OVID THERAPEUTICS INC.

STATISTICAL ANALYSIS PLAN

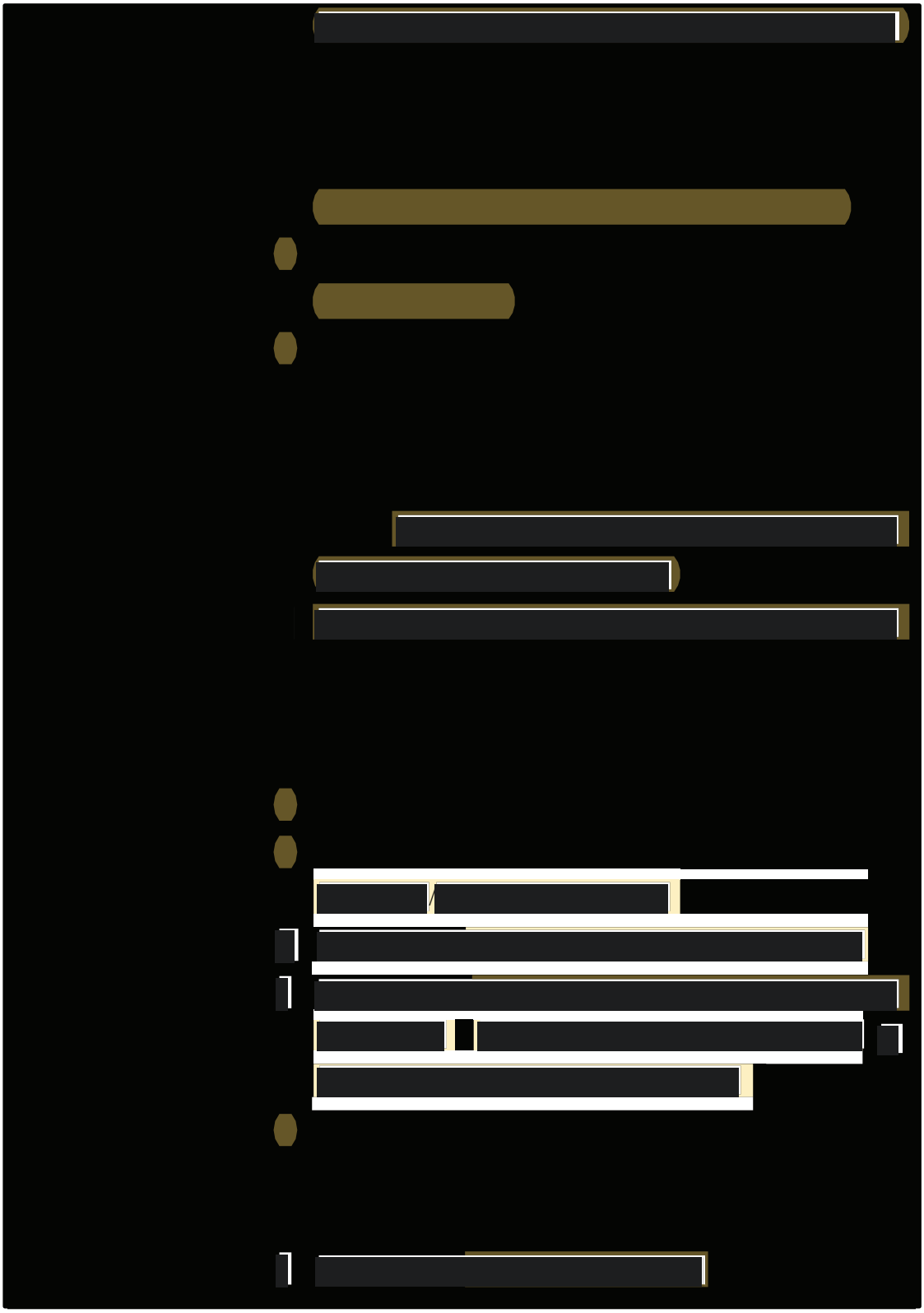
A MulticeNter, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group, Phase 3 Study to Evaluate the Efficacy and Safety of OV101 in PediaTric IndividUals with AngelmaN SyndromE (NEPTUNE)

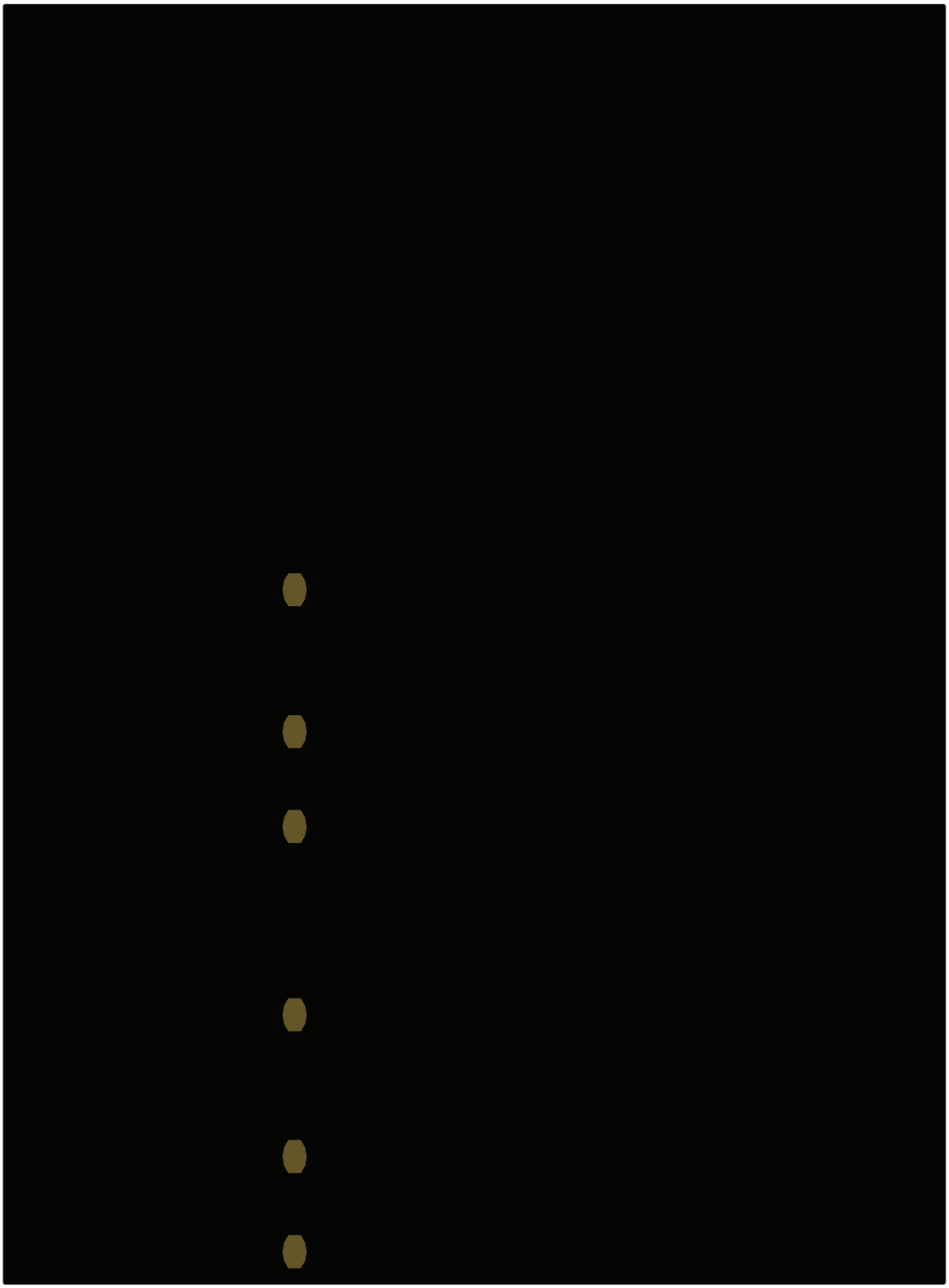
OV101-19-001

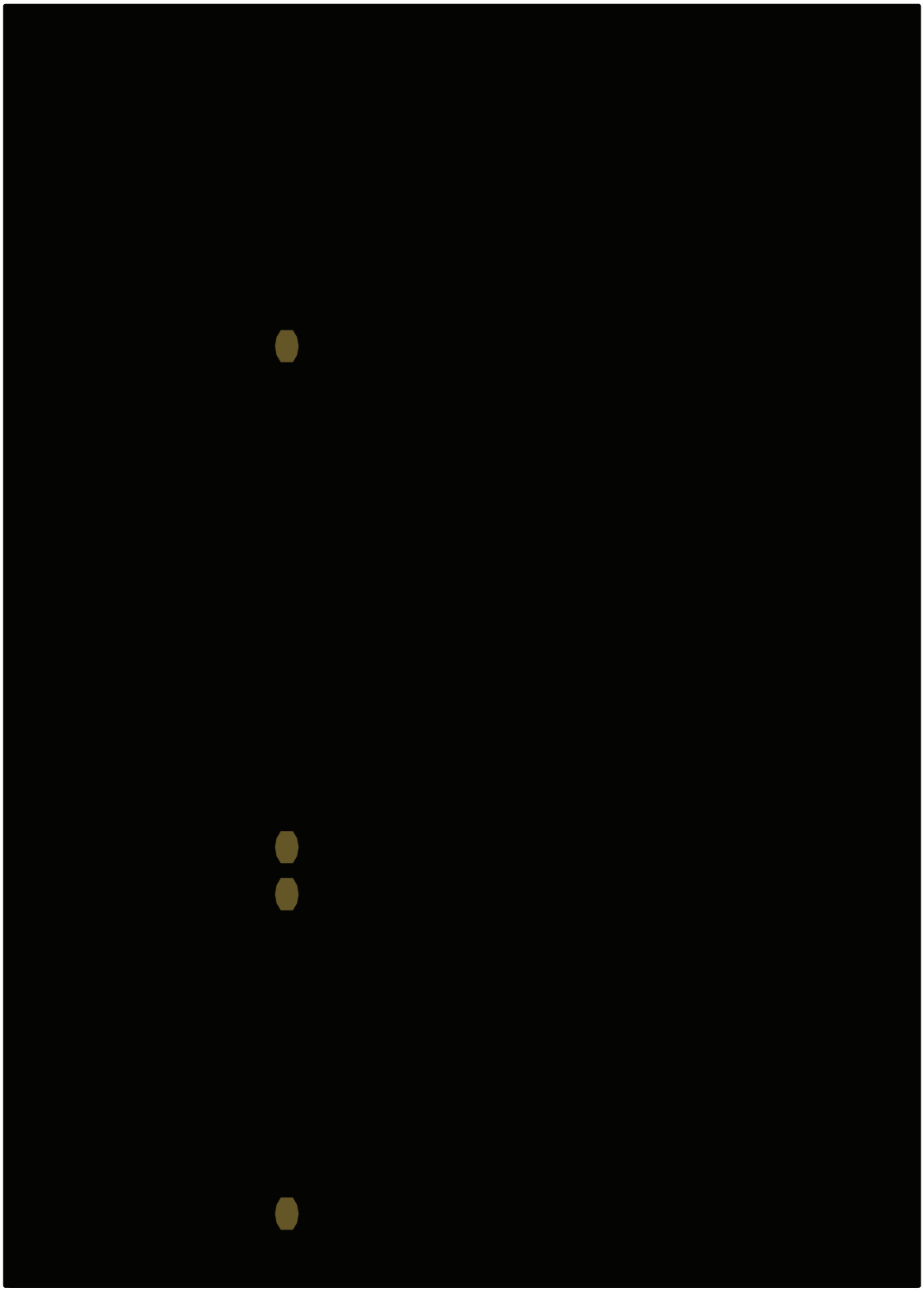
SAP Version:	Date of Statistical Analysis Plan:
Version 3.0	15Oct2020

SIGNATURE PAGE

DOCUMENT HISTORY







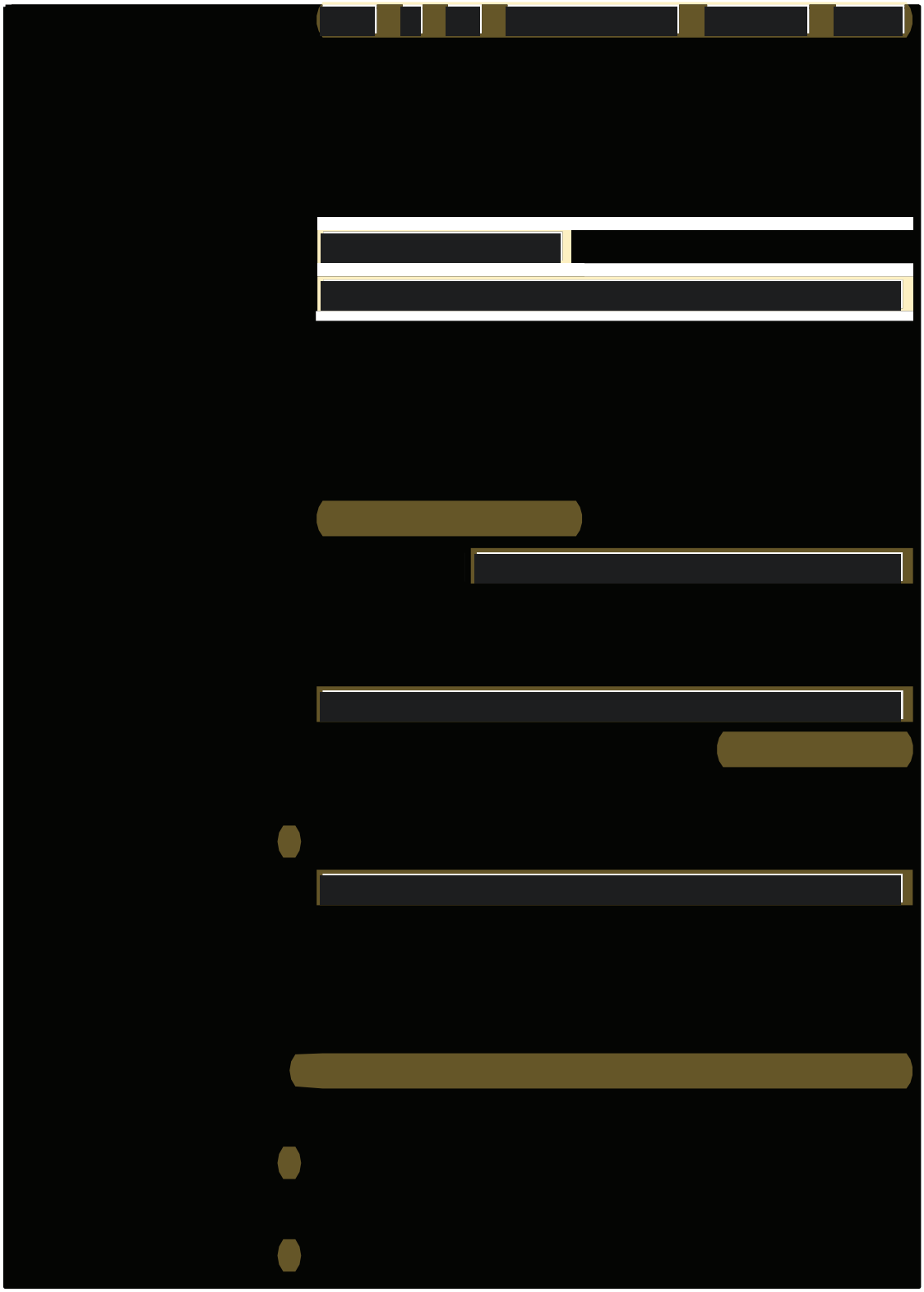


Table of Contents

LIST OF ABBREVIATIONS

Abbreviation	Full Term
ABC-I	Aberrant Behavior Checklist – Irritability Subscale
AE	adverse event
ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
AS	Angelman syndrome
ATC	Anatomic Therapeutic Class
BMI	Body Mass Index
CI	Confidence Interval
CGI-I-AS	Clinical Global Impressions-Improvement-Angelman Syndrome
CGI-S-AS	Clinical Global Impressions-Severity-Angelman Syndrome
[redacted]	
CSR	clinical study report
eCRF	electronic case report form
EOS	end of study
EOT	end of treatment
FAS	full analysis set
GSV	Growth Scale Value scores from Vineland Adaptive Behavior Scale, 3rd Edition
HLGT	high level group term
HLT	high level term
ICF	informed consent form

ICH	International Council for Harmonisation
IDMC	Independent Data Monitoring Committee
ITT	intent-to-treat set
IWRS	interactive web response system
LAR	legally acceptable representative
[redacted]	
MAR	missing at random
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple Imputation
MMRM	Mixed Model Repeated Measures
MNAR	missing not at random
[redacted]	
PPS	per protocol set
PT	preferred term
Q1	First Quartile
Q3	Third Quartile
QHS	each evening
REML	restricted maximum likelihood
SAE	serious adverse event(s)

SAP	statistical analysis plan
SD	Standard Deviation
SOC	system organ class
SS	safety set
TEAE	Treatment-Emergent Adverse Event
VABS-3	Vineland Adaptive Behavior Scale, 3rd Edition
WHO-DD	World Health Organization Drug Dictionary

1 INTRODUCTION

Study OV101-19-001 is a Phase 3, multicenter, double-blind, placebo-controlled, parallel-group study. The goals of the study are to evaluate the efficacy and safety of once each evening (QHS) OV101 (gaboxadol) compared to placebo in pediatric individuals with Angelman Syndrome (AS).

The statistical analysis plan (SAP) contains a detailed description of the data presentations and statistical analyses that will be included in the clinical study report (CSR) for Protocol OV101-19-001 version 4.0, 19 June 2020. The statistical methods and analyses described here are based on those presented in Section 7 of the study protocol.

This study is ongoing during the Coronavirus Disease 2019 (COVID-19) pandemic, and as a result, there are expected to be protocol deviations related to COVID-19, including virtual visits (visits conducted remotely via visiting nurse, audio only, or audio and video) instead of clinic visits and use of local labs. Data will be collected as closely as possible to the Schedule of Assessments (Table 1), and analysis will be performed using the observed data.

2 STUDY SUMMARY

2.1 Study Objectives

All efficacy study objectives will be assessed in the pediatric AS study population of subjects who are 4 to 12 years old. All [redacted] safety study objectives will be assessed in the pediatric AS study population of subjects who are 2 to 12 years old.

Primary (Efficacy) Objective:

- To evaluate the efficacy of OV101 versus placebo as assessed by the Clinical Global Impressions-Improvement-Angelman Syndrome (CGI-I-AS) score at Week 12.

Key Secondary Objectives:

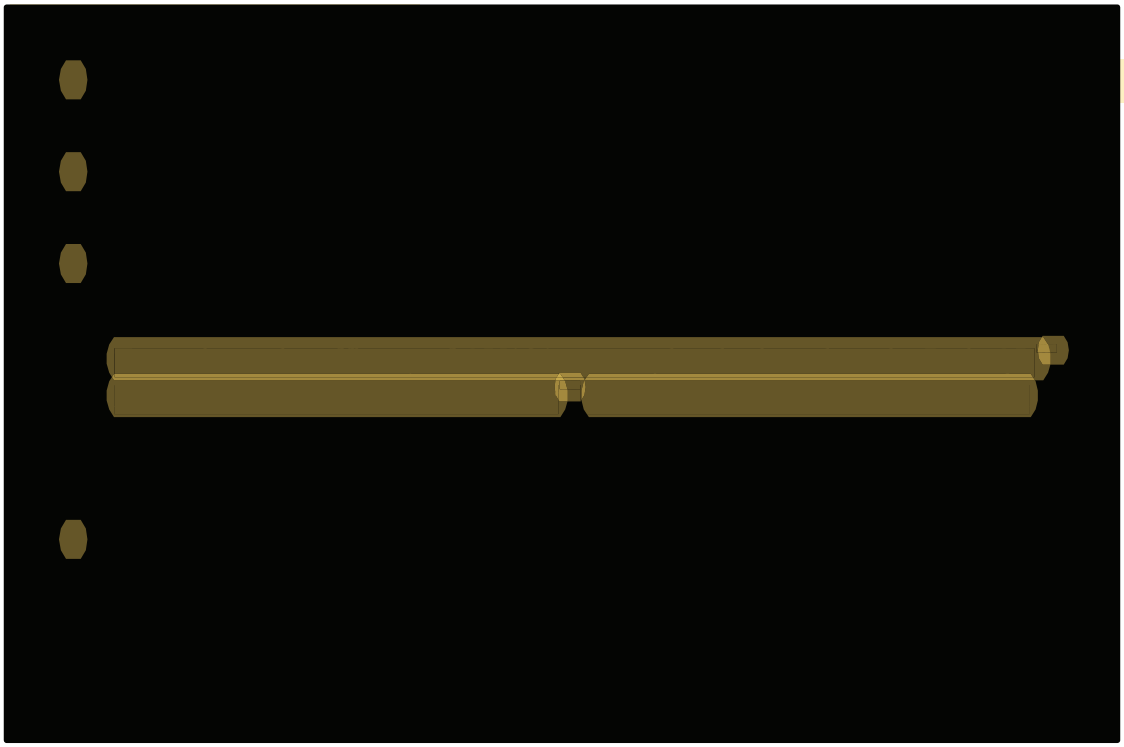
The purpose of the key secondary objectives is to fulfil a recommendation of the EMA for a responder analysis.

- To evaluate the efficacy of OV101 vs placebo as assessed by the proportion of subjects who experience any meaningful improvement on study treatment (defined as CGI-I-AS score of 1, 2, or 3 at Week 12).

- To evaluate the efficacy of OV101 vs placebo as assessed by the proportion of subjects who experience a response of much improved or very much improved (defined as CGI-I-AS score of 1 or 2 at Week 12).

Other Secondary (Efficacy) Objectives:

- To evaluate the efficacy of OV101 versus placebo on Communication, Socialization, Daily Living Skills, Motor Skills, and Maladaptive Behavior domains assessed by the Vineland Adaptive Behavior Scales, 3rd Edition (VABS-3).
- To evaluate the efficacy of OV101 versus placebo based on the Clinical Global Impressions-Severity-Angelman Syndrome (CGI-S-AS) Symptoms Overall score.
- To evaluate the relationships of CGI-S-AS Symptoms Overall and CGI-S-AS domains at baseline with CGI-I-AS at Week 12.



Safety Objectives:

- To evaluate the safety and tolerability of OV101 versus placebo in subjects 2 to 12 years old, including seizure diary data and assessment of suicidality.

2.2 Study Design

A brief summary of the main study design features and assessments to be performed is presented here. Refer to Section 3 of the study protocol for complete details.

This is a Phase 3, multicenter, double-blind, placebo-controlled, parallel-group study of OV101 treatment for 12 weeks in pediatric individuals with Angelman Syndrome. Subjects 4 to 12 years old will be randomized to either OV101 or placebo, while all subjects 2 to 3 years old will be treated with OV101. Subjects who meet all of the inclusion criteria, none of the exclusion criteria and who consent to participate in the study are eligible to be enrolled to complete a starting dose period for the first 5 study days and a maintenance period thereafter through Week 12. Subjects will take OV101 capsules orally once daily at bedtime, with dosing dependent on a subject's weight and tolerance of the study drug. Subjects will visit their study site at Weeks 6 and 12 for safety and efficacy evaluation. Efficacy evaluations are performed only in subjects 4 to 12 years of age. Safety evaluations are performed in all subjects 2 to 12 years of age. An end of study (EOS) phone safety visit will occur approximately 2 weeks after the last dose of study drug to assess safety and tolerability associated with discontinuation of treatment in subjects not rolling into the open label extension study (ELARA). Subjects eligible and willing to enroll into the ELARA study will have the NEPTUNE EOT visit also be their NEPTUNE EOS visit. A subject will be considered to have completed the study after completing the EOS visit.

2.2.1 Number of Subjects

Approximately 90 subjects 4 to 12 years old (inclusive) will be randomized into this study, with approximately 45 subjects in the placebo group and 45 subjects in the OV101 treatment group. The sample size determination is described in Section 7.3 of the Protocol. Approximately 5 subjects 2 to 3 years old (inclusive) will be included and treated with OV101.

The study will be conducted at approximately 15 study sites and will likely involve variable numbers of subjects at each site. Enrollment will be on a competitive basis but each site will be capped at 12 subjects (unless authorized by the Sponsor). Activation of study sites will be centrally controlled by interactive web response system (IWRS). Subject randomization will be deactivated for all sites when the planned number of subjects is met.

2.2.2 Randomization and Blinding Procedures

Eligible subjects are required to sign the informed consent form (ICF), complete the screening period and assessments, and meet all inclusion/exclusion requirements.

Eligible subjects 4 to 12 years old (inclusive) will be assigned in a 1:1 ratio on Day 1 to receive investigational product according to a computer-generated randomization schedule. Subjects will be assigned to treatment groups in sequential order as they qualify for the study, using a centrally located and managed IWRS. Randomization will be stratified by age categories: 9 to 12 years old (inclusive) and 4 to 8 years old (inclusive), and region: US and outside US, with at least 24 subjects per age category. Upon randomization, each subject will receive a subject randomization number and a drug kit assignment.

This is a double-blind, placebo-controlled study. The sponsor, investigators, study coordinators, pharmacists, and study subjects will be blinded to the study medication. To maintain blinding of the individual treatment assignment, the results of [redacted] will not be made available to any study personnel or subjects.

If knowledge of the treatment is needed to manage the subject's condition, the investigator will contact the IWRS to obtain the treatment assignment. If unblinding occurs for any reason, the time and reason for breaking the blind will be recorded on the electronic case report form (eCRF) and the sponsor must be notified within 24 hours.

A designated statistician who is not involved with the study operation will hold the treatment codes. The Independent Data Monitoring Committee (IDMC) will review blinded data, but will also have the ability to review unblinded data as needed to evaluate benefit risk during the study. For example, the unblinded treatment information can be provided to the IDMC to facilitate the evaluation of any clinically important increase in the rate of a serious suspected adverse reaction or to the designated safety contact when the treatment information is required to determine if an expedited safety report must be submitted to regulatory agencies. The IDMC and relevant unblinded personnel should prevent any risk of unblinding of data to the sponsor, investigators, study coordinators, pharmacists, and study subjects.

Eligible subjects 2 to 3 years old (inclusive) will be treated with OV101, and their treatment is not blinded.

2.2.3 Efficacy Assessments

Efficacy assessments will include the CGI-I-AS and CGI-S-AS scales, [redacted]

CGI-S-AS will be collected at Screening, Baseline, Weeks 6 and 12. CGI-I-AS will be assessed at Weeks 6 and 12. [redacted] will be obtained at Baseline, Week 6 and Week 12. The VABS-3 and [redacted] instruments will be administered at Baseline and Week 12.

2.2.3.1 Clinical Global Impressions Scales

The CGI-S-AS scale with AS-specific anchors will be used by the investigator to assess the severity of symptoms, and the CGI-I-AS scale will be used by investigators to assess improvement from baseline. The CGI-S-AS has been adapted to capture specific characteristics commonly present in the AS population. The CGI-I-AS captures clinical impression that reflects the rater's estimate of change from the initiation (Baseline) of treatment. Both the CGI-S-AS and the CGI-I-AS assessments will be conducted by a clinician rater with experience in AS and trained in CGI-S-AS and CGI-I-AS rating by the Sponsor. The rater should be the same rater throughout the study for a given subject (on the eCRF, the investigator will be identified as either the same or different since the last time the subject was seen). As an exception, a clinician sub-investigator, trained in CGI-S-AS and CGI-I-AS rating by the Sponsor, may be used. The Baseline CGI-S-AS will be accompanied by detailed Investigator Notes. Each CGI-I-AS rating at Week 6 and Week 12 will also be accompanied by detailed Investigator Notes. In order to reduce recall bias, within 24 hours prior to the Week 6 CGI-I-AS rating, the rater will review the CGI-S-AS rating and Investigator Notes from the baseline visit and within 24 hours prior to the Week 12 CGI-I-AS rating, the rater will review the CGI-S-AS rating and Investigator Notes from the baseline visit and CGI-I-AS rating and Investigator Notes from Week 6. If the CGI-I-AS assessment is missing at Week 6 or 12, the reason for its absence will be collected on the eCRF.

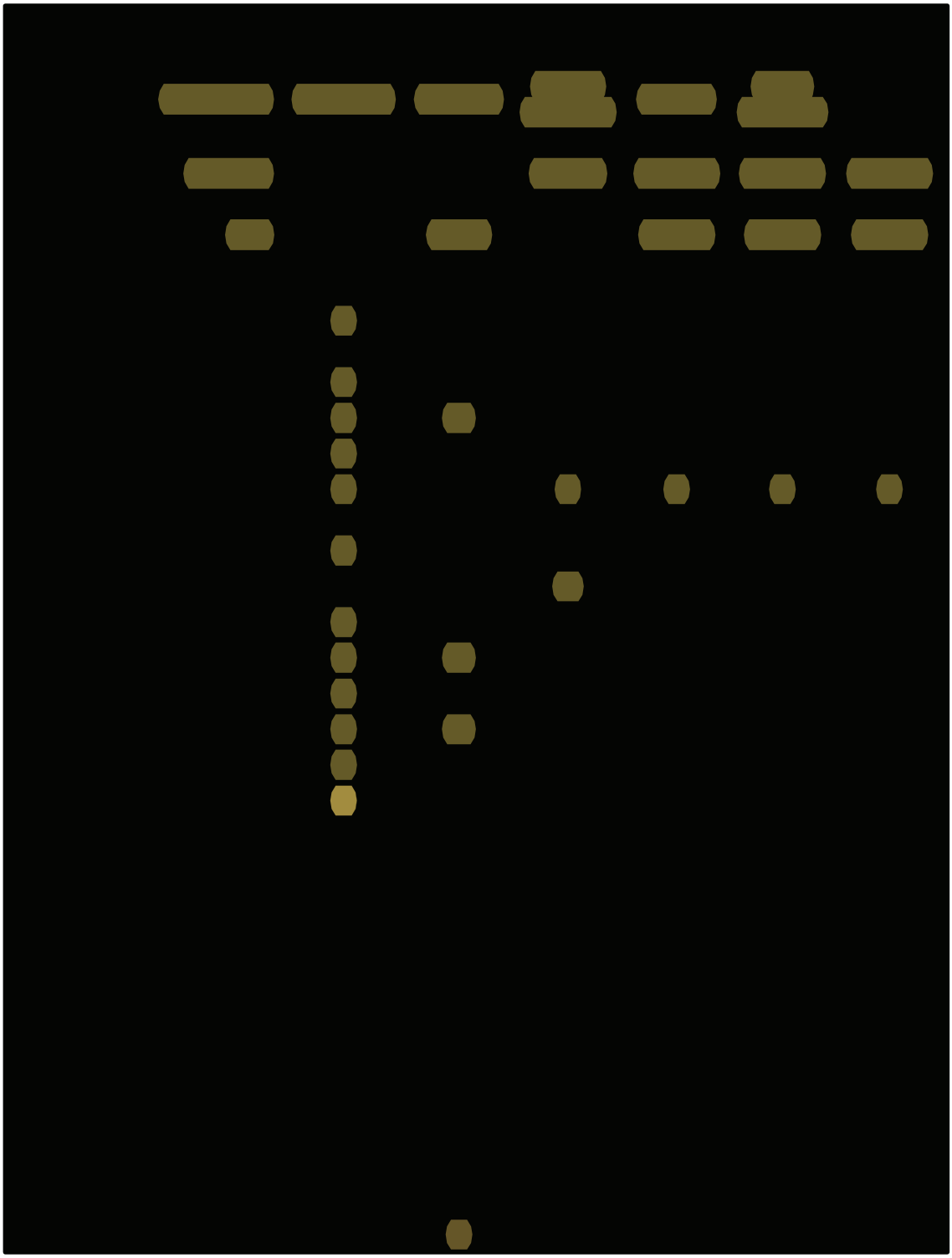
2.2.4 Safety Assessments

Adverse event (AE) occurrences will be recorded throughout the study. Adverse events will be assessed beginning on the date of signed informed consent and up to 14 days after the last dose of study drug (unless a subject decides to participate in the ELARA open-label extension study) and must be followed until resolution or for 14 days after the subject's last study drug dose, whichever comes first. If the investigator becomes aware of an SAE within 30 days after the subject's last dose of study drug or within 30 days after the last study visit or more than 30 days after the last dose of study drug if considered by the investigator to be related to study drug, the SAE must be reported.

Seizures in this subject population will be detected and assessed using the seizure diary. Diaries will be collected at the study site during the study and will be analyzed by the sponsor along with the reportable serious AEs (SAEs) in evaluating risks and benefits. The sponsor will report the SAE events of seizure that meet the criteria for reporting seizures as an AE (protocol section 6.4.1.1) in an aggregated, unblinded report at the end of the study.

Self-harm will be determined using the Aberrant Behavior Checklist – Irritability Subscale (ABC-I), which is one of the five subscales of the Aberrant Behavior

Checklist – Community (ABC-C). Additional safety assessments include clinical laboratory analyses, vital sign measurements, and physical examinations.



3 STATISTICAL METHODS

3.1 General Methods

3.1.1 Computing Environment

All statistical analyses will be performed using SAS® Version 9.4 or higher.

3.1.2 Reporting of Numerical Values

All clinical study data will be presented in subject data listings. Data summaries will be presented for all endpoints and will include descriptive statistics (number of subjects [n], mean, standard deviation [SD], first quartile [Q1], median, third quartile [Q3], minimum and maximum) by treatment group for continuous variables. Frequencies and percentages will be presented by treatment group for categorical and ordinal variables. If there are missing values, the number missing will be presented, but without a percentage. All percentages will be based on the number of subjects with non-missing values. Confidence intervals will be provided where appropriate.

Means, quartiles, and confidence intervals will be reported to one decimal place more than the data reported on the eCRF or by the laboratory/vendor. Standard deviations will be reported to two decimal places more than the data reported. Minimum and maximum will be reported to the same number of decimal places displayed on the eCRF or by the laboratory/vendor. P-values will be reported to 4 decimal places.

3.1.3 Statistical Testing and Significance Level

The primary efficacy analysis will be performed on the full analysis set (FAS) and the per protocol set (PPS), which are defined in Section 3.2. However, the PPS analysis will only be considered a sensitivity analysis.

Inferential treatment comparison of the primary efficacy endpoint, CGI-I-AS at Week 12, will be conducted at the 0.05 level of significance using a two-tailed test. If the null hypothesis of no treatment difference for the primary endpoint is rejected, the alpha (0.05) will be passed to the significance testing for the keysecondary efficacy endpoints (added to fulfill an EMA recommendation for responder analysis), CGI-I-AS response (at least minimally improved) at Week 12 and CGI-I-AS response (at least much improved) at Week 12, which will be tested using the corresponding 2-sided tests at 0.05 level of significance in a fixed sequence. If and only if the null hypothesis of no treatment difference based on CGI-I-AS response (at least minimally improved) is rejected at 0.05 level of significance, testing will continue and the null

hypothesis of no treatment difference based on CGI-I-AS response (at least much improved) will be tested at 0.05 level of significance.

No formal statistical testing of other secondary endpoints will be performed. The p-values presented for other secondary and tertiary endpoints will not be adjusted for multiplicity and will be considered descriptive in nature.

3.1.4 Baseline Value and Change from Baseline

Baseline value is defined as the latest non-missing value obtained prior to administration of first dose, [redacted] Change from baseline will be calculated by subtracting the baseline value from the post-dose assessment for each subject (i.e., post-dose – baseline).

3.1.5 Study Day

Day 1 will be defined as the day of first dose of investigational product during the Treatment Period. Study day for events prior to randomization and first dose will be determined as Study Day = (Event Date – First Dose Date). Study day for events on or after first dose will be determined as Study Day = (Event Date – First Dose Date) + 1.

3.1.6 Handling of Missing/Incomplete Values

3.1.6.1 Missing or Incomplete Efficacy Endpoints

To the extent possible, attempts will be made to minimize the amount of missing data through measures planned in the study conduct. Due to the scale of the study and the anticipated small number of missing data for the primary efficacy endpoint (Week 12 CGI-I-AS), no simple imputation for the missing data will be performed. However, if missing data are more than 3% for the primary efficacy endpoint, a sensitivity analysis will be performed. The missing data will be handled using a multiple imputation approach (including a tipping point analysis), as a part of the sensitivity analysis, as described in Section 3.8.6 of this SAP.

For each of the [redacted] endpoints within the other secondary endpoint set [redacted] if more than 5% but less than 20% of the subjects have missing values, sensitivity analyses will be performed where missing data will be handled using a multiple imputation approach, as described in Section 3.8.6 of this SAP.

For any other secondary or tertiary efficacy endpoints, the missing data will not be imputed and only the observed data will be used for analyses.

3.1.6.2 Missing and Incomplete Dates

Partial medication dates will be imputed with maximum conservatism when determining if a medication is concomitant. The rules used for imputation of dates in concomitant medication analyses are detailed in section 3.5 of this SAP.

AE start dates will be imputed but AE stop dates will not be imputed. Rules for handling incomplete or missing AE start dates are detailed in section 3.8.1 of this SAP.

No other baseline or safety data points will be imputed.

3.1.7 Methods of Pooling Data

Because of the large number of centers relative to the number of subjects, an analysis adjusted by study center is not planned. Similarly, it is not expected to have a sufficient number of subjects in each country involved in the study to adjust for country. However, descriptive analysis will be performed by country for the primary efficacy endpoint.

3.1.8 Analysis Visit Windows

Visit windows for study conduct are specified in Table 1: Schedule of Activities. Unless otherwise specified, the following analysis visit rules will be implemented for all endpoints [redacted] and seizure diary (see sections 3.8.3, 3.8.4, and 3.9.6 for analysis visit window computations). For all unscheduled visits and end of treatment (EOT) visits from subjects discontinuing treatment early (before Week 12), assessments will be mapped according to the analysis visit windows in Table 2 for primary summarization and analysis. If the visit does not fall within any of the analysis visit windows, then the visit will be considered missing at the specified time point for analysis purposes.

Table 2: Nominal (Targeted) Time Points and Analysis Visit Windows

Assessment	Nominal (Targeted) Time Point	Analysis Visit Window (Interval in Study Days)
Baseline Assessments	Study Day 1	≤ Day 1
Phone/Titration (Day 6)	Study Day 6	>Day 1 - ≤Day 8
Phone/Titration (Day 11)	Study Day 11	≥Day 9 - ≤Day 21
Week 6 (Study Day 43)	Study Day 43	≥Day 22 - ≤Day 64
Week 12 (Study Day 85)	Study Day 85	≥Day 65 - ≤Day 92
Week 14 (Study Day 99)	Study Day 99	≥Day 93

After mapping the data from unscheduled and end of treatment (EOT) visits from subjects discontinuing treatment early to the nominal time point, the following rules will apply if multiple records are available within a single analysis visit window unless otherwise specified for a particular analysis:

- If a record exists for the nominal time point, that record will be selected for analysis.
- If a record for the nominal time point does not exist and multiple records are available within a single analysis visit window,
    - The record closest to the planned assessment day will be selected for analysis
    - If 2 records are equidistant from the target day, then the earliest record will be selected (does not apply to the Week 12 or 14 visits).
- If a subject has no record in an analysis window, the subject will be considered missing at that time point.

For the primary and key secondary objectives, a supportive analysis will be conducted in which CGI-I-AS assessments from unscheduled visits and end of treatment (EOT) visits from subjects discontinuing treatment early (before Week 12) will be mapped according to the analysis visit windows in Table 3.

Table 3: Nominal (Targeted) Time Points and Alternative Analysis Visit Windows

Assessment	Nominal (Targeted) Time Point	Analysis Visit Window (Interval in Study Days)
Baseline Assessments	Study Day 1	≤ Day 1
Week 6 (Study Day 43)	Study Day 43	>Day 1 - ≤Day 43
Week 12 (Study Day 85)	Study Day 85	>Day 43

After mapping the data to the nominal time point, the following rules will apply if multiple records are available within a single analysis visit window unless otherwise specified for a particular analysis:

- If a record exists for the nominal time point, that record will be selected for analysis.
- If a record for the nominal time point does not exist and multiple records are available within a single analysis visit window,
    - The record closest to the planned assessment day will be selected for analysis
    - If 2 records are equidistant from the target day, then the earliest record will be selected (does not apply to the Week 12 visit).
- If a subject has no record in an analysis window, the subject will be considered missing at that time point.

In addition, for the primary and key secondary objectives, an additional supportive analysis will be conducted in which CGI-I-AS assessments will be analyzed according to the nominal visit time point (i.e. only records identified as baseline and records from the Day 43/Week 6 and Day 85/EOT/Week 12 visits will be included).

3.2 Analysis Populations and Subgroups

3.2.1 Definition of Analysis Populations

The following analysis sets, with the exception of the Intent-to-treat and Safety Set, will only include subjects 4 to 12 years old (inclusive) at screening. The baseline and safety analyses for subjects 2 to 3 years old (inclusive) at screening will be conducted separately.

Intent-to-Treat Analysis Set

The intent-to-treat set (ITT) will consist of all subjects who are randomized (subjects aged 4 to 12 years old (inclusive) or enrolled (subjects aged 2 to 3 years old (inclusive) and meet all inclusion and exclusion criteria), whether or not study drug is received. Subjects aged 4 to 12 years old (inclusive) will be analyzed according to their randomized treatment.

Full Analysis Set

The full analysis set (FAS) will include all subjects aged 4 to 12 years old (inclusive) who are randomized, receive at least 1 dose of study drug, and have at least 1 post-baseline assessment of the primary efficacy endpoint. The FAS will be used for efficacy analyses based on the randomized treatment for each subject.

Per Protocol Set

The per protocol set (PPS) is a subset of the FAS that includes subjects who complete the Week 12 CGI-I-AS assessment (with appropriate visit mapping as shown in Table 2) and have no major protocol deviations that are deemed to impact efficacy. These violations include but are not limited to violations of inclusion/exclusion criteria deemed to impact efficacy, inadequate study drug compliance over the dosing period as determined from the dosing information in the eCRF, usage of prohibited concomitant medications, and incorrect randomization. Additionally, other relevant protocol deviations, such as early termination or interruption of treatment, may be identified for review.

Prior to database lock and unblinding, there will be a meeting to review all protocol deviations to determine PPS exclusions. The threshold for percent compliance that would be considered inadequate and would result in exclusion from the PPS will also be determined in this meeting following review of the relevant compliance

information. The PPS will be used for efficacy analyses based on the randomized treatment for each subject.

Safety Set

The safety set (SS) will consist of all subjects in the ITT who receive at least 1 dose of study drug and will be used for safety analyses. The safety analyses will be based on the treatment that was actually administered to each subject and will be conducted separately for subjects 2 to 3 years old (inclusive) and subjects 4 to 12 years old (inclusive).

[redacted]

3.2.2 Definition of Subgroups

Subgroups will be defined by age group (randomization stratification variable: age 4 to 8 years versus age 9 to 12 years) and by region (randomization stratification variable: US versus outside US).

3.3 Analysis Endpoints

3.3.1 Primary Efficacy Endpoint

- The CGI-I-AS score at Week 12

3.3.2 Key Secondary Efficacy Endpoint

- CGI-I-AS response (at least minimally improved) defined as CGI-I-AS score of 1, 2, or 3 at Week 12
- CGI-I-AS response (at least much improved) defined as CGI-I-AS score of 1 or 2 at Week 12

The key secondary endpoints will be utilized to fulfill a recommendation of EMA for a responder analysis.

3.3.3 Other Secondary Efficacy Endpoints

[redacted]

- VABS-3 assessment scores, change from Baseline to Week 12
    - Total Score (Adaptive Behavior Composite Score)
    - Communications domain and its subdomains
    - Socialization domain and its subdomains
    - Daily Living Skills domain and its subdomains
    - Motor Skills domain and its subdomains
    - Maladaptive Behavior subdomains
- CGI-S-AS scores, change from Baseline to Week 12
    - The CGI-S-AS symptoms overall score
    - The CGI-S-AS domain scores

3.3.6 Safety Endpoints

- Incidence of adverse events
- Change in clinical laboratory parameters from Baseline to each post-baseline assessment
- Shift in clinical laboratory parameters from Baseline to each post-baseline assessment
- Change and shift in vital sign measurements from Baseline to each post-baseline assessment

- Change in suicidality assessment (ABC-I) from Baseline to each post-baseline assessment. The ABC-I will serve as a surrogate measure for suicidality
- Percent change in seizure frequency (collected in seizure diary) per 28 days from Baseline to each post-baseline time point

3.4 Subject Disposition and Evaluability

3.4.1 Subject Disposition

The number of subjects screened and who failed screening will be displayed. The number and percentage of subjects who are randomized, dosed, withdrawn from the study (including reasons for withdrawal), complete the study, the number of subjects with a Week 6 remote visit performed, and the number of subjects with a remote Week 12 visit performed will be displayed for subjects 4-12 years old (inclusive) by treatment group and overall for all subjects who signed informed consent. The number and percentage of subjects in the ITT, SS, FAS, PPS, [redacted] will also be displayed by treatment group and overall.

The number of subjects who screened, dosed, completed the study, withdrew from the study, and the number and percentage of subjects in the ITT and SS will be summarized separately for subjects 2 to 3 years old (inclusive). The number of subjects with a Week 6 remote visit performed, and the number of subjects with a remote Week 12 visit performed will also be included.

In addition, for subjects 4-12 years old (inclusive) and separately subjects aged 2 to 3 years old (inclusive), a summary table of subjects with at least one visit affected by COVID-19 will be presented. A separate subject listing of subject visits and assessments affected by COVID 19 will be provided.

A by-subject listing of study completion information will be presented. A by-subject listing of subjects who are screen failures, and the reason for screen failure, will also be presented.

3.4.2 Protocol Deviations

The number and percent of subjects with at least one major protocol deviation and the number and percent of subjects with at least one major deviation of each deviation type will be presented by treatment group and overall for the FAS. All relevant protocol deviations will be shown in a subject listing by treatment group. Subjects with major deviations identified as exclusions from the per-protocol analysis set will be indicated in a listing. In addition, deviations related to COVID-19 will be identified in a separate listing.

A separate listing of reasons for exclusion from the Per Protocol Set for subjects ages 4-12 years old (inclusive) will also be provided.

3.5 Demographics and Baseline Characteristics

3.5.1 Demographics

All demographic and baseline characteristics data will be summarized descriptively for subjects 4-12 years old (inclusive) by treatment group and overall for the SS, FAS, and PPS. A separate summary will be provided for subjects 2-3 years old (inclusive) in the SS.

Descriptive statistics will be displayed for age at screening, age at baseline, height, weight at baseline, and body mass index (BMI). Age at screening (collected as age at time of informed consent) is recorded at the time of randomization (or enrollment for subjects aged 2 to 3 years). In addition, if the age of the subject changed between signing of informed consent and the randomization (enrollment) visit, this information is also recorded. For such cases in which the age is reported to have changed, one year is added to age at screening to determine age at baseline. For all other subjects, age at baseline is set to the age at screening. Height, weight and BMI will be summarized with the data collected at the baseline visit.

Frequencies and percentages will be tabulated for age group (randomization stratification variable: age 4 to 8 years versus age 9 to 12 years) [for subjects 4-12 years old (inclusive) only], region, gender, race, ethnicity, study center, country and history of seizures. History of seizures will be categorized as follows:

- Yes, seizure within last two years (defined as seizures within the same year or the year before informed consent)
- Yes, but no seizure within the last two years
- Yes, but unknown status in last two years
- No

- Unknown.

All demographics and baseline characteristics will be presented by subject in a data listing.

3.5.2 Medical History

All medical history conditions will be coded using the Medical Dictionary of Regulatory Activities (MedDRA, Version 22) and will be classified by MedDRA system organ class (SOC) and preferred term (PT). Medical history will be tabulated separately for subjects 4-12 years old (inclusive) and subjects 2-3 years old (inclusive) by SOC and PT for each treatment group and overall for the SS. All medical and surgical history conditions will be included in subject data listings.

3.5.3 Seizure History

Seizure history will be included in a subject data listing for subjects in the SS, sorted by age group at screening (subjects 4-12 years old (inclusive) and subjects aged 2 to 3 years old (inclusive)).

3.6 Prior and Concomitant Medications

Prior medications are defined as medications that started before first study drug administration and either stopped before or continued after first study drug administration. Concomitant medications are defined as medications that are being taken while on study drug. Medications that are ongoing on the date of the first administration of study drug will be classified as both prior and concomitant. Any medication that cannot be confirmed as stopping before the start of study drug will be classified as both a prior and a concomitant medication.

Partial medication dates will be imputed with maximum conservatism when determining if a medication is concomitant. If the onset date is missing or partial, the date will be compared as far as possible with the date of first dose of study medication. If no month is present and it is in the same year as the date of first treatment, or the year and month are the same as the first treatment date, the onset date will be imputed as the date of first treatment. Otherwise it will be imputed as 01 January in the available year, or the first day of the month and year available. End dates will be imputed as 31 December in the available year, or the last day of the available month and year collected. The imputed start date must be less than or equal to the imputed stop date.

All medications will be coded with the WHO Drug Dictionary (WHO-DD B3 format, March 2019). Prior and concomitant medications will be summarized separately and the number and percentage of subjects in each treatment group who took at least one

prior (concomitant) medication as well as the number and percentage of subjects who took each type of medication will be summarized separately for subjects 4-12 years old (inclusive) and subjects 2-3 years old (inclusive) by Anatomic Therapeutic Class (ATC) Level 2, ATC Level 4, and Preferred Term (PT) for the SS. If a subject has more than one occurrence of the same PT, then the PT will be counted only once for that subject. Similarly, if a subject has more than one PT within ATC Level 4 or ATC Level 4 within ATC Level 2, then the subject will be counted only once in that ATC Level 4 or ATC Level 2.

3.6.1 Non-Benzodiazepine Concomitant Medication

A separate summary of non-benzodiazepine concomitant medications indicated for seizure therapy will also be presented. Non-benzodiazepine medications will be classified based on WHO-DD ATC Level 2, ATC Level 4 and PT, as appropriate, prior to unblinding at database lock. Full details of the classification will be provided in the analysis data specifications. The number and percentage of subjects in each treatment group who took at least one non-benzodiazepine medication for seizure therapy, only one non-benzodiazepine medication for seizure therapy, two non-benzodiazepine medications for seizure therapy, three or more non-benzodiazepine medications for seizure therapy, as well as the number and percentage of subjects who took each type of non-benzodiazepine medication will be summarized separately for subjects 4-12 years old (inclusive) and subjects 2-3 years old (inclusive) by Anatomic Therapeutic Class (ATC) Level 2, ATC Level 4, and Preferred Term (PT) for the SS. A similar summary will be presented for non-benzodiazepines indicated for sleep. Medications with an indication for seizure therapy and with an indication for sleep will also be classified prior to unblinding at database lock and detailed in the analysis data specifications.

3.6.2 Benzodiazepine Concomitant Medication

A summary of benzodiazepine concomitant medications by dose frequency (As Needed, Immediately (STAT), Regularly, Unknown, Other) will also be presented. Benzodiazepine medications will be classified based on WHO-DD ATC Level 2, ATC Level 4 and PT, as appropriate, prior to unblinding at database lock. Full details of the classification will be provided in the analysis data specifications. For each dose frequency, the number and percentage of subjects in each treatment group who took at least one benzodiazepine medication, only one benzodiazepine medication, two benzodiazepine medications, three or more benzodiazepine medications, as well as the number and percentage of subjects who took each type of benzodiazepine medication will be summarized separately for subjects 4-12 years old (inclusive) and subjects 2-3 years old (inclusive) by ATC Level 2, ATC Level 4, and PT for the SS. A similar summary will be presented for benzodiazepines indicated for seizure therapy

by dose frequency. In addition, a summary of benzodiazepines indicated for non-seizure therapy by dose frequency will also be presented.

3.6.3 Most Common Concomitant Medication

A table of the most common concomitant medications for subjects aged 4 to 12 years old (inclusive) and separately for subjects 2-3 years old (inclusive) will be presented. The most common concomitant medications are those with an incidence ≥10% overall. The number and percentage of subjects in each treatment group who took at least one of the most common concomitant medications as well as the number and percentage of subjects who took each type of medication will be summarized by ATC Level 2 for the SS.

All prior and concomitant medication data will be listed for subjects in the ITT, including the WHO Drug ATC 2 and ATC 4 class, PT term, the investigators verbatim description and the classification as prior, concomitant or both. Medications will be sorted by medication start date within subject. Rescue medications and anti-epileptic medications will be flagged. Separate listings will be provided for benzodiazepine and non-benzodiazepine medications.

Concomitant procedures will be presented in subject listings for subjects in the ITT.

3.7 Treatment Exposure and Compliance

3.7.1 Exposure to Study Treatment

Treatment duration or days on treatment (defined as Last dose date – First dose date + 1) will be described for the SS, separately for subjects aged 2 to 3 years old (inclusive) and 4 to 12 years old (inclusive), and for the FAS using summary statistics. Additionally, study drug administration method to be used throughout the study (capsule or powder) as recorded at baseline and the number of subjects with at least one missed dose will also be presented. A missed dose is defined as at least one day when study medication was not administered as reported on the medication diary, with a missing diary day considered as dose not administered. Treatment duration frequencies will be summarized for the intervals 1 to ≤5 days, 6 to ≤11, 12 to ≤42 days, 43 to ≤70, 71 to ≤84 days, 85 days or more by treatment group and overall. Furthermore, treatment duration frequencies will also be summarized for the same intervals described above by weight band at screening (9 to 16.99 kg, 17 to 24.99 kg, 25 to 34.99 kg, 35 to 44.99 kg, and 45 to 64.99 kg).

3.7.2 Compliance to Protocol Intended Dose

Treatment compliance will be summarized for the SS and FAS for subjects aged 4 to 12 years old (inclusive) and for the SS for subjects 2-3 years old (inclusive), by

treatment group over the entire treatment period. Study drug is dispensed at Baseline and Week 6. Compliance rates during the treatment period will be derived using the following formula:

100*((Total number of capsules taken) / (Expected number of capsules to be taken),

where "Total number of capsules taken" will be derived separately for the intervals Baseline to Week 6 and for Week 6 to Week 12/EOT and is calculated as the total number of capsules dispensed minus the total number of capsules returned from the corresponding drug kit. The total number of capsules taken on study will be derived as the sum of the total capsules taken from the two intervals. Total number of capsules taken will be missing for an interval based on drug dispensation if drug kit is dispensed but not returned or is lost. In such cases, total number of capsules taken for the interval will be derived based on the study medication diary, where the "Total number of capsules taken" will be derived as the sum of the total number of capsules reported to have been taken each study day in the interval, with missing diary dates treated as 0 capsules taken. Complete derivation details will be provided in the ADaM specifications.

Expected number of capsules to be taken is based on the date of first study medication dose and the date of last study medication dose and the subject's weight range band at screening that determines the number of capsules expected to be taken at each study week, following the expected daily dosing and titration schedule (see protocol section 5.1)

Descriptive statistics will be displayed for total number of capsules taken, expected number of capsules taken, and compliance rate. Compliance rates will also be presented using frequency and percentages for the categories 0% to <20%, 20% to 40%, 40% to <60%, 60% to <80%, 80% to <=100%, >100% to ≤120%, or >120% by treatment group and overall.

All study drug exposure and compliance information will be presented in subject listings. All study drug titrations, study drug accountability, study medication taken daily diary information, and study medication overdoses will also be presented in by-subject listings.

3.8 Efficacy Analyses

All efficacy analyses will be based on the FAS. Additionally, the analysis of the primary endpoint will be repeated on the PPS to assess the sensitivity of the results to major protocol violations. All efficacy variables will be summarized descriptively by treatment group and visit, including observed, change from baseline, and percent change from baseline, if appropriate. Sensitivity analyses to assess the impact of

missing primary endpoint data will be conducted on the FAS analysis set and are described below.

3.8.1 Primary Efficacy Endpoint Analyses

The CGI-I-AS scale assessed by the clinician will be collected at Week 6 and 12 visits. This is an ordinal variable with scores ranging from 1 (very much improved), 4 (no change), to 7 (very much worse) which will be analyzed as a continuous/ordinal outcome variable in these analyses. The normality of CGI-I-AS at Week 6 and Week 12 for each treatment group will be examined. The examination of CGI-I-AS at each visit to ascertain the degree of departure from normality will involve generating bar charts, histograms, box-plots and normal probability plots, to visually evaluate the skewness and the kurtosis of the CGI-I-AS Week 6 and Week 12 data.

Descriptive statistics will be displayed for CGI-I-AS total improvement score at each visit by treatment group for the FAS. Frequencies and percentages will also be tabulated for the number of subjects with each specific score, the number of subjects with a response of at least minimally improved, and the number of subjects with a response of at least much improved by treatment group and overall. A response of at least minimally improved is defined as a CGI-I-AS score of 1, 2, or 3. A response of at least much improved is defined as a CGI-I-AS score of 1 or 2.

A separate summary will also be displayed for CGI-I-AS total improvement score at each visit for each age group at randomization (4 to 8 years or 9 to 12 years) within treatment group and all age groups combined within treatment group. The 95% CIs will be calculated for the mean difference in the CGI-I-AS total improvement score by week and age group for OV101 versus placebo. Frequencies and percentages will also be tabulated for the number of subjects with each specific score, the number of subjects with a response of at least minimally improved, and the number of subjects with a response of at least much improved by age group within treatment group and all age groups combined within treatment group.

Descriptive statistics will also be displayed for CGI-I-AS total improvement score at each visit by treatment group, separately for each stratification region (US, outside US). An additional table displaying CGI-I-AS total improvement score at each visit by treatment group, separately for each country will also be presented.

In addition, a separate summary will also be displayed for CGI-I-AS total improvement score at each visit for each method of CGI-I-AS assessment. For this analysis, subjects will be grouped into the remote/phone visit collection group if any CGI-I-AS assessment is conducted remotely and into the on-site visit collection group if none of the CGI-I-AS assessments are conducted remotely.

As the primary analysis, a Mixed Model Repeated Measures (MMRM) Analysis of Variance (ANOVA) model will be fitted using restricted maximum likelihood (REML) for CGI-I-AS with fixed effects for visit (Weeks 6 and 12), randomized treatment, the visit-by-treatment interaction, age group (randomization stratification variable: age 4 to 8 years versus age 9 to 12 years), and region (randomization stratification variable: US versus outside US) for the FAS. An unstructured within subject-covariance structure is preferred and will be specified first. If the model converges, the Kenward-Roger approximation will be used to estimate the denominator degrees of freedom.

If the model with unstructured covariance fails to converge, the following covariance structures will be tested in this order: toeplitz with heterogeneity, autoregressive with heterogeneity by visit, compound symmetry with heterogeneous variances by visit, toeplitz, autoregressive, and compound symmetry without heterogeneous variances by visit. The first covariance structure that converges will be used for the analysis. If a structured covariance is used, then a robust sandwich estimator will be utilized for estimating the variance of the treatment effect estimate.

From the model, the Week 12 difference between least squares means of OV101 and placebo groups will be presented along with the corresponding 95% confidence interval and 2-sided *P*-value. This analysis will be repeated for the PPS. Furthermore, as an additional supportive analysis, this analysis will also be repeated based on method of CGI-I-AS assessment collection (remote/phone or on-site) as described above.

As supportive analysis, an ANOVA on the Week 12 data will be performed (observed case analysis) for the FAS and PPS with effects for randomized treatment, age group, and region.

As sensitivity analysis of the normality assumption for Week 12 CGI-I-AS data, a Wilcoxon rank sum exact test, with a continuity correction of 0.5, will be used for treatment group comparison of Week 12 CGI-I-AS scores for the FAS and PPS. The 2-sided exact p-value will be presented. The Hodges–Lehmann estimate of the median CGI-I-AS difference between OV101 and placebo groups, and its 95% confidence interval will also be presented.

If more than 3% of CGI-I-AS values at Week 12 are missing, a sensitivity analysis involving multiple imputation will be conducted. Additionally, if more than 5% of CGI-I-AS values at Week 12 are missing, a tipping point analysis will be also performed. Details are specified in the Section 3.8.6.

In addition to the supportive analyses mentioned above, an additional supportive analysis will be conducted in which CGI-I-AS assessments from unscheduled visits and end of treatment (EOT) visits from subjects discontinuing treatment early (before

Week 12) will be mapped using alternative visit window mapping (see Table 3 section 3.1.8).

For this analysis, descriptive statistics will be displayed for CGI-I-AS total improvement score at each visit by treatment group for the FAS. Frequencies and percentages will also be tabulated for the number of subjects with each specific score, the number of subjects with a response of at least minimally improved, and the number of subjects with a response of at least much improved by treatment group and overall.

An MMRM ANOVA methodology similar to the one described above will also be performed after mapping the CGI-I-AS assessments using the alternative visit mapping (Table 3). From the model, the Week 12 difference between least squares means of OV101 and placebo groups will be presented along with the corresponding 95% confidence interval and 2-sided *P*-value.

An additional supportive analysis will be conducted in which CGI-I-AS assessments will be analyzed according to the nominal visit time point as defined in section 3.1.8. Descriptive statistics will be displayed for CGI-I-AS total improvement score at each visit by treatment group for the FAS. An MMRM ANOVA analysis as described above will also be presented.

3.8.2 Key Secondary Efficacy Endpoint Analyses

A generalized logit model will be used for the analysis of key secondary endpoints: CGI-I-AS response (at least minimally improved) and CGI-I-AS response (at least much improved). The generalized logit models will include the response variable at each visit as the dependent variable, and fixed effects for visit (Week 6 and Week 12), randomized treatment, the visit-by-treatment interaction, age group (randomization stratification variable: age 4 to 8 years versus age 9 to 12 years), and region (randomization stratification variable: US versus outside US). An unstructured Cholesky working correlation for the binary outcome variable will be assumed for the model parameter estimation, and if the model fails to converge, the sequence of covariance structures specified in Section 3.8.1 will be used, until convergence is achieved. If there is no response (at least minimally improved or at least much improved) in either treatment arm at the Week 6 visit, the generalized logit model will be substituted with a logistic regression for the Week 12 CGI-I-AS response variable with effects for the randomized treatment and the two randomization stratification variables. From the model (generalized logit or logistic regression, as appropriate), the Week 12 odds ratio will be presented along with the corresponding 95% confidence interval and 2-sided *P* value.

Descriptive statistics for the observed response (either at least minimally improved or at least much improved) will also be displayed along with the 95% CI based on a

binomial distribution. In addition, the differences in observed percentages between treatment groups will be displayed, along with the corresponding 95% CI for differences in percentages and two-sided p-value.

Sensitivity analyses will be conducted based on method of assessment collection (at least one CGI-I-AS assessment conducted remotely or no CGI-I-AS assessment conducted remotely). Descriptive statistics and a subgroup analysis of the generalized logit analysis described above will be performed.

Additionally, supportive analyses will be performed for the descriptive analysis and generalized logit model with the visit mapping described in section 3.1.8 Table 3 and separately for visit mapping using nominal visits.

3.8.3 Other Secondary Efficacy Endpoint Analyses

For each subject, the average of daily values for each [redacted] will be calculated for the following visit windows:

Table 4: Analysis Visit Windows for [redacted]

Assessment	Analysis Visit Window (Interval in Study Days)
Baseline	< Day 1
Week 6	≥ (Week 6 Visit date – 14 days) to ≤ (Week 6 Visit Date + 7 days); or ≥ Day 22 to ≤ (End of Treatment – 1 day), for subjects who early terminate before Week 6
Week 12	≥ (Week 12 Visit date – 14 days) to (Week 12 Visit Date - 1 day); or ≥ Day 65 to ≤ (End of Treatment – 1 day), for subjects who early terminate between Week 6 and Week 12

An average measure will be calculated using all available data in the above analysis windows as follows:

1. Calculate average of all observed weekday values (AVGWKD).
2. Calculate average of all observed weekend values (AVGWKE).
3. Within a visit, if both AVGWKD and AVDWKE are non-missing, then the final average for the visit is AVGWKD*(5/7) + AVGWKE*(2/7).
4. Otherwise, the final average for the visit is set to the non-missing AVGWKD or AVGWKE.

An MMRM Analysis of Covariance (ANCOVA) model for the change from Baseline will be fitted for each [redacted] endpoint defined above with fixed effects for visit (Weeks 6 and 12), randomized treatment, the visit-by-treatment interaction, age group (randomization stratification variable: age 4 to 8 years versus age 9 to 12 years), region (randomization stratification variable: US versus outside US), and the corresponding Baseline as a covariate. An unstructured within-subject covariance structure will be specified as the first choice. If the model converges, the Kenward-

Roger approximation will be used to estimate the denominator degrees of freedom. If the model with unstructured covariance fails to converge, the following covariance structures will be tested in this order: toeplitz with heterogeneity, autoregressive with heterogeneity by visit, compound symmetry with heterogeneous variances by visit, toeplitz, autoregressive, and compound symmetry without heterogeneous variances by visit. The first covariance structure that converges will be used for the analysis. If a structured covariance is used, then a robust sandwich estimator will be utilized for estimating the variance of the treatment effect estimate. From the model, the Week 12 difference between least squares means of OV101 and placebo groups will be presented along with the corresponding 95% confidence interval and 2-sided *P*-value.

Based on the visit average values, if more than 5% but not more than 20% of any [redacted] secondary endpoint are missing at Week 12, a sensitivity analysis will be conducted. Details are specified in Section 3.8.6.

An additional sensitivity analysis may be performed if visit average values at Weeks 6 or 12 are based on less than 3 days of data. In this scenario, only subjects whose visit average values are based on a minimum of 3 days of data (either weekday or weekend) will be included.

The VABS-3 Total Score (Adaptive Behavior Composite Score), domain standard scores (Communication, Socialization, Daily Living Skills, and Motor Skills domains) and their subdomain raw scores, including the Maladaptive Behavior domain subscale scores, will be assessed at Baseline and Week 12 and analyzed as continuous outcome variables. When corresponding subdomain Growth Scale Value (GSV) scores or v-Scale scores are available, the GSV and v-Scale scores will also be analyzed as a sensitivity analysis.

Change from Baseline to Week 12 in VABS-3 standard/raw score will be analyzed using an ANCOVA model with effects for randomized treatment, age group (randomization stratification variable: age 4 to 8 years versus age 9 to 12 years), region (randomization stratification variable: US versus outside US), and the corresponding Baseline as a covariate. A test for a significant treatment-by-baseline interaction will be performed in a separate model (including the above terms plus the interaction term), using a 0.100 alpha level. If the treatment-by-baseline interaction is significant, then the ANCOVA assumptions are violated and an ANOVA model will be used instead. From the final model, the difference between least squares means of OV101 and placebo groups will be presented along with the corresponding 95% confidence intervals and 2-sided *P*-values.

When it is available, change from Baseline to Week 12 in VABS-3 GSV scores will be analyzed using the same method described above. Individual VABS-3 item scores and its shift from baseline at scheduled visits will be summarized in exploratory analysis.

Scatterplots to assess correlation between VABS-3 Total Score with CGI-S-AS overall at each visit (Baseline and Week 12), and change from Baseline to Week 12 in VABS-3 Total Score with change from Baseline to Week 12 CGI-S-AS will also be explored.

The CGI-S-AS assessment will be collected at Baseline, Week 6, and Week 12 and will be treated as a continuous outcome variable. An MMRM ANCOVA methodology similar to the one described previously for [redacted] variables will be used to analyze the change from Baseline in CGI-S-AS score.

The relationships between CGI-S-AS overall score at baseline and CGI-I-AS at Week 12 will be explored through calculation of Pearson correlation coefficient and the associated CI within each treatment group accompanied by the corresponding scatterplot to aid interpretation of the correlation. The relationships between each of CGI-S-AS domain scores at baseline and CGI-I-AS at Week 12 will also be explored using a similar approach.

3.8.5 Subgroups to be Analyzed

Descriptive statistics will be presented separately by age group (randomization stratification variable: age 4 to 8 years versus age 9 to 12 years) for the primary and the two key secondary endpoints. Observed values will be reported by treatment group. The difference in means or proportions for OV101 versus placebo will also be presented with the corresponding 95% confidence intervals. Similar subgroup analysis will be conducted for region (US versus outside US).

If for a given endpoint, the results in the two subgroups appear different from a clinical perspective, the appropriate model (MMRM, generalized logit, or logistic) could be constructed as described above, including an age-by-treatment interaction term. This interaction term will be tested at alpha level of 0.100 to assess whether the treatment effects are statistically significantly different for the two age subgroups.

3.8.6 Missing Data Handling and Sensitivity Analysis

All analyses involving imputation of missing efficacy data will be accompanied with a descriptive table showing the number and percentage of missing endpoint values by visit and treatment group. The reasons of missing, if captured, will be also displayed.

Primary Analysis – CGI-I-AS

The primary endpoint, CGI-I-AS scale assessed by the clinician will be collected at the Week 6 and 12 visits. This is an ordinal variable with scores ranging from 1 (very much improved), 4 (no change), to 7 (very much worse) which will be analyzed as a continuous outcome variable in the analyses.

For the primary analysis of CGI-I-AS, missing data will not be imputed. The primary analysis will be based on all available data at the mapped visits (Table 2) at Week 6

and Week 12 as captured in the clinical database, using a MMRM ANOVA model with fixed effects for visit (Weeks 6 and 12), randomized treatment, the visit-by-treatment interaction, and age group (randomization stratification variable: age 4 to 8 years versus age 9 to 12 years) and region (randomization stratification variable: US versus outside US). MMRM models implicitly adjust for missing data under a missing at random (MAR) assumption.

Sensitivity Analyses Using Multiple Imputation – CGI-I-AS

If more than 3% of CGI-I-AS values at Week 12 are missing, as a sensitivity analysis, missing data will be imputed using the multiple imputation (MI) procedure. The imputation strategy described in Ouyang (2017) will be implemented for this study. There are only two visits where CGI-I-AS is assessed, hence, there are 2 possible missing patterns in the CGI-I-AS data: 1) Week 6 assessment is missing and Week 12 assessment is present (non-monotone pattern); or 2) Week 12 assessment is missing but Week 6 assessment is present (monotone pattern). In general, the MAR assumption is reasonable for intermittent missing values before a subject is discontinued because the data before and after the intermittent missing value are known. In this case, if Week 6 CGI-I-AS is missing but Week 12 CGI-I-AS is assessed (missing data pattern 1), it is reasonable to consider the missing Week 6 data as MAR. However, missing CGI-I-AS at Week 12 after the subject is withdrawn from the trial may or may not be MAR. Multiple imputation will first be implemented under the MAR assumption. To assess the appropriateness of the MAR assumption, additional sensitivity analyses using a tipping point approach will be performed under scenarios assuming missing not at random (MNAR) for the missing data pattern 2 above.

Multiple Imputation Procedure under MAR assumption will be conducted as follows using SAS PROC MI:

1. Missing Week 6 CGI-I-AS data points will be imputed under MAR assumption using a logistic regression model for ordinal data with randomized treatment, age group, region, and CGI-S-AS Overall Score at Baseline as independent variables.

2. Missing Week 12 CGI-I-AS data points will be imputed under MAR assumption using a logistic regression model for ordinal data with randomized treatment, age group, region, CGI-S-AS Overall Score at Baseline, and the Week 6 CGI-I-AS as independent variables.

3. Thirty complete imputed datasets will be generated by repeating the above steps 30 times.

4. The primary MMRM analysis of CGI-I-AS described above will be performed for each of the 30 complete imputed data sets.

5. The estimated treatment effects will be combined across the 30 analyses using SAS PROC MIANALYZE.

If the treatment effect is significant in the analysis using the MI procedure under MAR, with significance level of 0.05 and more than 5% of CGI-I-AS values at Week 12 are missing, a tipping point analysis will be conducted.

The tipping point analysis will focus on the imputed CGI-I-AS values which may be MNAR (missing data pattern 2 above). These analyses will start with the 30 imputed datasets created under a MAR assumption as described above. For each dataset, the imputed values of CGI-I-AS for the OV101 group (with missing pattern 2) will be worsened by adding 1 (unless the current value is the highest level on the CGI-I-AS scale [7=very much worse]). The placebo group imputed values will remain unchanged. After the OV101 group's imputed Week 12 CGI-I-AS data have been "worsened" by 1 point for those with missing pattern 2, the MMRM model used in the primary analysis will be fitted to the data and the estimated treatment effects will be combined using SAS PROC MIANALYZE. This process will be repeated, at each step incrementally worsening by 1 the imputed values of CGI-I-AS for the OV101 group, until the treatment effect is no longer significant (p-value for the treatment comparison ≥0.05).

As a sensitivity analysis, the following rank-based imputation strategy described in Ouyang (2017) will also be implemented when more than 5% of CGI-I-AS values at Week 12 are missing. This analysis will include both subjects with a Week 12 CGI-I-AS and subjects who discontinued prior to the Week 12 assessment, without the need for multiple imputation methodology and the associated assumptions. This approach assumes that discontinuation from the trial is detrimental so it penalizes the treatment group with more early discontinuations. The subjects will be first divided into 2 sets according to their CGI-I-AS assessments: 1) those with a Week 12 CGI-I-AS; and 2) those with a Week 6 CGI-I-AS but no Week 12 CGI-I-AS. Within the two sets, subjects will be ranked separately (based on the Week12 CGI-I-AS for set 1 and the Week 6 CGI-I-AS for set 2). The lowest rank will be given to the largest CGI-I-AS within that set. The Wilcoxon Rank-Sum test will then be applied to this ranking instead of ranking solely based on the Week 12 CGI-I-AS for completers.

[redacted]

As described in Section 3.8.3, MMRM Analysis of Covariance (ANCOVA) models for the change from Baseline will be fitted for each [redacted] endpoint with fixed effects for visit (Weeks 6 and 12), randomized treatment, the visit-by-treatment interaction, age group, region, and the corresponding Baseline as a covariate. Missing data will not be imputed in the main analyses, but MMRM models implicitly adjust for missing data under a missing at random (MAR) assumption.

[redacted]

Based on visit average values, if more than 5% but not more than 20% of [redacted] data are missing at Week 12, as sensitivity analyses for each [redacted] secondary endpoint, missing data will be imputed using a multiple imputation procedure under a MAR assumption as follows:

1. Missing Week 6 [redacted] data points will be imputed under a MAR assumption using a linear regression model with randomized treatment, age group, region, and the corresponding baseline as independent variables.

2. Missing Week 12 [redacted] data points will be imputed under MAR assumption using a linear regression model with randomized treatment, age group, region, and the corresponding baseline and Week 6 [redacted] variable as independent variables.

3. Thirty complete imputed datasets will be generated by repeating the above steps 30 times.

4. The primary MMRM analysis of each [redacted] endpoint as described above will be performed for each of the 30 complete imputed data sets.

5. The estimated treatment effects will be combined across the 30 analyses using SAS PROC MIANALYZE.

If there is more than 20% of data missing for [redacted] endpoints, then baseline characteristics (including Baseline CGI-S-AS) will be compared for the subjects missing the specific [redacted] endpoint versus the subjects who are not missing the specific [redacted] endpoint, for both treatment groups combined. In this analysis, Baseline CGI-S-AS will be summarized as a continuous variable, with the focus on the mean if the normality assumptions are met for Baseline CGI-S-AS, and the focus on the median if Baseline CGI-S-AS is non-normal. Also, if the prevalence of missing is similar across treatment groups, CGI-I-AS at Week 12 will also be summarized by missing status and treatment group. These analyses will inform the interpretation of the descriptive analysis of the [redacted] endpoints.

3.9 Safety Analysis

All safety analyses will be based on the SS. Descriptive statistics will be used to summarize all safety endpoints by treatment group. Data summaries will be displayed for incidence of TEAEs, clinical laboratory variables, vital signs, body weight and body mass index, physical examinations, clinical assessment of suicidality (ABC-I) and seizure diary data.

Unless otherwise specified below, safety data for subjects 2 to 3 years old (inclusive) will only be presented in listings and will be displayed in by-subject graphs, as appropriate.

3.9.1 Adverse Events

All adverse events will be coded using the MedDRA Version 22.0 and will be classified by MedDRA SOC and PT. Analyses of adverse events will be performed using the SS.

AEs that started on or after the first dose of study medication or worsened after the first dose of study medication through 14 days after the last dose of study medication and 30 days after the last dose of study drug/last study visit for SAEs are considered treatment-emergent AEs (TEAEs).

For all summaries, SOCs and PTs within SOC will be presented by decreasing frequency of incidence, sorted first for the OV101 group and then for the placebo group.

Handling of Missing or Partial Start Dates for Adverse Events

Rules for handling incomplete or missing adverse event start dates are addressed below (stop dates will not be imputed).

In the unusual case that the month portion of an adverse event start date is missing but the day portion is not missing, the day portion will be assumed to be missing. Likewise, in the case where the year portion of an adverse event start date is missing but the month and/or day portion is not missing, the month and/or day portion will be assumed to be missing. All missing portion(s) of the date will be handled using the same rules:

- In the event that the day portion (and only the day portion) of the date is missing:
    - If the adverse event started in the same month and year as the first dose date, the adverse event start date will be assumed to be the first dose date.
    - Otherwise, the adverse event start date will be assumed to be the last day of the given month and year, e.g., XX–DEC-2005 would be 31-DEC-2005 where XX represents an unknown day.
- In the event that the day and month portions (and only the day and month portions) of the date are missing:
    - If the adverse event started in the same year as the first dose date, the adverse event start date will be assumed to be the first dose date.
    - Otherwise, the adverse event start date will be assumed to be last month and day of the given year, e.g., XX–XXX-2005 would be 31-DEC-2005 where XX represents an unknown day and XXX represents an unknown month.

- In the event that the day, month, and year portions of the adverse event start date are missing, it will be assumed to be the first dose date.
- If the adverse event start date has been imputed using the rules above, then the adverse event start date must be compared with the adverse event stop date to ensure the logical ordering of dates. If the imputed adverse event start date is after the non-missing adverse event stop date, then the imputed start date will be reset as the stop date.

3.9.1.1 Overall Summary of AEs

An overall summary of TEAEs will be presented by treatment group. The number and percentage of subjects who experienced at least one TEAE, at least one mild TEAE, at least one moderate TEAE, at least one severe TEAE, at least one TEAE leading to dose change/interruption, at least one TEAE leading to study drug discontinuation, at least one TEAE leading to study withdrawal, at least one treatment-related TEAE, at least one serious TEAE, at least one treatment-related serious TEAE, and a TEAE leading to death will be displayed. Treatment-related TEAEs are events with a study drug causality of 'possible', 'probable' or 'definite.'

Incidence of AEs will be summarized separately for subjects 2 to 3 years old (inclusive).

3.9.1.2 AE Incidences

The incidence of TEAEs will be presented by actual age group (4-8 versus 9-12 years old) and treatment group. The number and percentage of subjects who experienced at least one TEAE as well as the number and percentage of subjects who experienced each specific SOC and PT will be presented. For this analysis, if a subject has more than one occurrence of the same PT, then the PT will be counted only once for that subject. Similarly, if a subject has more than one occurrence of the same PT within a SOC, the subject will be counted only once in that SOC.

A summary table by PT only will be presented using the same methods as described above for the by SOC and PT summary table. Additionally, the incidence of TEAE within each specific SOC, HLGT, and PT and specific SOC, HLT, and PT will be presented. The incidence of TEAEs by SOC/PT and maximum severity will also be presented.

For subjects 2 to 3 years old, a summary table by SOC and PT, by PT, and the incidence of TEAEs by PT and maximum severity will be presented.

All TEAEs will be included in a subject listing.

3.9.1.3 Related AEs

The incidence of drug-related TEAEs will be presented by actual age group (4-8 versus 9-12 years old) and treatment group. Related TEAEs will be defined as those with a 'possible', 'probable' or 'definite' relationship to treatment. TEAEs with missing relationship will be counted as related. For this analysis, if a subject has more than one occurrence of the same PT, then the PT will be counted only once for that subject. Similarly, if a subject has more than one occurrence of the same PT within a SOC, the subject will be counted only once in that SOC.

A summary table of drug-related TEAEs will also be presented for subjects 2 to 3 years old. These tables will be repeated for drug-related serious AEs.

3.9.1.4 Deaths, Other Serious AEs and AEs That Led to Treatment Discontinuation

Summaries of deaths, other serious AEs, and AEs that led to treatment discontinuation will be presented by actual age group (4-8 versus 9-12 years old) and treatment group. A summary of serious AEs, and AEs that led to treatment discontinuation will also be presented for subjects 2 to 3 years old. Adverse events that led to study withdrawal will be shown in a listing.

3.9.1.5 AEs That Led to Study Drug Interruption

All adverse events that resulted in study drug interruption will be included in a by-subject listing.

3.9.1.6 AEs of Special Interest

Adverse events of special interest include the following categories:

- Sleep Disorders and Disturbances (HLGT)
- Seizures (Including Subtypes) (HLGT)

All AEs of special interest will be identified by using the above MedDRA PTs during a blinded review of the data.

The incidence of TEAEs of special interest by actual age group (4-8 versus 9-12 years old) and treatment group will be presented separately for each term. The number and percentage of subjects who experienced at least one TEAE of special interest as well as the number and percentage of subjects who experienced each specific SOC and PT will be presented. For this analysis, if a subject has more than one occurrence of the same PT, then the PT will be counted only once for that subject (e.g., subject incidence). Similarly, if a subject has more than one occurrence of the same PT within a SOC, the subject will be counted only once in that SOC.

3.9.2 Clinical Laboratory Evaluation

A list of clinical laboratory tests is given below:

Hematology	- Hematocrit - Hemoglobin - Red blood cell count - Mean corpuscular volume - Mean corpuscular hemoglobin - Mean corpuscular hemoglobin concentration - Reticulocytes - White blood cell count	- Neutrophils (% and count) - Lymphocytes (% and count) - Monocytes (% and count) - Eosinophils (% and count) - Basophils (% and count) - Platelet count - Prothrombin time - Partial thromboplastin time - International normalization ratio
Clinical Chemistry	- Albumin - Alkaline phosphatase - Blood urea nitrogen - Gamma-glutamyl transferase - Calcium - Creatinine - Glucose - Cholesterol, high-density lipoprotein [calculated] - Cholesterol, low-density lipoprotein [calculated] - Cholesterol, homogenous low-density lipoprotein - Triglycerides	- Magnesium - Phosphate - Potassium - Alanine aminotransferase - Aspartate aminotransferase - Lactate dehydrogenase - Sodium - Chloride - Total bilirubin - Direct bilirubin - Total protein - Uric acid - Creatine phosphokinase
Pregnancy Testing[1]	- Serum ß-hCG or urine pregnancy testing for female subjects who have experienced menarche	

[1] Pregnancy testing parameter will be listed only.

Continuous laboratory data will be examined for trends using descriptive statistics of observed values and changes from baseline to each planned post-baseline visit. These data will also be categorized as low, normal or high based on the reference ranges of the central/local laboratory. Frequencies and percentages will be presented for the shifts in these categories from baseline to each post-baseline visit (i.e., low to normal, low to high, high to low, etc.) for all applicable laboratory parameters listed above. The analyses will be performed using the SS and presented by treatment group.

A separate listing for all clinically significant laboratory abnormalities will be provided.

3.9.3 Vital Signs and Other Physical Findings

Vital signs (temperature, heart rate, respiratory rate, and systolic and diastolic blood pressure), weight, and body mass index will be summarized descriptively at Baseline and all post-Baseline study visits by treatment group. Values will be categorized

based on reference ranges by age group as normal, low abnormal, and high abnormal (see section 4.1.8). Shift from Baseline to all post-Baseline study visits for these categories will be summarized descriptively by treatment group.

All vital sign results will be included in a subject listing.

3.9.4 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) will be performed.

Whether or not a physical examination was conducted with date performed will be included in a subject listing.

3.9.5 Assessment of Suicidality

ABC-Irritability (ABC-I) subscale scores (Aman et al., 1985) will be summarized descriptively at Baseline and all post-Baseline study visits by treatment group. Change from Baseline and percent change from Baseline to all post-baseline study visits will also be summarized descriptively by treatment group. The Aberrant Behavior Checklist – Community (ABC-C) Baseline Record Form version 23Apr2018 and the Follow Up Record Form version 20Apr2018 will be implemented.

All ABC-I assessment results will be included in a subject listing.

3.9.6 Seizure Diary Data

The number of subjects reporting a new type of seizure during treatment and the type of new seizures reported during treatment will be summarized. Additionally, the seizure frequency per 28 days (normalized) will be calculated for the following study periods: Baseline (Screening to Day -1), Week 6 (Day 1 to Week 6 visit date or EOT before Day 43), and Week 12 (Week 6 visit date + 1 day to Week 12 visit date or EOT between Day 44 and Day 85). Percent change in seizure frequency per 28 days from Baseline to all post-baseline study visits also will be summarized descriptively by treatment group.

All seizure diary results will be included in a subject listing.

3.11 Interim Analysis

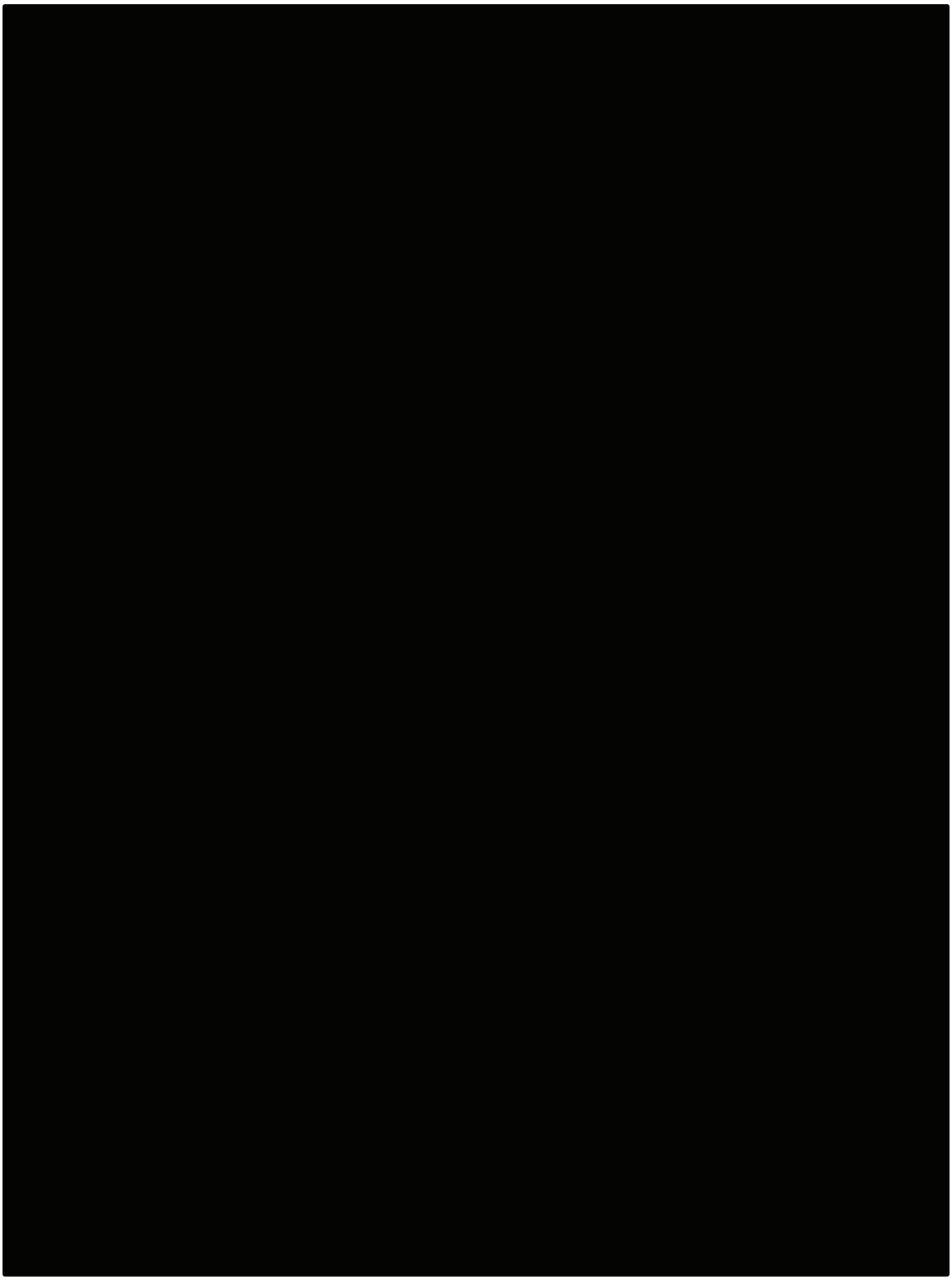
No interim analyses of efficacy data are planned for this study.

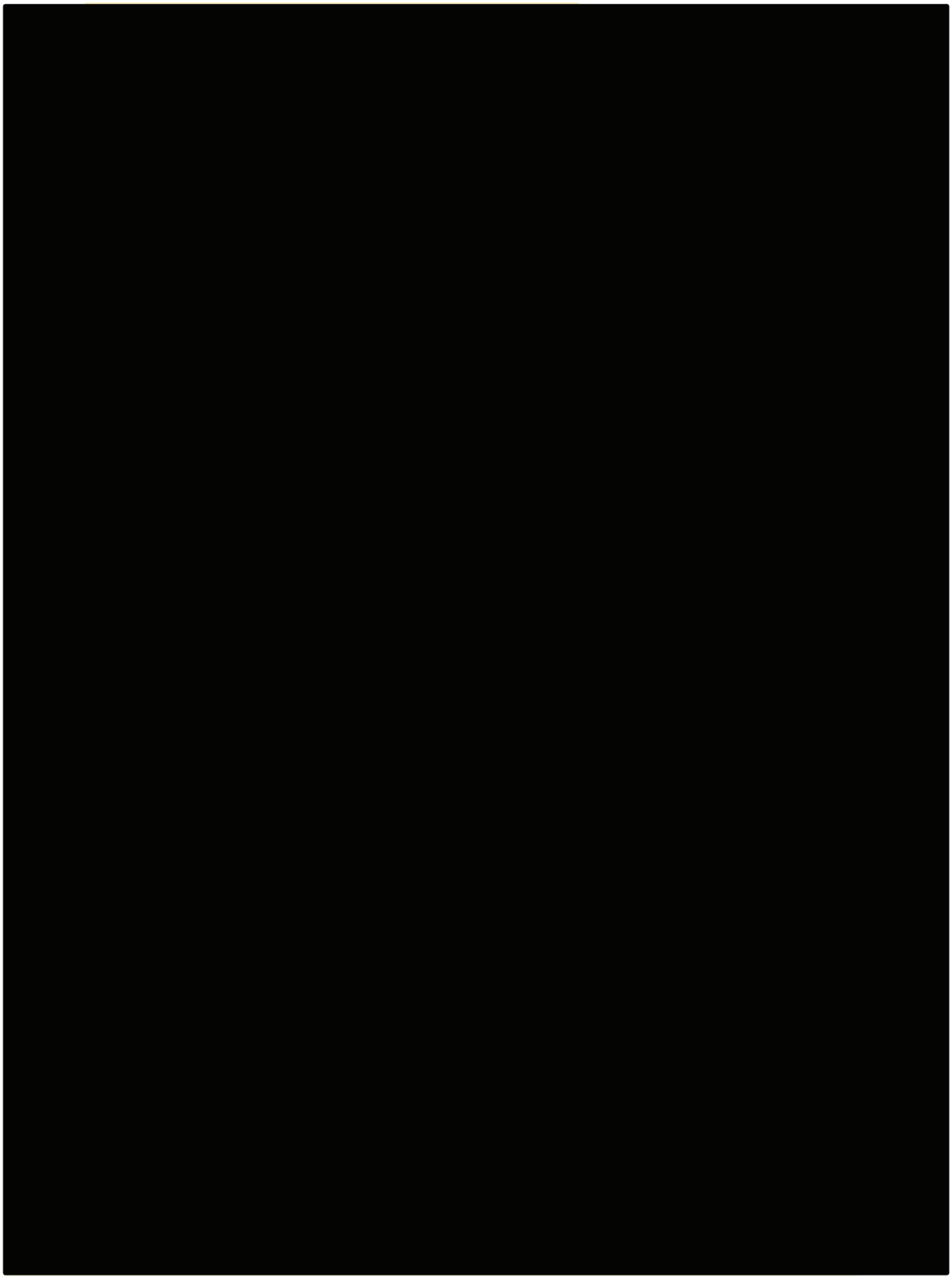
4 DATA HANDLING

4.1 Derived Endpoints and Data Handling

4.1.1 Handling of Repeated Assessments

If multiple results are reported for an assessment at the same post-baseline visit, then the first record will be used in analysis.



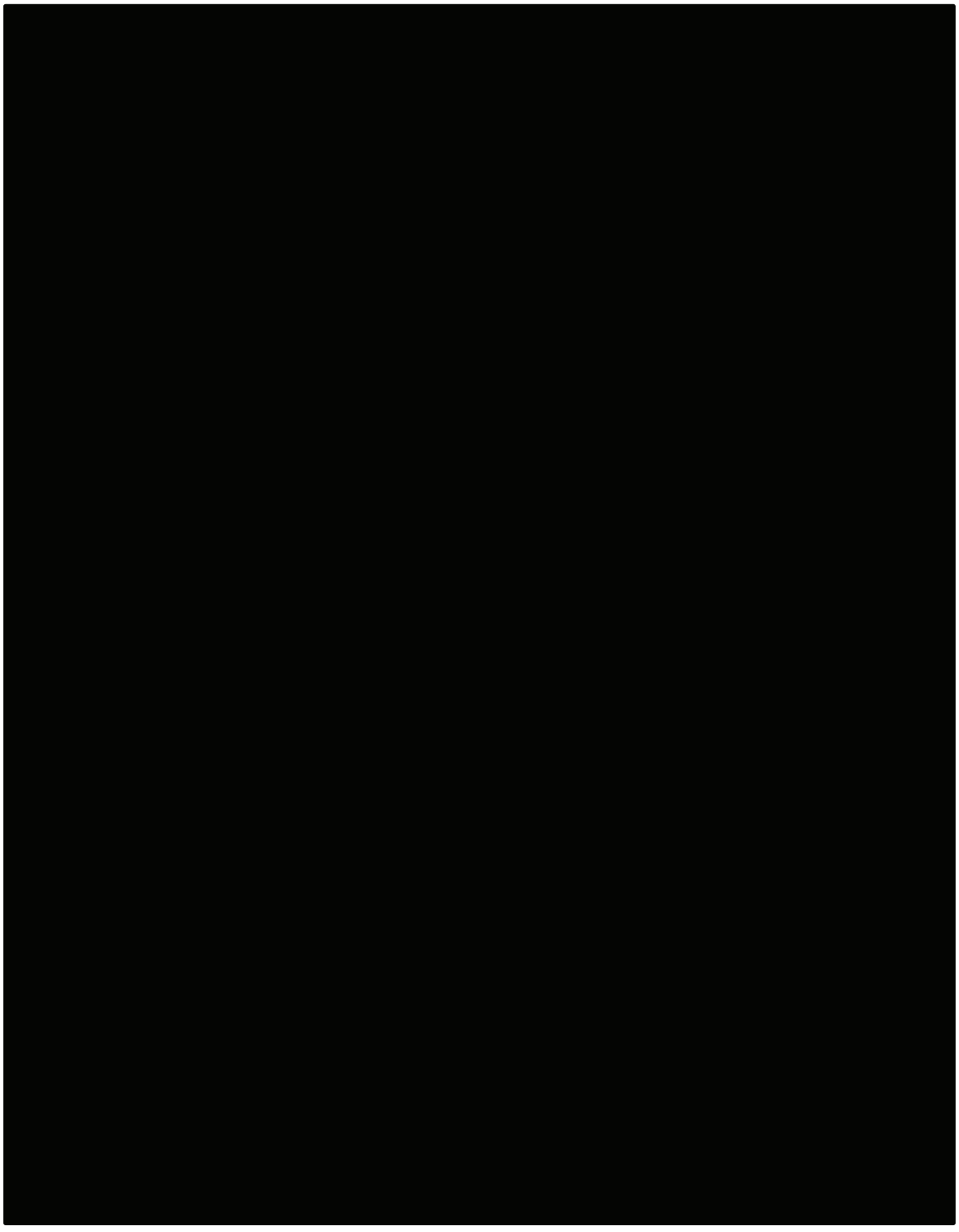


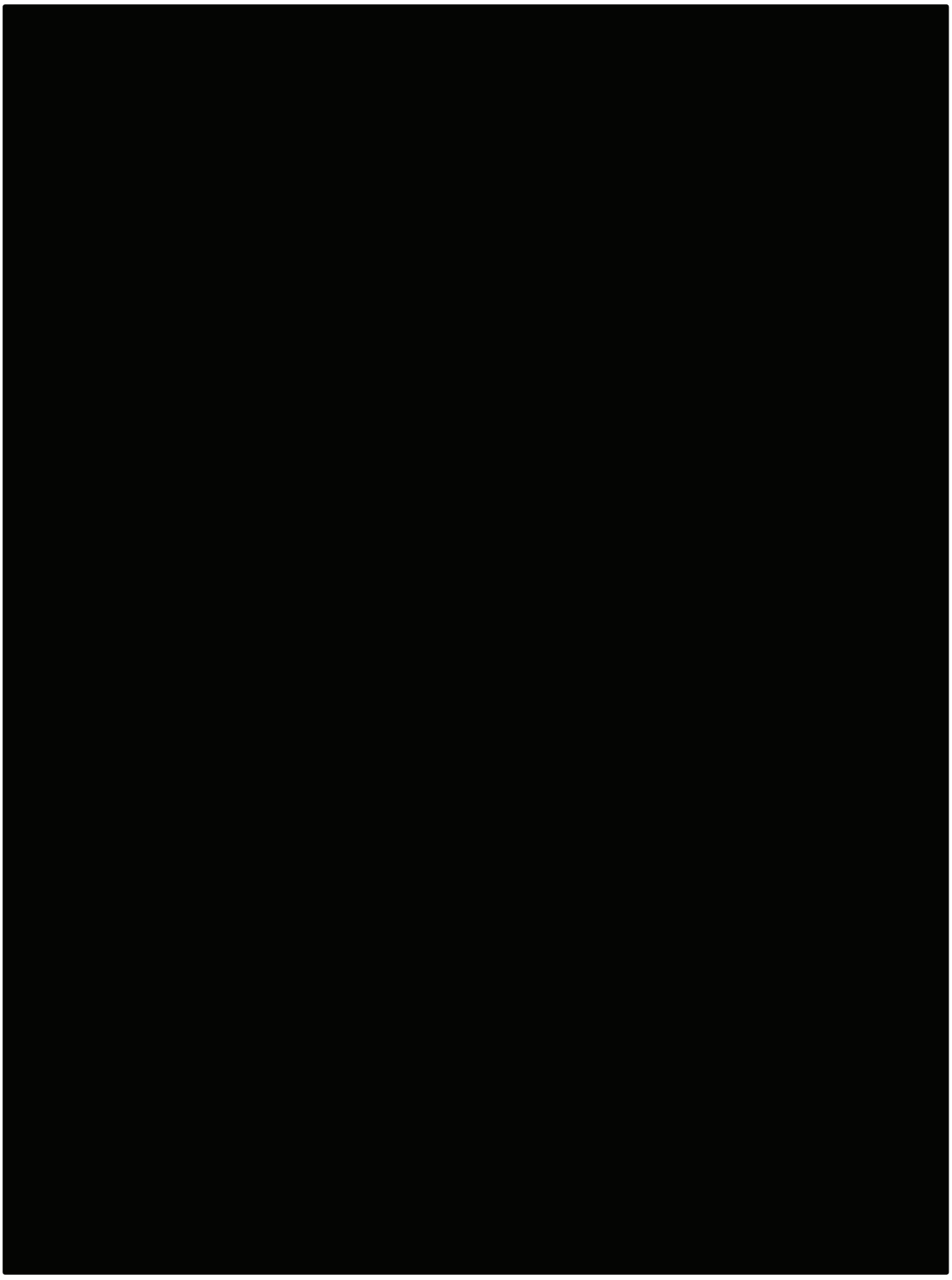
[redacted]

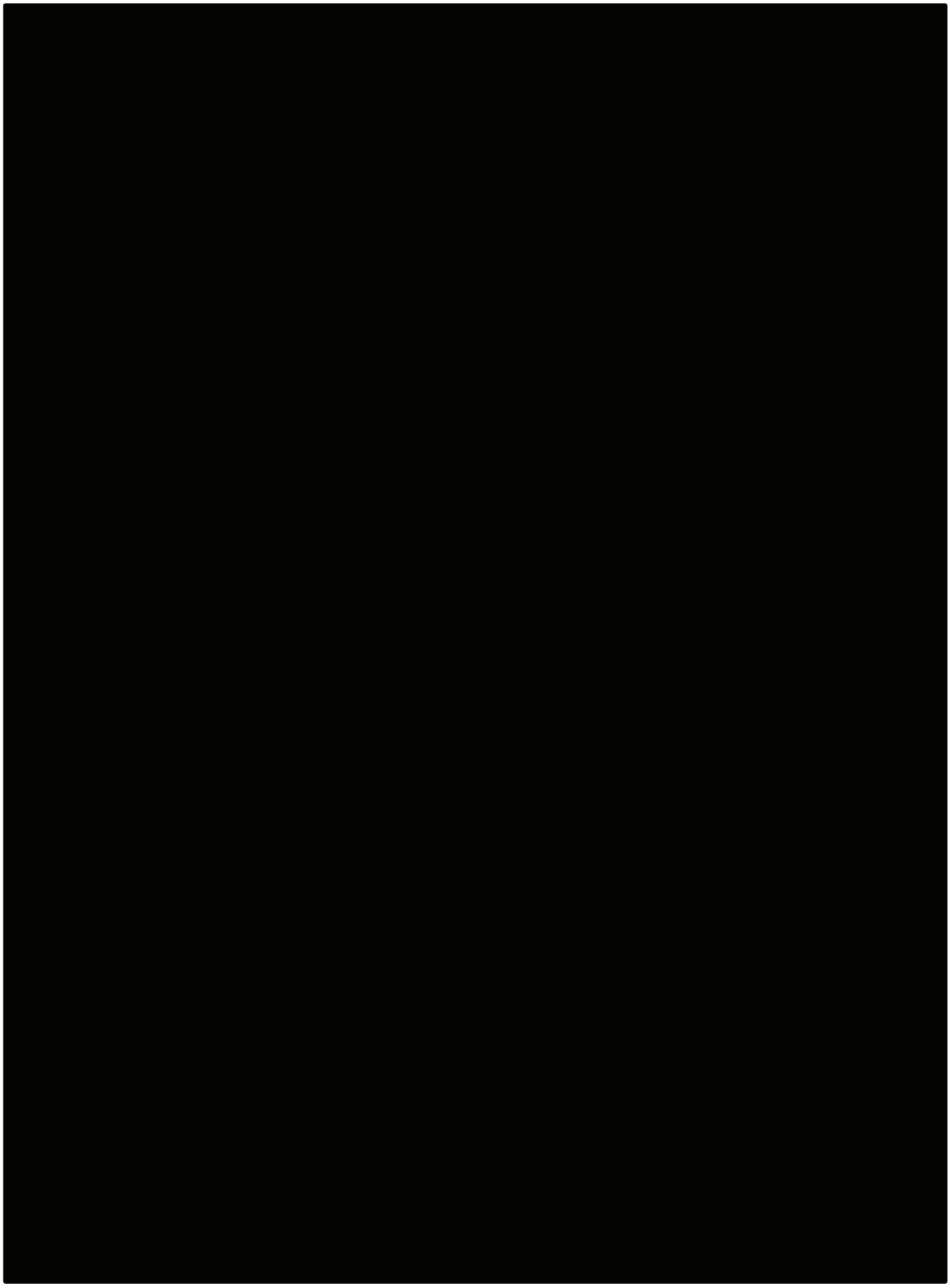
4.1.5 Vineland Adaptive Behavior Scales, 3rd Edition

The VABS-3 is a standardized psychometric instrument designed to measure personal and social skills needed for everyday living (Sparrow et al., 2005; Sparrow et al., 2016). The VABS-3 is administered via a caregiver using a semi-structured interview format and assesses four main adaptive domains (with associated subdomains): communication (receptive, expressive and written), socialization (interpersonal relationships, play and leisure, coping skills), daily living skills (personal, domestic, community), and motor skills (fine and gross). It also includes a maladaptive behavior scale for identifying behavior problems in individuals ages 3 through adulthood. The scale has been found to have good internal consistency, test-retest reliability and validity (Sparrow et al. 2016). The VABS-3 comprehensive interview form will be used to evaluate subjects on the Communication, Socialization, Daily Living Skills, Motor Skills, and Maladaptive Behavior domains to assess their overall adaptive functioning. QGlobal software will be used to calculate the summary scores. The motor skills domain standard score is calculated only through age 9. Higher scores for regular domain and subdomain scores (Communication, Socialization, Daily Living Skills, Motor Skills and 11 subdomains) indicate better adaptive functioning. Higher scores for maladaptive (internalizing and externalizing scores) indicate more problem behaviors. This assessment is an interview with the caregiver by a trained qualified rater, and for each study site, the same rater is to be used throughout the study.

[redacted]







4.1.7 ABC-Irritability Subscale (ABC-I)

The ABC-I will be used to assess suicidality. The entire ABC-C will be administered although only the ABC-I will be used to assess the potential for self-harm.

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

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

For missing data, the following upper limits will apply (number of missing items tolerated for each subscale before discarding the data for that subscale): (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, the subscale will not be computed. If the subscale has the required minimum number of items, the 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.

There is no total score for adding all the subscales together.

4.1.8 Vital Sign Reference Ranges

Reference ranges of vital signs by age group are described in the following table from The Harriet Lane Handbook: A manual for Pediatric House Officers, Twentieth Edition, 2015. A value on the boundary is considered normal; therefore, using the systolic blood pressure example for 2 to 3 years, a value of 105 mmHg would be considered normal, and a value of 106 mmHg abnormal.

Age Group	2 to 3 years	4 to 8 years	9 to 12 years	13+ years
Systolic 95th percentile (mm hg)	Max 105 Min 74	Max 110 Min 78	Max 120 Min 90	Max 129 Min 90
Diastolic 95th percentile (mm hg)	Max 60	Max 70	Max 80	Max 80
Heart rate (beats per min)	80-135	60-130	60-110	60-100
Respiratory Rate (breaths per min)	22-30	18-24	16-22	14-20
ECG (seconds)	PR 0.1-0.14 QRS > 0.07 QTc 0.37-0.44	PR 0.11-0.16 QRS > 0.08 QTc 0.37-0.44	PR 0.12-0.17 QRS > 0.09 QTc 0.37-0.44	PR 0.12-0.20 QRS > 0.10 QTc Males $\leq$ 0.45, Females $\leq$ 0.46

4.2 Computation of Derived Data

4.2.1 Time Points and Duration

Standard calculations used across many assessments:

- Changes from baseline = Post Baseline value – Baseline value
- Percentage change from baseline = [(Post Baseline value – Baseline value)/Baseline Value]*100.

4.2.2 Other Data Derivations

Body Mass Index BMI = Weight in Kg / (Height in meter)^2

5 PROGRAMMING SPECIFICATIONS

5.1 Analysis Specifications

Sample SAS code for the MMRM ANCOVA analysis is as follows:

```
proc mixed data=temp method=reml;
      class subject visit treat agegrp region;
      model endpoint = visit treat agegrp region baseline
      visit*treat ;
      repeated visit / subject=subject type=covstruc;
      lsmeans visit*treat / diff=all cl alpha=0.05;
run;
```

Use the following sequence for choosing a covariance structure, covstruc.

Use the first covariance structure that converges:

SAS Name for Covariance Structure	Covariance Description
UN	Unstructured
TOEPH	Heterogeneous Toeplitz
ARH(1)	Heterogeneous Autoregressive(1)
CSH	Heterogeneous Compound Symmetry
TOEP	Toeplitz
AR(1)	Autoregressive(1)
CS	Compound Symmetry

If the model converges with an unstructured covariance matrix, the Kenward-Roger approximation will be used to estimate the denominator degrees of freedom. If a structured covariance is used for the model, then a robust sandwich estimator will be utilized for estimating the variance of the treatment effect estimate.

For the MMRM ANOVA model, baseline is excluded from the model statement.

Sample SAS code for the generalized logit model analysis is as follows:

```
PROC GLIMMIX DATA = xxx;
   CLASS subject visit treat agegrp region;
   MODEL endpoint = visit treat agegrp region visit*treat /
      DIST= b link=logit;
   random intercept / subject=subject type = covstruc;
   LSMEANS visit*treat / diff exp cl or;
RUN;
```

Use the sequence for choosing a covariance structure, covstruc, described above with the unstructured covariance replaced by an unstructured Cholesky covariance.

Sample SAS code for the logistic analysis is as follows:

```
PROC LOGISTIC DATA = xxx;
   CLASS treat agegrp region;
   MODEL endpoint = treat agegrp region;
   LSMEANS treat / diff cl or;
RUN;
```

5.2 General Specifications

It is suggested that computer-generated table output adhere to the following specifications.

1. Unless otherwise specified, all computer-generated tables and listings should be produced in landscape mode using SAS® ODS to create RTF output which can be imported by Microsoft® Word in table format. All output should have the following two-line header at the upper left margin:

   ```
   Ovid Therapeutics Inc.
   OV101-19-001
   ```

   and the following header at the upper right margin:

   ```
   Page x of y
   ```

2. Each table should be identified by in a sequential numeric order, and the table designation should be centered above the title. A decimal system within the numeric numbering (i.e., x.y and x.y.z) should be used to identify tables, listings and figures with related contents. The title is centered in initial capital characters and should include the analysis set analyzed (e.g. SS). The title and table designation are single-spaced, but are separated from the table by at least a double space.

   ```
                    Table No.

            First Line of Title
   Second Line of Title (if needed)
        Analysis Set Analyzed
   ```

3. Column headings should be in proper case.
4. For variables with numeric values, include “unit” in column heading when appropriate.
5. Footnotes should be single spaced, but separated by at least a double space from the bottom line of the table. The notes are aligned vertically by the left vertical

border of the table. All output should have at least the footnote about the program name and date of the program run.

```
[1] Footnote 1
[2] Footnote 2
[3] Footnote 3

PROGRAM: program file name                                DDMMMYYYY HH:MM
```

6. Unless specified otherwise, all data listings should be sorted by subject number with the study center, and by visit date within subject, where appropriate.
7. For tables that summarize categorical (discrete) data, a Missing category should be added to any parameters for which information is not available for any subject.
8. Unless otherwise specified, the estimated mean, median, first and third quartiles (Q1 and Q3, respectively) for a set of values should be printed out to one more decimal place than the raw (observed) data and rounded appropriately. Standard errors (or standard deviations) should be printed out to two additional decimal places than the raw (observed) data and rounded appropriately. For example, for age (with raw data in whole years):

```
n                    xx
Mean (SD)       xx.x (xx.xx)
Median              xx.x
Q1, Q3           xx.x, xx.x
Min, Max           xx, xx
```

9. The p-values will be printed in the tables rounded appropriately to 4 decimal places and formatted as '0.xxxx'. P-values less than 0.0001 will be formatted in the tables as '<0.0001'.
10. All fractional numeric values should be printed with a zero to the left of the decimal point (e.g., 0.12, 0.3).
11. Unless otherwise specified, percentage values should be printed with one digit to the right of the decimal point (e.g., 12.8, 5.4).
12. Missing descriptive statistics due to non-estimability in tables, as well as missing data in subject listings should be represented as either a hyphen (“-“) with a corresponding footnote (“ - = unknown or not evaluated”), or as “N/A” with the footnote “N/A = not applicable” whichever is appropriate.
13. Dates printed as a result in the table, listing, or graph should be printed in SAS DATE9. format (“DDMONYYYY”: 01 JUL 2002). Missing portions of dates should be represented on subject listings as dashes (-- JUL 1999). Dates that are missing because they are not applicable for the subject should be listed as “N/A”, unless otherwise specified.

6 REFERENCES

Aman MG, Singh NN, Stewart AW, Field CJ. The Aberrant Behavior Checklist: A behavior rating scale for the assessment of treatment effects. American Journal of Mental Deficiency. 1985;89(5):485–491.

ICH E9 Guideline: Statistical Principles for Clinical Trials, 5 February 1998.

ICH-E9 (R1) Guideline: Addendum: Estimands and Sensitivity Analysis in Clinical Trials, 16 June 2017.

Kieser M, Friede T. Simple procedures for blinded sample size adjustment that do not affect the type I error rate. Statistics in Medicine 2003; 22:3571-3581.

Ouyang J, Carroll KJ, Koch G, Li J. Coping with missing data in phase III pivotal registration trials: Tolvaptan in subjects with kidney disease, a case study. Pharmaceutical Statistics. 2017; 16: 250–266.

[redacted]

Sparrow, S. S., D. V. Cicchetti and D. A. Balla (2005). Vineland II: Vineland Adaptive Behavior Scales. Circle Pines, MN, American Guidance Service.

Sparrow, S. S., D. V. Cicchetti and C. A. Saulnier (2016). Vineland Adaptive Behavior Scales, third edition (Vineland-3). Circle Pines, MN, American Guidance Service.

[redacted]