AEVI GENOMIC MEDICINE

PROTOCOL NUMBER: AEVI-001-ADHD-202 (Part A)

A Multicenter, 2-Part, 6-Week, Double-blind, Randomized, Placebo-controlled, Parallel-design Study to Assess the Efficacy and Safety of AEVI-001 in Children and Adolescents (Ages 6-17 Years) with Attention Deficit Hyperactivity Disorder and with or without Copy Number Variants in Specific Genes Implicated in Glutamatergic Signaling and Neuronal Connectivity

STATISTICAL ANALYSIS PLAN

Final Version 2.0
24 September 2018

APPROVAL SIGNATURES

[redacted]

[redacted]
Consultant Statistician
Aevi Genomic Medicine, Inc.
435 Devon Park Drive, Suite 715
Wayne, PA 19087

Date

[redacted]

[redacted]
Chief Scientific Officer
Aevi Genomic Medicine, Inc.
435 Devon Park Drive, Suite 715
Wayne, PA 19087

Date

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ADHD	Attention deficit hyperactivity disorder
ADHD-RS-5	Attention Deficit Hyperactivity Disorder Rating Scale Version 5
AE	Adverse event
ATC	Anatomical Therapeutic Chemical
BMI	Body mass index
C-SSRS	Columbia Suicide Severity Rating Scale
CBCL	Child Behavior Checklist
CGI	Clinical Global Impressions
CGI-I	Clinical Global Impression – Improvement
CGI-S	Clinical Global Impression – Severity
CI	Confidence interval
CNV	Copy number variation
Conners-3-P(S)	Conners 3rd Edition – Parent Short Form
CRO	Contract research organization
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
ECG	Electrocardiogram
eCRF	Electronic case report form
ET	Early termination
FDA	Food and Drug Administration
ICH	International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use
IP	Investigation product
IQ	Intelligence quotient
IRT	Interactive response technology
ITT	Intent-to-treat
KBIT-2	Kaufman Brief Intelligence Test, Second Edition
LOCF	Last observation carried forward
LS	Least squares
MedDRA	Medical Dictionary for Regulatory Affairs
MINI-KID	Mini International Neuropsychiatric Interview for Children and Adolescents
MMRM	Mixed model repeated measures
n	Number of subjects

PP	Per protocol
PR Interval	Interval between the P and R waves on an electrocardiogram tracing
QRS Duration	Width of the QRS complex from beginning to end on an electrocardiogram tracing
QT Interval	Interval between the Q and T waves on an electrocardiogram tracing
QTcB Interval	QT interval corrected for heart rate using the Bazett formula
QTcF Interval	QT interval corrected for heart rate using the Fridericia formula
RR Interval	Interval from the beginning of a QRS complex to the beginning of the next QRS complex
SAP	Statistical analysis plan
SD	Standard deviation
SI	International System of Units
SOC	System organ class
TEAE	Treatment emergent adverse event

1 OVERVIEW

This Statistical Analysis Plan (SAP) describes the planned analysis and reporting for Aevi Genomics Medicine Protocol AEVI-001-ADHD-202 (A Multicenter, 2-Part, 6-Week, Double-blind, Randomized, Placebo-controlled, Parallel-design Study to Assess the Efficacy and Safety of AEVI-001 in Children and Adolescents [Ages 6-17 Years] with Attention Deficit Hyperactivity Disorder and with or without Copy Number Variants in Specific Genes Implicated in Glutamatergic Signaling and Neuronal Connectivity), Final Version 7.0 dated 25 July 2018. This SAP is for Part A of the study only; a separate SAP will be prepared for Part B.

The statistical analyses and summaries described in this SAP will follow the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) guideline for Statistical Principles for Clinical Trials (ICH E9 Expert Working Group, 1999) and will form the basis of the results sections of the clinical study report (CSR) in accordance with the ICH guideline for Structure and Content of Clinical Study Reports (ICH Harmonised Tripartite Guideline E3, 1995). This SAP will be finalized and signed-off before database lock and unblinding of treatment codes. Any changes to the analyses that are not included in this SAP will be documented in the CSR.

2 STUDY DETAILS

2.1 Study objectives

2.1.1 Primary objectives

The primary objective of each part of this study is:

- To evaluate the efficacy of AEVI-001 compared with placebo in children and adolescents (6-17 years of age inclusive) with Attention Deficit Hyperactivity Disorder (ADHD) and with or without copy number variants (CNVs) in specific genes implicated in glutamatergic signaling and neuronal connectivity as measured by the change in ADHD Rating Scale Version 5 (ADHD-RS-5) total score.

2.1.2 Secondary objectives

The key secondary objective of each part of this study will be:

- To evaluate the efficacy of AEVI-001 compared with placebo using a global clinical measure of improvement, the Clinical Global Impression – Improvement (CGI-I) at Visit 8/Early Termination (ET) (Week 6/ET).

Additional secondary objectives of each part of the study include:

- To evaluate AEVI-001 compared with placebo on response at Visit 8/ET (Week 6/ET), based on the following parameters:
  - A $\geq$30% reduction from baseline in ADHD-RS-5 total score.
  - Improved on the CGI-I which includes the categories of 1 (very much improved) and 2 (much improved).
  - Both a $\geq$30% reduction in ADHD-RS-5 total score and improved (1 [very much improved] and 2 [much improved]) on the CGI-I.
- To evaluate AEVI-001 compared with placebo on remission at Visit 8/ET (Week 6/ET), based on the following parameters:
  - An ADHD-RS-5 total score $\leq$18.
  - A CGI-I of 1 (very much improved).
  - Both an ADHD-RS-5 total score $\leq$18 and a CGI-I of 1 (very much improved).
- To evaluate the efficacy of AEVI-001 compared with placebo using a global clinical measure of severity, the Clinical Global Impression – Severity (CGI-S) at Visit 8 (Week 6).

- To evaluate the safety and tolerability of AEVI-001 compared with placebo based on occurrence of treatment-emergent adverse events (TEAEs), vital signs, clinical laboratory results, electrocardiogram (ECG) results, and the Columbia-Suicide Severity Rating Scale (C-SSRS).

2.1.3 Exploratory objectives

Exploratory objectives of each part of the study include:

- To evaluate the efficacy of AEVI-001 compared with placebo on behavior as measured by the change from baseline at Visit 8 (Week 6) on the Conners 3rd Edition–Parent Short Form (Conners 3–P(S)).

2.2 Study design

This is a 2-part, 6-week, double-blind, dose-optimization, parallel-group study in children and adolescents (ages 6-17 years) with ADHD with and without CNVs in specific genes implicated in glutamatergic signaling and neuronal connectivity (see Figure 1 for the study design of both parts). Only subjects who had genetic testing performed during another Aevi study (i.e., MDGN-NFC1-ADHD-001, MDGN-NFC1-ADHD-101, or AEVI-001-ADHD-002) and who consented to future research were considered for participation in this study. Part A will include subjects determined to have one of 8 specific gene mutation(s) implicated in glutamatergic signaling and neuronal connectivity. Part B will be conducted and will assess subjects who do not have CNVs in any of the 272 specific gene mutation(s) implicated in glutamatergic signaling and neuronal connectivity. The list of excluded mutations includes the 8 gene mutations studied in Part A and an additional 264 mutations identified in an antecedent study. Once subjects are confirmed as eligible for each part of the study, they will be randomized to one of two treatment groups (AEVI-001 or placebo) in a 1:1 ratio. The randomization will be stratified by age (6 to 12, 13 to 17 years old) and part (A, B) with each part of the study analyzed separately.

During both parts of the study, randomized subjects will receive investigational product (IP) at the Baseline Visit (Visit 2; Day 0) and will begin taking IP at a dose of 100 mg twice daily on Day 1, once in the morning upon awakening (between 7:00 AM to 9:00 AM) and once in the mid-afternoon (between 3:00 PM to 5:00 PM). Dosing will be optimized weekly ($\pm$2 days) to 100 mg, 200 mg, or 400 mg twice daily, as appropriate, over the first 4 weeks of treatment (Dose-optimization Period), based on clinical response and tolerability. If the subject tolerates a dose well, the dose will be maintained for an additional 2 weeks (Dose Maintenance Period). Visits during the Dose-optimization and Dose Maintenance Periods are to be conducted every 7 days ($\pm$2 days). Visits conducted after randomization should be scheduled relative to the Baseline Visit (Visit 2). The primary assessments of safety and efficacy will be performed weekly during the 6-week treatment period. A follow-up telephone call will be performed 7 days ($+$2 days) after the last dose of IP. The complete Schedule of Events is included in Appendix A.

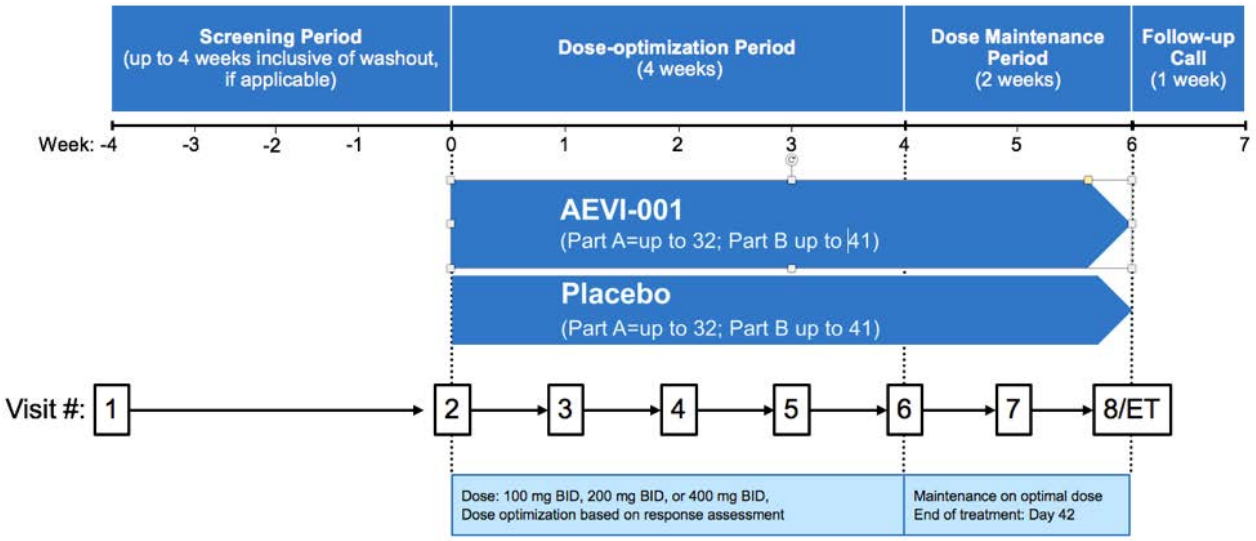
2.3 Determination of sample size

For Part A of the study, 42 unique subjects are planned to be randomized to yield approximately 34 subjects who provide post-baseline efficacy data on the primary endpoint. Assuming an effect size of 1.0, this number will yield approximately 80% power to detect a significant difference in mean change from baseline for the ADHD-RS-5 total score at Visit 8 (Week 6) between the treatment groups, based on a 2-tailed test with $\alpha = 0.05$ and 1:1 randomization. In Part A of the study, an interim analysis will be performed when approximately 75% of randomized subjects have completed the study. The sample size could be increased to 64 randomized subjects.

3 ANALYSIS SETS

An analysis data set will be created for each efficacy and safety assessment. In addition to reported data values, each analysis data set will include derived variables as specified in Section 4 (e.g., baseline, total, and change from baseline values). All analysis data sets will also include key study- and subject-level information (including but not limited to: study identifier, study site identifier, subject identifier, age, sex, race, randomization number, analysis set eligibility flags, randomized treatment, actual treatment received, study treatment start and stop dates, reason for discontinuation [if applicable], and age strata).

Figure 1: Study design (Parts A and B)



All data analyses will be performed using at least one of the following analysis sets. Subject eligibility for each analysis set will be finalized before unblinding of the data, where applicable.

3.1 Intent-to-treat analysis set

The intent-to-treat (ITT) analysis set includes all subjects who are randomized and dispensed study drug at the Baseline Visit (Visit 2). Subjects will be categorized according to randomized treatment. This analysis set will be used for all demographics, baseline characteristics, and prior medication analyses.

3.2 Safety analysis set

The safety analysis set will include all subjects who are randomized and take at least one dose of randomized study drug during this trial. Subjects will be categorized according to actual treatment received. This analysis set will be used for all concomitant medication, study drug exposure, and safety analyses.

3.3 Full analysis set

The full analysis set will include all randomized subjects who took at least one dose of randomized study drug and have a valid baseline ADHD-RS-5 assessment and at least 1 valid post baseline ADHD-RS-5 assessment. Subjects will be categorized according to randomized treatment. This analysis set will be used for all efficacy analyses.

3.4 Per-protocol analysis set

The per-protocol (PP) analysis set is a subset of the full analysis set which will include data from subjects who do not have any significant protocol deviations. Subjects will be categorized according to actual treatment received. This analysis set will be used for sensitivity analyses to examine the robustness of the ITT results for the primary efficacy variable and key secondary efficacy variables.

4 DEFINITION OF STUDY VARIABLES

4.1 Subject disposition

The subject number, previous subject number from the Aevi study where genetic testing was performed, informed consent date, screening date, randomization number, and randomization date will be recorded in the Interactive Response Technology (IRT) system and transferred to the electronic case report form (eCRF).

In addition, screen failure date, reason for screen failure, study completion status, the date of study completion or ET, date of last dose, reason for discontinuation, follow-up completion status, date of follow-up telephone call, and date of last contact will be recorded in the eCRF. The reason for screen failure will be recorded as "Does not meet inclusion/exclusion criteria"; "Withdrawal of consent"; "Lost to follow up"; or "Other". Study/follow-up completion status will be recorded as "Yes" or "No". The reason for discontinuation will be recorded as "Adverse Event"; "Lack of efficacy"; "Major protocol deviation"; "Withdrawal by subject"; "Lost to follow up"; "Other"; or "Identity not confirmed".

4.2 Protocol deviations

The failure to meet any inclusion/exclusion criteria and the specific criteria not met will be recorded in the eCRF.
All subject data will be reviewed for the occurrence of protocol deviations; identified protocol deviations will be captured by the contract research organization (CRO) and categorized as follows:

- Informed consent/assent form
- Inclusion/exclusion criteria
- Concomitant medication
- Study assessments/procedures
- Investigational product
- Compliance
- Overdose, misuse, and abuse
- Medication error
- Visit window
- Identity confirmation
- Other

In addition to the above categories, all deviations will also be classified as "Major" or "Minor". Major deviations include any deviation from the protocol that could potentially affect the scientific soundness of the research plan or the rights, safety, or welfare of human subjects. The following deviations will be considered "Major", but are not considered all-inclusive as others may be added during the study.

- Failure to obtain written informed consent and assent (if required) from the parent or legally authorized representative and subject, respectively.
- Randomization of a subject who did not meet inclusion criteria.
- Randomization of a subject who met exclusion criteria.
- Use of a prohibited medication on $\geq$3 consecutive days during the double-blind treatment period (Visit 2 through Visit 8/ET).
- Protocol defined screening/baseline evaluations needed to assess eligibility were not performed or required results were not obtained prior to randomization.

- Failure to confirm the subject's identity prior to randomization.
- Medication error related to IP dispensation (e.g., incorrect dose, incorrect blister card, etc.).
- Overall noncompliance with IP dosing defined as $<80\%$ or $>120\%$.
- Subject overdoses, abuses, and/or misuses the study drug.
- Site failure to follow steps/guidelines/procedures established in the safety monitoring plan such that the safety of the subject or other subjects in the study is potentially endangered.

The Sponsor project team will review all protocol deviations and their major/minor classifications prior to database lock. Subjects with confirmed "Major" protocol deviations will be excluded from the PP analysis set (see Section 3.4).

4.3 Demographics and other baseline characteristics

4.3.1 Age

The date of consent and date of birth will be recorded in the IRT system. In addition, age at informed consent in years and age stratum will be derived within the IRT system. Age stratum will be derived at randomization based on the subject's age at informed consent. Table 1 shows the stratum codes and descriptions assigned to subjects in Part A. Prior to analysis, all age-related data recorded or derived in the IRT system will be transferred into the eCRF.

Table 1: Age stratum codes and descriptions for Part A

Stratum code	Stratum description
1	Age at informed consent 6 to 12 years old, Part A
2	Age at informed consent 13 to 17 years old, Part A

4.3.2 Sex

Sex will be recorded in the IRT system and transferred into the eCRF as "Male" or "Female".

4.3.3 Ethnicity

Ethnicity will be recorded in the eCRF as "Hispanic or Latino"; "Not Hispanic or Latino"; "Not Reported"; or "Unknown".

4.3.4 Race

Race will be recorded in the eCRF as "American Indian or Alaska Native"; "Asian"; "Black or African American"; "Native Hawaiian or Other Pacific Islander"; "White"; or "Other".

4.3.5 Height

Height (in) will be recorded in the eCRF.

4.3.6 Weight

Weight (lb) will be recorded in the eCRF.

4.3.7 Body mass index

Body mass index (BMI; kg/m^2) will be calculated as weight (lb) divided by height squared (in^2) and multiplied by 703.

4.3.8 Copy number variant type

Saliva samples will have been processed for identification of genetic mutations; these samples could have been collected during participation in this study or another AEVI-001 study (i.e., the MDGN-NFC1-ADHD-001, MDGN-NFC1-ADHD-101, AEVI-001-ADHD-002 studies). Only those subjects determined to have a copy number variant (CNV) in one of the 8 genes implicated in glutamatergic signaling and neuronal connectivity will be eligible for Part A of this study. The presence of individual CNVs of interest will be determined using a microarray based technology from Illumnia (Illumnia, San Diego) and will be recorded in a separate CNV identification dataset.

An indicator for the CNTN4 CNV will be calculated as:

- IF the subject has at least one CNV type of "CNTN4", THEN set the CNTN4 CNV indicator equal to 1.
- OTHERWISE set the CNTN4 CNV indicator equal to 0.

4.3.9 Kaufman Brief Intelligence Test, Second Edition

The Kaufman Brief Intelligence Test, Second Edition (KBIT-2; Kaufman & Kaufman, 2004) is designed to measure both the verbal and nonverbal intelligence in children and adults (ages 4 years, 0 months through 90 years, 11 months) using a brief, individual format. The KBIT-2 includes 3 subtests; 2 classified as verbal and 1 as nonverbal. Administration of the KBIT-2 yields a Verbal score, Nonverbal score, and intelligence quotient (IQ) Composite score.

The KBIT-2 Verbal, Nonverbal, and IQ Composite standard scores will be recorded in the eCRF.

4.3.10 Child Behavior Checklist

The Child Behavior Checklist (CBCL; Achenbach & Rescorla, 1996) was designed to address the problem of defining child behavior problems. The checklist assesses the behavioral problems and social competencies of children as reported by parents using a standardized format. The CBCL can be self-administered or administered by an interviewer and consists of items related to behavior problems. There are social competency items used to obtain parent reports of the amount and quality of their child's participation in specific tasks and activities. In addition, there are 120 syndrome scale items. Only 118 of these syndrome scale items will be assessed in this study; items 56H and 113 will not be assessed. Each syndrome scale item is scored on a 3-point scale ranging from 0 = "not true (as far as I know)" to 2 = "very true or often true".

Each CBCL competence scale response and syndrome scale item score will be recorded in the eCRF. Various CBCL syndrome scale and problem scores will be calculated as described in Achenbach & Rescorla (1996). For each calculated syndrome scale and problem score, standard scores will also be derived according to the subject's age (6 to 10 years, 11 to 18 years) and sex using normative samples of CBCL scores. All CBCL standard scores are scaled so that a score of 50 is average (with a standard deviation of 10 points) for the subject's respective age and sex categories. Standard scores will be calculated using the Group 2 scoring spreadsheet (i.e., CBCGroup2-ProbScales.xls). Higher scores indicate greater problems.

4.3.10.1 CBCL syndrome scale scores

Eight empirically-based CBCL syndrome scale scores will be calculated as follows:

- Anxious/Depressed syndrome score = sum of items 14, 29, 30, 31, 32, 33, 35, 45, 50, 52, 71, 91, and 112
- Withdrawn/Depressed syndrome score = sum of items 5, 42, 65, 69, 75, 102, 103, and 111
- Somatic Complaints syndrome score = sum of items 47, 49, 51, 54, 56a, 56b, 56c, 56d, 56e, 56f, and 56g
- Social Problems syndrome score = sum of items 11, 12, 25, 27, 34, 36, 38, 48, 62, 64, and 79
- Thought Problems syndrome score = sum of items 9, 18, 40, 46, 58, 59, 60, 66, 70, 76, 83, 84, 85, 92, and 100
- Attention Problems syndrome score = sum of items 1, 4, 8, 10, 13, 17, 41, 61, 78, and 80

- Rule-Breaking Behavior syndrome score = sum of items 2, 26, 28, 39, 43, 63, 67, 72, 73, 81, 82, 90, 96, 99, 101, 105, and 106
- Aggressive Behavior syndrome score = sum of items 3, 16, 19, 20, 21, 22, 23, 37, 57, 68, 86, 87, 88, 89, 94, 95, 97, and 104

Standard scores will be derived for each syndrome scale (as described in Section 4.3.10). In addition, CBCL syndrome scale clinical range categories will be calculated for each syndrome scale as:

- IF the syndrome scale standard score is <65, THEN set the clinical range category equal to "Normal".
- IF the syndrome scale standard score is ≥ 65 and ≤ 69, THEN set the clinical range category equal to "Borderline".
- IF the syndrome scale standard score is ≥ 70, THEN set the clinical range category equal to "Clinical".

According to the CBCL scoring manual (see Achenbach & Rescorla, 1996, p. 90), these clinical range categories correspond to the following cutoff criteria: scores below the 93rd percentile are considered "Normal", scores between the 93rd to 97th percentile are "Borderline" clinical, and any scores above the 97th percentile are in the "Clinical" range.

4.3.10.2 CBCL problem scores

In addition to syndrome scores, CBCL problem scores will be derived. The Internalizing Problems score will be calculated as the sum of the Anxious/Depressed, Withdrawn/Depressed, and Somatic Complaints syndrome scores; the Externalizing Problems score will be calculated as the sum of the Rule-Breaking and Aggressive Behavior syndrome scores; the Total Problems score will be derived as the sum of all syndrome scale item scores.

Standard scores will be derived for each problem score (as described in Section 4.3.10). In addition, CBCL problem score clinical range categories will be calculated for each problem score as:

- IF the problem standard score is <60, THEN set the clinical range category equal to "Normal".
- IF the problem standard score is ≥ 60 and ≤ 63, THEN set the clinical range category equal to "Borderline".
- IF the problem standard score is ≥ 64, THEN set the clinical range category equal to "Clinical".

According to the CBCL scoring manual (see Achenbach & Rescorla, 1996, p. 96), the "Borderline" clinical range category is defined as scores between the 84th to 90th percentiles, inclusive.

4.4 Baseline disease characteristics

4.4.1 Mini International Neuropsychiatric Interview for Children and Adolescents

The date of the Mini International Neuropsychiatric Interview for Children and Adolescents (MINI-KID) assessment and those diagnostic criteria met will be recorded in the eCRF. Only the ADHD subtype (past 6 months) and the medical, organic, drug cause ruled out criteria are required; all other MINI-KID criteria will be recorded only if present. The ADHD (past 6 months) criteria will be recorded as "Combined Presentation", "Predominantly Inattentive Presentation", or "Predominantly Hyperactive Type Presentation". The medical, organic, drug cause ruled out criteria will be recorded as "Yes", "No", or "Uncertain".

4.4.2 Age at ADHD diagnosis/Time since ADHD diagnosis

Date of ADHD diagnosis will be recorded in the eCRF. Age at ADHD diagnosis will be calculated in years as the date of ADHD diagnosis minus the date of birth. Time since ADHD diagnosis will be calculated in years as the date of consent minus the date of ADHD diagnosis.

4.4.3 Age at ADHD Onset

Age at ADHD Onset in years will be recorded in the eCRF.

4.5 Prior and concomitant medications

For each prior or concomitant medication, the medication name, indication, dose, frequency, route, start date, ongoing indicator, and stop date (for those medications not ongoing) will be recorded in the eCRF. Each medication will be coded using the World Health Organization Drug Dictionary (WHODrug) Enhanced version September 2016.

All medications will be classified according to the study period in which its use occurred (i.e., Prior, Concomitant, or Post-treatment). Prior medications are those that started and ended before the initiation of study drug. Concomitant medications are (i) those that started before initiation of study drug and continued after initiation of study drug or (ii) those that started after the initiation of study drug; any medications that started on the same day as the initiation of study drug will be considered concomitant. Post-treatment medications are those that started after completion of study drug.

If the start date (or end date) of a medication is completely unknown (i.e., the day, month, and year are all missing) or only the day is known, then the start date (or end date) will not be imputed. Unless the end date is before the start date of study drug, the medication will be considered concomitant.

For a partial start date of medication, the following conventions will be used for imputing the start date:

- If the year is present and the month and day are missing, then the month and day will be set to January 1.
- If the year and day are present and the month is missing, then the month will be set to January.
- If the year and month are present and the day is missing, then the day will be set to the first day of the month.
- If the imputed start date of medication is after the non-imputed end date of medication, then the start date will be set to the end date of medication.

For a partial end date of medication, the following conventions will be used for imputing the end date:

- If the year is present and the month and day are missing, then the month and day will be set to December 31.
- If the year and day are present and the month is missing, then the month will be set to December.
- If the year and month are present and the day is missing, then the day will be set to the last day of the month.

4.6 Study drug exposure variables

Drug accountability will be performed at each visit during the treatment period. At each visit that study drug is dispensed to subjects, the dose level, investigator dose assessment, quantity dispensed, intended first dose date, and intended first dose AM/PM will be recorded in the IRT system. At each visit that study drug is returned, the quantity remaining, intended last dose date, intended last dose AM/PM, percent compliance, and an overdose flag will be recorded in the IRT system. The percent compliance for the prior dosing period will be derived as the number of capsules taken divided by the expected number of capsules taken. For percent compliance calculations, capsules not returned will be assumed taken.

- Dose level will be recorded as “Level 1 (100mg or placebo bid)”, “Level 2 (200mg or placebo bid)”, or “Level 3 (400mg or placebo bid)”.
- Intended first and last dose AM/PM values will be recorded as “AM Only”, “PM Only”, or “AM and PM”.
- Investigator dose assessment will be recorded as “Increase the dose”, “Maintain the Dose”, or “Decrease the dose”.
- Overdose flag will be recorded as “Yes” or “No”.

Prior to analysis, all study drug exposure data recorded or derived in the IRT system will be transferred into the eCRF.

Optimized dose will be calculated as the dose level dispensed at Week 4 (Visit 6). Total study drug exposure (in mg) will be calculated as the sum of all administered doses not returned. Duration of total exposure (in days) is calculated as date of last dose − date of first dose + 1.

Overall percent compliance for the entire treatment period will also be derived. To calculate overall percent compliance, a weight for each weekly percent compliance is required; this weight value will be calculated as the number of doses expected during that week divided by the total number of doses expected during the subject's participation in the study. The total number of doses expected will be calculated as:

- IF the latest intended last dose AM/PM value is "AM Only", THEN the total number of doses expected is equal to (the latest intended last dose date − the earliest intended first dose date + 0.5) × 2
- IF the latest intended last dose AM/PM value is "AM and PM", THEN the total number of doses expected is equal to (the latest intended last dose date − the earliest intended first dose date + 1) × 2

Overall percent compliance will be calculated as:

$$\sum_{i=1}^{6} (\text{Percent Compliance})_i \times (\text{Weight})_i, \text{ where } i \text{ is the study week of dosing.}$$

In addition, overall compliance categories will be calculated as:

- $<80\%$
- $\geq 80\%$ and $\leq 120\%$
- $>120\%$

4.7 Efficacy variables

For each efficacy assessment, study day will be calculated as: the date of assessment − the date of first dose of study drug + 1. As deviations are expected in the number of days from the date of first dose (i.e., study day 1) to the study day that planned assessments actually occur, visit windows will be used to derive visit numbers for use in efficacy analyses. Efficacy variable visit windows and their associated derived visit numbers and labels are defined in Table 2.

Table 2: Visit windows for efficacy variables

Derived visit number	Derived visit label	Visit window (days)
1	Visit 1	≤ -1
2	Visit 2 (Week 0)	0
3	Visit 3 (Week 1)	1 to 11
4	Visit 4 (Week 2)	12 to 18
5	Visit 5 (Week 3)	19 to 25
6	Visit 6 (Week 4)	26 to 32
7	Visit 7 (Week 5)	33 to 39
8	Visit 8 (Week 6)	40 to 50

For all efficacy variables, baseline will be defined as the last assessment prior to the first dose of IP. Change from baseline values will be calculated as the assessment value minus the baseline value. For the derivation of any total or subscale score, if at least 80% of the items needed to calculate a score have valid responses, then that score will be computed as the sum of all non-missing items divided by the percentage of non-missing items in that scale or subscale. If fewer than 80% of the items have valid responses, then that score will be considered as missing.

For some efficacy variables, the last observation carried forward (LOCF) method of imputation will be used to calculate additional LOCF efficacy variables. For LOCF variables, a derived visit number of 99 and a visit label of "LOCF" will be assigned.

4.7.1 Attention Deficit Hyperactivity Disorder Rating Scale Version 5

The Attention Deficit Hyperactivity Disorder Rating Scale Version 5 (ADHD-RS-5; DuPaul et al., 2016) is a parent reported/clinician-rated scale designed to reflect the symptomatology of ADHD based on the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) criteria. The ADHD-RS-5 is comprised of 18 frequency items and 12 impairment items. Each frequency item will be scored on a scale from 0 = "Never or rarely" to 3 = "Very often". Each impairment item will be scored on a scale from 0 = "No Problem" to 3 = "Severe Problem". Each ADHD-RS-5 item score will be recorded in the eCRF.

4.7.1.1 ADHD-RS-5 total score

The ADHD-RS-5 total score will be calculated as the sum of the 18 frequency item scores. The total score ranges from 0 to 54. Higher scores indicate greater symptom severity.

The change from baseline in the ADHD-RS-5 total score is the primary efficacy variable and will be calculated at each post-baseline timepoint. In addition, the percent reduction in ADHD-RS-5 total score will be calculated at each post-baseline timepoint as the change from baseline value divided by the baseline value.

4.7.1.2 ADHD-RS-5 response

An indicator of ADHD-RS-5 response will be calculated at each post-baseline timepoint as:

- IF the percent reduction in ADHD-RS-5 total score is $\geq$30%, THEN set the ADHD-RS-5 response indicator equal to 1.
- IF the percent reduction in ADHD-RS-5 total score is $<$ 30%, THEN set the ADHD-RS-5 response indicator equal to 0.

In addition to deriving ADHD-RS-5 response at each timepoint, an LOCF ADHD-RS-5 response indicator will be calculated.

4.7.1.3 ADHD-RS-5 remission

An indicator of ADHD-RS-5 remission will be calculated at each post-baseline timepoint as:

- IF the ADHD-RS-5 total score is $\leq$18, THEN set the ADHD-RS-5 remission indicator equal to 1.
- IF the ADHD-RS-5 total score is $>$18, THEN set the ADHD-RS-5 remission indicator equal to 0.

In addition to deriving ADHD-RS-5 remission at each timepoint, an LOCF ADHD-RS-5 remission indicator will be calculated.

4.7.1.4 ADHD-RS-5 Inattention subscale score

The ADHD-RS-5 Inattention subscale score will be calculated as the sum of the first 9 individual item scores. The Inattention subscale score ranges from 0 to 27. Higher scores indicate greater symptom severity. The change from baseline in ADHD-RS-5 Inattention subscale score will be calculated at each post-baseline timepoint.

4.7.1.5 ADHD-RS-5 Hyperactivity–Impulsivity subscale score

The ADHD-RS-5 Hyperactivity–Impulsivity subscale score will be calculated as the sum of the next 9 individual item scores (i.e., from items 10 to 18). The Hyperactivity–Impulsivity subscale score ranges from 0 to 27. Higher scores indicate greater symptom severity. The change from baseline in ADHD-RS-5 Hyperactivity–Impulsivity subscale score will be calculated at each post-baseline timepoint.

4.7.2 Clinical Global Impression

The Clinical Global Impression (CGI; Guy, 1976) scale is a clinician-rated scale consisting of 3 items. Only the CGI – Severity (CGI-S) and CGI – Improvement (CGI-I) items will be assessed in this study. The CGI-S item is rated on a 7-point scale from 1 = "Normal, not at all ill" to 7 = "Among the most severely ill patients". The CGI-I item is rated

on a 7-point scale from 1 = "Very much improved" to 7 = "Very much worse". Each CGI item score will be recorded in the eCRF.

A CGI-S or CGI-I score of "0" will be considered as missing as this score denotes "Not assessed".

4.7.2.1 CGI-I response

An indicator of CGI-I response will be calculated as:

- IF the CGI-I score is $\leq$2 (i.e., "very much improved" or "much improved"), THEN set the CGI-I response indicator equal to 1.
- IF the CGI-I score is $>$2, THEN set the ADHD-RS-5 response indicator equal to 0.

In addition to deriving CGI-I response at each timepoint, an LOCF CGI-I response indicator will be calculated.

4.7.2.2 CGI-I remission

An indicator of CGI-I remission will be calculated as:

- IF the CGI-I score is equal to 1 (i.e., "very much improved"), THEN set the CGI-I response indicator equal to 1.
- IF the CGI-I score is $>$1, THEN set the CGI-I response indicator equal to 0.

In addition to deriving CGI-I remission at each timepoint, an LOCF CGI-I remission indicator will be calculated.

4.7.2.3 CGI-S score

The change from baseline in CGI-S will be calculated at each post-baseline timepoint.

4.7.3 Combined ADHD-RS-5/CGI-I response and remission

4.7.3.1 Combined ADHD-RS-5/CGI-I response

An indicator of combined ADHD-RS-5/CGI-I response will be calculated as:

- IF the percent reduction in ADHD-RS-5 total score is $\geq$30% AND the CGI-I score is $\leq$2 (i.e., "very much improved" or "much improved"), THEN set the combined ADHD-RS-5/CGI-I response indicator equal to 1.
- OTHERWISE, set the ADHD-RS-5/CGI-I response indicator equal to 0.

In addition to deriving combined response at each timepoint, an LOCF combined ADHD-RS-5/CGI-I response indicator will be calculated.

4.7.3.2 Combined ADHD-RS-5/CGI-I remission

An indicator of combined ADHD-RS-5/CGI-I remission will be calculated as:

- IF the ADHD-RS-5 total score is $\leq$18 AND the CGI-I score is equal to 1 (i.e., "very much improved"), THEN set the combined ADHD-RS-5/CGI-I remission indicator equal to 1.
- OTHERWISE, set the combined ADHD-RS-5/CGI-I remission indicator equal to 0.

In addition to deriving combined remission at each timepoint, an LOCF combined ADHD-RS-5/CGI-I remission indicator will be calculated.

4.7.4 Conners 3rd Edition – Parent Short Form

The Conners 3rd Edition – Parent Short Form [Conners 3-P(S); Conners, 2008] is a parent completed assessment consisting of 45 items to assess the severity of ADHD and other common co-morbid disorders in children and adolescents aged 6 to 18 years. Only the 43 multiple choice items will be assessed in this study (i.e., the final two open ended questions will not be assessed). The Conners 3-P(S) consists of 6 content subscales: Inattention, Hyperactivity/Impulsivity, Learning Problems, Executive Functioning, Aggression, and Peer Relations. Each item from these content subscales will be scored on a scale from 0 = "Not true at all (Never, Seldom)" to 3 = "Very much true (Very often, Very Frequently)". Additional Positive Impression and Negative Impression items will be scored on a scale from 0 to 1. Each Conners 3-P(S) item score will be recorded in the eCRF.

4.7.4.1 Conners 3-P(S) Inattention subscale score

The Conners 3-P(S) Inattention subscale score will be calculated as the sum of the item 17, item 27, item 30, item 34, and item 41 scores. The Inattention subscale score ranges from 0 to 15. Higher scores indicate greater symptom severity. The change from baseline score will be calculated at each applicable post-baseline timepoint.

4.7.4.2 Conners 3-P(S) Hyperactivity/Impulsivity subscale score

The Conners 3-P(S) Hyperactivity/Impulsivity subscale score will be calculated as the sum of the item 3, item 5, item 7, item 13, item 24, and item 28 scores. The Hyperactivity/Impulsivity subscale score ranges from 0 to 18. Higher scores indicate greater symptom severity. The change from baseline score will be calculated at each applicable post-baseline timepoint.

4.7.4.3 Conners 3-P(S) Learning Problems subscale score

The Conners 3-P(S) Learning Problems subscale score will be calculated as the sum of the item 8, item 10, item 25, item 36, and item 39 scores. The Learning Problems subscale score ranges from 0 to 15. Higher scores indicate greater symptom severity. The change from baseline score will be calculated at each applicable post-baseline timepoint.

4.7.4.4 Conners 3-P(S) Executive Functioning subscale score

The Conners 3-P(S) Executive Functioning subscale score will be calculated as the sum of the item 1, item 15, item 20, item 32, and item 35 scores. The Executive Functioning subscale score ranges from 0 to 15. Higher scores indicate greater symptom severity. The change from baseline score will be calculated at each applicable post-baseline timepoint.

4.7.4.5 Conners 3-P(S) Aggression subscale score

The Conners 3-P(S) Aggression subscale score will be calculated as the sum of the item 14, item 19, item 21, item 23, and item 26 scores. The Aggression subscale score ranges from 0 to 15. Higher scores indicate greater symptom severity. The change from baseline score will be calculated at each applicable post-baseline timepoint.

4.7.4.6 Conners 3-P(S) Peer Relations subscale score

The Conners 3-P(S) Peer Relations subscale score will be calculated as the sum of the item 4, item 6, item 18, item 38, and item 43 scores. The Peer Relations subscale score ranges from 0 to 15. Higher scores indicate greater symptom severity. The change from baseline score will be calculated at each applicable post-baseline timepoint.

4.8 Safety variables

For all safety variables, baseline will be defined as the last assessment prior to the first dose of IP. Change from baseline values will be calculated as the assessment value minus the baseline value. For each safety assessment, study day will be calculated as: the date of assessment − the date of first dose + 1.

4.8.1 Adverse events

An adverse event (AE) is defined as any untoward medical occurrence in a patient or clinical investigation subject administered a pharmaceutical product that does not necessarily have a causal relationship with the product. An

AE can therefore be any unfavorable and unintended sign (including a new, clinically important abnormal laboratory finding), symptom, or disease, temporally associated with the product, whether related to the product.

AEs are collected from the time of the informed consent is signed until the follow-up call is completed. For each AE, the following data will be recorded in the eCRF: verbatim text, start date, stop date, outcome, investigator's assessment of severity, assessment of seriousness, relationship to IP, action taken with IP, and discontinuation due to AE. The verbatim text will be coded by using Medical Dictionary for Regulatory Affairs (MedDRA) version 19.

- Outcome will be recorded as "FATAL"; "NOT RECOVERED/NOT RESOLVED"; "RECOVERED/RESOLVED"; "RECOVERED/RESOLVED WITH SEQUELAE"; "RECOVERING/RESOLVING"; or "UNKNOWN".
- Investigator's assessment of severity will be recorded as "MILD"; "MODERATE"; or "SEVERE".
- Assessment of seriousness will be recorded as "N" or "Y".
- Relationship to IP will be recorded as "NOT RELATED"; "POSSIBLY RELATED"; or "PROBABLY RELATED".
- Action taken with IP will be recorded as "DOSE NOT CHANGED"; "DRUG WITHDRAWN"; "NOT APPLICABLE"; "UNKNOWN"; "DOSE INCREASED"; "DOSE REDUCED"; or "DRUG INTERRUPTED".
- Discontinuation due to AE will be recorded as "N" or "Y".

A flag for related AEs will be calculated as:

- IF the relationship to IP is "POSSIBLY RELATED" or "PROBABLY RELATED", THEN set the related AE flag equal to "Y".
- OTHERWISE, set the related AE flag equal to "N".

A flag for AEs leading to study drug withdrawal will be calculated as:

- IF the action taken with IP is "DRUG WITHDRAWN", THEN set the AE leading to study drug withdrawal flag equal to "Y".
- OTHERWISE, set the AE leading to study drug withdrawal flag equal to "N".

Treatment emergent AEs (TEAEs) are defined as those AEs occurring or worsening after the first dose of IP and within 3 days of the last dose of IP. If an AE start date is completely missing (i.e., in which the day, month, and year are all unknown), then the AE start date will be set to the date of first dose of IP. For a partial AE start date, the following conventions will be used for imputing the AE start date:

- When the year is present and the month and day are missing:
  - If the year of AE start $=$ the year of first dose of IP, then the month and day will be set to the month and day of first dose of IP.
  - If the year of AE start $<$ the year of first dose of IP, then the month and day will be set to December 31st.
  - If the year of AE start $>$ the year of first dose of IP, then month and day will be set to January 1st.
- When the year and day are present and the month is missing:
  - If the year of AE start $=$ the year of first dose of IP, then the month will be set to the month of first dose of IP.
  - If the year of AE start $<$ the year of first dose of IP, then the month will be set to December.
  - If the year of AE start $>$ the year of first dose of IP, then the month will be set to January.
- When the month and year are present and the day is missing:
  - If the year of AE start $=$ the year of first dose of IP and

- the month of AE start = the month of first dose of IP, then the day will be set to the day of first dose of IP.
- the month of AE start < the month of first dose of IP, then the day will be set to the last day of the month.
- the month of AE start > the month of first dose of IP, then the day will be set to the 1st day of the month.

– If the year of AE start < the year of first dose of IP, then the day will be set to the last day of the month.

– If the year of AE start > the year of first dose of IP, then the day will be set to the 1st day of the month.

- If the imputed AE start date is after the AE stop date, then the AE start date will be set to the AE stop date.

For each AE, a treatment emergent flag will be calculated using the recorded AE start date (or imputed start date, where applicable).

- IF the date of first dose of IP $\leq$ AE start date $\leq$ date of last dose of IP + 72 hours, THEN set the treatment emergent flag equal to "Y".

4.8.2 Laboratory test results

Blood and urine samples will be collected by site staff for clinical laboratory testing. Urine pregnancy tests will be performed at the site. All other samples will be analyzed at the central laboratory. For all samples collected, the following data will be recorded in the eCRF: test category (i.e., Hematology, Serum Chemistry, Urinalysis, Serum Pregnancy Test, Urine Pregnancy Test, or Urine Drug Screen), collection date, and the accession number. For urine pregnancy tests, the result will also be recorded in the eCRF as "POSITIVE" or "NEGATIVE".

Collected laboratory samples will then be processed at the central laboratory for the calculation of reported values for the laboratory tests shown in Table 3. Central laboratory data will include test category, lab test name, reported result, reported unit, reported reference range (i.e., low and high), International System of Units (SI) result, SI unit, SI reference range, and reference range indicator. The reference range indicator will be recorded as "NORMAL" or "ABNORMAL" for character results and as "NORMAL", "LOW", or "HIGH" for numeric results. Data from the central laboratory will be combined with the laboratory eCRF data prior to analysis.

Change from baseline values to each post-baseline assessment will be calculated for all continuous laboratory tests.

Table 3: Laboratory test categories and tests assessed

Test category	Tests assessed
Hematology	hemoglobin, hematocrit, red blood cell count, red blood cell indices, mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration, mean corpuscular volume, platelet count (or estimate), white blood cell count including differential
Serum chemistry	albumin, total bilirubin, total protein, calcium, alkaline phosphatase, alanine aminotransferase, aspartate aminotransferase, blood urea nitrogen, creatinine, creatine kinase, glucose, sodium, potassium, chloride, bicarbonate, lactate dehydrogenase, uric acid, thyroid stimulating hormone[1], T4[1]
Urinalysis	pH, specific gravity, blood, glucose, protein, ketones
Pregnancy test[2]	serum, urine
Urine drug screen	amphetamines, barbiturates, benzodiazepines, cocaine, opiates, phencyclidine, cannabinoids, propoxyphene, methadone

[1] Screening only
[2] Females only

4.8.3 Vital sign results

The following vital sign data will be recorded in the eCRF: assessment date, height, weight, temperature, temperature measurement method, pulse rate, respiratory rate, systolic blood pressure (BP), and diastolic BP. The temperature measurement method will be recorded as "Oral" or "Tympanic".

Change from baseline values to each post-baseline assessment will be calculated for pulse rate, respiratory rate, systolic BP, and diastolic BP. Potentially clinically important (PCI) flags will be derived for systolic BP, diastolic BP, pulse rate, respiratory rate, and change in body weight (see Table 4).

Table 4: Potentially clinically important criteria for vital sign results

Vital sign measure	Lower PCI criterion	Upper PCI criterion
Pulse rate	$\leq$50 bpm	$\geq$130 bpm
Systolic BP	$<$90 mmHg	$>$130 mmHg
Diastolic BP	$<$50 mmHg	$>$85 mmHg
Change in Systolic BP	decrease $\geq$30 mmHg	increase $\geq$30 mmHg
Change in Diastolic BP	decrease $\geq$20 mmHg	increase $\geq$20 mmHg
Change in body weight	decrease $\geq$7%	increase $\geq$7%

BP = blood pressure; PCI = potentially clinically important

4.8.4 Electrocardiogram results

Standard 12-lead ECGs will be performed after the subject has been supine for approximately 5 minutes. All ECGs will be performed using the equipment supplied by the central ECG vendor. For all ECGs performed, the following data will be recorded in the eCRF: assessment date and time and the investigator interpretation of the ECG. The investigator assessment will be recorded as "Normal", "Abnormal - Not clinically significant", "Abnormal - Clinically significant", or "Unable to be evaluated".

Electronic ECG tracings will be analyzed by the central ECG vendor for the calculation of reported values for the following ECG tests: heart rate, RR interval, PR interval, QRS duration, QT interval, corrected QT interval using the Bazett formula (QTcB), corrected QT interval using the Fridericia formula (QTcF), and evaluator interpretation. ECG

vendor data will include ECG test name, reported result/finding, reported unit, standard result/finding, standard unit, and lead location. The evaluator interpretation result will be recorded as "NORMAL", "ABNORMAL", or "UNABLE TO EVALUATE". Data from the ECG vendor will be combined with the ECG eCRF data prior to analysis.

Change from baseline values to each post-baseline assessment will be calculated for heart rate, RR interval, PR interval, QRS duration, QT interval, QTcB interval, and QTcF interval. In addition, potentially clinically important flags will be calculated for heart rate, PR interval, QRS duration, QT interval, QTcB interval, QTcF interval, change from baseline in heart rate, and change in QT, QTcB, and QTcF intervals (see Table 5).

Table 5: Potentially clinically important criteria for ECG results

ECG measure	Lower PCI criterion	Upper PCI criterion
Heart rate	$\leq$50 bpm	$\geq$130 bpm
PR interval		$\geq$180 msec
QRS duration		$\geq$100 msec
QT interval		$\geq$450 msec
QTcB interval		$\geq$450 msec
QTcF interval		$\geq$450 msec
Change in heart rate	$\leq$50 bpm & decrease $\geq$5 bpm	
Change in QTcB		increase $\geq$30 msec
Change in QTcF		increase $\geq$30 msec

ECG = electrocardiogram; PCI = potentially clinically important

4.8.5 Physical exam findings

The following physical exam data will be recorded in the eCRF: assessment date, body system, result, abnormal findings, and clinically significant flag.

- Result will be recorded as "NORMAL", "ABNORMAL", or "NOT DONE".
- Clinically significant flag will be recorded as "N" or "Y".

The abnormal findings and clinically significant flag variables are only recorded when the result is "ABNORMAL".

4.8.6 Columbia–Suicide Severity Rating Scale

The Columbia–Suicide Severity Rating Scale (C-SSRS; available at www.cssrs.columbia.edu) is a semi-structured interview that captures the occurrence, severity, and frequency of suicide-related thoughts and behaviors during the assessment period. The interview includes definitions and suggested questions to solicit the type of information needed to determine if a suicide-related thought or behavior has occurred. The C-SSRS has a "Baseline" version that will be completed at the Screening Visit (Visit 1) and a "Since Last Visit" version that will be completed at Visit 2 through Visit 8/ET. Both versions contain all 11 categories of suicidal ideation and behavior defined in the Food and Drug Administration (FDA) guidance document on suicidal behavior and ideation (FDA, 2012). Each category is assessed using a single question requiring a "Yes" or "No" response with follow-up questions for any "Yes" responses. The assessment date and each C-SSRS response will be recorded in the eCRF.

In addition, three C-SSRS composite indicator variables will be calculated as follows:

4.8.6.1 C-SSRS suicidal ideation indicator

A composite indicator of C-SSRS suicidal ideation will be calculated as:

- IF the response to any one of the five suicidal ideation questions is "Yes", THEN set the suicidal ideation indicator equal to 1.
- OTHERWISE, set the suicidal ideation indicator equal to 0.

4.8.6.2 C-SSRS suicidal behavior indicator

A composite indicator of C-SSRS suicidal behavior will be calculated as:

- IF the response to any one of the five suicidal behavior questions is "Yes", THEN set the suicidal behavior indicator equal to 1.
- OTHERWISE, set the suicidal behavior indicator equal to 0.

4.8.6.3 C-SSRS suicidal ideation or behavior indicator

A composite indicator of C-SSRS suicidal ideation or behavior will be calculated as:

- IF the response to any one of the 5 suicidal ideation questions or any of the 5 suicidal behavior questions is "Yes", THEN set the suicidal ideation or behavior indicator equal to 1.
- OTHERWISE, set the suicidal ideation or behavior indicator equal to 0.

5 ANALYSIS METHODS

All personnel involved in the conduct and analysis of the study will remain blinded to individual subject treatment assignments until the study has been completed, the data have been checked for accuracy with any errors corrected, and all analysis sets have been identified. Analysis will be carried out using SAS® (version 9.4 or newer, SAS Institute Inc, Cary, NC). Templates for the summary tables, figures, and subject data listings will be available separately. Data listings will include all data recorded for all enrolled subjects; where included in listings, data entered for screen failure subjects will be flagged.

Descriptive statistics for continuous data will include number of subjects (n), mean, standard deviation (SD), median, minimum, and maximum. Summaries of change from baseline variables will include only subjects who have both a baseline value and corresponding value at the timepoint of interest. Descriptive statistics for categorical data will include frequency and percentage. Calculation of percentages will exclude missing data as a category. Where appropriate, descriptive statistics may be presented with 95% confidence intervals (CIs). For analyses by visit (or derived visit; see Section 4.7), only the latest assessment for each visit will be used.

The minimum and maximum will be reported with the same degree of precision (i.e., the same number of decimal places) as the recorded data. Measures of location (e.g., mean and median) will be reported to 1 degree of precision more than the recorded data, and measures of spread (e.g., standard deviation) will be reported to 2 degrees of precision more than the recorded data.

Any changes to the analyses that are not included in this SAP will be documented in the CSR.

5.1 Summary of subject disposition, demographics, and baseline characteristics

5.1.1 Subject disposition

All subjects who provide informed consent will be included in disposition summaries. The number of subjects screened, screen failed, randomized, completed, and discontinued from the study, as well as reason for discontinuation, will be summarized by treatment group and overall. The number of subjects in each analysis set will also be summarized. All disposition data will listed by subject.

5.1.2 Demographics and other baseline characteristics

Descriptive summaries of the demographic and other baseline characteristics will be provided by treatment group and overall using the ITT analysis set. These tabulations will include the following variables:

- Demographics (age, age strata, sex, ethnicity, race, height, weight, and BMI)
- CNV type

- KBIT-2 Verbal, Nonverbal, and IQ Composite standard scores
- CBCL syndrome and problem standard scores and clinical range categories

All demographics and other baseline characteristics data will be listed by subject.

5.1.3 Baseline disease characteristics

Descriptive summaries of baseline disease characteristics will be provided by treatment group and overall using the ITT analysis set. These tabulations will include the following variables:

- ADHD subtype
- Age at ADHD diagnosis
- Time since ADHD diagnosis
- Age at ADHD onset
- Baseline ADHD-RS-5 total and subscale scores
- Baseline CGI-S score

All baseline disease characteristics data will be listed by subject.

5.1.4 Medical history

All medical history data will be listed by subject.

5.1.5 Prior medications and behavioral therapies

The incidence of prior medication use will be summarized by treatment group and overall and by preferred term using the ITT analysis set. Each prior medication reported more than one time will only be counted once per subject for each preferred term. All prior medication use and behavior therapy data will be listed by subject.

5.2 Protocol deviations

Major protocol deviations will be summarized by treatment group and overall using the ITT analysis set. All inclusion/exclusion criteria and protocol deviation data (including all major and minor deviations) will be listed by subject.

5.3 Primary efficacy analysis

The change from baseline in ADHD-RS-5 total score will be analyzed with mixed model repeated measures (MMRM) methods using the full analysis set. The model will include derived visit (see Section 4.7), age stratum, treatment group, and treatment-by-visit interaction as fixed factors and baseline ADHD-RS-5 total score as a covariate. Robust variance estimates for the fixed effects will be used for testing treatment differences. Within subject variability will be modeled using an unstructured (i.e., TYPE = UN) covariance pattern. In case of convergence problems, a banded Toeplitz (i.e., TYPE = TOEP(5)) structure for the covariance pattern will be used. The comparison of interest will be the difference between the AEVI-001 and placebo treatment groups at Visit 8. Model based point estimates (i.e., least square [LS] means for each treatment group and the treatment difference), 95% confidence interval for the difference, p-value, and effect size will be reported for each derived visit. Effect size will be calculated as the LS mean difference divided by an estimate of the pooled standard deviation.

Similar MMRM methods will be used to test for treatment differences in each CNTN4 CNV subgroup (see Section 4.3.8 for definition of CNTN4 CNV indicator). To control the overall type I error rate for these subgroup comparisons, a hierarchical testing procedure will be used with comparisons of the primary efficacy variable tested in the following fixed sequence:

- Full analysis set
- Full analysis set – CNTN4 CNV subgroup
- Full analysis set – Non-CNTN4 CNV subgroup

Supportive MMRM analyses of the primary endpoint will also be performed using the PP analysis set.

Descriptive statistics for the ADHD-RS-5 total score and change from baseline in ADHD-RS-5 total score will be presented by treatment group and derived visit. A similar descriptive statistics display will also be presented by CNTN4 CNV indicator and age stratum.

5.4 Secondary efficacy analyses

Logistic regression methods will be used for the analysis of the LOCF ADHD-RS-5 response indicator using the full analysis set. The logistic regression model will include terms for age stratum, treatment group, and baseline ADHD-RS-5 total score. The comparison of interest will be the difference between the AEVI-001 and placebo treatment groups at end-of-treatment. Model based point estimates (i.e., odds ratios), 95% confidence intervals, and p-value will be reported.

Similar logistic regression models will be used for the LOCF CGI-I remission, LOCF combined ADHD-RS-5/CGI-I response, and LOCF combined ADHD-RS-5/CGI-I remission indicators using the full analysis set. For the CGI-I response/remission analyses, baseline CGI-S score will be included in the model. For the combined response/remission analyses, baseline ADHD-RS-5 total and CGI-S scores will be included in the model.

MMRM methods similar to those described in Section 5.3 will be used for the analysis of the following change from baseline values: ADHD-RS-5 inattention subscale, ADHD-RS-5 hyperactivity–impulsivity subscale, and CGI-S, Conners 3-P(S) inattention subscale, Conners 3-P(S) hyperactivity–impulsivity subscale, Conners 3-P(S) learning problems subscale, Conners 3-P(S) executive function subscale, Conners 3-P(S) aggression subscale, and Conners 3-P(S) peer relations subscale. For each model, the baseline score of the respective subscale used in the model will be used as a covariate. For the models including Conners 3-P(S) subscale scores, a heterogeneous first order autoregressive (i.e., TYPE = ARH(1)) covariance structure will be used in cases of convergence problems. As with the primary efficacy analysis, the comparison of interest for each model will be the difference between the AEVI-001 and placebo treatment groups at Week 6. Model based point estimates (i.e., LS means for each treatment group and the treatment difference), 95% confidence interval for the difference, p-value, and effect size will be reported for each derived visit.

Subgroup analyses by ADHD subtype (see Section 4.4.1) will be performed for select secondary efficacy variables. No multiplicity correction to the reported p-values will be made for any secondary efficacy analysis. Supportive analyses of secondary efficacy endpoints may be performed using the PP analysis set.

Descriptive statistics for all scores, change from baseline scores, and response/remission indicators will be summarized by treatment group and derived visit. All ADHD-RS-5, CGI, and Conners 3-P(S) data will be listed by subject and visit.

5.5 Safety analyses

For all safety endpoints, descriptive statistics will be presented by treatment group and visit (where applicable) using the safety analysis set. In addition, some safety endpoints may be further summarized by optimized dose (see Section 4.6). In summaries of change from baseline safety variables, only subjects with both baseline and post baseline data will be included. No formal hypothesis tests will be performed for any safety endpoint.

5.5.1 Concomitant medications and study drug exposure

5.5.1.1 Concomitant medications

The incidence of concomitant medication use will be summarized by treatment group and preferred term (see Section 4.5 for definition of concomitant medications). Each concomitant medication reported more than one time will only be counted once per subject for each preferred term. All concomitant medication data will be listed by subject.

5.5.1.2 Study drug exposure

Optimized dose, total study drug exposure, and duration of total exposure will be summarized by treatment group. All study drug exposure data will be listed by subject.

5.5.1.3 Study drug compliance

Overall percent compliance will be summarized by treatment group. In addition, the frequency and percentage of subjects in each compliance category (see Section 4.6 for definition of overall compliance indicator) will be tabulated by treatment group. All compliance data will be listed by subject.

5.5.2 AEs

The overall incidence of subjects having at least one AE in each of the following categories will be summarized by treatment group:

- Any TEAE
- Any SAE
- Any AE leading to death
- Any TEAE leading to study drug withdrawal
- Any related TEAE

All AE data will be listed by subject. Separate listings by subject will presented for all SAEs, AEs leading to death, and TEAEs leading to study drug withdrawal.

5.5.2.1 TEAEs

The incidence of TEAEs will be summarized by treatment group, system organ class (SOC), and preferred term. Each subject will be counted only once per SOC and preferred term. An additional display may be produced by treatment group and preferred term.

5.5.2.2 SAEs

The incidence of SAEs will be summarized by treatment group, SOC, and preferred term. Each subject will be counted only once per SOC and preferred term.

5.5.2.3 TEAEs leading to study drug withdrawal

The incidence of TEAEs leading to study drug withdrawal will be summarized by treatment group, SOC, and preferred term. Each subject will be counted only once per SOC and preferred term.

5.5.2.4 TEAEs by maximum relationship to study drug

The incidence of TEAEs will be summarized by maximum relationship to study drug, treatment group, SOC, and preferred term. Each subject is counted only once per preferred term and most related category reported.

5.5.2.5 TEAEs by maximum severity

The incidence of TEAEs will be summarized by maximum severity to study drug, treatment group, SOC, and preferred term. Each subject is counted only once per preferred term and most severe category reported.

5.5.3 Laboratory test results

Descriptive statistics for all reported values and change from baseline values will be summarized by test category, treatment group, and visit. Only the original assessment for each visit will contribute to descriptive statistics; repeat or unscheduled assessments will not be included in the calculation of descriptive statistics by visit. All laboratory test

results data (including pregnancy test and urine drug screen data and any repeat or unscheduled assessments) will be listed by subject and visit.

5.5.4 Vital sign results

Descriptive statistics for all reported values and change from baseline values will be summarized by treatment group and visit. Only the original assessment for each visit will contribute to these summaries; repeat or unscheduled assessments will not be included in the calculation of descriptive statistics by visit. The number and percentage of subjects with PCI values at any time will be tabulated by treatment group. Summaries of PCI values will include all assessments regardless of whether an original, repeat, or unscheduled assessment. All vital sign results data will be listed by subject and visit.

5.5.5 ECG results

Descriptive statistics for all reported values and change from baseline values will be summarized by treatment group, and visit. Only the original assessment for each visit will contribute to these summaries; repeat or unscheduled assessments will not be included in the calculation of descriptive statistics. The number and percentage of subjects with PCI values at any time will be tabulated by treatment group; the number and percentage of subjects with abnormal ECG findings will be presented similarly. Summaries of PCI values and abnormal ECG findings will include all assessments regardless of whether an original, repeat, or unscheduled assessment. All ECG results data will be listed by subject and visit.

5.5.6 Physical exam findings

All physical exam findings data will be listed by subject and visit.

5.5.7 C-SSRS

The incidence of suicidal ideation/behavior at any time will be summarized by treatment group for each C-SSRS ideation, behavior, and ideation/behavior indicator. All C-SSRS data will be listed by subject and visit.

5.6 Interim analysis

In Part A of the study, an interim analysis will be performed when approximately 75% of randomized subjects have completed the study. This analysis will focus on the magnitude of placebo response and will determine whether the enrollment target should be increased to a total of 64 randomized subjects.

The interim analysis will require unblinding of subject-level data and will be conducted by an independent statistician having no other involvement in the analysis or conduct of this study. The interim analysis will be limited to the evaluation of ADHD-RS-5 data (i.e., the primary outcome). To maintain the integrity of the experimental blind, the independent statistician will only share qualitative results with the sponsor (e.g., “increase the sample size” or “do not increase the sample size”).

Details of the interim analysis will be provided in a separate Interim Analysis Plan.

6 CHANGES TO PLANNED ANALYSES

6.1 Changes to analyses specified in protocol

Not applicable.

6.2 Changes to approved prior versions of the SAP

Table 6: Changes to Final Version 1.0 dated 18 April 2018

SAP section	Prior text	Change to text	Rationale
List of Abbreviations		SAP = Statistical analysis plan	Previously omitted abbreviation added.
Section 1 – Overview	Final Version 5.0 dated 01 March 2018	Final Version 7.0 dated 25 July 2018	Reference updated to current version of the protocol.
Section 2.2 – Study design	If Part A of the study is positive (e.g., meets the primary endpoint), Part B will be conducted and will assess subjects who do not have CNVs in any of the specific gene mutation(s) implicated in glutamatergic signaling and neuronal connectivity.	Part B will be conducted and will assess subjects who do not have CNVs in any of the 272 specific gene mutations(s) implicated in glutamatergic signaling. This list of excluded mutations includes the 8 gene mutations studied in Part A and an additional 264 mutations identified in an antecedent study.	Change made to be consistent with current version of the protocol. Protocol Version 7.0 removed text requiring Part A to be positive prior to starting Part B and added text to clarify the mutations to be included in Part B.
Section 4.6 – Study drug exposure variables	$(\text{Weight})_i \times 100$	$(\text{Weight})_i$, where i is the study week of dosing.	Multiplication by 100 removed from the overall percent compliance formula as the Percent Compliance variable is already on a percent scale; text added to define i.
Section 4.7 – Efficacy variables (Table 2)	–28 to –1	$\leq$–1	Visit window for derived visit number 1 updated to allow derivation for visits occurring prior to Day –28.
Section 4.7 – Efficacy variables (Table 2)	$\geq$40	40 to 50	Visit window for derived visit number 8 updated to restrict derivation for visits occurring within a few days of stopping study drug.
Section 5.4 – Secondary efficacy analyses		Subgroup analyses by ADHD subtype (see Section 4.4.1) will be performed for select secondary efficacy variables.	Subgroup analyses added for secondary efficacy variables by ADHD subtype.

7 REFERENCES

Achenbach, T. M. & Rescorla, L. A. (1996). *Manual for the ASEBA School-Age Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Conners, C. K. (2008). *Conners 3rd Edition*. North Tonawanda, NY: Multi-Health Systems.

DuPaul, G. J., Power, T. J., Anastopoulos, A. D., & Reid, R. (2016). *ADHD Rating Scale-5 for Children and Adolescents Checklists, Norms, and Clinical Interpretation*. New York, NY: Guildford Press.

FDA (2012). Guidance for Industry Suicidal Ideation and Behavior: Prospective Assessment of Occurrence in Clinical Trials. https://www.fda.gov/downloads/drugs/guidances/ucm225130.pdf.

Guy, W. (Ed.). (1976). *ECDEU Assessment Manual for Psychopharmacology*. Rockville, MD: U.S. Department of Health, Education, and Welfare.

ICH (1995). ICH Harmonised Tripartite Guideline: Structure and Content of Clinical Study Reports E3. Step 4, 30 November 1995. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E3/E3_Guideline.pdf.

ICH E9 Expert Working Group (1999). Statistical Principles for Clinical Trials. *Statistics in Medicine*, *18*, 1905–1942.

Kaufman, A. S. & Kaufman, N. L. (2004). *Kaufman Brief Intelligence Test Second Edition*. Bloomington, MN: Pearson, Inc.

APPENDIX A Schedule of Events (Parts A and B)

Source: Table 1 from Protocol Version 7.0, 25 July 2018

Table 1: Schedule of Events (Parts A and B)

	Screening Visit	Washout Period[1]	Baseline Visit	Treatment Period						Follow-up Period[6]
				Dose-optimization Phase				Dose Maintenance Phase		
Visit	**1**	**No Visit**	**2**	**3**	**4**	**5**	**6**	**7**	**8/ET**	**No Visit**
Assessment Week	**-4**		**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Assessment Day	**-28**		**0**	**7**	**14**	**21**	**28**	**35**	**42/Any**	**+7 days from last dose**
Visit Window[2]				**±2 days**	**±2 days**	**±2 days**	**±2 days**	**±2 days**	**±2 days**	**+2 days**
Informed Consent/Assent	X									
Inclusion/Exclusion Review	X[5]	X	X							
Genotype result available	X									
Randomization			X							
M.I.N.I. International Neuropsychiatric Interview for Children and Adolescents (M.I.N.I. Kid)	X									
CBCL			X							
KBIT-2	X									
Demographics	X									
Medical and Medication History	X									
Physical Exam[3]	X								X	
Weight	X[5]		X	X	X	X	X	X	X	
Height	X[5]								X	
Calculate BMI	X									
Vital Signs[4]	X[5]		X	X	X	X	X	X	X	
12-lead ECG	X[5]		X	X	X	X	X	X	X	
Hematology & Clinical Chemistry	X[5]		X	X			X		X	
Thyroid Stimulating Hormone and T4	X[5]									
Identity Confirmation[7,8]	X									

	Screening Visit	Washout Period[1]	Baseline Visit	Treatment Period						Follow-up Period[6]
				Dose-optimization Phase				Dose Maintenance Phase		
Visit	**1**	**No Visit**	**2**	**3**	**4**	**5**	**6**	**7**	**8/ET**	**No Visit**
Assessment Week	**-4**		**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Assessment Day	**-28**		**0**	**7**	**14**	**21**	**28**	**35**	**42/Any**	**+7 days from last dose**
Visit Window[2]				**±2 days**	**±2 days**	**±2 days**	**±2 days**	**±2 days**	**±2 days**	**+2 days**
Urine Drug Screen	X[5]		X	X	X	X	X	X	X	
Urinalysis	X[5]		X	X			X		X	
Serum pregnancy test[9]	X[5]								X	
Urine pregnancy test[9]			X	X	X	X	X	X		
Investigator Dose Assessment				X	X	X	X			
ADHD-RS-5			X	X	X	X	X	X	X	
CGI-S	X		X	X	X	X	X	X	X	
CGI-I				X	X	X	X	X	X	
Conners 3-P(S)	X		X				X		X	
C-SSRS Baseline Version	X									
C-SSRS Since Last Visit Version			X	X	X	X	X	X	X	
Study Drug Dispensed			X	X	X	X	X	X		
Study Drug Collected				X	X	X	X	X	X	
IWRS	X		X	X	X	X	X	X	X	
Compliance Calculation				X	X	X	X	X	X	
Adverse Event	X	X	X	X	X	X	X	X	X	X
Concomitant Medications	X	X	X	X	X	X	X	X	X	X

Abbreviations: ADHD-RS-5 = Attention Deficit Hyperactivity Disorder Rating Scale Version 5; BMI = body mass index; CBCL = Child Behavior Checklist; CGI-I = Clinical Global Impression of Improvement; CGI-S = Clinical Global Impression of Severity; Conners 3-P(S): Conners 3rd Edition–Parent Short form; C-SSRS = Columbia Suicide Severity Rating Scale; ECG = electrocardiogram; ET = early termination; IWRS = interactive web response system.

[1] For subjects currently taking a nonstimulant medication for ADHD and/or or clonidine taken at night for sleep, the call should take place to allow for a 14-day washout prior to the Baseline Visit (Visit 2). For subjects receiving a stimulant medication for ADHD, the call should take place to allow for a 5-day washout prior to the Baseline Visit (Visit 2). Subjects who do not require a washout must still be contacted by phone prior to the Baseline Visit (Visit 2) being conducted.

[2] After Day 0 (Visit 2), visits at which AEVI-001 is administered should occur every 7 ± 2 days. These visits should be scheduled relative to Day 0 (Visit 2), which is the baseline visit.

[3] A complete physical examination (excluding genitourinary examination) will be performed at Visit 1 and at the completion of exposure (Visit 8/ET).

[4] Includes blood pressure, pulse rate, and respiratory rate will be measured after the subject has been in a sitting position for approximately 5 minutes. Oral or tympanic temperature will be collected at the Screening Visit (Visit 1) only.
[5] If more than 28 days have elapsed between the Screening Visit (Visit 1) and the Baseline Visit (Visit 2), clinical laboratory tests, serum pregnancy test, vital signs, height, weight, urine drug test, and ECG must be repeated and the results available and reviewed prior to proceeding with the Baseline Visit (Visit 2). Applicable inclusion and exclusion criteria must be reassessed using the results received.
[6] The follow-up call will take place 7 days (+ 2 days) after the subjects last dose of investigational product. If applicable, a subject should not be restarted on ADHD medication until the day after completion of the follow-up call.
[7] A saliva sample will be collected for identity confirmation of subjects with a previously banked saliva sample from the AEVI-001-ADHD-002 study. This requirement only applies to subjects who participated virtually in the AEVI-001-ADHD-002 study. If a subject participated in the MDGN-NFC1-ADHD-001 study, the MDGN-NFC1-ADHD-101 study or was screened on site in the AEVI-001-ADHD-002 study, a saliva sample is not required to be collected.
[8] If required, identity confirmation test result to be reviewed upon receipt to determine the subject's ability to be randomized.
[9] For all female subjects.