STATISTICAL ANALYSIS PLAN

NCT Number:	NCT03325881
Study Title:	A Phase 3, Randomized, Double-blind, Multi-center, Placebo-controlled, Fixed-Dose, Efficacy, and Safety Study of SHP465 in Children Aged 6-12 Years with Attention-Deficit/Hyperactivity Disorder (ADHD)
Study Number:	SHP465-309
SAP Version:	Version 2
SAP Version Date:	16 August 2018

STATISTICAL ANALYSIS PLAN

Protocol No.:	SHP465-309
Protocol Title:	A Phase 3, Randomized, Double-blind, Multi-center, Placebo-controlled, Fixed-Dose, Efficacy, and Safety Study of SHP465 in Children Aged 6-12 Years with Attention-Deficit/Hyperactivity Disorder (ADHD)
Drug:	SHP465, mixed salts of a single-entity amphetamine
Sponsor:	Shire Development LLC 300 Shire Way, Lexington, MA 02421
Version No. and Date	Version 2.0, 16 Aug 2018

Version No.	Document History Description of Update	Author(s)	Effective Date
1.0	Final for release	PPD	18 Dec 2017
2.0	Added the interim analysis results regarding the blinded sample size re-estimation Added the handling for unintentional urine drug screen Applied a number of editorial changes	PPD	16 Aug 2018

Updates between Version 1.0 and Version 2.0

Section #	Changes
1	Added administrative changes to study protocol
2.5.1	Added a section for updated sample size based on blinded sample size re-estimation
5.4	Clarified definition of the full analysis set by adding the valid baseline ADHD-RS-5 total score as a requirement.
9	Clarified the exposure and compliance analyses by age group
11.5	Added 2 sensitivity analyses
11.7	Clarified the subgroup analyses
12.1.1	Added “TERMINAL INSOMNIA” to the analysis for insomnia.
12.2 (Table 3)	Modified the PCI criteria for Mean Corpuscular Hemoglobin (MCH)
12.6	Clarified TEAE summaries by age group
18	Added handling of the urine drug screen

TABLE OF CONTENTS

LIST OF IN-TEXT TABLES

LIST OF IN-TEXT FIGURES

ABBREVIATIONS

ADHD	attention-deficit/hyperactivity disorder
ADHD-RS-5	Attention-deficit/Hyperactivity Disorder Rating Scale-5
AE	adverse event
ANCOVA	Analysis of covariance
BMI	body mass index
CGI-I	Clinical Global Impressions–Improvement
CGI-S	Clinical Global Impressions–Severity of Illness
CI	confidence interval
CMH	Cochran–Mantel–Haenszel
CRO	contract research organization
C-SSRS	Columbia-Suicide Severity Rating Scale
CSHQ	Children's Sleep Habits Questionnaire
DINFC	date of informed consent
DOB	date of birth
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
ECG	electrocardiogram
eCRF	electronic case report form
ET	early termination
FAS	full analysis set
FDA	Food and Drug Administration
FoTA	Final on-Treatment Assessment
IWRS	interactive web response system
MAR	missing at random
MCH	mean corpuscular hemoglobin
MCHC	mean corpuscular hemoglobin concentration
MCV	mean corpuscular volume
MedDRA	Medical Dictionary for Regulatory Activities
MINI-KID	Mini International Neuropsychiatric Interview for Children and Adolescents
MMRM	mixed-effects model for repeated measures
MNAR	missing not at random
PCI	potentially clinically important
PSQ	Post Sleep Questionnaire
QT	QT interval of the cardiac cycle
QTcB	Bazett's corrected QT interval
QTcF	Fridericia's corrected QT interval
RBC	red blood cells
REML	restricted maximum likelihood
SAE	serious adverse event
SAP	statistical analysis plan
SD	standard deviation
SOC	system organ class
TEAE	treatment-emergent adverse event
US	United States
WBC	white blood cells
WHO	World Health Organization

1. INTRODUCTION

This statistical analysis plan (SAP) provides a technical and detailed elaboration of the statistical analyses of efficacy and safety data as described in the protocol dated 07 September 2017, the protocol administrative change #1 dated 20 Oct 2017, and the protocol administrative change #2 dated 28 Nov 2017. Specifications for tables, figures, and listings are contained in a separate document.

2. STUDY DESIGN

2.1 General Study Design

This study is a Phase 3 randomized, double-blind, multi-center, parallel-group, placebo-controlled, fixed-dose, efficacy, and safety study in which children aged 6-12 with attention-deficit/hyperactivity disorder (ADHD) will be randomized at baseline (Visit 2) in a 1:1 ratio to SHP465 6.25 mg or placebo for 4 weeks of double-blind treatment as outlined in Table 1.

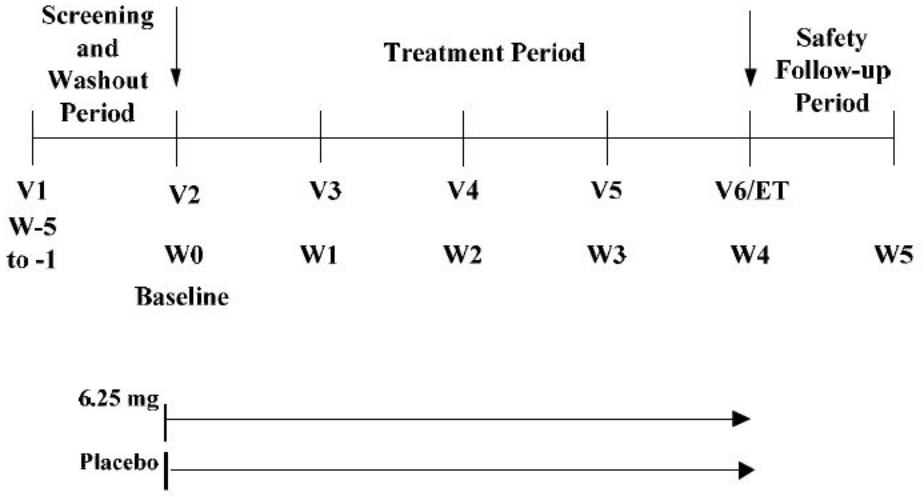
The study will have up to 3 periods as outlined below and in Figure 1:

- Screening and washout period (up to 32 days)
- Treatment period (4 weeks)
- Safety follow-up period (1 week)

Approximately 60 subjects will be randomized in this study to evaluate the efficacy and safety of SHP465 6.25 mg, administered as a daily morning dose in the treatment of children aged 6-12 years with ADHD. Approximately 50% of the total number of subjects will be aged 6-8 years and approximately 50% of subjects will be aged 9-12 years. Approximately 25% of all subjects will be female.

This study will be conducted at approximately 45 sites in the United States (US).

Figure 1 Study Design Flow Chart



ET=early termination; V=visit; W=week

2.2 Randomization

Subject numbers are assigned to all subjects as they consent to take part in the study. Within each site (numbered uniquely within a protocol), this number is assigned to subjects according to the sequence of presentation for study participation.

The actual treatment given to individual subjects is determined by a randomization schedule. Subjects who meet all inclusion criteria and none of the exclusion criteria will be randomized in a 1:1 ratio to either SHP465 6.25 mg or placebo. Randomization will be stratified to ensure approximately 50% of subjects in each age group (6-8 years and 9-12 years), and to facilitate balance of treatment allocation within each age group. The study is not powered to detect within age stratum treatment effects.

The randomization number represents a unique number corresponding to investigational product allocated to the subject, once eligibility has been determined. Individual subject treatment is automatically assigned by the interactive web response system (IWRS).

Investigational product packaging identification numbers, separate from randomization numbers/unique identifiers, may also be assigned to subjects for specific treatment assignment as dictated by the study. In these cases, the same investigational product packing identification number may not be assigned to more than 1 subject.

2.3 Blinding

This is a double-blind, placebo-controlled study. All investigational and reference product (SHP465 6.25 mg or placebo) will appear identical to protect the study blind.

Data that may potentially unblind the treatment assignment (i.e., treatment allocation, and investigational product preparation/accountability data) will be handled with special care during the data cleaning and review process. These data will be handled in such a way that, prior to unblinding, any data that may unblind study team personnel will be presented as blinded information or otherwise will not be made available. If applicable, unblinded data may be made available to quality assurance representatives for the purposes of conducting independent drug audits.

The treatment assignment must not be broken during the study except in emergency situations where the identification of the investigational product is required for further treatment of the subject. The investigator should contact the medical monitor as soon as possible after the investigator has been unblinded.

In the event that the treatment assignment is broken, the date, the signature of the person who broke the code and the reason for breaking the code will be recorded on the IWRS and the source documents. Upon breaking the blind, the subject will be withdrawn from the study but should be followed up for safety purposes. Any code breaks that occur must be reported to the contract research organization (CRO) and to Shire. Code-break information is held by the pharmacist/designated person at the site and by the CRO medical monitor/designated person for

the study. There will be a provision for unblinding to ensure adequate treatment of the subject in the case of an emergency.

2.4 Schedule of Assessments

Table 1 below presents a schedule of study assessments.

Table 1 Schedule of Assessments

	Screening/Washout[a]		Baseline	Treatment Period				Follow-up
Visit Number[b]	**1**	**Phone call**	**2**	**3**	**4**	**5**	**6/ET[c]**	**Phone call**
Study Week	**-5 to -1**	**-1**	**0**	**1**	**2**	**3**	**4**	**5**
Study Day[b]	**–32 to –3**	**-7**	**0**	**7**	**14**	**21**	**28**	**35**
Informed consent/assent	✓							
Inclusion/exclusion criteria	✓	✓[d]	✓[d]					
Subject demography	✓							
MINI-Kid	✓							
Medical history[e] includes prior medications and procedures	✓							
Randomization			✓					
ADHD-RS-5[f]			✓	✓	✓	✓	✓	
CGI-S[f]			✓					
CGI-I[f]				✓	✓	✓	✓	
Urine drug screen	✓		✓[g]					
Serum pregnancy test[k]	✓		✓[g]					
Urine pregnancy test[k]			✓				✓	
Physical examination	✓		✓[g]					
Height[h]	✓		✓	✓	✓	✓	✓	
Weight[h]	✓		✓	✓	✓	✓	✓	
Vital signs[i,j]	✓		✓	✓	✓	✓	✓	
Clinical laboratory test	✓		✓[g]				✓	
Electrocardiogram (12-lead)[l]	✓[i]		✓[g,i]		✓		✓	
C-SSRS baseline version	✓							
C-SSRS since last visit version			✓	✓	✓	✓	✓	
PSQ[f]			✓	✓	✓	✓	✓	
CSHQ[f]			✓	✓	✓	✓	✓	
Adverse events	✓	✓	✓	✓	✓	✓	✓	✓
Concomitant medications	✓	✓	✓	✓	✓	✓	✓	✓

Table 1 Schedule of Assessments

	Screening/Washout[a]		Baseline	Treatment Period				Follow-up
Visit Number[b]	**1**	**Phone call**	**2**	**3**	**4**	**5**	**6/ET[c]**	**Phone call**
Study Week	**-5 to -1**	**-1**	**0**	**1**	**2**	**3**	**4**	**5**
Study Day[b]	**–32 to –3**	**-7**	**0**	**7**	**14**	**21**	**28**	**35**
Subject check to remain in study[f]			✓	✓	✓	✓		
IWRS	✓		✓	✓	✓	✓	✓	
Study drug dispensed			✓	✓	✓	✓		
Study drug capsules returned				✓	✓	✓	✓	

ADHD-RS-5=Attention-deficit/Hyperactivity disorder-Rating Scale 5 Child, Home Version; BMI=body mass index; CGI=Clinician's Global Impression; C-SSRS=Columbia-Suicide Severity Rating Scale; ET=early termination; IWRS=interactive web response system; MINI=Mini International Neuropsychiatric Interview (MINI) Kid version for Children and Adolescents; PSQ=Post Sleep Questionnaire; CSHQ=Children's Sleep Habits Questionnaire.

[a] Following successful screening, a site representative will contact the subject's parent/LAR to instruct the subject on discontinuing any prohibited medication for the washout period.

[b] Visit windows are with respect to baseline (Visit 2) and ±2 days during the treatment period and +2 days for the safety follow-up phone call.

[c] Subjects who terminate early will undergo the evaluations listed for Visit 6.

[d] Inclusion/exclusion criteria will be reviewed during the washout phone call and at baseline (Visit 2).

[e] Medical history will include all lifetime psychiatric and nonpsychiatric medications and procedures.

[f] Whenever possible, the same individual should complete/rate consistently the following scales and questionnaires as appropriate: ADHD-RS-5, CGI-S/I, and subject check to remain in the study; and the same caregiver/LAR for PSQ and CSHQ. Include assessment of decreased appetite.

[g] If > 32 days have elapsed since the screening evaluation was completed at Visit 1, then the following evaluations must be repeated at baseline (Visit 2): vital signs, serum pregnancy test, urine drug screen, clinical laboratory evaluations, and ECGs in triplicate. The physical exam will be abbreviated with a review of the following body systems: general appearance, respiratory, and cardiovascular.

[h] Height and weight will be measured without shoes and with light clothing using a calibrated stadiometer for height and calibrated scale for weight.

[i] Vital signs include oral or tympanic temperature, pulse, sitting blood pressure, and respiration rate. The subject will have been seated for a minimum of 3 minutes before blood pressure, pulse and respiration rate measurements are taken. Measurement of blood pressure and pulse will be collected 3 times (with approximately 2 minutes in between each collection). The average of each set of 3 measurements will be used to determine continued participation in the study.

[j] Blood pressure and pulse rate will be measured at each study visit. Temperature and respiration rate will be measured at screening (Visit 1) and/or baseline (Visit 2) only.

[k] For females of child-bearing potential only.

[l] Electrocardiograms will be recorded in triplicate with approximately 3 minutes in between each collection at screening only and at baseline only if >32 days have elapsed since the screening evaluation was completed at Visit 1.

2.5 Determination of Sample Size

The primary efficacy endpoint is defined as the change from baseline in the Attention-deficit/Hyperactivity Disorder Rating Scale Version 5 (ADHD-RS-5) total score at Visit 6 (Week 4). The baseline ADHD-RS-5 total score is the last value obtained prior to taking the first dose of investigational product, usually at Visit 2.

The null hypothesis is that there is no difference in mean change from baseline at Visit 6 (Week 4) in ADHD-RS-5 total score between SHP465 6.25 mg and placebo, and the alternate hypothesis is that there is a difference in mean change from baseline at Visit 6 (Week 4) in ADHD-RS-5 total score between SHP465 6.25 mg and placebo.

The sample size was estimated for the primary comparison of SHP465 at 6.25 mg with placebo by using nQuery Advisor 7.0. To detect an assumed difference of 11.9 for the change from baseline in the ADHD-RS-5 total score between the SHP465 at 6.25 mg treatment group and the placebo group with the assumed common standard deviation (SD) of 14, 26 completers per group are needed to provide 85% power for a 2-sided t-test with α level of 0.05. This yields a total of 52 subjects (26 on active treatment and 26 on placebo). Taking into account an expected post-randomization dropout rate of 15%, the randomization target is set at 60 subjects in total.

The assumptions of a treatment difference of 11.9 and a standard deviation of 14.0 corresponds to an effect size of 0.85 which is low compared with that in recently completed study (SHP465-305) of children and adolescents aged 6-17 years with ADHD treated with SHP465. The SHP465-305 study yielded an effect size of 0.93 in children aged 6-12 years (101 subjects).

A blinded interim analysis at the late stage of the trial (when approximately 75% of all randomized patients have either completed or discontinued from the study) is needed for a possible sample size increase in case of an underestimated variability postulated at the design stage (see Section 15).

The final total number of subjects randomized between the 2 groups will be calculated in the blinded interim analysis, based on the estimate of the pooled variance, as detailed in Section 15. The number of subjects could remain at 60, or could potentially be as high as 90, which corresponds to the 97.5th percentile of the distribution of the variance estimator based on the true standard deviation of 14.0. Note that the total of 90 subjects is not considered a cap. As the final sample size is data driven, a higher number, though unlikely, is possible.

2.5.1 Sample Size Update

Following the SAP Section 2.5 and the protocol Section 9.6, a blinded interim analysis for sample size re-estimation was performed based on 47 randomized subjects as of 19 March 2018 (the interim cohort), among which 45 were in the FAS and 43 completed the study. Data of the interim cohort were extracted on 27 April 2018, after the collection of all their primary efficacy data.

The pooled standard deviation for Visit 6 (Week 4) used for sample size re-estimation was calculated using a pooled common variance that was estimated by using blinded cumulative primary efficacy data in the interim cohort. Then the final pooled standard deviation was

calculated as stated in Section 15 and Appendix 1. The interim analysis results showed a re-estimated pooled standard deviation of 13.29 which is less than 14 and therefore an increase to the sample size was not required.

2.6 Multiplicity Adjustments for Type I Error Control

In order to protect the study-wide Type I error at the 2-sided 0.05 for testing across the primary and the key secondary hypotheses, the Fixed-Sequence Test procedure will be used (see Section 11 for details).

3. OBJECTIVES

3.1 Primary Objective

The primary objective of this study is to evaluate the efficacy of SHP465 at 6.25 mg compared to placebo as a daily morning dose in children 6-12 years of age (inclusive at the time of consent) diagnosed with ADHD. The primary measure of efficacy will be the clinician-administered ADHD-RS-5 Child Home Version total score.

3.2 Secondary Objectives

Key Secondary:

- To assess the efficacy of SHP465 at 6.25 mg compared to placebo using a global clinical measure of improvement, the Clinical Global Impression of Improvement (CGI-I) scale.

Secondary:

- To evaluate the safety and tolerability of SHP465 at 6.25 mg based on the occurrence of treatment-emergent adverse events (TEAEs), evaluation of vital signs (systolic and diastolic blood pressure and pulse), weight, height, body mass index (BMI), clinical laboratory and electrocardiogram (ECG) results, sleep assessment (Post Sleep Questionnaire [PSQ] and Child's Sleep Habits Questionnaire [CSHQ]), and responses to the Columbia Suicide Severity Rating Scale (C-SSRS).

4. ESTIMANDS

4.1 Primary Estimand

The primary estimand is the mean change from baseline to Week 4 in ADHD-RS-5 total score between active (SHP465 6.25 mg) and placebo groups.

- Population: 6-12 year old children with ADHD defined through inclusion and exclusion criteria as stated in the protocol
- Variable: change from baseline to Week 4 in ADHD-RS-5 total score
- Intercurrent event: had patients not discontinued treatment
- Population-level summary: difference in mean change of ADHD-RS-5 total score from baseline to Week 4 between active (SHP465 6.25 mg) and placebo groups

4.2 Secondary Estimand

The key secondary estimand is the mean CGI-I value at Week 4 between active (SHP465 6.25 mg) and placebo groups.

- Population: 6-12 year old children with ADHD defined through inclusion and exclusion criteria as stated in the protocol
- Variable: CGI-I value at Week 4
- Intercurrent event: had patients not discontinued treatment
- Population-level summary: difference in mean CGI-I value at Week 4 between active (SHP465 6.25 mg) and placebo groups

5. SUBJECT POPULATION SETS

The following subject sets are applicable to this study:

5.1 Screened Set

The Screened Set will consist of all subjects who have provided informed consent.

5.2 Randomized Set

The Randomized Set will consist of all subjects in the Screened Set for whom a randomization number has been assigned.

5.3 Safety Set

The Safety Set will consist of all subjects in the Randomized Set who have taken at least 1 dose of investigational product.

5.4 Full Analysis Set

The Full Analysis Set (FAS) will consist of all subjects in the Safety Set who have had the baseline ADHD-RS-5 total score and at least 1 post-dose ADHD-RS-5 total score.

6. SUBJECT DISPOSITION

A listing of all Screen Failures (i.e., subjects who were screened but not randomized) will be presented along with reasons for screen fail and details of any adverse events (AEs).

The number of subjects included in each analysis set (i.e., Screened, Randomized, Safety, FAS) will be summarized by treatment group (SHP465 6.25 mg or Placebo) and overall, except for the Screened Set, which will be summarized only overall.

The number and percentage of subjects who completed and prematurely discontinued during the Treatment period will be presented for each treatment group and overall for the Safety Set. Reasons for premature discontinuation from the Treatment period, as recorded on the end of study discontinuation page of the electronic case report form (eCRF), will be summarized (number and percentage) by treatment group for the Safety Set. The subjects who completed the study are those that completed the final scheduled visit (Visit 6) at the end of the Treatment period.

In addition, the number and percentage of subjects who completed the study visits will be presented by visit for each treatment group and overall for the Safety Set.

All subjects who prematurely discontinued during the Treatment period will be listed by discontinuation reason for the Randomized Set.

Follow-up information, new (post-treatment) AEs and new (post-treatment) medication, will be listed for the Safety Set.

7. PROTOCOL DEVIATIONS

Protocol deviations and violations will be recorded by the site separately from the clinical database. The CRO/Shire will classify the protocol deviations and violations per the agreed protocol violation and deviation management plan. The Shire study team will review the protocol deviations and their classification throughout the study and before treatment unblinding and database lock.

Decisions of the review will include:

- Accuracy of protocol deviations and violations categorization.

For any criteria for protocol deviations that can be completely implemented by a computer program, the detailed algorithm will be agreed upon. Details of such algorithms will be included in the derived dataset specifications and finalized before treatment unblinding.

Confirmed protocol violations and protocol deviations will be documented in the Protocol Violation/Deviation tracker for the study. Protocol deviations/violations will be summarized by category and site for each treatment group (SHP465 6.25 mg and Placebo) and overall, for the Safety Set. Protocol deviations/violations will be listed for the Randomized Set.

8. DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS

8.1 Demographic and Other Baseline Characteristics

Descriptive summaries of demographic and baseline characteristics will be presented by treatment group (SHP465 6.25 mg and Placebo) and overall for the Safety Set and FAS.

Demographics and baseline characteristics will be examined to assess the comparability of the treatment groups at baseline, where baseline is defined as the last assessment prior to the first administration of the investigational product.

The following demographic characteristics will be summarized in the following order in the tables: age, sex, ethnicity, race, weight, height and body mass index (BMI). In addition, other baseline characteristics such as years since diagnosis of ADHD, ADHD Subtype, baseline ADHD-RS-5 total score, and baseline CGI-S will be summarized.

Age, weight, height, BMI, years since diagnosis of ADHD, ADHD-RS-5 total score will be summarized as continuous variables using number of subjects, mean and SD, median, and minimum and maximum values, while age (categorical), sex, ethnicity, race, and Baseline ADHD Subtype and CGI-S will be summarized as categorical variables using number of subjects and percentages for each category. All summarized values will be taken from the Screening Visit (Visit 1), except for baseline ADHD-RS-5 total score, baseline CGI-S, height, and weight. In addition, baseline weight and baseline height will be summarized in both conventional and international system of units.

Age will be calculated as the difference between date of birth (DOB) and date of informed consent (DINFC), truncated to months, using the following:

- Age = floor((intck('month',DOB,DINFC) - (day(DINFC) < day(DOB))) / 12)

BMI will be calculated using one of the following:

- Weight in pounds * 703 / (Height in inches)^2
- Weight in kilograms / (Height in meters)^2

Number of years since diagnosis of ADHD will be based on the date of informed consent. Demographics and baseline characteristics will be listed for the Safety Set.

In addition, demographics and baseline characteristics tables will be repeated by age group (6-8 years and 9-12 years).

8.2 Medical History

Medical history will be collected at the Screening Visit (Visit 1) and will be listed for the Safety Set.

The psychiatric diagnosis will be established with Mini International Neuropsychiatric Interview for Children and Adolescents (MINI-KID) at the Screening Visit (Visit 1) and will be listed for the Safety Set.

ADHD non-medication treatment history, behavioral therapy history, collected at the Screening Visit (Visit 1) will be listed for the Safety Set.

9. EXTENT OF EXPOSURE AND TREATMENT COMPLIANCE

9.1 Exposure to Investigational Product

Exposure to double-blind investigational product for the Safety Set will be summarized by visit for each treatment group and age group. The following statistics will be calculated by visit for the investigational product:

- Days of dosing: number of days on which dose information was available and non-zero
- Total dose (mg) for SHP465: total number of capsules taken × dose level
- Average daily dose (mg/day) for SHP465: the total dose / total days of dosing
- Cumulative dose (mg) for SHP465: sum of total dose since start of treatment
- Duration of exposure in weeks: the total days of dosing / 7

An appropriate statistical summary will be applied to present for each of the above statistical quantities.

Subjects will be categorized in 1 week increments by their overall duration of exposure in days and will be presented using subject counts and percentages.

In addition, person-time (overall total exposure in days) will be derived for SHP465. It is calculated as total number of days in which SHP465 was taken for each subject (total days of dosing for the Treatment period), and then sum over all subjects for the Treatment period.

All dosing information, for example, total dose, average daily dose and cumulative dose, will be listed.

If a subject is lost to follow up without returning the leftover investigational product and without providing subsequent safety information, then the subject's dose information will be treated as missing. In particular, for subjects who are lost in the follow up after the Baseline Visit (Visit 2), it will not be assumed that any of the investigational products has been ingested unless there is a post-baseline safety assessment. Subjects who do not return the investigational product but return to the site for the following visit will have the dose return information entered into the database on the scheduled visit.

9.2 Measurement of Treatment Compliance

Investigational product dosing compliance during the treatment period is defined as the total number of capsules actually taken by a subject during that period divided by the total number of capsules expected to be taken during the same period multiplied by 100. The total number of capsules actually taken is calculated by the total number of capsules dispensed minus the number of capsules returned. If a bottle is not returned, the number of capsules returned for that bottle will be imputed to zero. The number of capsules expected to be taken is calculated as the number of days the subject was in the treatment period multiplied by the number of capsules to be taken per day during the treatment period. Compliance rates will be calculated separately for SHP465 6.25 mg and placebo.

Summary statistics for investigational product compliance will be presented by treatment group (SHP465 6.25 mg and Placebo) and by age group during the treatment period for the Safety Set. The investigation product treatment compliance will be categorized as <80%, 80-120%, or >120%. The category 80-120% is considered compliant. Categorical data will be presented for the Safety Set.

Calculated compliance data for each subject will be listed.

10. PRIOR AND CONCOMITANT MEDICATION

Version WHODRUG March 2017 of the WHO drug dictionary will be used to classify prior and concomitant medications by preferred term.

Prior medication is defined as any medication taken prior to the date of first dose of investigational product.

Concomitant medication is defined as any medication with a start date prior to the date of the first dose of investigational product and continuing after the first dose of investigational product or with a start date between the dates of the first dose and last dose of investigational product, inclusive.

Any medication with a start date after the date of the last dose of investigational product (post-treatment) will not be considered a concomitant medication.

Both prior and concomitant medication usage will be summarized separately by the number and percentage of subjects receiving each medication within each preferred term by treatment group (SHP465 6.25 mg and Placebo) and overall for the Safety Set. Medications can be counted both as prior and concomitant medication. Multiple medication usage by a subject in the same category will be counted only once for the purpose of the summaries.

For ADHD medication history collected at the Screening Visit (Visit 1), version WHODRUG March 2017 of the World Health Organization (WHO) drug dictionary will be used to classify ADHD medications by preferred term. ADHD medication history will be summarized separately by the number and percentage of subjects receiving each medication within each preferred term by treatment group (SHP465 6.25 mg and Placebo) for the Safety Set. ADHD medication history will be listed for the Safety Set.

All prior, concomitant and post-treatment medications will be listed for the Safety Set.

11. EFFICACY ANALYSES

All efficacy analyses will be performed using the FAS. All statistical tests will be 2-sided hypothesis tests performed at the 0.05 level of significance. All confidence intervals (CI) will be 2-sided 95% CIs, unless stated otherwise. Only post-baseline assessments from Visit 3 (Week 1) up to Visit 6 (Week 4) that were collected on or before 2 days after the last dose of double-blind investigational product will be analyzed.

11.1 Primary Efficacy Endpoint and Analysis

The primary efficacy endpoint is defined as the change from baseline of the ADHD-RS-5 total score at Visit 6 (Week 4). Baseline ADHD-RS total score is defined as the last valid ADHD-RS-5 total score assessment prior to taking the first dose of double-blind investigational product, usually at Visit 2 (Week 0).

The primary efficacy analysis will be conducted on the FAS for the change from baseline for the ADHD-RS-5 total score, including all assessments from Visit 3 (Week 1) up to Visit 6 (Week 4). The primary efficacy endpoint will be analyzed by using the linear mixed-effects model for repeated measures (MMRM) with treatment group (SHP465 6.25 mg and Placebo), visit, age group (6-8 years vs. 9-12 years) and the interaction of treatment group with visit as factors, baseline ADHD-RS-5 total score as a covariate, and the interaction of baseline ADHD-RS-5 total score with visit adjusted in the model.

The REML (restricted maximum likelihood) method of estimation, which is default in SAS, will be utilized along with the Kenward-Roger method for estimating the covariance matrix and degrees of freedom. The model will use an unstructured covariance type.

The primary contrast of interest will be at Visit 6 (Week 4) for SHP465 6.25 mg compared with placebo.

The null and alternative hypotheses of the primary efficacy analysis are:

- Null hypothesis: There is no difference in mean change from baseline at Visit 6 (Week 4) in ADHD-RS-5 total score between SHP465 6.25 mg and placebo;
- Alternative hypothesis: There is a difference in mean change from baseline at Visit 6 (Week 4) in ADHD-RS-5 total score between SHP465 6.25 mg and placebo.

Descriptive statistics will be displayed for each treatment group and visit, as well as the least squares means and standard errors resulting from the MMRM. To compare SHP465 6.25 mg and placebo, the difference of least squares means, the corresponding 95% CI, and the effect size (Cohen's *d*) at Visit 6 (Week 4), defined as difference of the least squares means divided by the square root of the variance estimate at Visit 6 obtained from the SAS-estimated $\boldsymbol{R}$ matrix, will be displayed (Cohen, 1988). The p-value and effect size will only be presented for Visit 6 (Week 4). A corresponding line graph of the least squares means of change from baseline and associated 95% CI in ADHD-RS-5 total score by visit and treatment group will be presented.

The observed and change from baseline in ADHD-RS total score will be summarized at each applicable visit using the number of subjects, mean, SD, median, minimum, and maximum values.

The items in the ADHD-RS-5 will also be grouped into 2 subscales: hyperactivity/impulsivity (9 items for hyperactivity/impulsivity) and inattentiveness (9 items for inattentiveness). Each subscale will be analyzed using a similar model as the primary efficacy endpoint and summarized descriptively at each visit by treatment group. A corresponding line graph of the least squares means of change from baseline and associated 95% CI in ADHD-RS-5 subscale scores by visit and treatment group will be presented.

11.2 Key Secondary Efficacy Endpoint(s) and Analysis

The key secondary efficacy endpoint, CGI-I score, will be analyzed using the same analysis method (MMRM) as for the primary efficacy endpoint for the Full Analysis Set. The baseline CGI-S score will be used as the covariate. The primary contrast of interest will be at Visit 6 (Week 4) for SHP465 6.25 mg compared with placebo.

The CGI-I will be summarized descriptively at each applicable visit using the number of subjects, mean, SD, median, minimum, and maximum values.

As a supportive analysis, the key secondary efficacy measurement will also be analyzed using the proportion of subjects with an "improved" CGI-I measurement at Visit 6/ET (Week 4/ET). The CGI-I categories will be dichotomized into 2 categories; "very much improved" and "much improved" classified as "improved" (CGI-I score of 1 or 2) and all other assessed categories grouped together as "not improved" (CGI-I score of >2). If missing data exist at the Visit 6 (Week 4) visit, the visit will be imputed by carrying forward the last post-baseline observation value. The dichotomized CGI-I will be analyzed by comparing SHP465 6.25 mg and placebo on the FAS for the proportion of “improved” subjects using a Cochran-Mantel-Haenszel test stratified by age group and CGI-S value at baseline.

The observed and dichotomized CGI-I values will be summarized at each applicable visit using number of subjects and percentages as well.

A corresponding bar chart showing the percentage of subjects improved by visit and treatment group will be presented. CGI-I values and CGI-S values will be listed for FAS.

11.3 Adjustments for Multiplicity for the Primary and Key Secondary Endpoint

In order to protect the study-wide Type I error at the 2-sided 0.05 for testing across the primary and the key secondary hypotheses, the Fixed-Sequence Test procedure will be applied. The hypotheses will be tested in the order of the primary (ADHD-RS-5 Total Score) and then the secondary (CGI-I), if significant for the primary, each at the 2-sided 0.05 significance level. Both tests in the sequence are based on the MMRM.

11.4 Sensitivity Analysis Based on Missing Not at Random Assumptions

The primary efficacy analysis (MMRM) relies on the assumption that the missing data mechanism follows the MAR (missing at random) scenario. It is assumed that the probability of missing data is unrelated to the value of the variable, after controlling for other variables in the data. The likelihood-based MMRM analysis is considered an appropriate method for the statistical analysis under MAR assumptions.

The following 2 sensitivity analysis models will be used to examine robustness of the primary analysis results. The sensitivity analysis models assume different MNAR mechanisms and are within the pattern-mixture model framework.

Model 1 - Placebo multiple imputations

Rationale: The imputations are based on the distribution of placebo group responses over time. The underlying assumption is that a subject on SHP465 6.25 mg treatment with missing data follows the distribution of the placebo responses, i.e., the means and the intra-subject correlations based on the placebo responses will apply.

The model is implemented in 3 steps: 1. imputations, 2. analysis of complete data sets, and 3. inference. Sample SAS codes are provided in Section 22.

Step 1: Imputations

A total of 200 sets of posterior mean and co-variance estimates are extracted from the SAS MI procedure using the available non-missing placebo data. One hundred of the posterior sets will be applied to the SHP465 6.25 mg treatment group, the other 100 applied to the placebo group. One set of imputations for all missing values will be generated based on each variation of posterior estimates. All 100 sets for imputations within a treatment group will be ordered from 1 to 100, and combined between the SHP465 6.25 mg treatment group and placebo, for a total of 100 completely imputed data sets.

Step 2: Analysis of complete data sets

The primary endpoint will be analyzed for each of the 100 complete data sets with imputed data using an ANCOVA with treatment and age group as the factor and the baseline value as a covariate.

Step 3: Inference

The LS mean difference estimates will be averaged and the associated standard errors will be summarized based on within-imputation and between-imputation variance using the SAS MIANALYZE procedure to yield a final estimate with associated 95% CI and p-value.

Model 2 - Multiple imputations with penalties applied to dropouts

Rationale: The underlying assumption is that subjects who drop out perform worse than MAR by a penalty.

The model is implemented in 3 steps: 1. imputations and application of penalty, 2. analysis of complete data sets, and 3. inference. Sample SAS codes for steps 1a and 1b are provided in Section 22.

Step 1a: Imputations

Missing data will be multiply imputed for 100 times based on a treatment specific, multivariate normal distribution of the response over time using the SAS MI procedure with treatment in the BY statement. This step is based on the MAR assumption.

Step 1b: Application of Penalty

A penalty will then be applied to the multiply imputed values at Visit 6 (Week 4). The penalty will be a fraction of the estimated SD for the primary endpoint (the square root of the estimated element for Visit 6 of the co-variance matrix ***R*** from the primary MMRM model): (0*SD), (0.25*SD), (0.5*SD), (0.75*SD), and (1*SD).

Step 2 (analysis of complete data sets) and Step 3 (inference) are the same as Steps 2 and 3 respectively for Model 1.

11.5 Other Sensitivity Analysis

Other sensitivity analyses will be used to examine 2 different issues in this study:

1. Robustness of the key secondary efficacy analysis (MMRM) results for the CGI-I will be examined using the 2 sensitivity analysis models described in Section 11.4.

2. Impact of potential unblinding of 10 subjects due to unintentional urine drug screen performed erroneously after the final on-treatment assessments (see Section 18.1 for further details)

To evaluate whether the potential unblinding modified the efficacy results, sensitivity analyses of the primary efficacy (ADHD-RS-5 total score) and key secondary efficacy (CGI-I) analyses will be performed, in which these 10 subjects will be excluded from the FAS. The results between the primary and key secondary efficacy analyses and the sensitivity analyses will be examined.

11.6 Other Secondary Efficacy Endpoint(s) and Analysis

No other secondary efficacy endpoints are defined for this study.

11.7 Exploratory Efficacy Endpoint(s) and Analysis

The primary efficacy endpoint, change from baseline of the ADHD-RS-5 total score at Visit 6 (Week 4), will be analyzed separately for age group (6-8 years and 9-12 years), sex group (male and female), race group (white and non-white), and ethnicity group, using the same analysis method (MMRM) as for the primary efficacy endpoint, with treatment group, visit, age group (6-8 years vs. 9-12 years) and the interaction of treatment group with visit as factors, baseline ADHD-RS-5 total score as a covariate, and the interaction of baseline ADHD-RS-5 total score with visit adjusted in the model. The age group will not be included as a factor for the subgroup

analysis for age group. ADHD-RS subscales will also be analyzed separately for age group (6-8 years and 9-12 years) using a similar model as the primary efficacy endpoint.

Descriptive statistics will be displayed for each treatment group within each age group, sex group, race group, and ethnicity group as well as the least squares means and standard errors resulting from the MMRM. To compare SHP465 6.25 mg and placebo, the difference of least squares means and the corresponding 95% CI, and the effect size (defined as difference of the least squares means divided by the estimated SD from the unstructured covariance matrix) will be displayed. The effect size will only be presented for Visit 6 (Week 4). In addition, the ADHD-RS total score will be summarized at each visit by treatment group for each age group (6-8 years and 9-12 years), sex group (male and female), race group (white and non-white), and ethnicity group. ADHD-RS subscales will also be summarized at each visit by treatment group and age group (6-8 years and 9-12 years).

The key secondary efficacy endpoint, CGI-I score, will be analyzed separately for each age group (6-8 years and 9-12 years), sex group (male and female), race group (white and non-white), and ethnicity using the same analysis method (MMRM) as for the primary efficacy endpoint with treatment group, visit, and the interaction of treatment group with visit as factors, and baseline CGI-S score used as a covariate, and interaction of the baseline CGI-S score with the visit adjusted in the model.

12. SAFETY ANALYSES

Safety data will be analyzed for the Safety Set. Safety endpoints include the occurrence of TEAEs, vital signs (systolic and diastolic blood pressure, pulse, temperature, and respiration rate), height and weight, ECG results, C-SSRS, PSQ, and CSHQ. For each safety endpoint, the last value collected before the first dose of double-blind investigational product will be used as baseline for all analyses of that safety endpoint. Except for AE data, only post-baseline assessments from Visit 2 (Week 0) up to Visit 6 (Week 4) that were collected on or before 2 days after the last dose of double-blind investigational product will be analyzed. A Final on-Treatment Assessment (FoTA) will be defined as the last valid assessment obtained after Baseline and whilst on investigational product (on or before 2 days after the last dose date).

12.1 Adverse Events

Adverse events will be coded using Version 18.0 of Medical Dictionary for Regulatory Activities (MedDRA).

An AE (classified by preferred term) that occurs during the Treatment period will be considered a TEAE if it has a start date on or after the first dose of double-blind investigational product or if it has a start date before the date of the first dose of double-blind investigational product, but increases in severity on or after the date of the first dose of double-blind investigational product. An AE that occurs after more than 3 days of the last dose of double-blind investigational product will not be counted as a TEAE. If more than 1 AE with the same preferred term is reported before the date of the first dose of double-blind investigational product, then the AE with the greatest severity will be used as the benchmark for comparison to the AEs occurring during the Treatment period under the preferred term.

An overall summary of the number of subjects with TEAEs in each treatment group will be presented, including the number and percentage of subjects with any TEAEs, serious TEAEs, TEAEs related to investigational product, TEAEs leading to drug withdrawal, severe TEAEs and serious TEAEs leading to death.

The number and percentage of subjects reporting TEAEs in each treatment group will be tabulated by system organ class (SOC) and preferred term; by SOC, preferred term, and severity; and by SOC, preferred term, and relationship to investigational product. If more than 1 AE occurs with the same preferred term for the same subject, then the subject will be counted only once for that preferred term using the most severe and most related occurrence for the summarization by severity and by relationship to investigational product.

The incidence of common ($\geq$ 2% of subjects in any treatment group) will be summarized by preferred term and treatment group, sorted in decreasing frequency for subjects on SHP465 6.25 mg, and then for placebo. The incidence of TEAEs, serious TEAEs, and TEAEs leading to withdrawal will be summarized by SOC, preferred term and treatment group, sorted in decreasing frequency for subjects on SHP465 6.25 mg, and then for placebo.

For the Safety Set, listings will be presented for all AEs, TEAEs related to investigational product, TEAEs leading to drug withdrawal, and serious TEAEs.

12.1.1 Adverse Events of Special Interest

Adverse events of special interest will include psychiatric events of interest (categorized as psychosis/mania, suicidal, aggression and other events) listed in Table 2. Adverse event terms of special interest will be searched across the TEAE data and summarized with the total exposure (in subject-days).

Table 2 Psychiatric Events

Category	Preferred Term/Verbatim term
Signs and/or symptoms of psychosis/mania	Hallucination *(any type, including visual, auditory, tactile, mixed, etc.)*
	Delusion (*any type including somatic, persecutory, grandeur, reference)*
	Schizophrenia *(any type)*
	Psychotic disorder
	Transient psychosis
	Acute psychosis
	Paranoia
	Childhood psychosis
	Schizophreniform disorder
	Schizoaffective disorder
	Catatonia
	Mania
	Hypomania
Suicidal ideation and behavior	Depression suicidal
	Gun shot wound
	Intentional self-injury
	Non-accidental overdose
	Overdose
	Self injurious behavior
	Self injurious ideation
	Self-mutilation
	Suicidal ideation
	Suicidal attempt
	Completed suicide
Aggression and violent behavior	Aggression
	Anger
	Hostility
	Homicidal ideation
	Sexual offense
	Murder
	Imprisonment

Table 2 Psychiatric Events

Category	Preferred Term/Verbatim term
Miscellaneous psychiatric events (include events with serious outcome only)	Abnormal behavior
	Agitation
	Amnesia
	Confusional state
	Depressed mood
	Depression
	Disorientation
	Emotional disorder
	Emotional distress
	Feeling abnormal
	Memory impairment
	Mood altered
	Mood swings
	Personality change
	Thinking abnormal
	Anxiety
	Fearfulness
	Phobia
	Panic attack
	Sleep disturbance
	Tics
	Obsessive or compulsive behavior
	Trichotillomania

A separate listing will be provided for AEs of special interest.

Adverse events of particular interest include insomnia (including AEs with preferred terms of 'INSOMNIA', 'INITIAL INSOMNIA', 'MIDDLE INSOMNIA', and 'TERMINAL INCOMNIA'), weight decrease and decreased appetite. The preferred terms for these 3 AEs will be reviewed prior to database lock to determine which will be used to identify them.

For each of these 3 AEs, the following information will be presented by treatment group:

- Number and percentage of subjects reporting the TEAE
- Number of TEAEs
- Summary of onset day of first TEAE: The onset day of first TEAE is calculated as (onset date of first TEAE-date of first dose) + 1 for subjects who experienced the TEAE
- Summary of duration of TEAE while on study drug: The duration of each event is the number of days from the onset of the TEAE, while on study drug during SHP465 6.25 mg, until the earlier of the end date of the TEAE or the date of last dosing + 3 days (if the date of last dosing is missing, the date of last day on study will be used). TEAEs that either overlap in time, or are adjacent in time will be merged into one TEAE only for the purposes of calculating duration of event. If a subject has multiple TEAEs that

are not overlapping or adjacent in time, the durations will be averaged for that subject. Summary statistics are then based on the averaged subject durations. The calculated duration only includes the duration of TEAEs.

- Number and percentage of mild, moderate and severe TEAEs: Overlapping TEAEs with different preferred terms are counted as multiple TEAEs.
- Summary of the number of TEAEs per subject: Overlapping TEAEs with different preferred terms are counted as multiple TEAEs.
- Number and percentage of TEAEs that:
    - Resolved while on study drug
    - Ongoing
    - Dose interrupted
    - Dose withdrawn (and other action categories related to dose)
    - TEAEs leading to discontinuation
    - Number of subjects who discontinued investigational product due to the event

Also, for each of these 3 AEs, the following information will be presented by week:

- Number of subjects in the study
- Number and percentage of subjects with the TEAE
- Number of TEAEs
- Number and percentage of subjects who discontinued due to the TEAE
- Number and percentage of mild, moderate and severe TEAEs. If a subject has more than one TEAE in the same week, then the worst severity is counted.

Where, Week 1 includes TEAEs that start after Visit 2 (Baseline) through Visit 3 (Week 1), Week 2 includes TEAEs that start after Visit 3 (Week 1) through to Visit 4 (Week 2), Week 3 includes TEAEs that start after Visit 4 (Week 2) through to Visit 5 (Week 3), and Week 4 includes TEAEs that start after Visit 5 (Week 3) through to Visit 6 (Week 4).

In addition, for each subject, all occurrences of each of these 3 AEs will be listed separately by treatment group, subject, start date and stop date. The information presented will include: subject, sex, age, race, ethnicity, preferred term/adverse event, start date, stop date, onset day, study visit, last dose date, duration/imputed duration of TEAE (days), SAE, severity, effect on dosing and treatment for AE.

12.2 Clinical Laboratory Variables

Descriptive statistics for clinical laboratory values (in SI units) and changes from Baseline to Visit 6/ET as well as shift tables from Baseline to Visit 6/ET for quantitative variables will be presented by treatment group for the following clinical laboratory variables:

Biochemistry and Endocrinology

Total cholesterol
Aspartate aminotransferase
Phosphorus
Alanine aminotransferase
Sodium
Alkaline phosphatase
Potassium
Gamma-glutamyl transferase
Thyroid-stimulating hormone
Free thyroxine
Calcium
Uric acid
Blood urea nitrogen
Total bilirubin
Creatinine
Glucose
Albumin
Total protein
Lactate dehydrogenase

Hematology

Hemoglobin
Hematocrit
Red blood cells
Platelet count
White blood cell count – total and differential
Mean corpuscular hemoglobin
Mean corpuscular hemoglobin concentration
Neutrophils
Lymphocytes
Monocytes
Eosinophils
Basophils
Bands
Mean corpuscular volume

Urinalysis

Glucose
Specific gravity
Blood
Ketones
Protein
Bilirubin
pH
Urobilinogen
Color
Leukocyte esterase
Nitrate

Microscopic examination will be conducted if protein and/or blood is/are detected during urinalysis. At a minimum the microscopic examination will consist of red blood cells, white blood cells, casts, and bacteria.

Clinical laboratory test values are potentially clinically important (PCI) if they meet either the low or high PCI criteria listed in Table 3. The number and percentage of subjects with post-baseline PCI values will be tabulated. The percentages will be calculated relative to the number of subjects with available baseline values and at least 1 post-baseline assessment. The numerator is the total number of subjects with at least 1 post-baseline PCI value. A supportive listing of subjects with post-baseline PCI values will be provided including the subject number, baseline, and post-baseline values.

Table 3 Criteria for Potentially Clinically Important Laboratory Tests

Parameter	PCI Criteria
Biochemistry	
Cholesterol-H	>300mg/dl** (>7.8mmol/L)
Glucose, serum	<55mg/dl or >160mg/dl (<3.1mmol/L or >8.9mmol/L)
Bilirubin, total	>1.5 x ULN
Transaminase, SGOT, AST	>3 x ULN
Transaminase, SGPT, ALT	>3 x ULN
Gamma Glutamyl Transpeptidase (GGT)	>3 x ULN
Alkaline phosphatase	>2.5 x ULN
Lactate dehydrogenase (LDH)	>3 x ULN*
Sodium	<130mEq/L (grade 3) or >150mEq/L (mmol/L)
Potassium, serum/plasma	<3mEq/L (grade 3) or >5.5mEq/L (mmol/L)
Calcium	<8mg/dl or >11.5mg/dl (<2mmol/L or >2.9mmol/L)
Phosphorus, inorganic	<2.5mg/dl or >7.0mg/dl** (<0.8mmol/L or >2.3mmol/L)
Uric acid, serum	>10mg/dl (>0.6mmol/L or >594.8µmol/L)
Thyroid Stimulating Hormone (TSH) ~	< 0.35 µIU/ml or > 6 µIU/mL (0.35mIU/L or >6mIU/L)
Blood Urea nitrogen (BUN)	>2.5 x ULN* (or alternatively >30mg/dl**, >10.7mmol/L)
Creatinine, serum	>1.5 x ULN (or alternatively >2mg/dl**, >176.8µmol/L)
Albumin	<3g/dl (<30g/L)
Total protein, plasma or serum	<5g/dl* or >9g/dl* (<50g/L or >90g/L)
Hematology	
Hemoglobin (He.)	<10g/dl (<100g/L)
Hematocrit	<32% or >45% (<0.32 or >0.45 Fraction of 1)
Red Blood Cells (RBC)	male: <2.5 x10$_6$/µL; female: <2.0 x10$_6$/µL (male: <2.5 x 10^12/L; female: <2.0 x 10^12/L)
Mean Corpuscular Hemoglobin (MCH)	< LLN or > 32pg
Mean Corpuscular Volume (MCV)	< 70fL or > 110fL
Mean Corpuscular He. Concentration (MCHC)	< 28g/dL or > 41g/dL (<280g/L or >410g/L)
Platelet count	<75 x 10$_9$/L or >500 x 10$_9$/L
White Blood Cell Count (WBC)	<3 x 10$_9$/L or >16 x 10$_9$/L
Bands	> 5%
Neutrophils	< 15%
Eosinophils	> 10%
Basophils	> 15%
Lymphocytes	> 80%
Monocytes	> 40%
ANC (%neutrophils x WBC count)	<1.3x10e9/L
Urinalysis	
Glucose	Positive Value (excluding trace)
Blood	Positive Value (excluding trace)
Ketones	Positive Value (excluding trace)
Protein	Positive Value (excluding trace)
Bilirubin	Positive Value (excluding trace)

Table 3 Criteria for Potentially Clinically Important Laboratory Tests

Parameter	PCI Criteria
Leukocyte Esterase	Positive Value (excluding trace)
Nitrite	Positive Value (excluding trace)

*The NCI has not specified a value, Shire physicians have agreed on lab values provided.

** Values taken from the Reviewer Guidance, Conducting a Clinical Safety Review of a New Product Application and Preparing a Report on the Review, Table 7.1.7.3.2.1 pp 70-72. US DHHS FDA CDER, February 2005.

~These values were taken from Clinical Laboratory Diagnostics, ed. L. Thomas, MD, Frankfurt, Germany, 1998.

ULN: Upper limit of normal value provided by the laboratory; LLN: Lower limit of normal value provided by the laboratory

All laboratory data will be listed for the Safety Set.

12.3 Vital Signs (Including Height and Weight)

The averaged values of the 3 measurements collected for blood pressure and pulse at each visit will be used for all summaries and determination of PCIs.

Descriptive statistics for vital signs (sitting systolic and diastolic blood pressure, pulse, height, weight, and BMI) and their changes from baseline at each post-baseline visit and FoTA will be presented by treatment group. In the case of multiple assessments at post-baseline visits, the first readable value at the visit will be used in the summary tables.

Additionally, figures with the mean change from baseline ±SD of the vital signs values (sitting systolic and diastolic blood pressure, and pulse) will be presented by treatment group and visit.

Height, weight, and BMI will also be converted to percentile values based on the subject's age and sex at each visit and summarized categorically (<5th, 5th to <95th, and ≥95th) at each visit and at the FoTA. Post-baseline shifts in height, weight, and BMI percentile category (<5th, 5th to <95th, and ≥95th) from baseline will be summarized at each visit and at the FoTA. In addition, z-scores for height, weight, and BMI will be based on the subject's age at each visit and will be summarized categorically (<-2, ≥-2 to <-1, ≥-1 to <1, ≥1 to <2, and ≥2) at each visit and at the FoTA. Percentiles and z-scores will be derived using the CDC growth charts (Kuczmarski et al, 2002).

Vital sign values will be considered PCI if they meet the criteria listed in Table 4. The number and percentage of subjects with PCI post-baseline values will be summarized by visit and treatment group. The percentages will be calculated relative to the number of subjects with baseline and at least 1 post-baseline assessment. The numerator is the total number of subjects with at least 1 PCI post-baseline vital sign value. A supportive listing of subjects with post-baseline PCI values will be provided including the subject number, site, baseline, and post-baseline PCI values. A listing of all AEs for subjects with post-baseline PCI vital sign values will also be provided. In addition, for pulse, the number and percentage of subjects with at least 1 post-baseline value that fall within the categories >50-<100, 100-<110 and ≥110 beats/minute, will also be summarized by treatment group.

Measurements of oral or tympanic temperature, sitting respiratory rate will be performed at the Screening Visit (Visit 1) only.

Table 4 Criteria for Potentially Clinically Important Laboratory Tests

Vital Sign Parameter	PCI Criteria
Weight Loss (kg)	≥7% of weight (determined at baseline)
Systolic BP (mmHg), seated	>120* >120* on 2 or more consecutive visits >10 increase from baseline >10 increase from baseline on 2 or more consecutive visits >120* and increase >10 from baseline >120* and increase >10 from baseline on 2 or more consecutive visits
Diastolic BP (mmHg), seated	>80* >80* on 2 or more consecutive visits >10 increase from baseline >10 increase from baseline on 2 or more consecutive visits >80* and increase >10 from baseline >80* and increase >10 from baseline on 2 or more consecutive visits
Pulse (beats/minute)	<=50* <=50* and decrease >15 from baseline >=100* >=100* on 2 or more consecutive visits >=100* and increase >15 from baseline >=100* and increase >15 from baseline on 2 or more consecutive visits
Temperature (tympanic)	>38.0°C (100.4°F)
Temperature (oral)	>37.5°C (99.5°F)
Respiratory Rate (breaths per minute)	<10 or >24

~Given the pharmacological effect of the drugs used in ADHD and based on experience gained in previous ADHD trials, the focus is on rise of blood pressure only.

*The NCI has not specified a value, Shire physicians have agreed on values provided.

Pulse rate (beats per minute) will be considered abnormal if a value is lower than the 1st percentile or higher than the 99th percentile cut-off values listed in Table 5.

Table 5 Normal Range of Pulse Rate (beats/min)

Age	1st Percentile	99th Percentile
	Males	
6-8 years	59	114
9-11 years	56	110
12 years	52	108
	Females	
6-8 years	61	117
9-11 years	58	113
12 years	54	110

Source: National health statistics report: Resting Pulse Rate Reference Data for Children, Adolescents, and Adults: United States, 1999–2008
https://www.cdc.gov/nchs/data/nhsr/nhsr041.pdf

The number and percentage of subjects with abnormal pulse rate values for at least 1 post-baseline assessment will be tabulated by treatment group for the lower than the 1st percentile and higher than the 99th percentile cut-off values, respectively, for males and females separately. The denominator will be the number of subjects in the Safety Set (per gender) with at least 1 post-baseline pulse measurement. The numerator is the total number of subjects with at least 1 abnormal pulse record. A supportive listing of subjects with abnormal pulse rate values will be provided.

In addition, following the US Department of Health and Human Services standards, blood pressure values higher than the 95th percentile determined by sex, age, and height percentiles will be considered abnormal. The cut-off values are presented in Table 6.

Table 6 95th Percentile Cut-off Values of Blood Pressure Norms by Sex, Age, and Height Percentiles

	Systolic Blood Pressure (mmHg)								Diastolic Blood Pressure (mmHg)						
Age (y)	Percentile of Height								Percentile of Height						
	Boys														
	5%	10%	25%	50%	75%	90%	95%		5%	10%	25%	50%	75%	90%	95%
6	109	110	112	114	115	117	117		72	72	73	74	75	76	76
7	110	111	113	115	117	118	119		74	74	75	76	77	78	78
8	111	112	114	116	118	119	120		75	76	77	78	79	79	80
9	113	114	116	118	119	121	121		76	77	78	79	80	81	81
10	115	116	117	119	121	122	123		77	78	79	80	81	81	82
11	117	118	119	121	123	124	125		78	78	79	80	81	82	82
12	119	120	122	123	125	127	127		78	79	80	81	82	82	83
	Girls														
6	108	109	110	111	113	114	115		72	72	73	74	74	75	76
7	110	111	112	113	115	116	116		73	74	74	75	76	76	77
8	112	112	114	115	116	118	118		75	75	75	76	77	78	78
9	114	114	115	117	118	119	120		76	76	76	77	78	79	79
10	116	116	117	119	120	121	122		77	77	77	78	79	80	80
11	118	118	119	121	122	123	124		78	78	78	79	80	81	81
12	119	120	121	123	124	125	126		79	79	79	80	81	82	82

Source: National Heart Lung and Blood Institute; May 2004
http://www nhlbi nih.gov/guidelines/hypertension/child_tbl.htm

The number and percentage of subjects with any post-baseline blood pressure values higher than the 95th percentile (exclusive) will be tabulated by treatment group and by age group. The denominator will be the number of subjects in the Safety Set (per gender) with at least 1 post-baseline blood pressure measurement. The numerator will be the total number of subjects with at

least 1 blood pressure value higher than the 95th percentile. A supportive listing of all subjects with abnormal blood pressure values will be provided.

All vital signs data will be listed for the Safety Set.

12.4 Electrocardiogram (ECG)

Electrocardiogram results and change from baseline will be summarized at each visit and at the FoTA by treatment group. If there are multiple assessments for the interpretation at any visit, the worst interpretation will be used in the summary tables. For the interval parameters at post-baseline visits, the first readable result will be used if there are multiple assessments.

For the ECG interval parameters, baseline is defined as the average of all valid ECG measurements as the last assessment obtained prior to the first dose of the investigational product. For the ECG interpretation, baseline will be the ECG with the worst interpretation obtained at the baseline Visit (Visit 2), providing this is prior to the first dose of the investigational product.

PCI criteria will be applied to all ECG data at each visit including any repeat/unscheduled assessments and presented by visit and treatment group. PCI criteria are defined in Table 7.

Table 7 Criteria for Potentially Clinically Important ECG Values

ECG Parameter	Outlier Criteria
ECG result	Shift from a normal baseline to an abnormal finding or from an abnormal baseline to a new abnormal finding
Heart rate	≤50 beats/minute* or ≥100 beats/minute*
PR interval (increase)	≥200 msec*
QRS interval	≥120 msec*
QTcF/QTcB	≥450 msec* and <480 msec* ≥480 msec* and <500 msec* ≥ 500msec*
QT/QTc from Baseline	≥30 msec* and <60 msec* ≥60 msec*

*The NCI has not specified a value, Shire physicians have agreed on values provided.
ECG=electrocardiogram; QTcB=Bazett's corrected QT interval; QTcF=Fridericia's corrected QT interval

The number and percentage of subjects with post-baseline PCI values will be tabulated. The percentages for the observed value criteria will be calculated relative to the number of subjects with at least 1 post-baseline assessment available per parameter and visit. The percentages for the change from baseline criteria will be calculated relative to the number of subjects with available baseline and at least 1 post-baseline assessment per parameter and visit. The numerator is the total number of subjects with at least 1 PCI post-baseline ECG value.

Additionally, a summary table showing the number and percentage of subjects with the Central Reader's assessment of the ECG result as normal, abnormal or unable to evaluate, and the investigator's assessment of abnormal ECG results as not clinically significant or clinically significant at each visit and at the FoTA by treatment group will be produced. A shift table

showing the change in evaluation from baseline to the FoTA by treatment group will also be produced.

Listings of ECG data including the central reader's assessment and investigator's interpretation by individual subject will be produced. Separate listings will be produced for subjects with ECG results meeting the PCI criteria. Data from unscheduled visits will be listed, but not summarized.

12.5 Other Safety Variables

12.5.1 Columbia-Suicide Severity Rating Scale (C-SSRS)

The C-SSRS 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 occurred.

The C-SSRS contains two required items pertaining to suicidal ideation, four required items pertaining to suicidal behavior, and one required item pertaining to non-suicidal self-injurious behavior. There are eight additional suicidal ideation items and two additional suicidal behavior items which are completed in cases of positive responses for other items, as well as two items for suicide and suicide behavior present during the interview. Most items are rated on a dichotomous scale (yes or no) or 3- or 5-point Likert scale. In addition, the total number of attempts (including interrupted and aborted attempts) is recorded. In the event of a positive categorical response the interviewer can provide text or narrative that further describes the thought or behavior.

Two versions of the C-SSRS are used in this study:

- The "Baseline" version will be administered at the Screening Visit (Visit 1) and will be completed for all subjects.
- The "Since Last Visit" version will be completed for all subjects at all study visits after the Screening Visit (Visit 1).

Listings of the C-SSRS data will be provided for subjects with a positive response.

12.5.2 Post Sleep Questionnaire (PSQ)

The PSQ is a 7-item questionnaire typically used to assess sleep quality with pharmacologic treatment. The questionnaire collects data on average time to sleep, sleep latency, frequency of interrupted sleep, duration of interrupted sleep, total sleep time and sleep quality over the last week. The PSQ will be completed by the parent/LAR with the subject and the responses will be reviewed by the clinician during the study visit. The PSQ will be completed at Baseline (Visit 2), and each visit through Visit 6/ET.

The PSQ results will be summarized by treatment and visit. Continuous variables, for example, time to go to sleep and time spent awake, will be summarized using number of subjects, mean and standard deviation, median, and minimum and maximum values. Categorical values, for example, quality of sleep, will be summarized using number of subjects and percentages for each category.

All PSQ data will be listed for the Safety Set.

12.5.3 Children's Sleep Habit Questionnaire (CSHQ)

The Child's Sleep Habits Questionnaire (CHSQ) consists of 33 items for scoring and several extra items intended to provide administrators with other potentially useful information about respondents. The instrument evaluates the child's sleep based on behavior within 8 different subscales: bedtime resistance, sleep-onset delay, sleep duration, sleep anxiety, night wakings, parasomnias, sleep-disordered breathing, and daytime sleepiness.

1. Bedtime resistance (sum of the responses for Goes to bed at same time, Falls asleep in own bed, Falls asleep in other's bed, Needs parent in room to sleep, Struggles at bedtime and Afraid of sleeping alone)
2. Sleep-onset delay (Falls asleep in 20 minutes item)
3. Sleep duration (sum of the responses for Sleeps too little, Sleeps the right amount and Sleeps same amount each day)
4. Sleep anxiety (sum of the responses for Needs parent in room to sleep, Afraid of sleeping in the dark, Afraid of sleeping alone and Trouble sleeping away)
5. Night wakings (sum of the responses for Moves to other's bed in night, Awakes once during night and Awakes more than once)
6. Parasomnias (sum of the responses for Wets the bed at night, Talks during sleep, Restless and moves a lot, Sleepwalks, Grinds teeth during sleep, Awakens screaming, sweating and Alarmed by scary dream)
7. Sleep-disordered breathing (sum of the responses for Snores loudly, Stops breathing and Snorts and gasps)
8. Daytime sleepiness (sum of the responses for Wakes by himself, Wakes up in negative mood, Others wake child, Hard time getting out of bed, Takes long time to be alert, Seems tired, Watching TV and Riding in car)

Total sleep disturbance score is sum of the 8 subscale scores minus Needs parent in room to sleep and Afraid of sleeping alone (as Needs parent in room to sleep and Afraid of sleeping alone are included in two subscales (1 and 4) and need to be included once).

Total sleep disturbance score, each subscale score and each individual item will be summarized using descriptive statistics (n, mean, SD, minimum, median, and maximum) at each visit (and Visit 6/ET) by treatment group. All CSHQ data will be listed for the Safety Set.

12.6 Exploratory Safety Endpoints

Safety data for TEAEs, vital signs, ECGs, clinical laboratory test results, PSQ, and CSHQ will be descriptively summarized by treatment within each age group (6-8 years and 9-12 years). TEAEs will also be descriptively summarized for overall AEs, AEs by SOC and PT, frequently occurring AEs within each sex group (male and female), race group (white and non-white), and ethnic group.

13. CLINICAL PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES

Pharmacokinetic and pharmacodynamics data will not be collected for this study.

14. OTHER ANALYSES

No other analyses are planned for this study.

15. INTERIM ANALYSIS

In order to maintain sufficient study power to detect a clinically meaningful treatment effect for the primary efficacy endpoint, a blinded interim analysis at the late stage of the trial (when approximately 75% of all randomized subjects have either completed or discontinued from the study) will be performed to reassess the sample size in case of an underestimated variability postulated at the design stage.

Using the cumulative real data, a blinded, pooled analysis of both treatment groups for estimating variability will be conducted. If the re-estimated pooled standard deviation is larger than the 14 postulated at the design stage, the final total number of subjects to be enrolled will be calculated using the re-estimated pooled standard deviation together with the assumed treatment difference of 11.9 (Friede et al, 2006). If the re-estimated pooled standard deviation is smaller than 14, the sample size will not be adjusted. Using the pooled common variance estimated to do a blinded sample size re-estimation does not inflate the Type I error. The details for estimating the pooled SD are included in Appendix 1.

16. DATA MONITORING/REVIEW COMMITTEE

An external independent Data Monitoring Committee (DMC) was set up to review the data pertaining to safety and tolerability of the study therapy for the duration of this program, which includes studies SHP465-112, SHP465-308, and SHP465-309. The DMC will review the data pertaining to safety and tolerability of the study therapy. Active surveillance for insomnia and decreased appetite leading to weight loss are included. Confidentiality of the unblinded DMC analyses is a critical concern and to address this, an unblinded independent reporting team will be identified within a CRO. The independent reporting team will have no involvement in the conduct of the study. Further details regarding the DMC can be found in the DMC charter, which will be available before the administration of investigational product.

17. COMPUTER METHODS

Statistical analyses will be performed using Version 9.3 (or newer) of SAS® on a suitably qualified environment.

18. CHANGES TO ANALYSES SPECIFIED IN PROTOCOL

18.1 Handling of the Unintentional Urine Drug Screen

At the end of this study, after completing the final on-treatment efficacy and safety assessments, urine drug screens were erroneously performed for one subject at last visit (Visit 6) and 9 subjects who were rolled over to the long-term safety extension study (SHP465-308). Thus, the site and vendor personnel might have been unblinded to treatment drug information of these subjects although a quick action was taken to remove and restrict access to the urine drug screen results. To assess any compromise to data integrity following the potential unblinding for the ten (10) subjects, Shire and PPD study teams performed a thorough investigation of data collected for study primary efficacy, secondary efficacy and safety analyses, including eCOA data collected on the ERT devices and data entered into the EDC.

This investigation showed no patterns of data changes or data integrity concerns which would be indicative of actions based on unblinded information. It was concluded that there was no evidence of the introduction of bias that would affect the aggregate data reported. In addition, Shire respects the value for each subject participant in the study, therefore decided to include all subjects in the final study analysis.

As a conservative measure to evaluate whether the potential unblinding modified the efficacy results due to the unintentional urine drug screen, sensitivity analyses of the primary efficacy (ADHD-RS-5 total score) and key secondary efficacy (CGI-I) analyses will be performed, in which these 10 subjects were excluded from the FAS (see Section 11.5).

19. DATA HANDLING CONVENTIONS

19.1 General Data Reporting Conventions

Continuous variables will be summarized using the following descriptive statistics: n, mean, median, SD, minimum, and maximum. Unless specified otherwise, summary statistics will be presented to 1 more significant digit than the raw data. The minimum and maximum values will be presented to the same number of decimal places as the raw data; the mean and median will be presented to 1 more decimal place than the raw data; and the SD and standard error will be presented to 2 more decimal places than the raw data.

Categorical and count variables will be summarized by the number of subjects (n) and the percent of subjects in each category. Percentages will be presented to 1 decimal place.

P-values will generally be presented to 4 decimal places; values less than 0.001 will be presented as <0.001.

19.2 Derived Efficacy Endpoints

The ADHD-RS-5 total score will be calculated as the sum of the individual items from the ADHD-RS-5. The ADHD-RS-5 is an 18-item questionnaire that requires the respondent to rate the frequency of occurrence of ADHD symptoms as defined by Diagnostic and Statistical Manual of Mental Disorders criteria. Each item is scored on a 4-point scale ranging from 0 (never or rarely) to 3 (very often) with total scores ranging from 0 to 54. As specified in Section 11.1, the 18 items will also be grouped into 2 subscales: hyperactivity/impulsivity (9 items for hyperactivity/impulsivity) and inattentiveness (9 items for inattentiveness), with subscale scores ranging from 0-27.

A missing individual item in the ADHD-RS-5 is imputed as follows (DuPaul et al., 2016): if only 1 single item is missing in a given subscale, the mean score for all other items in the subscale for the specific visit is imputed as the score rounded up to the nearest integer for the missing score. The total score is computed as the sum of the imputed subscale scores. If more than 1 item is missing in a subscale then the subscale score and total score would be missing.

19.3 Association of Early Termination Assessments to Scheduled Visits

For purposes of reporting early termination assessments during the study, each early termination visit will be assigned the next nominal visit number after the last completed visit. This rule applies to both efficacy variables (e.g., ADHD-RS-5 total score) and safety variables (e.g., vital signs) that are analyzed and/or summarized by visit.

19.4 Repeated or Unscheduled Assessments of Safety Parameters

If a subject has repeated assessments before the start of double-blind investigational product, then the results from the final assessment made prior to the start of double-blind investigational product will be used as baseline.

If a subject has repeated assessments at any post-baseline visit, the first recorded assessment will be assigned to the visit for generating descriptive statistics. For vital signs (sitting systolic and diastolic blood pressure, and pulse), this will be the average of the 3 measurements from the first recorded assessment.

If final on-treatment safety assessments are repeated or unscheduled, the last post-baseline assessment will be used as the final on-treatment assessment for generating descriptive statistics. For vital signs (sitting systolic and diastolic blood pressure, and pulse), this will be the average of the 3 measurements from the last post-baseline assessment.

However, all post-baseline assessments will be used for PCI value determination and all assessments will be presented in the data listings.

All repeat and unscheduled visits will be listed.

19.5 Missing Dispensed or Returned Date or Number of Capsules of Investigational Product

When the date of the last dose of double-blind investigational product is missing for a subject in the Safety Set, all efforts should be made to obtain the date from the investigator. If it is still missing after all efforts, then the last visit date when investigational product was returned will be used in the calculation of investigational product exposure.

19.6 Missing Date Information for Prior or Concomitant Medications

For prior or concomitant medications, incomplete (i.e., partially missing) start date and/or stop date will be imputed. When the start date and the stop date are both incomplete for a subject, impute the start date first.

19.6.1 Incomplete Start Date

The following rules will be applied to impute the missing numerical fields. If the stop date is complete and the imputed start date is after the stop date, then the start date will be imputed using the stop date.

Missing day and month

- If the year of the incomplete start date is the same as the year of the date of the first dose of double-blind investigational product, then the day and month of the date of the first dose of double-blind investigational product will be assigned to the missing fields.
- If the year of the incomplete start date is before the year of the date of the first dose of double-blind investigational product, then 31 December will be assigned to the missing fields.
- If the year of the incomplete start date is after the year of the date of the first dose of double-blind investigational product, then 01 January will be assigned to the missing fields.

Missing month only

- The day will be treated as missing and both month and day will be replaced according to the above procedure.

Missing day only

- If the month and year of the incomplete start date are the same as the month and year of the date of the first dose of double-blind investigational product, then the day of the date of the first dose of double-blind investigational product will be assigned to the missing day.
- If either the year is before the year of the date of the first dose of double-blind investigational product or if both years are the same but the month is before the month of the date of the first dose of double-blind investigational product, then the last day of the month will be assigned to the missing day.
- If either the year is after the year of the date of the first dose of double-blind investigational product or if both years are the same but the month is after the month of the date of the first dose of double-blind investigational product, then the first day of the month will be assigned to the missing day.

19.6.2 Incomplete Stop Date

The following rules will be applied to impute the missing numerical fields. If the date of the last dose of double-blind investigational product is missing, then replace it with the last visit date. If the imputed stop date is before the start date (imputed or non-imputed start date), then the imputed stop date will be equal to the start date.

Missing day and month

- If the year of the incomplete stop date is the same as the year as of the date of the last dose of double-blind investigational product, then the day and month of the date of the last dose of double-blind investigational product will be assigned to the missing fields.
- If the year of the incomplete stop date is before the year of the date of the last dose of double-blind investigational product, then 31 December will be assigned to the missing fields.
- If the year of the incomplete stop date is after the year of the date of the last dose of double-blind investigational product, then 01 January will be assigned to the missing fields.

Missing month only

- The day will be treated as missing and both month and day will be replaced according to the above procedure.

Missing day only

- If the month and year of the incomplete stop date are the same as the month and year of the date of the last dose of double-blind investigational product, then the day of the date of the last dose of double-blind investigational product will be assigned to the missing day.
- If either the year is before the year of the date of the last dose of double-blind investigational product or if both years are the same but the month is before the month of the date of the last

dose of double-blind investigational product, then the last day of the month will be assigned to the missing day.

- If either the year is after the year of the last dose of double-blind investigational product or if both years are the same but the month is after the month of the date of the last dose of double-blind investigational product, then the first day of the month will be assigned to the missing day.

19.7 Missing Date Information for Adverse Events

For AEs, incomplete (i.e., partially missing) start dates will be imputed and will follow the same rules as in Section 19.6.1. Incomplete stop dates will not be imputed.

19.8 Missing Severity Assessment for Adverse Events

If severity is missing for an AE starting prior to the date of the first dose of double-blind investigational product, then a severity of "Mild" will be assigned. If the severity is missing for an AE starting on or after the date of the first dose of double-blind investigational product, then a severity of "Severe" will be assigned. The imputed values for severity assessment will be used for incidence summaries, while the actual values will be used in data listings.

19.9 Missing Relationship to Investigational Product for Adverse Events

If the relationship to investigational product is missing for an AE starting on or after the date of the first dose of double-blind investigational product, a causality of "Related" will be assigned. The imputed values for relationship to double-blind investigational product will be used for incidence summaries, while the actual values will be presented in data listings.

19.10 Character Values of Clinical Laboratory Variables

The actual values of clinical laboratory variables as reported in the database will be presented in data listings. No coded values (e.g., when a character string is reported for a numerical variable) are necessary.

20. REFERENCES

Cohen J, 1988. *Statistical Power Analysis for the Behavioral Sciences,* Second Edition. Hillsdale, NJ: Lawrence Earlbaum Associates.

DuPaul GJ, Power TJ, Anastopoulos AD, Reid R, 2016. *ADHD Rating Scale-5 for Children and Adolescents: Checklists, Norms, and Clinical Interpretation*, New York, NY: Guilford Press.

Friede T, Kieser M, 2006. *Sample Size Recalculation in Internal Pilot Study Designs: A Review.* Biometrical Journal 48 (4): 537-55.

Kuczmarski RJ, Ogden CL, Guo SS, et al., 2002. 2000 CDC growth charts for the United States: Methods and development. National Center for Health Statistics, *Vital Health Stat* 11(246).

21. TABLE OF CONTENTS FOR FIGURES, TABLES, AND LISTINGS

Table of contents for figures, tables and listings will be documented separately.

22. APPENDICES

22.1 Appendix 1 Algorithms for Sample Size Re-estimation

Based on the rationale that the total variance is similar to the within group variance as long as the treatment effect is not too large, a pooled common variance can be estimated by using blinded cumulative primary efficacy data that are collected during the study, up to the time of sample size re-estimation. The pooled common variance can be used along with the pre-specified treatment difference to re-estimate the sample size. Below are steps in details:

- Ignoring treatment difference and applying the formula below to estimate the pooled sample variance, which is an unbiased estimate of the variance.

$$S^2_{pooled} = \frac{1}{n_1 - 1} \sum_{i,j} (X_{ij} - \bar{X})^2,$$

  where n_1 is total number of subjects who either completed or discontinued from the study at the time point of the sample size re-estimation. $i=1,2$; $j=1,\cdots,n_{1i}$ and $n_1 = n_{11} + n_{12}$.

- Let Δ^* be the alternative hypothesis of treatment mean difference that is pre-specified as 11.9, for which the study is powered. Hence, a common variance that can be used for the sample size re-estimation is:

$$S^2_{final} = S^2_{pooled} - \frac{n_1}{4(n_1 - 1)} \Delta^{*2}$$

- Applying S^2_{final} and the assumed true mean treatment effect difference together, an updated effect size can be calculated. Using the effect size, the sample size can be re-estimated (Friede et al, 2006).

Furthermore, the blinded sample size re-estimation only allow upwards adjustments of the initially (at the study design stage) planned sample size, and using the pooled common variance estimated above to do a blinded sample size-estimation does not inflate the Type I error rate (Friede et al, 2006).

22.2 Appendix 2 Sample SAS Codes

Mixed Model Repeated Measures (MMRM)

```
Proc mixed  /* select derived data set */ ;
      Class trt subjid visit agegp;
      Model chg = trt visit agegp trt*visit base base*visit / ddfm=kr;
      Repeated visit / subject=subjid type=UN;
      /* LSMean, Estimate, and/or LSMestimate statement(s) */
      /* Data output or object delivery statement(s) */
Run;
```

Mixed Model Repeated Measures (MMRM) for subgroup analysis

```
Proc mixed  /* select derived data set */ ; Class trt
      subjid visit;
      Model chg = trt visit trt*visit base base*visit / ddfm=kr; Repeated visit /
      subject=subjid type=UN;
      by subgroup;
      /* LSMean, Estimate, and/or LSMestimate statement(s) */
      /* Data output or object delivery statement(s) */
 Run;
```

Analysis of Covariance (ANCOVA)

```
Proc glm /* select derived data set */ ;
      Class trt ;
      Model chg = trt agegp base/alpha=.05;
      Lsmeans trt/stderr pdiff;
      /* Data output or object delivery statement(s) */
Run;
```

Missing Not at Random (MMNAR) Sensitivity Analysis

Model 1: Placebo Multiple Imputation

```
* Step 1;
proc mi data=unimputed out=discard nimpute=200 seed=309 noprint;
        where treatment=1; * where 1 stands for placebo;
        var baseline chy1-chy4;
        mcmc outest=posteriors; * here we pick up the PLA posteriors;
run;

* 100 posteriors for the placebo, rest for SHP465 6.25 mg;
data posteriors(type=est);
        set posteriors;
```

```
            if 1<=_imputation_<=100 then do; * assign placebo posteriors ;
                  treatment=1;
            end;
          if 100<_imputation_<=200 then do; * assign posteriors to SHP465 6.25 mg;
                  treatment=2;
                 _imputation_=_imputation_-100;
           end;
  run;
  proc sort data=unimputed;
     by treatment;
  run;

  proc mi data=unimputed out=imputed;
     by treatment;
     var baseline chy1-chy4;
     mcmc inest=posteriors; * here we use the placebo posteriors;
  run;


  * Step 2;
data endpoint;
     set imputed;
     endpoint=chy4;
run;

  proc sort data=endpoint;
     by _imputation_;
  run;

  proc glm data=endpoint;
     by _imputation_;
     class treatment agegp;
     model endpoint = baseline treatment agegp / solution;
     lsmeans treatment;
     estimate 'SHP465 6.25 mg vs placebo' treatment -1  1;
     ods output estimates=est;
  run;

  * Step 3;
  proc sort data=est;
     by label_imputation_;
  run;

  proc mianalyze data=est;
     by label;
     modeleffects estimate;
     stderr stderr;
  run;
```

Model 2: Multiple Imputations with Penalties Applied to Dropouts

```
* Note: dataset "unimputed" has one row per subject, repeated measures are structured in columns;

* Step 1;
data unimputed;
    set unimputed;
    missing=(Chy4 =.);
run;

proc sort data=unimputed;
    by treatment;
run;

proc mi data=unimputed out=imputed nimpute=&nimpute seed=309 noprint;
    by treatment;
    var baseline chy1-chy4;
run;

%let SD=10;
* This will be replaced by the estimated SD[chy5] from the MMRM co-variance matrix R (use R or RCORR option in
REPEATED statement under PROC MIXED);
%let penalty=0.25*&SD;
* Repeat this step for various values (0.25, 0.5, 0.75 and 1) as per SAP;

data imputed;
    set imputed;
    if missing then chy4=chy4+ &penalty; * + or -, depending on the measure;
run;

* Step 2 (analysis of completed data sets) and Step 3 (inference) are the same as Steps 2 and 3 respectively for
Model 1.
```