Ovid Therapeutics, Inc.

OV101-17-001

A Phase 2a, Randomized, Double-Blind, Parallel-Group, Proof-of-Concept Study Evaluating the Safety, Tolerability, and Efficacy of OV101 in Fragile X Syndrome

06MAR2020

Statistical Analysis Plan

Version *3.0*

Prepared by:

[redacted]
Senior Biostatistician II

[redacted]

Approved by

_____ Date: ___/___/____
[redacted]
Senior Manager, Biostatistics
[redacted]

_____ Date: ___/___/____
[redacted]
VP, Biostatistics and Data Management
[redacted]

_____ Date: ___/___/____
[redacted]
VP, Clinical Development
[redacted]

TABLE OF CONTENTS

List of Abbreviations and Definition of Terms

Abbreviation	Definition
ABC-C	Aberrant Behavior Checklist-Community
ADAMS	Anxiety, Depression, and Mood Scales
ADHD	Attention Deficit Hyperactivity Disorder
AE	Adverse Event
AM	Morning dose only
AS	Angelman Syndrome
ATC	Anatomical Therapeutic Class
BASC-3 PRQ	Behaviour Assessment System for Children, 3rd Edition, Parenting Relationship Questionnaire
BID	Twice a Day
CGI	Clinical Global Impressions
CGI-I	Clinical Global Impressions-Improvement
CGI-S	Clinical Global Impressions-Severity
CI	Confidence Interval
Conner's 3	Conner's 3rd Edition
CSHQ	Children's Sleep Habits Questionnaire
CSR	Clinical Study Report
DSM-5	Diagnostic and Statistical Manual of Mental Disorders
eCRF	Electronic Case Report Form
[redacted]	[redacted]
EOS	End of Study
EOT	End of Treatment
[redacted]	[redacted]
FAS	Full Analysis Set
FDA	US Food and Drug Administration
FXS	Fragile X Syndrome
GABA	Gamma-aminobutyric Acid

GCP	Good Clinical Practice
ICF	Informed Consent Form
ICH	International Council for Harmonisation
IEC	Independent Ethics Committee
IQ	Intelligence Quotient
IRB	Institutional Review Board
IWRS	Interactive Web Response System
KiTAP	Test of Attentional Performance for Children
KO	Knockout
LAR	Legally Acceptable Representative
LOCF	Last Observation Carried Forward
MedDRA	Medical Dictionary for Regulatory Activities
mg	Milligram
mmHg	Millimeter of mercury
MMRM	Mixed Model with Repeated Measures
PD	Pharmacodynamic(s)
PedsQL	Pediatric Quality of Life Inventory
PGI-I	Parent Global Impressions-Improvement
PGI-S	Parent Global Impressions-Severity
[redacted]	[redacted]
PT	Preferred Term
Q1	1st Quartile
Q3	3rd Quartile
QD	Once a Day
RBANS	Repeatable Battery for the Assessment of Neuropsychological Status
RBS-R	Repetitive Behavior Scale-Revised
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SAS	Statistical Analysis System

SB-5	Stanford-Binet-5
SD	Standard Deviation
SI	International System
SOC	System Organ Class
SOL	Sleep Onset Latency
TEAE	Treatment-Emergent Adverse Event
TID	Three Times a Day
TST	Total Sleep Time
VABS-3	Vineland Adaptive Behavior Scale, 3rd Edition
WASO	Wake Time After Sleep Onset

1. INTRODUCTION

Fragile X syndrome (FXS) is the most common inherited form of intellectual disability, with a prevalence of 1 in 4000 males and 1 in 8000 females (Crawford et al 2001). It is also the most common known genetic cause of autism; approximately 25% to 30% of individuals with FXS meet the full diagnostic criteria for autism, and approximately 5% of children with autism carry the fragile X mutation (Hall et al 2008; Harris et al 2008). Males with FXS have moderate to severe intellectual disability, and females with FXS range from normal to moderate impairment (Hagerman et al 2010).

OV101 is the first highly selective, extrasynaptic gamma-aminobutyric acid (GABA) receptor agonist that binds as an orthosteric agonist to the α4- and δ-containing subunit of extrasynaptic GABA receptors. The mechanism of action of OV101 is unique among GABAergic agents, including benzodiazepines, zolpidem and other zolpidem-like drugs, neurosteroids, and drugs that act on GABA metabolism or uptake.

The use of OV101 for treatment of individuals with FXS is based on studies in a relevant mouse model of the disease, the *Fmr1* knockout (KO) mouse, which lacks the fragile X mental retardation protein due to a disruption in its *FMR1* gene. *Fmr1* KO mice have decreased GABA synthesis, decreased presynaptic GABA release, and decreased extrasynaptic levels of GABA, resulting in decreased tonic inhibition. OV101 was shown to restore tonic inhibition in the amygdala, reduce sensory hypersensitivity and motor hyperactivity, and improve pre-pulse inhibition (signal to noise ratio) in *Fmr1* KO mice. Cogram et al recently presented data that showed full normalization of abnormal repetitive, social, and anxiety-like behaviors with OV101 in another FXS KO mouse model, the *Fmr1* KO2 mouse. Taken together, these data support the potential of OV101 to treat the core behavioral deficits noted in individuals with FXS. Importantly, the ability of OV101 to potentiate tonic inhibition is unlike any other GABAergic agent, including benzodiazepines, zolpidem, other zolpidem-like drugs, neurosteroids, and drugs that act on GABA metabolism or uptake.

Phase 2 and Phase 3 studies in adult subjects with primary insomnia demonstrated that OV101 is effective in restoring classical sleep parameters (sleep induction and sleep maintenance) and slow wave sleep, resulting in an improvement in the quality and restorative effects of sleep.

This Statistical Analysis Plan (SAP) describes the data-handling and statistical procedures to be used for the statistical analysis and reporting of efficacy and safety data collected under the Ovid Protocol OV101-17-001 Amendment #2 (March 27, 2019). The methods of analysis in this SAP expand on statistical considerations identified in the protocol; where considerations are substantially different, they will be identified as such in this document. This SAP has been developed and finalized prior to locking the clinical database for the primary analysis. Any additional analyses required to supplement the analyses specified in this SAP will be considered exploratory and will be identified in the Clinical Study Report (CSR).

The SAP was written in accordance with the recommendations outlined in the International Conference on Harmonisation (ICH) E9 Guideline entitled “Guidance for Industry: Statistical Principals for Clinical Trials” and the most recent ICH-E3 Guideline entitled “Guidance for Industry: Structure and Content of Clinical Study Reports.”

2. OBJECTIVES

2.1. Primary Objective

- To evaluate the safety and tolerability of OV101 over 12 weeks of treatment across different daily dosing regimens in adolescent and adult male subjects with FXS.

2.2. Secondary Objectives

- To evaluate changes in behavior during 12 weeks of treatment with OV101 across different daily dosing regimens in adolescent and adult male subjects with FXS

2.3. Exploratory Objectives

- To evaluate changes in sleep, adaptive functioning, and quality of life across different daily dosing regimens of OV101 treatment over 12 weeks in adolescent and adult male subjects with FXS

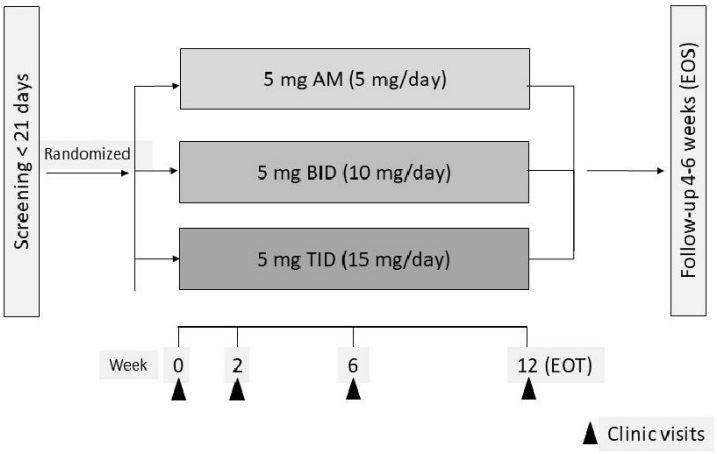
- To explore the relationships among study endpoints (e.g., behavior, sleep), where appropriate

3. INVESTIGATIONAL PLAN

3.1. Overall Study Design and Plan

This is a Phase 2a, randomized, double-blind, parallel-group study to evaluate the safety, tolerability, and efficacy of OV101 administered daily as 3 different dosing regimens: 5 mg/day (5 mg QD), 10 mg/day (5 mg BID), and 15 mg/day (5 mg TID) in adolescent and adult male subjects with FXS. Subjects with a clinically and molecularly confirmed diagnosis of FXS will be eligible for inclusion in the study (Section 4.1 of the protocol). The study will comprise a screening period of up to 3 weeks; a baseline visit on Day 1 for treatment randomization (enrollment), baseline assessments, and first dose of study drug; clinic visits for safety and efficacy assessments over a 12-week treatment period, including clinic visits at Week 2, Week 6, and Week 12 (end-of-treatment [EOT]); and an end-of-study (EOS) follow-up visit occurring any time between the ends of Week 16 and Week 18 (Figure 1). The schedule of study procedures is presented in Section 13.5.

Figure 1 Study Schematic



For each subject, after informed consent is obtained (Section 9.3 of the protocol), screening assessments will be completed within the 3 weeks preceding Baseline. At Baseline, subjects 13 to 22 years old (inclusive) who meet all eligibility criteria (Section 4.1 of the protocol) will be enrolled on Day 1 in a 1:1 target ratio between adolescents (13 to17 years old, inclusive) and adults (18 to 22 years old, inclusive). A sufficient number of subjects will be enrolled to ensure that at least 30 evaluable subjects complete 12 weeks of treatment.

Subjects will be randomly assigned in a 1:1:1 ratio to 1 of 3 treatment groups (5 mg QD, 5 mg BID, and 5 mg TID), stratified by age group (adolescent/adult) to achieve balance among treatment arms. There is no placebo group in this study. All subjects will take study drug TID (morning, afternoon, and evening), OV101 in the morning and OV101 or placebo in the afternoon and evening throughout the treatment period (including Baseline), as described in Section 5.2 and Section 5.4.1 of the protocol. There will not be any dose adjustments permitted during the study.

After random assignment at Baseline, subjects will receive the assigned package of blinded study treatment according to the treatment code for the first 6 weeks of treatment and will take their first dose of study drug at the clinic. They will receive another assigned package of study treatment at the Week 6 clinic visit.

Each subject (assisted by a legally acceptable representative [LAR]/caregiver) will complete paper sleep diaries and wear actigraphs over the 7-day periods immediately preceding Baseline and the Week 2, Week 6, EOT, and EOS clinic visits. The Baseline, Week 2, Week 6, and EOT visits will be full-day clinic visits.

For each subject, after informed consent is obtained, the maximum duration of individual subject participation will be 21 weeks. A subject will have completed the study when he completes the EOS visit. The end of the study is defined as the date when the last subject completes his last visit (includes follow-up visit).

Safety information will be collected during phone calls at the end of Weeks 1, 4, 8, and 10 (on Days 8, 29, 57, and 71, respectively), in addition to during every clinic visit (screening, baseline, Week 2, Week 6, EOT, and EOS). At the investigator's discretion, subjects may be evaluated at unscheduled clinic visits for reasons related to subject safety. At unscheduled visits, subjects will be queried about AEs, changes in concomitant medications, and suicidality, and safety laboratory assessments may be conducted. Periodic interim blinded review of safety data will be performed as part of routine pharmacovigilance activities. No interim statistical analysis is planned.

3.2. Study Endpoints

3.2.1 Primary Safety Endpoints

Treatment-emergent adverse events (TEAEs), treatment-related TEAEs, TEAEs leading to study discontinuation, and serious adverse events (SAEs) will be evaluated as primary safety endpoints.

3.2.2 Secondary Efficacy Endpoint

The efficacy parameters of the continuous secondary endpoints will be the mean changes from baseline in ABC-C total and subscale scores; in ABC-C factor score for FXS; in the ADAMS subscale scores; in the RBS-R total and subscale scores; in Conner's 3 subscale scores; in the Short Sensory Profile-2 total and subscale scores; and in the CGI-S and CGI-I scores.

3.2.3 Exploratory Endpoints

The exploratory efficacy endpoints are not listed in order of importance.

3.2.3.1 Caregiver-Completed Endpoints

- Change from baseline in the BASC-3 PRQ subscale scores
- Change from baseline in PGI-S and PGI-I scores
- Change from baseline in PedsQL total and subscale scores
- Change from baseline in the CHSQ total and subscale scores
- Change from baseline in sleep diary parameters, including how long it took the subject to fall asleep (sleep onset latency [SOL]), the amount of time the subject was awake during the night (WASO), the total amount of sleep obtained (total sleep time [TST]), and daytime sleepiness, defined as duration of napping in the daytime.

3.2.3.2 Clinician-Completed Endpoints

- Change from baseline in top 3 concerns visual analog scale
- Change from baseline in the VABS-3 overall composite and subscale scores
- Change from baseline in Pediatric Sleep CGI-Severity and Pediatric Sleep CGI-Improvement scores

3.2.3.3 Direct Subject Assessment Endpoints

- Actigraphy parameters for sleep, including SOL, WASO, sleep fragmentation, number of nocturnal awakenings, TST, sleep efficiency, and daily activity level.
- Mean change from baseline in KiTAP subtest scores
- Mean change from baseline in RBANS subtest scores

3.2.3.4 [redacted]

[redacted]

3.3. Treatments

OV101 is formulated as the monohydrate form (4,5,6,7-tetrahydro-isoxazolo[5,4-c]-pyridin-3-ol, monohydrate) in a white size 2 capsule. The capsule contains 5.645 mg of OV101 monohydrate, which corresponds to 5.0 mg of OV101 (Table 1).

Table 1 Pharmaceutical Composition of OV101 for the 5-mg Capsule

Component	Function	Amount (mg)
OV101[a]	Active compound	5.645
Microcrystalline cellulose	Diluent	138.6
Magnesium stearate	Lubricant	0.725
Size 2 hard-gelatin capsule	-	60

a Delivered as the zwitterion monohydrate. One gram of the anhydrous free zwitterion is equivalent to 1.129 g of the zwitterion monohydrate.

Placebo capsules will be identical in appearance to the capsules containing OV101 and have the same excipient ingredients but will not contain the active compound. Metrics Contract Services (Greenville, NC) will manufacture the OV101 and placebo capsules.

Subjects will take all doses orally (assisted by an LAR/caregiver, if necessary) TID (morning, afternoon, evening), at approximately the same times each day, for the duration of the 12-week treatment period. The morning and afternoon doses should be taken with light low-fat food (a glass of low-fat milk would qualify as food), and the afternoon dose should be taken at least 6 hours after the morning dose. The evening dose should be taken 30 minutes before bedtime. Capsules may be opened and the contents sprinkled onto 1 teaspoon of low-fat semiliquid food (e.g., applesauce, yogurt, pudding) for ingestion, but this approach must be followed consistently throughout the study. The capsule contents must not be placed directly in liquid. The LAR/caregiver must document specifically how each capsule was taken (e.g., swallowed whole or taken in 1 teaspoon of semiliquid food).

3.4. Dose Adjustment/Modification

There will not be any dose adjustments permitted during the study.

4. GENERAL STATISTICAL CONSIDERATIONS

All statistical analyses will be performed using SAS® Version 9.3 or higher.

All clinical study data will be presented in subject data listings. In by-treatment group descriptive presentations, the 3 randomized treatment groups will be presented separately as well as overall.

Descriptive statistics (number of subjects, mean, standard deviation (SD), median, Q1, Q3, minimum, and maximum) will be presented by treatment group for continuous variables and their changes from baseline. In addition, 95% CIs for the mean change from baseline will be provided for continuous variables. Graphical displays will be utilized to investigate trends over time and dose and the relationship between certain variables and differences between age groups. Ordinal variables with sufficient numerical categories (e.g., 7-point scales of PGI-S and CGI-S) and subscale total scores from questionnaires will be treated as continuous variables.

Categorical (not treated as continuous) endpoints (including AE incidence and binary endpoints) will be summarized by frequency and percentage. Frequencies and percentages will be presented for the shifts in these categories (e.g., low to normal, low to high, high to low) from baseline to each post-baseline assessment time point for scales with categorical variables. If there are missing values, the number missing will be presented (as a category), but without a percentage.

For all analyses, 'Baseline' refers as the last non-missing observation on or before the date of first dose of study medication. Change from baseline will be calculated by subtracting baseline value from post-baseline assessment for each subject.

If there are multiple records within a pre-specified visit window the non-missing visit closest to the scheduled visit day will be included in the summary tables.

The following reporting conventions apply generally to tables, listings and figures:

- All mean and median values will be formatted to one more decimal place than the measured value. Standard deviation values will be formatted to two more decimal places than the measured value. Minimum and maximum will be formatted to the same decimal as the measured value;
- All percentages will be rounded to one decimal place. The number and percentage of responses will be presented in the form xx (xx.x), where the percentage is in the parentheses. In the case the numerator is equal to the denominator, the percentage should be presented as (100) instead of (100.0). For TEAEs, the number, percentage, and event incidence will be presented in the form of xx (xx.x) [E], where E is the number of events;

- All listings will be sorted for presentation in order of randomized treatment group, study center, subject, and date of procedure or event;
- All analysis and summary tables will have the analysis population sample size for each treatment group in the column heading (i.e., number of subjects);
- P-values will be rounded to 4 decimal places. P-values that round to 0.0000 will be presented as "<0.0001" and p-values that round to 1.0000 will be presented as ">0.9999";
- The day of first dose of study drug, OV101, will be defined as Day 1. For day calculation, if date is before first dose date then day=date - first dose date, if date is after first dose date then day=date - first dose date + 1;
- Listings of all data will identify the age group (adolescent/adult) of each subject.

4.1. Visits and Day Ranges

Since it is not always possible for all study participants to come in for their clinic visits on the exact day specified in the protocol schedule, the 'Week' of a subject visit will be defined by the following relative day ranges.

Visit	Target	Study Days
Baseline	Day 1	Day 1
Phone 1	Day 8	Days 2 to 10
Week 2	Day 15	Days 11 to 25
Phone 4	Day 29	Days 26 to 32
Week 6	Day 43	Days 33 to 53
Phone 8	Day 57	Days 54 to 64
Phone 10	Day 71	Days 65 to 74
Week 12 (EOT)	Day 85	Days 75 to 93
Week 16 (EOS)	Day 113	Days 94 to 127

4.2. Sample Size

A sufficient number of subjects (36 estimated) will be enrolled to ensure that 30 evaluable subjects complete 12 weeks of treatment. Subjects will be randomly assigned to the 3 OV101 total daily dosing regimens in a 1:1:1 ratio, stratified by age group (adolescents/adults), with a target ratio of 1:1 between adolescents and adults.

The main emphasis in reporting the data will be on descriptive statistics and estimation with 95% confidence intervals (CIs).

For the objective of estimating the overall subject incidence of TEAEs (for the 3 OV101 total daily dose arms combined), the precision varies with true incidence, e.g., a true incidence of 20% would be estimated with a 13.9% precision (half-width of the 2-sided 95% CI, assuming N = 30). Table 2 contains the estimated precision for different incidence rates, using the Wilson method of calculating a confidence interval for a binomial proportion.

Table 2 Incidence Rates and Corresponding Precision

True TEAE Incidence Rate	10%	13.3%	20%	23%	30%
Expected n with a TEAE (N=30)	3	4	6	7	9
Expected proportion, p	0.100	0.133	0.200	0.233	0.300
95% CI	(3.5%,25.6%)	(5.3%,29.7%)	(9.5%,37.3%)	(11.8%,40.9%)	(16.7%,47.9%)
Precision (half-width of CI)	11.1%	12.2%	13.9%	14.6%	15.6%

Note: All CIs are 2-sided and 95% CIs are calculated using Wilson's method.
Abbreviations: n, number of subjects observed; N, total number of evaluable subjects in the study (all treatment groups); TEAE, treatment-emergent adverse event.

4.3. Stratification, Randomization, and Blinding

An interactive web response system (IWRS) will be used to administer the randomization schedule. The study will be performed in a double-blind manner, with the subjects, LAR/caregivers, investigators, study personnel, monitors, and study sponsor blinded to the identity of all study drug.

Subjects who meet all eligibility criteria will be randomly assigned at Baseline to the 5 mg/day, 10 mg/day, or 15 mg/day OV101 treatment group in a 1:1:1 allocation ratio. It is anticipated that approximately 36 male subjects will be enrolled (to obtain 30 evaluable subjects), while maintaining a 1:1 target ratio between adolescents (13 to17 years old, inclusive) and adults (18 to 22 years old, inclusive). The randomization schedule will be stratified by age (13 to 17, inclusive; 18 to 22, inclusive). It will also use an appropriate block size, which will not be revealed.

4.4. Analysis Set

Randomized Analysis Set: The randomized analysis set will consist of all randomly assigned subjects. This will be used for summaries of subject disposition, demographics, and other baseline characteristics.

Full-Analysis Set (FAS): The FAS will consist of all randomly assigned subjects who receive at least 1 dose of study drug and have at least 1 postbaseline efficacy assessment. Following

the intent-to-treat principle, subjects will be analyzed according to the treatment they are assigned to at randomization, regardless of compliance or any deviations from the study protocol. All efficacy evaluations will be carried out on the FAS.

Safety Set: The safety set will consist of all subjects who receive at least 1 dose of study drug. All analyses using the safety set will group subjects according to treatment actually received. All safety evaluations will be carried out on the safety set.

5. SUBJECT DISPOSITIONS

5.1. Dispositions

The number and percentage of subjects included in the analysis sets (safety, randomized, and full analysis sets) will be presented overall and by treatment group. Tables will indicate number of subjects who completed the study, number and percentage of subjects who completed the treatment, and the number of subjects who discontinued study and treatment for any of the following reasons:

- Does not meet I/E criteria
- Adverse Event
- Lab Results Clinically Significant
- Symptoms not consistent with protocol
- Lost to Follow-up
- Physician Decision
- Withdrawal by Subject/Withdraw Assent
- Death
- Other

These tables will be summarized by treatment group and overall. Disposition data will be summarized and provided in subject listings. The other reasons will be specified in the subject listings.

5.2. Protocol Deviations

A deviation from the protocol is an unintended or unanticipated departure from the procedures or processes approved by the sponsor and the Institutional Review Board (IRB)/ Independent Ethics Committee (IEC) and agreed to by the investigator. A significant deviation occurs when there is non-adherence to the protocol by the subject or investigator that results in a significant, additional risk to the subject. Significant deviations can include non-adherence to inclusion or exclusion criteria, enrolment of the subject without prior sponsor approval, or non-adherence to Food and Drug Administration (FDA) regulations or International Conference on Harmonisation (ICH) Good Clinical Practice (GCP) guideline. Relevant significant protocol deviations, with impacts to key safety and efficacy endpoints, will be summarized in a table by treatment group and overall. All protocol deviations will be displayed in a listing with significant and relevant deviations indicated.

6. DEMOGRAPHIC AND BASELINE CHARACTERISTICS

6.1. Demographic

Demographic characteristics include age, race, ethnicity, and enrollment by study center. Summary statistics by treatment group and overall will include counts and percentages for discrete variables, and number of subjects, mean, SD, median, minimum and maximum for continuous variables.

The following variables will be summarized by treatment group:

1) Continuous variable: age (years)
2) Categorical/discrete baseline and demographic variables: age [adolescent (13 to 17, inclusive), adult (18 to 22, inclusive)], ethnicity (Hispanic or Latino, Not Hispanic or Latino), race (American Indian or Alaska Native, Asian, Black or African-American, Native Hawaiian or Other Pacific Islander, White or Caucasian, or Other), enrollment by study center

6.2. Baseline Characteristics

Baseline characteristics include baseline body weight (kg), height (cm), alcohol use, Intelligence Quotient (IQ) by Stanford-Binet-5, Diagnostic and Statistical Manuel of Mental Disorders (DSM-5) autism spectrum disorder criteria and Molecular FXS test.

Summary statistics by treatment group and overall will include counts and percentages for discrete variables, and mean, SD, median, minimum and maximum for continuous variables.

6.3. Medical History

6.3.1 General Medical History

General medical history will be coded by MedDRA Version 20.1 or higher and summarized by SOC and PT for each treatment group as well as overall. Subject medical history including specific details will be presented in a listing. Body system will be presented in descending order for all subjects (descending order by counts in the Total column).

6.4. Inclusion and Exclusion Criteria

All inclusion/exclusion information on screen fail subjects will be included in a by-subject listing. The listing will include whether all criteria were satisfied. For subjects who did not satisfy the criteria, the criterion number will be listed with the deviation, along with whether an exception was obtained.

7. TREATMENTS AND MEDICATIONS

7.1. Prior and Concomitant Medications and Procedures

Therapies, procedures, and medications started and stopped prior to the start of study treatment are referred to as prior therapies and prior medications, respectively. Therapies, procedures, and medications that started prior to treatment and continued on treatment as well as those that started during treatment are referred to as concomitant therapies and concomitant medications. In order to define whether a medication with missing start or stop dates is a concomitant medication, refer to the following additional criteria.

- if both the start and stop dates of a particular medication are missing, that medication will be considered concomitant;
- if the start date of a medication is missing and the stop date of that medication falls on or after the dose date, that medication will be considered concomitant;
- if the start date of a medication is missing and the stop date of the medication is prior to the dose date, that medication is considered not concomitant;
- if the start date of a medication is prior to the dose date and the stop date of the medication is missing, that medication is considered concomitant.

Prior and concomitant medications will be coded using the World Health Organization Drug Dictionary (WHO-DD) version December 2017 or later and will be classified by Anatomical Therapeutic Class (ATC) level 4 and preferred term (PT). For the presentation of concomitant medications, the ATC level 4 terms will be sorted alphabetically, and within ATC level 4 term, the PT will be used and presented by decreasing total frequency overall.

The number and percentage of subjects who took at least 1 medication during the double-blind period as well as the number and percentage of subjects who took each type of medication will be presented for each treatment group and overall.

Prior and concomitant medications and procedures will be presented in separate listings.

7.2. Study Treatments

7.2.1 Extent of Exposure

Total capsules taken will be derived as the total number of capsules dispensed minus the total number of capsules returned and the **duration of exposure** will be computed as the last dose date – first dose date +1 and will be presented in categories by age group. The duration of exposure will be categorized as follows: fewer than 1 day, 1 to < 8 days, 8 to < 15 days, 15 to < 43 days, 43 to < 85 days, and 85 days or more.

Subject listing of study treatment exposure will be provided.

7.2.2 Treatment Compliance

Treatment compliance of OV101 capsules will be summarized by treatment group over the entire treatment period. Compliance of placebo capsules will not be determined. Treatment compliance will be summarized by treatment group and overall.

Compliance (%) = (Actual number of OV101 capsules/ Expected number of OV101 capsules) *100%.

The actual number of active OV101 capsules taken will be derived as the sum of: 1) number of capsules dispensed at Baseline minus number of capsules returned at Week 6, and 2) number of capsules dispensed at Week 6 minus number of capsules returned at EOT.

The expected number of OV101 capsules will be derived by multiplying the daily dose (1, 2, or 3 active OV101 capsules for the 5, 10, or 15 mg groups, respectively) by duration of exposure (see Section 7.2.1).

The number of capsules dispensed and the number remaining will be tracked and recorded by individual bottle (labeled “Morning,” “Afternoon,” or “Evening”).

A subject is compliant if overall compliance is >=80%.

Dosing error specification (missed dose, capsule not taken in usual manner, extra capsules taken, wrong dose taken and others), time of dosing error (morning, afternoon, evening), and reasons for missed dose (e.g., forgot, subject refused, subject had side effects so parent decided to give a drug holiday and other reasons) for all dispensed capsules, both active or placebo, will be listed.

8. EFFICACY ANALYSES

The primary endpoint of the study is a safety endpoint and not an efficacy endpoint.

8.1. Secondary Efficacy Analysis

Descriptive statistics will be presented for continuous variables and their changes from baseline.

Exploratory modelling will be conducted on the efficacy variables using an MMRM to further investigate any trends evident in the descriptive statistics and to inform a future study. In case of exploratory modelling, the MMRM analysis will be performed on the change from baseline or the post-baseline result (depending on the endpoint) with age and baseline (if appropriate) as covariates, daily dosing regimen, visit, and age group as fixed effects, and visit by daily dosing regimen interaction. An unstructured covariance structure will be assumed. From this model, the least squares mean change from baseline and the corresponding 95% CI will be presented for each daily dosing regimen and the total (all regimens) at each post-baseline visit. The least squares mean difference and the corresponding 95% CI will be presented for each pairwise comparison between the 3 daily dosing regimens at each week. The least squares mean changes from baseline and corresponding 95% CIs will be plotted at each time point by daily dosing regimens.

In the case of a statistically significant age group main effect ($P<0.1$) in the MMRM or age effects deemed evident in descriptive statistics, the mean changes with 95% CI described above will be estimated for each age group separately as well as overall; this will be achieved by inclusion of an additional term (age by week by daily dosing regimen interaction) in the MMRM. In this case, these additional estimates and 95% CI will be tabulated and plotted.

8.1.1 Aberrant Behavior Checklist-Community (ABC-C)

The Aberrant Behavior Checklist-Community (ABC-C) is a 58-item questionnaire completed by the LAR/caregiver that assesses a wide range of behaviors, including irritability, lethargy/social withdrawal, inappropriate speech, hyperactivity, stereotypic behavior, and social avoidance. Each item is rated on a scale of 0 to 3 (“not at all a problem” to “the problem is severe in degree”). The ABC-C will be scored 2 ways: according to the original method used to assess a population with a variety of disabilities (Aman et al 1985) and according to a method adapted for FXS that has a 6-factor structure (Sansone et al 2012).

The original method includes 5 subscales with the sub scores derivation:

SUBSCALES	**Sum of Questions**
Irritability	2, 4, 8, 10, 14, 19, 25, 29, 34, 36, 41, 47, 50, 52, 57
Lethargy/Social Withdrawal	3, 5, 12, 16, 20, 23, 26, 30, 32, 37, 40, 42, 43, 53, 55, 58
Stereotypic Behavior	6, 11, 17, 27, 35, 45, 49
Hyperactivity	1, 7, 13, 15, 18, 21, 24, 28, 31, 38, 39, 44, 48, 51, 54, 56
Inappropriate speech	9, 22, 33, 46

The method adapted for FXS subscales will be derived as follows:

SUBSCALES	**Sum of Questions**
Irritability	2, 4, 7, 8, 10, 14, 18, 19, 21, 24, 29, 34, 36, 41, 47, 50, 52, 57
Hyperactivity	1, 13, 15, 31, 38, 39, 44, 48, 54
Lethargy/Social Withdrawal	12, 20, 23, 25, 28, 32, 37, 40, 43, 51, 53, 55, 56, 58
Social Avoidance	5, 16, 30, 42
Stereotypic Behavior	6, 11, 17, 35, 45, 49
Inappropriate speech	9, 22, 33, 46

The mean change of ABC-C total and subscale scores according to the original method and according to the method adapted for FXS, at Week 2, Week 6, and Week 12 from baseline will be analyzed as described in Section 8.1.

For ABC-C, the maximum amount of missing data that can be tolerated varies by subscale in original method. The following upper limits of missing items tolerated for each subscale are: (I) Irritability (15-item scale): 3 items; (II) Lethargy/Social Withdrawal (16-item scale): 3 items; (III) Stereotypic Behavior (7-item scale): 2 items; (IV) Hyperactivity/Noncompliance (16-item scale): 3 items; (V) Inappropriate Speech (4-item scale): 1 item. If more items than the stated upper limit have been left blank (unscored) by the rater, then the associated subscale score is not computed at all; instead, list the data as missing for those subscales.

When the amount of missing data is tolerated, subscale scores will be prorated as follows: (a) Take the total number of items on the subscale and divide this by the number of completed items. This will result in a number larger than 1.00. (b) Multiply that number by the total score for that subscale. (c) This becomes the new total score for this subscale for the given subject. For example, values of 13 items out of 15 Irritability questions are applicable with a sum of 23, then the prorated score of Irritability is (15/13)*23=26.54.

For subscales in FXS method, if the traditional subscales can be derived then the corresponding FXS subscale scores will be calculated. For FXS Social Avoidance subscale (4-item scale), if more than 1 item is missing then the subscale will be missing. Assuming values of 14 items out of 18 FXS-Irritability questions are applicable with a sum of 25, as the original Irritability is calculated, the FXS-Irritability score is prorated as (18/14)*25=32.14.

Subjects with total score of 16 or lower at the Baseline visit will be excluded from analysis of ABC-C, while sensitivity analysis with all subjects involved will be conducted.

The ABC-C total score will be calculated only when all associated subscales scores are derived and not missing.

8.1.2 Anxiety, Depression, and Mood Scales (ADAMS)

The Anxiety, Depression, and Mood Scales (ADAMS) is a caregiver-completed report that screens comprehensively for anxiety and depression in persons with intellectual disability and consists of 28 items grouped into 5 subscales that assess the frequency and severity of manic/hyperactive behavior, depressed mood, social avoidance, general anxiety, and obsessive/compulsive behavior (Esbensen et al 2003).

Sub scores will be obtained as follows:

SUBSCALES	**Sum of Questions**
Manic/Hyperactive behavior	3, 4, 12, 17, 22
Depressed mood	5, 9, 10, 14, 18, 23, 28
Social avoidance	2, 6, 13, 19, 21, 25, 27
General anxiety	1, 3, 7, 11, 15, 24, 26
Obsessive/compulsive behavior	8, 16, 20

Mean change from baseline to Week 2, Week 6, and Week 12 on the ADAMS score and subscale scores will be analyzed as described in Section 8.1. Missing data convention from the provider will be used for missing assessments and scores. Otherwise, missing data from the assessment will be kept as missing.

8.1.3 Repetitive Behavior Scale-Revised (RBS-R)

The Repetitive Behavior Scale-Revised (RBS-R) is a caregiver-completed, 43-item questionnaire assessing a variety of repetitive behaviors with 6 behavior subscales: stereotyped, ritualistic, self-injurious, compulsive, restricted, and sameness (Bodfish et al 2000).

Overall total RBS-R raw score will be obtained by getting the sum of items 1-43. The number of items endorsed (defined as the number of items rated 1, 2, or 3) and overall will be derived and summarized. Missing data convention from the provider will be used for missing assessments and scores. Otherwise, missing data from the assessment will be kept as missing.

Subscale scores will be computed as follows:

SUBSCALES	Sum of Questions
Stereotyped Behavior	1-6
Self-Injurious Behavior	7-14
Compulsive Behavior	15-22
Ritualistic Behavior	23-28
Sameness Behavior	29-39
Restricted Behavior	40-43

Change from baseline to Week 2, Week 6, and Week 12 on the RBS-R total and subscale raw scores and number of items endorsed will be analyzed as described in Section 8.1.

8.1.4 Conner's 3rd Edition

The Conner's 3rd Edition (Conner's 3; North Tonawanda Multi-Health System, North Tonawanda, NY) is designed to assess attention deficit hyperactivity disorder, and it will be used to collect LAR/caregiver observations of subject behavior. The assessment of behavior in neurodevelopmental disorders, such as FXS, frequently relies upon proxy report. The long form of Conner's 3 is needed to assess inconsistency (reporter reliability) and will be administered at Screening only. The short form of Conner's 3 will be administered as an outcome assessment at the baseline, Week 2, Week 6, Week 12, and EOS visits.

The Inconsistency Index/Response Style Analysis is a categorical variable obtained using two variables obtained at the Screening visit using the long form of Conner's 3: Inconsistency Index Raw Score and Inconsistency Index Number of Differentials ≥ 2. If the Inconsistency Index Raw Score ≥ 7 and Inconsistency Index Number of Differentials ≥ 2, then there is a possible inconsistent response style. These data will be summarized and listed.

The Raw sub scores for Content scales will be provided in the eCRF. Content scales include inattention, hyperactivity/impulsivity, learning problems, executive functioning, defiance/aggression, peer relations. The mean change of content subscale scores at Week 2,

Week 6, and Week 12 from baseline will be analyzed as described in Section 8.1. Subjects with inconsistent response style at Screening visit were excluded. Missing data convention from the provider will be used for missing assessments and scores. Otherwise, missing data from the assessment will be kept as missing.

8.1.5 Short Sensory Profile-2

The Short Sensory Profile-2 is a caregiver questionnaire to evaluate a subject's sensory processing patterns in the context of home, school, and community-based activities.

Sensory Processing Raw Total score is the sum of items 1-14 and Behavioral Responses Raw Total score is the sum of items 15-34. Raw scores of the subscales will be derived as follows and will be analyzed as described in Section 8.1:

SUBSCALES	Sum of Questions
Seeking	6-8, 11, 14, 31, 32
Avoiding	16-20, 22-24, 26
Sensitivity	1-5, 21, 25, 28, 29, 33
Registration	9, 10, 12, 13, 15, 27, 30, 34

Change from baseline in Sensory Processing Raw Total score, Behavioral Responses Raw Total score and subscale scores at Week 2, Week 6, and Week 12 will be analyzed as described in Section 8.1. Missing data convention from the provider will be used for missing assessments and scores. Otherwise, missing data from the assessment will be kept as missing.

8.1.6 Clinical Global Impressions Scales

The CGI-S scale will be used by the investigator to assess the severity of symptoms, and the CGI-I scale will be used by investigators to assess improvement from baseline. The CGI-S has been adapted to capture specific characteristics commonly present in the FXS population. Both CGI-S and CGI-I will be conducted by a clinician with experience in FXS.

A dichotomous responder variable will be defined whether or not there is improvement in any CGI-I. Response is defined as improvement in any of the CGI-I survey questions, with improvement indicated by a response of "Very much improved", "Much improved", and "Minimally improved" compared to time of enrollment. The CGI-I responder endpoint will be compared between treatments (OV101 5mg vs. OV101 10mg; OV101 5mg vs. OV101 15mg; OV101 10mg vs. OV101 15mg) at week 12 and exact test will be used to test for the difference of the treatments.

CGI severity (CGI-S) and CGI improvement (CGI-I) scores as well as by-domain CGI-S sub scores will be provided in the eCRF Domains includes: Anxiety, ADHD, Communication/Connectedness, Repetitive and Restrictive Behavior, Disruptive Behavior, and Activities of Daily Living.

The mean change of CGI-S score and by-domain CGI-S sub scores at Week 2, Week 6, and Week 12 from baseline will be analyzed as described in Section 8.1.

The frequencies of CGI-I categories at Week 2, Week 6, and Week 12 will be summarized.

Responder analysis will be performed to test difference in CGI-I between treatments with exact test at Week 12 using observed data. Additionally, a sensitivity analysis will be conducted at Week 12 using last observation carried forward (LOCF) method for missing data.

8.2. Exploratory Efficacy Analysis

8.2.1 Caregiver-Completed Endpoints

8.2.1.1 Behavior Assessment System for Children, 3rd Edition (BASC-3 PRQ)

The Behavior Assessment System for Children, 3rd Edition, Parenting Relationship Questionnaire (BASC-3 PRQ) provides information on the relationship between a caregiver and a child across multiple dimensions (Pearson, New York, NY). It yields the following subscale scores: attachment, communication, discipline practices, involvement, parent confidence, satisfaction with school, and relation frustration. The questionnaire is written at a third grade level and also calculates a number of response validity scores, such as consistency and response pattern, to detect careless or exaggerated responding. It includes normative samples for both female and male raters that are closely matched to the US Census population estimates.

Content Domains raw scores (Attachment, Communication, Discipline, Involvement, Parenting Confidence, Satisfaction with School, and Relationship Frustration) at Week 6 and Week 12 will be summarized. Missing data convention from the provider will be used for missing assessments and scores. Otherwise, missing data from the assessment will be kept as missing.

Validity Index Summary Raw Scores and Category Assignment (F Index, D Index, Response Pattern Index, and Consistency Index) will be summarized.

8.2.1.2 Parent Global Impressions Scales

The Parent Global Impressions-Severity (PGI-S) scale will be completed by the LAR/caregiver to assess the severity of symptoms, and the Parent Global Impressions-Improvement (PGI-I) scale will be used by the LAR/caregiver to assess improvement from baseline. The PGI-S has been adapted to capture specific characteristics commonly present in the FXS population.

Severity of Participant's condition will be collected in the eCRF and the change from baseline at Week 2, Week 6, and Week 12 will be analyzed as described in Section 8.1.

Improvement of Participant's condition will be collected in the eCRF and the frequencies at Week 2, Week 6, and Week 12 will be summarized. Missing data convention from the provider will be used for missing assessments and scores. Otherwise, missing data from the assessment will be kept as missing.

8.2.1.3 Pediatric Quality of Life Inventory

The Pediatric Quality of Life Inventory (PedsQL) generic core scales and Multidimensional Fatigue Scale will be administered in this study (Varni et al 2004). The generic core scales include 23 items covering the domains of physical, emotional, social, and school functioning. The Multidimensional Fatigue Scale contains 18 items measuring general fatigue, sleep/rest fatigue, and cognitive fatigue.

PedsQL Generic Core and Multidimensional Fatigue transformed scores, mean scores for the domains, and total scores will be derived. Each item will be reverse scored and linearly transformed to a 0-100 scale, so that higher scores indicate better Health-Related Quality of Life. Items will be transformed as follows: 0=100, 1=75, 2=50, 3=25, and 4=0.

Transformed scores will be used to calculate the 3 mean domain scores, and total score on the multidimensional fatigue scale as follows:

DOMAIN	**Mean Score***
General Fatigue	Mean of 6 transformed items under 'General Fatigue'
Sleep/Rest Fatigue	Mean of 6 transformed items under 'Sleep/rest Fatigue'
Cognitive Fatigue	Mean of 6 transformed items under 'Cognitive Fatigue'
Total	Mean of all transformed items

* If more than 50% of the items in the scale are missing, the mean domain score should not be computed. If more than 50% of the items are completed, mean of the completed items in a domain scale will be imputed.

Transformed scores will be used to calculate the 4 mean domain scores, total score, and Psychosocial Health Summary Score on the generic core scale as follows:

DOMAIN	**Mean Score***
Physical Functioning	Mean of 8 transformed items under 'Physical Functioning'
Emotional Functioning	Mean of 5 transformed items under 'Emotional Functioning'
Social Functioning	Mean of 5 transformed items under 'Social Functioning'
School Functioning (for Teens ages, 13-18) or Work/Studies Functioning (for Young Adults, ages 18-22)	Mean of 5 transformed items under 'School Functioning" or 'Work/Studies Functioning'

Psychosocial Health Summary Score	Mean of transformed items under 'Emotional Functioning', 'Social Functioning', and "School Functioning" or 'Work/Studies Functioning' mean scores
Total	Mean of all transformed items

* If more than 50% of the items in the scale are missing, the mean domain score should not be computed. If more than 50% of the items are completed, mean of the completed items in a domain scale will be imputed.

Change from baseline in the mean total score and mean scores of the domains for both PedsQL generic core and multidimensional will be analyzed at week 6 and EOT.

8.2.1.4 Children's Sleep Habits Questionnaire

The Children's Sleep Habits Questionnaire (CSHQ) is a caregiver-completed assessment of sleep behaviors in children and adolescents (Owens et al 2000). The abbreviated version of this questionnaire includes 33 items that assess sleep habits and possible difficulties with sleep. The LAR/caregiver will be asked to recall the subject's sleep behaviors over the preceding week.

The CHSQ consists of 8 domains with their sub scores and will be derived as follows:

Domains	**Derivation**
Bedtime Resistance	sum of items 1R, 3R, 4, 5, 6, and 8
Sleep Onset Delay	item 2R
Sleep Duration	sum of items 9, 10R, and 11R
Sleep Anxiety	sum of items 5, 7, 8, and 21
Night Wakings	sum of items 16, 24, and 25
Parasomnias	sum of items 12, 13, 14, 15, 17, 22, and 23
Sleep Disordered Breathing	sum of item 18, 19, and 20
Daytime Sleepiness	sum of items 26R, 27, 28, 29, 30, 31, 32, and 33

R- reversed scored.

Analysis as described in Section 8.1 will be used to analyze the change from baseline in CHSQ domains at Week 2, Week 6, and Week 12.

Subscale scores of CSHQ are not calculated if >= 50% of the subscale answers are missing. When < 50% of subscale answers are missing, the subscale score will be prorated following the steps described in Section 8.1.1. For example, 4 out of 6 Bedtime Resistance questions are answered with a sum of 9, then the prorated score of Bedtime Resistance is (6/4)*9=13.5.

CSHQ total will be calculated if all subscales are not missing.

8.2.1.5 Sleep Diary

A paper sleep diary will be completed by each subject's LAR/caregiver, on behalf of the subject. Endpoints assessed by the sleep include how long it took the subject to fall asleep

(sleep onset latency [SOL]), the amount of time the subject was awake during the night (WASO), the total amount of sleep obtained (total sleep time [TST]), daytime sleepiness, defined as duration of napping in the daytime, and number of times subject woke up at night. The sleep diary will be completed each day during the 7 days immediately preceding each clinic visit after screening.

Sleep Parameters	**Derivation**
Sleep Onset Latency (SOL) in minutes	[(First night: time to sleep – time subject went to bed) + (Second night: time to sleep – time subject went to bed) + (Third night: time to sleep – time subject went to bed) + (Fourth night: time to sleep – time subject went to bed) + (Fifth night: time to sleep – time subject went to bed) + (Sixth night: time to sleep – time subject went to bed) + (Seventh night: time to sleep – time subject went to bed)]/7 (or number of days with non-missing values) Note: For sleep time, consider the time closest on or after the time the subject went to bed on the same day
WASO (minutes)	Sum (time to sleep – time to wake from the previous log), considering all records between the time subject went to bed until subject woke up, then take the weekly average using the number of days with non-missing values.
TST (hours)	The weekly average, using the number of days with non-missing values, of total sleep time, converting the unit into hours.
Daytime Sleepiness (minutes)	The weekly average, using the number of days with non-missing values, of (time to sleep – time to wake), considering only records on the same day before subject went to bed
Number of times subject woke in the nights (counts)	The weekly average, using the number of days with non-missing values, of number of records from WASO considering all records between the time subject went to bed until subject woke up.

Weekly average of sleep parameters will be calculated with missing data excluded.

The baseline values of actigraphy parameters will be obtained from data collected within 7 days immediately preceding Day 1.

Change from baseline on the sleep parameters at Week 2, Week 6, and Week 12 will be analyzed as described in Section 8.1.

Missing data convention from the provider will be used for missing assessments and scores. Otherwise, missing data from the assessment will be kept as missing.

8.2.2 Clinician-Completed Endpoints

8.2.2.1 Caregiver Top 3 Concerns Visual Analog Scale

In Part 1 of the visual analog scale (VAS), LARs/caregivers will generate their top 3 FXS-associated concerns for the subject and rate these 3 concerns using a 10-cm visual analog scale, with troublesome behaviors anchored at one end with the descriptor "worst ever" and at the other end with the descriptor "no problem at all". The investigator will then map each LAR/caregiver-generated concern onto one of the predefined FXS domains (i.e. Anxiety, ADHD, Repetitive/Stereotypic Behavior, Disruptive Behavior, Communication, Activities of Daily Living, or Socialization and Other) to which each behavior or symptom applies. The top 3 FXS-associated concerns selected at Baseline will be fixed and assessed at Week 2, Week 6 and Week 12.
Part 2 of VAS will be used to rate three FXS-specific domains: Speech and Communication, Socialization, and Cognition.

At a single visit, it is possible that one subject could have more than one concerns within one domain as these are based on parent report. In this case, the averaged value within one FXS-specific domain per subject will be applied in analysis.

Change from baseline in 1) value of the top 3 FXS-associated concerns, in 2) value of FXS domains after mapping of concerns, and in 3) FXS-specific domains (Speech and Communication, Socialization, Cognition) for the subject will be summarized by Week 2, Week 6, Week 12, as described in Section 8.1.

Missing data convention from the provider will be used for missing assessments and scores. Otherwise, missing data from the assessment will be kept as missing.

8.2.2.2 Vineland Adaptive Behavior Scale

The Vineland Adaptive Behavior Scale, 3rd Edition (VABS-3) caregiver interview form will be used to evaluate communication, socialization, and daily living skills of subjects to assess their overall adaptive functioning. This assessment is an interview of the caregiver by a trained qualified rater.

Change from baseline in standard VABS domain scores (consisting of adaptive behavior composite, communication domain, daily living skills domain and socialization domain) and raw scores and v-scores for subdomains (Receptive, Expressive, Written, Personal Care, Domestic, Community, Interpersonal Relationships, Play and Leisure Skills, and Coping Skills) collected in the eCRF will be summarized and analyzed for Week 6 and Week 12, as described in Section 8.1.

Missing data convention from the provider will be used for missing assessments and scores. Otherwise, missing data from the assessment will be kept as missing.

Missing data convention from the provider will be used for missing assessments and scores. Otherwise, missing data from the assessment will be kept as missing.

8.2.2.3 Pediatric Sleep Clinical Global Impressions Scales

The Pediatric Sleep CGI scales were recently developed to measure pediatric insomnia in autism spectrum disorder and have been validated against the CSHQ and actigraphy (Malow et al 2016). Questions include the subject's ability to fall asleep and remain sleeping independently (i.e., apart from LAR/caregivers); bedtime resistance; sleep onset delay; night awakening; LAR/caregiver satisfaction with their child's current sleep patterns; family functioning, as affected by their child's current sleep patterns; and clinician's overall concern with the child's sleep.

Pediatric Sleep CGI severity questions will be analyzed for Week 2, Week 6, and Week 12, as described in Section 8.1.

Missing data convention from the provider will be used for missing assessments and scores. Otherwise, missing data from the assessment will be kept as missing.

8.2.3 Direct Subject Assessment Endpoints

8.2.3.1 Actigraphy

Actigraphy is a noninvasive method of monitoring human rest/activity cycles. A small actigraph unit is worn to measure daytime and sleep activity. The actigraph unit stores motion data by clock time, and the data must be uploaded to an actigraphy database for analysis. The actigraphy data (including parameters such as: SOL, WASO, TST, NA, sleep efficiency, daily activity level, and sleep fragmentation) will be compared to the sleep diary data. The derivation of these parameters will be as follows:

Sleep Parameters	Derivation
Sleep Onset Latency (SOL)	Mean of Onset Latency (minutes), duration from the beginning of the rest period to the start of sleep onset, reported when interval type = sleep
Total Sleep Time (TST) during night	Take Sum of sleep time (minutes) where interval type = ‘sleep’ within each day then the Mean of this sum for an average score per day. Convert its unit to hours in analysis.
Total Sleep Time (TST) during day	Take Sum of sleep time (minutes) where interval type = ‘Active’ within each day then the Mean of this sum for an average score per day. Convert its unit to hours in analysis.

Wake After Sleep Onset (WASO)	Mean of WASO when interval type = sleep
Nocturnal Awakenings (NA)	Mean of # Wake bouts when interval type = sleep
Sleep Efficiency	Mean of Efficiency (%) when interval type = sleep. Efficiency is defined as total sleep time as a fraction of the rest period (total time in bed).
Daily Activity Level	Take Sum of Total AC (counts) within each day then the Mean of this sum for an average score per day
Sleep Fragmentation	Sum of Fragmentation (%+%) when interval type = sleep

The baseline values of actigraphy parameters will be obtained from data collected within 7 days immediately preceding Day 1.

The change from baseline in the subject actigraphy parameters will be analyzed for Week 2, Week 6, and Week 12, as described in Section 8.1.

Missing data convention from the provider will be used for missing assessments and scores. Otherwise, missing data from the assessment will be kept as missing.

8.2.3.2 Test of Attentional Performance for Children

The KiTAP is an automated computer-based assessment of attentional performance developed and normed for a pediatric population. The KiTAP is suitable for assessment of function in individuals of various ages, including children, adolescents, and adults, with FXS (Knox et al 2012). The task displays an "enchanted castle" animation and measures 8 aspects of attention: alertness, distractibility, flexibility, Go/NoGo, virtual scanning, vigilance, sustained attention, and divided attention. Only 4 of the 8 subtests will be administered in this study: alertness, distractibility, Go/NoGo, and flexibility.

The change from baseline in the KiTAP (for the 4 sub-tests presented in the CRF: alertness, distractibility, flexibility, Go/NoGo) and their corresponding parameters (mean, stddev, correct, omissions, anticipations for alertness; mean, stddev, correct, errors for flexibility, and mean, stddev, correct, errors, omissions for Go/NoGo parameters; distractor mean, distractor correct, distractor errors, distractor omissions, no distractor mean, no distractor correct, no distractor errors, no distractor omissions, total mean, total correct, total errors, total omissions for distractibility parameter) will be analyzed for Week 6 and Week 12, as described in Section 8. 1.

If value of some of KiTAP parameter is missing from the Baseline visit, the value from the Screening visit can be used for analysis. Missing data convention from the provider will be

used for missing assessments and scores. Otherwise, missing data from the assessment will be kept as missing.

8.2.3.3 Repeatable Battery for the Assessment of Neuropsychological Status

The RBANS provides a brief individually administered battery to assess cognitive function in individuals 12 years of age and older (Randolph et al 1998). The full battery consists of 12 subtests to assess immediate memory, visuospatial abilities, language, attention, and delayed memory. In this study, 6 subtests of the RBANS will be used: list learning (immediate memory), list recall (delayed memory), list recognition (delayed memory index), story memory (immediate memory index), story recall (delayed memory index), and picture naming (language index). Each subtest can be administered in 5 minutes or less.

Subtests total scores will be derived in the eCRF and analysis as described in Section 8.1 will be used to analyze the change from baseline in the RBANS subset scores for Week 6 and Week 12.

If some subtest score of RBANS is missing from the Baseline visit, the subtest score from the Screening visit can be used for analysis. Missing data convention from the provider will be used for missing assessments and scores. Otherwise, missing data from the assessment will be kept as missing.

8.2.3.4 [redacted]

[redacted]

8.2.4 [redacted]

[redacted]

8.2.5 [redacted]

[redacted]

8.3. [redacted]

[redacted]

9. SAFETY ANALYSES

All safety data analyses will be conducted by using the Safety Set.

9.1. Adverse Events

9.1.1 Incidence of Adverse Events

An AE is defined as any untoward medical occurrence in a subject enrolled into this study regardless of its causal relationship to study drug. Subjects will be instructed to contact the investigator at any time after randomization if any symptoms develop.

A Treatment-Emergent Adverse Event (TEAE) is defined as any event AE not present before exposure to study drug or any event already present that worsens in either intensity or frequency after exposure to study drug and occurred within 30 days of last dose. TEAEs will be reported in tables regardless of seriousness.

For listings, any events after pre-specified range above will be flagged in the outputs.

The subject-incidence and event-incidence of TEAEs, TEAEs leading to study discontinuation, treatment-related TEAEs, serious TEAEs, severity of TEAEs, TEAEs with outcome of death, and seizure TEAEs will be summarized by treatment and overall. For summaries of severe or treatment-related AEs, for a given subject, the highest severity or relationship for a specific PT will be considered. Adverse events with missing start dates will be considered treatment emergent. All AEs will be coded using the most current Medical Dictionary for Drug Regulatory Affairs (MedDRA, version 20.1 or higher) and will be summarized by System Organ Class (SOC) and Preferred Term (PT). For each SOC and PT, a subject will be counted only once for subject-incidence tabulations. In addition, TEAE will be summarized by System Organ Class (SOC), High Level Group Term (HLGT), High Level Term (HLT), and Preferred Term (PT). For each SOC, HLGT, HLT, and PT, a subject will be counted only once for subject-incidence tabulations. The Wilson 95% CIs will be displayed for the proportion of subjects experiencing TEAEs as a total across the 3 dosing regimens. Additional overall AEs (including non-TEAEs) by PTs will also be provided. Detailed listings of AEs, SAEs, treatment-related AEs, and TEAEs leading to study discontinuation will be provided.

9.1.2 Relationship of Adverse Events to Study Drug

A TEAE is considered treatment-related if there is any causal relationship between the study medication. If the relationship of a TEAE is missing or the TEAE is possibly or probably related, the TEAE is reported as related. The number and percentage of subjects experiencing treatment-related TEAEs and number of events will be summarized.

9.1.3 Severity of TEAEs

A summary of TEAEs by severity will be presented in a table. The severity that will be presented represents the most extreme severity captured on the eCRF page. The possible severities are 'Mild,' 'Moderate,' and 'Severe'. In the TEAE severity table, if a subject reported multiple occurrences of the same TEAE, only the most severe TEAE is presented. TEAEs that are missing severity will be presented in tables as 'Severe' but will be presented in the data listing with a missing severity.

9.1.4 Serious TEAEs

A serious treatment-emergent adverse event is any untoward medical occurrence that, at any dose, results in death, life-threatening, requires hospitalization or prolongation of existing hospitalization, hospitalization for elective treatment of a pre-existing condition that did not worsen from baseline is not considered an AE, results in disability/incapacity, and congenital anomaly/birth defect.

Serious TEAEs will be categorized and presented by SOC and PT in the same manner to that described in Section 9.1.1.

9.1.5 TEAEs Leading to Study Discontinuation

The number and percentage of subjects with TEAEs leading to study discontinuation and number of events will be categorized and presented by SOC and PT in the same manner to that described in Section 9.1.1.

9.1.6 TEAEs with Outcome of Death

The number and percentage of TEAEs with outcome of death and number of events will be tabulated by treatment and overall. Adverse events leading to death will be categorized and presented by SOC and PT in the same manner to that described in Section 9.1.1.

9.1.7 Seizure TEAEs

The number and percentage of subjects with seizure TEAEs and number of events will be categorized and presented by SOC and PT in the same manner to that described in Section 9.1.1.

9.2. Clinical Laboratory Evaluations

All laboratory results are from PPD central laboratory and only standard results in SI unit will be summarized in the outputs.

Descriptive statistics for laboratory values at each time point and change from Baseline values to each post-baseline time points in clinical chemistry and hematology results will be summarized.

Shift tables for laboratory parameters will be presented to show the change of normality from baseline to each post-baseline time point. For each continuous laboratory parameter, results will be classified as low, normal, or high relative to the parameter's reference range.
Listings of subjects with abnormal results will be provided.

9.2.1 Hematology

The parameters to be analyzed for hematology includes hematocrit, hemoglobin, red blood cell count with indices (mean corpuscular volume, mean corpuscular hemoglobin, and mean corpuscular hemoglobin concentration), reticulocytes, white blood cell count and differential (neutrophils, lymphocytes, monocytes, eosinophils, and basophils) reported as percentages and absolute values, and platelets (platelet count, prothrombin time and partial thromboplastin time, international normalization ratio).

9.2.2 Clinical Chemistry

The parameters to be analyzed for clinical chemistry are albumin, alkaline phosphatase, blood urea nitrogen, gamma-glutamyl transferase, calcium, creatinine, glucose, cholesterol (high-density lipoprotein and low-density lipoprotein [calculated] and homogenous low-density lipoprotein), triglycerides, phosphate, potassium, alanine aminotransferase, aspartate aminotransferase, lactate dehydrogenase, sodium, chloride, bilirubin (total, direct), total protein, uric acid, and creatine phosphokinase.

9.2.3 Urinalysis

The parameters to be analyzed for urinalysis include macroscopic (pH, specific gravity, glucose, protein, ketones, nitrates, blood) and microscopic, reported only if present (red blood cells/high-power field, white blood cells/high-power field, casts, epithelial cells, crystals, granulation). These parameters will be summarized by treatment in shift tables comparing the results at each scheduled post-baseline visit with those at the baseline visit.

9.3. Vital Signs

Vital signs (temperature in °C, pulse rate [beats per minute], respiratory rate [breaths per minute], and systolic and diastolic blood pressure [mmHg]), weight (kg), and height (cm) will be summarized descriptively at Baseline and all post-baseline study time points by treatment group and overall. Change from baseline and post-baseline study time points will be summarized descriptively by treatment group and overall.

Data listings will be provided.

9.4. Physical Examination

Physical examinations (general appearance; skin; head, ear, eye, nose, and throat; neck; lymph node; chest; heart; abdominal cavity; limb; central nervous system; and musculoskeletal) at baseline and all post-baseline study time points will be summarized by treatment group and overall.

Shift table and data listings will be provided.

9.5. Clinical Assessment of Suicidality

The number and percentage of subjects on each of the following questions will be summarized in a table by treatment and overall.

- Suicidal ideation: Has the subject wished he were dead or wished he could go to sleep and not wake up?
- Suicidal behavior: Has the subject made a suicide attempt? Has the subject done anything to harm himself?
- Self-injurious behavior: Has the subject engaged in non-suicidal self-injurious behavior?

10. INTERIM ANALYSIS

No interim statistical analysis will be performed. Periodic interim blinded review of safety data will be performed as part of routine pharmacovigilance activities.

11. OTHER STATISTICAL ISSUES

11.1. Significance Levels

The primary objective of this study is safety and no power calculations for the efficacy analyses have been planned in the sample size justification.

No adjustment will be made for multiple testing, all p-values will be treated as exploratory and should be interpreted with caution.

All testing will be two-sided and conducted at 0.05 level of significance except for interactions, which will be tested at the 0.100 significance level.

11.2. Handling of Missing Data

If a baseline assessment is missing, then the last screening assessment close to the baseline will be used as the baseline assessment.

The maximum amount of missing data that can be tolerated and derivation of prorating subscale scores have been specified in Section 8.1.1.

Last observation carried forward (LOCF) method will be applied to impute responder (binary) endpoints, including CGI-I responder, in associated sensitivity analyses (see Section 8.1.6).

For CSHQ, rule of handling missing data and the derivation of prorating domain sub scores have been specified in Section 8.2.1.4.

In general, missing data convention from the provider will be used for missing assessments and scores. Otherwise, missing data from the assessment will be kept as missing (if it is not specified).

12. CHANGES COMPARED TO THE PROTOCOL

In Section 7.8.5 of the Protocol, p-value of 0.05 was applied as the level of significance in testing the effect of age group in MMRM. However, a p-value of 0.10 for age related interaction term was used in the model selection in the SAP.

13. REFERENCES

Aman MG, Singh NN, Stewart AW, et al. Psychometric characteristics of the Aberrant Behavior Checklist. Am J Ment Defic. 1985;89(5):492-502.

Bodfish JW, Symons FJ, Parker DE, et al. Varieties of repetitive behavior in autism: comparisons to mental retardation. J Autism Dev Disord. 2000;30(3):237-43.

Cogram P, Deacon RJ, During MJ, et al. Gaboxadol normalizes behavioral abnormalities in a mouse model of fragile X syndrome. Poster presented at: 18th International Fragile X and Related Neurodevelopmental Disorders Workshop; 2017 Oct 12-16; Saint-Alexis-des-Monts, Quebec.

Crawford DC, Acuña JM, Sherman SL. *FMR1* and the fragile X syndrome: human genome epidemiology review. Genet Med. 2001;3(5):359-71.

Esbensen AJ, Rojahn J, Aman MG, et al. Reliability and validity of an assessment instrument for anxiety, depression, and mood among individuals with mental retardation. J Autism Dev Disord. 2003;33(6):617-29.

Ethridge LE, White SP, Mosconi MW, et al. Reduced habituation of auditory evoked potentials indicate cortical hyper-excitability in fragile X syndrome. Transl Psychiatry. 2016;6:e787.

Hagerman R, Hoem G, Hagerman P. Fragile X and autism: intertwined at the molecular level leading to targeted treatments. Mol Autism. 2010;1:12.

Hall SS, Lightbody AA, Reiss AL. Compulsive, self-injurious, and autistic behavior in children and adolescents with fragile X syndrome. Am J Ment Retard. 2008;113(1):44-53.

Harris SW, Hessl D, Goodlin-Jones B, et al. Autism profiles of males with fragile X syndrome. Am J Ment Retard. 2008;113(6):427-38.

Knox A, Schneider A, Abucayan F, et al. Feasibility, reliability, and clinical validity of the Test of Attentional Performance for Children (KiTAP) in fragile X syndrome (FXS). J Neurodev Disord. 2012;4:2.

Malow BA, Connolly HV, Weiss SK, et al. The Pediatric Sleep Clinical Global Impressions Scale–a new tool to measure pediatric insomnia in autism spectrum disorders. J Dev Behav Pediatr. 2016;37(5):370-6.

Owens J, Spirito A, McGuinn M. The Children's Sleep Habits Questionnaire (CSHQ): psychometric properties of a survey instrument for school-aged children. Sleep. 2000;23(8):1-9.

Randolph C, Tierney MC, Mohr E, et al. The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): preliminary clinical validity. J Clin Exp Neuropsychol. 1998;20(3):310-9.

Sansone SM, Widaman KF, Hall SS, et al. Psychometric study of the aberrant behavior checklist in fragile X syndrome and implications for targeted treatment. J Autism Dev Disord. 2012;42(7):1377-92.

Varni JW, Sherman SA, Burwinkle TM, et al. The PedsQL™ Family Impact Module: preliminary reliability and validity. Health Qual Life Outcomes. 2004;2:55.

14. APPENDICES

14.1. Imputation Algorithm for Partial and Missing Dates

Adverse Event, Concomitant Procedures and Concomitant Medications

- Partial/Missing Start Date
  - Missing day - Impute the 1st of the month unless month is same as month of first dose of study drug then impute first dose date
  - Missing day and month – impute 1st January unless year is the same as first dose date then impute first dose date
  - Completely missing – impute first dose date unless the end date suggests it could have started prior to this in which case impute the 1st January of the same year as the end date. When imputing a start date ensure that the new imputed date is sensible i.e. is prior to the end date of the AE or med.
- Partial/Missing End Date
  - Missing day - Impute the last day of the month unless month is same as month of last dose of study drug then impute last dose date
  - Missing day and month – impute 31st December unless year is the same as last dose date then impute last dose date
  - Completely Missing – need to look at whether the AE/medication is still ongoing before imputing a date and also when it started in relation to study drug. If the ongoing flag is checked then assume that AE is still present / medication is still being taken (i.e. do not impute a date). If the AE/medication has stopped and start date is prior to first dose date then impute the 1st dose date, if it started on or after first dose date then impute a date that is after the last dose date.

14.2. Schedule of Study Procedures

Visit Name	**Screening**	**Baseline**	**Phone 1**	**Week 2**	**Phone 4**	**Week 6**	**Phone 8**	**Phone 10**	**EOT**	**EOS**
Visit Window (days)	**-[a]**	**-**	**±2**	**±3**	**±2**	**±3**	**±2**	**±2**	**±3**	**+14**
Day[b]	**-21 to 1**	**1**	**8**	**15**	**29**	**43**	**57**	**71**	**85**	**113**
Week	**-**	**0**	**1**	**2**	**4**	**6**	**8**	**10**	**12**	**16[c]**
Obtained informed consent and assesnt	X									
Verify Inclusion/exclusion criteria	X	X								
Obtain medical history	X	X								
Conduct clinical assessment[d] for FXS	X									
Stanford-Binet-5	X									
Review for DSM-5 autism spectrum disorder criteria	X									
Molecular FXS Test[e]	X									
[redacted]										
[redacted]										
Clinical assessment of suicidality	X	X	X	X	X	X	X	X	X	X
Physical examination	X	X		X		X			X	X
Vital signs	X	X		X		X			X	X
Clinical laboratory tests	X	X[f]		X		X			X	
Randomization assignment		X								
Adverse event inquiry	X	X	X	X	X	X	X	X	X	X
Concomitant medication inquiry	X	X	X	X	X	X	X	X	X	X
Dispense medication diary		X				X				
Review/collect medication diary, assess compliance				X		X			X	

Dispense study drug		X				X				
Collect unused study drug						X			X	
Caregiver-Completed Measures										
Aberrant Behavior Checklist-Community		X		X		X			X	X
Anxiety, Depression, and Mood Scales		X		X		X			X	X
Repetitive Behavior Scale-Revised		X		X		X			X	X
Conner's 3rd Edition	X (LF)	X (SF)		X (SF)		X (SF)			X (SF)	X (SF)
Short Sensory Profile-2		X		X		X			X	X
Children's Sleep Habits Questionnaire		X		X		X			X	X
Sleep diary issued[g]	X	X		X		X			X	
Sleep diary collected and reviewed		X		X		X			X	X
BASC-3 Parenting Relationship Questionnaire	X	X		X		X			X	
Parent Global Impressions-Severity	X	X		X		X			X	X
Parent Global Impressions -Improvement				X		X			X	X
Pediatric Quality of Life Inventory		X				X			X	X
Clinician-Completed Measures										
Pediatric Sleep CGI structured sleep history	X[h]	X[i]		X[i]		X[i]			X[i]	X[i]
Pediatric Sleep CGI-Severity	X	X		X		X			X	X
Pediatric Sleep CGI-Improvement		X		X		X			X	X
Caregiver top 3 concerns visual analog scale		X		X		X			X	X
Vineland Adaptive Behavior Scale, 3rd Edition		X				X			X	
CGI-Severity	X	X		X		X			X	X
CGI-Improvement				X		X			X	X
Subject-Completed Measures										
Actigraph issued	X	X		X		X			X	
Actigraph and actigraphy data collected[j]		X		X		X			X	X

KiTAP	X	X				X			X	X
RBANS	X	X				X			X	X
[redacted]										

Abbreviations: BASC-3, Behavior Assessment System for Children, 3rd Edition; CGI, clinical global impressions; DSM-V, Diagnostic and Statistical Manual of Mental Disorders;
EOS, end of study; EOT, end of treatment; FXS, fragile X syndrome; KiTAP, Test of Attentional Performance for Children; LF, long form; [redacted] RBANS, Repeatable Battery for the Assessment of Neuropsychological Status; SF, short form.
Note: At the investigator's discretion, subjects may be evaluated at unscheduled clinic visits for reasons related to subject safety. At unscheduled visits, subjects will be queried
about AEs, changes in concomitant medications, and suicidality, and safety laboratory assessments may be conducted.
a Screening is to be conducted for up to 21 days (inclusive) before Day 1; subjects who meet all entry criteria can be randomly assigned at any time during the screening period.
b There is no Day 0. The screening period can start up to 21 days before Day 1.
c The EOS visit may occur 4 to 6 weeks after the EOT visit, between the end of Week 16 and the end of Week 18 for subjects who do not stop treatment early.
d Signs and symptoms of FXS.
e Only if no written evidence for molecular diagnosis is available in the subject's records.
f If the baseline visit occurs within 10 days of the clinical laboratory tests during screening, clinical laboratory testing does not need to be performed at the baseline visit.
g The sleep diary is to be completed over the 7-day periods immediately preceding Day 1 and the Week 2, Week 6, EOT, and EOS clinic visits; a calendar will be issued to ensure that the diary information collection starts on the appropriate days.
h At the initial screening visit, the clinician will complete parts A, B, and C of the structured sleep history form.
i At clinic visits after the screening visit, the clinician will complete only part A of the structured sleep history form.
j The actigraph is to be worn and actigraphy data are to be collected for the 7-day periods immediately preceding Day 1 and the Week 2, Week 6, EOT, and EOS clinic visits; a calendar will be issued to ensure that the subject starts wearing the actigraph on the appropriate days.