from baseline will be calculated as the post-baseline value minus the baseline value. Only the numeric part in laboratory values that contain non-numeric qualifiers, such as less than (<) a certain value or greater than (>) a certain value, will be used in the summary statistics.

Laboratory values and change from baseline in these values will be listed and summarized for patients in the Safety Population by group (chemistry, hematology or urinalysis) and by visit. Listings of patients with positive drug screen or pregnancy results will be provided.

Descriptive summaries of continuous laboratory values and change from baseline in these values will include the number of observations, mean, SD, median, minimum and maximum values at each time point. Data will be summarized in SI units. Shift tables comparing baseline classification based on reference range (i.e., normal, low, or high) to the worst classification recorded post-baseline will be presented. In the case that a patient has both "low" and "high" post-baseline results for the same laboratory parameter, the patient will be counted under both "low" and "high". A patient will be counted under the "normal" category post-baseline only if all post-baseline results are categorized as "normal".

For non-numeric urinalysis data, a shift table comparing baseline results (negative, trace or positive) to the maximum post-baseline result will be presented.

Listings of laboratory results for each patient will be presented. Values outside of the laboratory's reference range will be flagged in the listing. A by-patient listing of laboratory values outside the normal reference range for the parameter will be produced.

6.3. Vital Signs, Physical, Neurological and Developmental Examination Findings

Vital Signs

Vital signs measured in this study include systolic and diastolic blood pressure, heart rate, respiratory rate, and body temperature. Baseline for vital signs will be the values obtained at the last assessment prior to the first dose of IP, typically Visit 2, Day 1. Change from baseline is calculated as the post-baseline value minus the baseline value. Missing values will not be imputed.

Absolute values and changes from baseline in vital signs at each visit will be summarized in the Safety Population using descriptive statistics (n, mean, SD, median, minimum, and maximum). A by-patient, by-visit listing of vital signs collected and height and weight recorded will be provided.

All vital signs included the parameter values and overall interpretations will be listed. Changes from baseline physical examination findings will also be listed individually by patient.

The observed data at baseline and change from baseline for each measurement day will be summarized for each parameter with descriptive statistics. Plots of mean vital signs values (and standard error bars) over week 4 will be produced.

Physical Examination Findings

Physical examinations will be conducted at all clinic visits and findings will be recorded as normal, abnormal but not clinically significant, and abnormal and clinically significant. Abnormal changes in baseline physical examinations findings will be summarized using counts and percentages of patients in the Safety Population. Abnormal findings noted at baseline will be listed for each patient. Any changes in physical examination findings at post-baseline visits relative to the previous visit will be listed.

Neurological Examination Findings

Neurological examinations will be conducted at all clinic visits and will include evaluation of cranial nerves, motor function, sensory function, reflexes and coordination/cerebellar function. Findings will be recorded as normal, abnormal but not clinically significant, and abnormal and clinically significant. The number and percentage of patients with abnormal neurological findings will be summarized by parameter. Abnormal neurological findings at baseline will be listed for each patient. Any changes in neurological findings at post-baseline visits relative to the previous visit will be listed.

Developmental Examination Findings

Developmental examinations will be conducted on pediatric patients 2 to 17 years of age, inclusive. Developmental parameters assessed include speech and language skills, motor skills and social skills. The number and percentage of pediatric patients will be summarized by each parameter.

6.4. Electrocardiogram (ECG)

Safety ECG measurements will be collected throughout the study. Baseline ECG values are those collected at Visit 2 (Day 1) prior to the first dose of IP.

ECG parameters include heart rate, PR interval, RR interval, QRS interval, QT (uncorrected) interval, QTcF interval, and QTcB interval. In addition, the overall interpretation of the investigator will be documented as normal, abnormal but not clinically significant (NCS), or abnormal and clinically significant (CS). All clinically significant abnormal findings will be reported as AEs also. Absolute values and change from baseline in post-baseline values will be summarized. By-patient listings of ECG data and overall interpretations will be provided.

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

The Columbia-Suicide Severity Rating Scale (C-SSRS) is a patient questionnaire that evaluates suicidal ideation and behaviors. The results (yes/no) from the 5 questions under suicidal ideation and the 5 questions under suicidal behavior will be summarized at baseline and post baseline using percentages and counts. The denominator for percentages will be the number of subjects with a C-SSRS assessment at baseline or at any time post baseline. The post-baseline by-question summary will list a subject as "yes" if they selected yes at any time after baseline.

The C-SSRS suicidal behavior and suicidal ideation scores recorded at baseline and any changes in post-baseline scores from the previous visit will be summarized descriptively by visit, including the number and percentage of subjects reporting any suicidal behavior and any suicidal ideation as defined in Table 5. Adult and pediatric C-SSRS data will be summarized separately.

C-SSRS data collected, including scores for suicidal ideation, intensity of ideation and actual suicide attempts, will be listed by patient and visit.

Table 5: C-SSRS Categories for Analysis

Category	C-SSRS Item response is "YES"
Suicidal behavior	Preparatory acts or behavior Aborted attempt Interrupted attempt Actual attempt Complete suicide
Suicidal ideation	Wish to be dead Non-specific active suicidal thoughts Active suicidal ideation with any methods (not plan) without intent to act Active suicidal ideation with some intent to act, without specific plan Active suicidal ideation with specific plan and intent
Non-suicidal self-injurious behavior	Non-suicidal self-injurious behavior

6.6. Pregnancies

Any reported positive pregnancy results or reported pregnancies will be listed.

6.7. Tanner Staging

The Tanner scale is a scale of physical development in children, adolescent and adults. Patients will be rated as Tanner I, Tanner II, Tanner III, Tanner IV or Tanner V at baseline and end of the study. The number and percentage of patients in each category will be summarized and listed by visit.

7. APPENDIX

The following analysis will be included for the submission to EMA.

- The primary endpoint analysis will be based on the maintenance period only.
- All secondary endpoints analyses will be based on the maintenance period only.

8. SUMMARY OF MAJOR CHANGES IN THE PLANNED ANALYSES

The ITT definition is updated in this SAP.

- Intent-to-treat (ITT) Set – Includes all randomized participants, who dosed and had at least one post-baseline efficacy assessment. Participants will be summarized within the treatment group to which they were randomized. This set will be used for efficacy analyses.

In protocol, it was:

- All randomized participants. Participants will be summarized within the treatment group to which they were randomized. This population will be used for efficacy analyses.

Signature Page for VV-CLIN-000496 v3.0

Approval Task Task Verdict: Approved	[redacted] Biometrics 20-Aug-2024 19:42:19 GMT+0000

Approval Task Task Verdict: Approved	[redacted] Clinical 21-Aug-2024 00:05:24 GMT+0000

Approval Task Task Verdict: Approved	[redacted] Biometrics 21-Aug-2024 05:15:00 GMT+0000

Signature Page for VV-CLIN-000496 v3.0