STATISTICAL ANALYSIS PLAN

Protocol No.:	SPD489-347
Protocol Title:	A Phase 3, Randomized, Double-blind, Multicenter, Parallel-group, Placebo-controlled, Fixed-dose Safety and Efficacy Study of SPD489 Compared with Placebo in Preschool Children Aged 4-5 Years with Attention-deficit/Hyperactivity Disorder
Drug:	SPD489, Lisdexamfetamine dimesylate
Sponsor:	Shire Development LLC and International Affiliates 300 Shire Way, Lexington, MA 02421
Version No. and Date	Version 1.1; Date: 07 June 2017

TABLE OF CONTENTS

LIST OF ABBREVIATIONS 5

1. INTRODUCTION 6

2. STUDY DESIGN 7

2.1 General Study Design 7
2.2 Randomization 8
2.3 Blinding 8
2.4 Schedule of Assessments 10
2.5 Determination of Sample Size 13
2.6 Interim Analysis and Multiplicity Adjustments for Type I Error Control 14
2.6.1 Waiving of Interim Analysis 14
2.6.2 Interim Analysis 14
2.7 Decision Tree of Efficacy Analyses 16

3. OBJECTIVES 17

3.1 Primary Objective 17
3.2 Secondary Objectives 17

4. SUBJECT POPULATION SETS 18

4.1 Screened Set 18
4.2 Randomized Set 18
4.3 Safety Analysis Set 18
4.4 Full Analysis Set 18
4.5 Dose-response Analysis Set 18

5. SUBJECT DISPOSITION 20

6. PROTOCOL DEVIATIONS 21

7. DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS 22

7.1 Demographic and Other Baseline Characteristics 22
7.2 Medical History 22

8. EXTENT OF EXPOSURE AND TREATMENT COMPLIANCE 23

8.1 Exposure to Investigational Product 23
8.2 Measurement of Treatment Compliance 23

9. PRIOR AND CONCOMITANT MEDICATION 25

10. EFFICACY ANALYSES 26
10.1 Primary Efficacy Endpoint and Analysis 26
10.2 Sensitivity Analyses for the Primary Efficacy – Missing Not at Random 27
10.3 Key Secondary Efficacy Endpoint and Analysis 29
10.4 Dose-response Analyses 29
10.4.1 Dose-response Candidate Curves 29
10.4.2 Dose-response Relationship Evaluations 30
10.4.3 Minimum Effective Dose 31
10.4.4 Additional Analyses 31
10.5 Supportive Analyses 32
10.5.1 Subgroup Analysis 32
10.5.2 Other Analyses 32

11. SAFETY ANALYSES 33
11.1 Adverse Events 33
11.2 Clinical Laboratory Variables 34
11.3 Vital Signs 36
11.4 Electrocardiogram (ECG) 38
11.5 Other Safety Variables 39
11.6 Children's Sleep Habits Questionnaire and Sleep Diary 39
11.7 Columbia-Suicide Severity Rating Scale 40
11.8 Follow-up Phone Call 41

12. CLINICAL PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES 42

13. OTHER ANALYSES 43

14. INTERIM ANALYSIS 44

15. DATA MONITORING/REVIEW COMMITTEE 45

16. COMPUTER METHODS 46

17. CHANGES TO ANALYSES SPECIFIED IN PROTOCOL 47

18. DATA HANDLING CONVENTIONS 48
18.1 General Data Reporting Conventions 48
18.2 Derived Efficacy Endpoints 48
18.3 Repeated or Unscheduled Assessments of Safety Parameters 49
18.4 Missing Date of Investigational Product 49
18.5 Missing Date Information for Prior or Concomitant Medications 49

18.6 Incomplete Start Date 49
18.7 Incomplete Stop Date 50
18.8 Missing Date Information for Adverse Events 51
18.9 Missing Severity Assessment for Adverse Events 51
18.10 Missing Relationship to Investigation Product for Adverse Events 51
18.11 Character Values of Clinical Laboratory Variables 51

19. REFERENCES 52

20. APPENDIX 53
20.1 Appendix 1 Algorithms for Deviation of the Pooled Common Variance 53
20.2 Appendix 2 Sample SAS Codes 54

LIST OF TABLES AND FIGURES

Figure 1: Study Design Flow Chart 7
Table 1: Schedule of Assessments 9
Table 2: Criteria for Potentially Clinically Important Laboratory Tests 35
Table 3: Criteria for Potentially Clinically Significant Vital Signs 37
Table 4: Criteria for Potentially Clinically Important ECG Values 38
Table 5: Examples for Coding of Special Character Values for Clinical Laboratory Variables 51

LIST OF ABBREVIATIONS

ADHD	attention deficit/hyperactivity disorder
ADHD-RS-IV	ADHD Rating Scale-IV
AE	adverse event
BMI	body mass index
CDC	Centers for Disease Control and Prevention
CGI-I	Clinical Global Impression – Global Improvement
CGI-S	Clinical Global Impression – Severity of Illness
CI	confidence interval
CSHQ	Children's Sleep Habits Questionnaire
C-SSRS	Columbia Suicide Rating Scale
DINFC	date of informed consent
DMC	data monitoring committee
DOB	date of birth
ECG	electrocardiogram
ET	early termination
FAS	Full Analysis Set
K-SADS-PL	Kiddie-Schedule for Affective Disorders and Schizophrenia for School-age Children-Present and Lifetime Version-Diagnostic Interview
LS	least squares
MAR	missing at random
MED	minimum effect dose
MMRM	mixed-effects model for repeated measures
MNAR	missing not at random
PCI	potentially clinically important
PT	preferred term
SD	standard deviation
SE	standard error
SOC	system organ class
TEAE	treatment-emergent adverse event

1. INTRODUCTION

This statistical analysis plan provides a technical and detailed elaboration of the statistical analyses of efficacy and safety data as described in the final study protocol. Specifications for tables, figures, and listings will be provided in a separate document.

2. STUDY DESIGN

2.1 General Study Design

This is a Phase 3, randomized, double-blind, multicenter, parallel-group, placebo-controlled, fixed-dose study to evaluate the safety and efficacy of SPD489 compared to placebo, administered as a daily morning dose in the treatment of preschool children 4-5 years of age, diagnosed with attention deficit/hyperactivity disorder (ADHD).

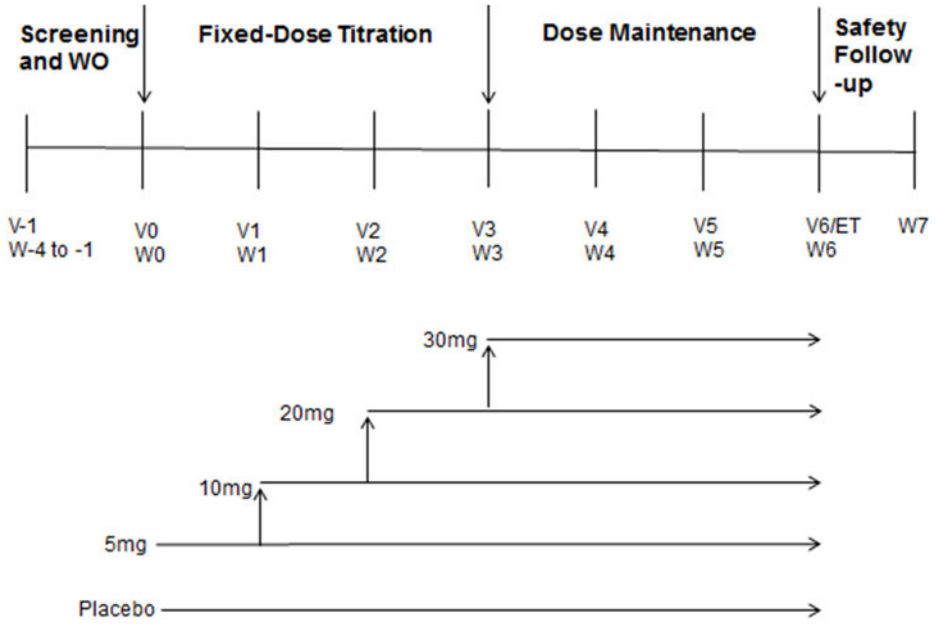
A randomized design is used to avoid baseline characteristics selection bias for different treatment groups on the part of the participants and investigators. Blinded treatment is used to reduce potential bias during data collection and evaluation of clinical results. The placebo arm is used to control the subject's expectation that a treatment would have an effect, to optimize sensitivity, to maintain the scientific rigor of the study, and to validate the study internally.

The study has 4 periods: 1) Screening and Washout; 2) Fixed-dose titration; 3) Dose Maintenance; and 4) Safety Follow-up. The duration of the double-blind evaluation period (Fixed-dose titration and Dose-maintenance Periods) will be 6 weeks.

The study will be conducted in up to 60 sites in the North America. The subject's maximum duration of participation is expected to be approximately 11 weeks. The study will be completed in approximately 24 months. Subjects will be randomized at baseline (Visit 0) in a 5:5:5:5:6 ratio to SPD489 5, 10, 20, 30mg or placebo group. Approximately 25% of the subjects enrolled will be female. Subjects are titrated to the randomly assigned dose strength of the investigational product (refer to SPD489-347 Protocol, Section 6.2.3).

SPD489 will be provided in dose strengths of 5, 10, 20 and 30 mg, and matching placebo will also be provided. Visits and phases are presented graphically in Figure 1.

Figure 1: Study Design Flow Chart



ET = early termination, V = visit, W = week, WO = washout.

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.

Individual subject treatment is automatically assigned by the interactive voice/web response system. 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.

Approximately 245 subjects will be screened to randomize approximately 195 subjects. Subjects who meet all the inclusion criteria and none of the exclusion criteria will be randomized in a 5:5:5:5:6 ratio to SPD489 5, 10, 20, 30 mg or placebo, using the interactive voice/web response system.

2.3 Blinding

This is a double-blind study. All investigational product (SPD489 5, 10, 20, 30 mg or placebo) is over-encapsulated and appears identical, in order to protect the blinding of the study.

Data that may potentially unblind the treatment assignment (ie, investigational product serum concentrations, antibodies to investigational product, 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 for an individual subject is required for further treatment of the subject. Prior to unblinding, and if the situation allows, the Investigator should first contact the Medical Monitor.

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 in the interactive voice/web response system and the source documents. Upon breaking the blind, the subject will be withdrawn from the study, but should be followed up for safety purposes.

2.4 Schedule of Assessments

Table 1: Schedule of Assessments

Period	Screening and Washout			Dose-titration			Dose-maintenance			Safety Follow-up
Visit[a]	-1 (Screening)	Washout Telephone Call	0 (Baseline)	1	2	3	4	5	6/ ET	Telephone Call
Assessment Week	-4 to -1	-1	0	1	2	3	4	5	6	7
Assessment Day	-28 to -1		0	7	14	21	28	35	42	49
Informed consent/assent	✓									
Psychiatric evaluation (utilizing the K-SADS-PL)	✓									
Peabody Picture Vocabulary Test	✓									
Inclusion/exclusion criteria	✓	✓[b]	✓[b]							
Demographics	✓									
Randomization			✓							
Medical and medication history[c]	✓									
Physical examination	✓		✓[d]						✓	
Vital signs[e]	✓		✓	✓	✓	✓	✓	✓	✓	
Height[f]	✓								✓	
Body weight[f]	✓		✓	✓	✓	✓	✓	✓	✓	
Calculate BMI[g]	✓									
Clinical laboratory tests[h, i]	✓		✓[d]						✓	
12-lead ECG[j]	✓		✓[d]	✓	✓	✓	✓	✓	✓	
ADHD-RS-IV Preschool Version[k]			✓	✓	✓	✓	✓	✓	✓	
CGI-S[l]			✓							
CGI-I[k]				✓	✓	✓	✓	✓	✓	
CSHQ[k]			✓	✓	✓	✓	✓	✓	✓	

Table 1: Schedule of Assessments

Period	Screening and Washout			Dose-titration			Dose-maintenance			Safety Follow-up
Visit[a]	-1 (Screening)	Washout Telephone Call	0 (Baseline)	1	2	3	4	5	6/ ET	Telephone Call
Assessment Week	**-4 to -1**	**-1**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Assessment Day	**-28 to -1**		**0**	**7**	**14**	**21**	**28**	**35**	**42**	**49**
Sleep diary [l]			✓	✓	✓	✓	✓	✓	✓	
C-SSRS [k]	✓		✓	✓	✓	✓	✓	✓	✓	
Suitability of subject to remain in study [m]			✓	✓	✓	✓	✓	✓		
Investigator dose assessment				✓	✓	✓	✓	✓		
Access IWRS	✓		✓	✓	✓	✓	✓	✓	✓	
Investigational product distributed			✓	✓	✓	✓	✓	✓		
Investigational product returned				✓	✓	✓	✓	✓	✓	
Investigational product compliance				✓	✓	✓	✓	✓	✓	
Concomitant medications	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Adverse events	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

ADHD-RS-IV=ADHD Rating Scale-IV; BMI=body mass index; CGI-I=Clinical Global Impressions-Global Improvement; CGI-S=Clinical Global Impressions-Severity of Illness; CSHQ=Children's Sleep Habits Questionnaire; C-SSRS=Columbia-Suicide Severity Rating Scale; ECG=Electrocardiogram; ET=early termination; IWRS= Interactive Web Response System; K-SADS-PL=Kiddie-Schedule for Affective Disorders and Schizophrenia for School-age Children-Present and Lifetime Version-Diagnostic Interview

[a] Visit windows are ±3 days [in reference to the baseline visit (Visit 0) date] during the Fixed-dose titration and Dose-maintenance Periods. Safety Follow-up Telephone Call window is +3 days.

[b] Inclusion/exclusion criteria must be reviewed at the Washout Telephone Call and at the baseline visit (Visit 0).

[c] Medication history will include a lifetime history of pharmacologic and non-pharmacologic therapies for ADHD.

[d] An abbreviated physical examination, 12-lead ECG, and all clinical laboratory tests must be repeated and results reviewed by the investigator prior to the baseline visit (Visit 0) if more than 30 days have elapsed since the safety measurements at the screening visit (Visit -1) were collected. Results must be obtained and reviewed prior to determining eligibility.

[e] Includes oral or tympanic temperature, sitting blood pressure, pulse, and respiratory rate. Measurements of blood pressure and pulse will be performed at each visit to the site. Measurements of temperature and respiratory rate will be performed at the screening visit (Visit -1) and Visit 6/ET only. Blood pressure, pulse, and respiratory rate will be

Table 1: Schedule of Assessments

Period	Screening and Washout			Dose-titration			Dose-maintenance			Safety Follow-up
Visit[a]	-1 (Screening)	Washout Telephone Call	0 (Baseline)	1	2	3	4	5	6/ ET	Telephone Call
Assessment Week	-4 to -1	-1	0	1	2	3	4	5	6	7
Assessment Day	-28 to -1		0	7	14	21	28	35	42	49

determined after subjects have remained seated for a minimum of 5 minutes. Measurement of blood pressure and pulse will be collected 3 times (with approximately 2 minutes in between each collection) using a manual cuff. The average of each set of 3 measurements will be used to determine continued participation in the study. Blood pressure measurements should be taken prior to the collection of blood draws.

[f] Height and weight to be measured without shoes.

[g] BMI will be recorded at Screening for eligibility. BMI will be derived from height and weight during the course of the study for analysis.

[h] Clinical laboratory tests will include hematology, chemistry, endocrinology, and urinalysis.

[i] Patients will have the option for blood draws to be collected by a home health care professional at their home. Home draws have a ±1 day window from the study visit date, with the exception of the 6/ET home draw which has a -1 day window.

[j] Three ECGs, taken approximately 5 minutes apart, are to be collected so that appropriate baseline intervals are established. Results must be reviewed prior to determining eligibility.

[k] Scales to be completed by same rater with input from the same parent/LAR whenever possible.

l Sleep diary to be completed by parent(s)/LAR. Sleep diaries will be dispensed at Visits 0-5 and collected at Visits 1-6.

[m] C-SSRS “Lifetime Recent” version completed at the screening visit (Visit -1). C-SSRS “Since Last Visit” version completed for all subsequent visits.

2.5 Determination of Sample Size

Approximately 245 subjects will be screened to randomize approximately 195 subjects in a 5:5:5:5:6 ratio to SPD489 5, 10, 20, 30 mg or placebo to achieve 156 completers for the study (30 in each active treatment group and 36 in the placebo group) and 85% power for the primary efficacy analysis at a 2-sided 0.05 significance level.

The sample size planned at study initiation is estimated based on the primary comparison between SPD489 10, 20, 30 mg pooled together (excluding the 5mg) compare with placebo on the primary efficacy endpoint, in a group sequential design with 1 interim analysis using the Lan-DeMets alpha spending function with O'Brien-Fleming boundary for the primary efficacy endpoint. Assumptions for the calculation include the true mean difference of 8.4 with the common standard deviation (SD) of 14 for the primary efficacy endpoint, for an effect size of 0.6, and a dropout rate of 20%.

In order to maintain sufficient study power to detect a clinically meaningful treatment effect for the primary efficacy endpoint, a blinded sample size re-estimation is planned. If the study continues beyond the proposed interim analysis, or if the interim analysis is waived, then the sample size re-estimation will be performed when approximately 75% of the 195 subjects have either completed or discontinued from the study. In this case, cumulative primary efficacy data will be used to estimate a pooled common SD to ensure the variability hypothesized at the design stage is not underestimated. Together with the assumed treatment difference of 8.4, the final total number of subjects to be enrolled will be calculated using the re-estimated pooled SD (Friede and Kieser 2006). Step by step details for estimating the pooled SD are included in the Appendix 1. Using the pooled common variance estimated to do a blinded sample size re-estimation does not inflate the Type I error.

The final total sample size could potentially be as high as 218 completers, which corresponds to the 97.5% percentile of the distribution of the estimator on the assumed true common SD of 14 postulated at the design stage. If the re-estimated pooled common SD is larger than 14, the sample size will be increased. Otherwise, the sample size will remain the same. Note that the total of 218 completers is not considered a cap. As the final sample size is data driven, a higher number, though unlikely, is possible.

The Type-I error level for the sample size re-estimation will be adjusted based on the Lan-DeMets alpha spending function with O'Brien-Fleming boundary when applicable.

The dose-response relationship that is measured by the ADHD-RS-IV Preschool Version Total Score change from baseline will be evaluated separately. Assuming the maximum difference in change from baseline in ADHD-RS-IV Preschool Version Total Score between 0 mg (placebo) and SPD489 (5, 10, 20 or 30 mg) is 13.0 points, and has a standard deviation of 14 for the change, then in order to detect a plausible dose-response curve at 85% power and a significance level of 0.05 (2-sided) using MCP-Mod with equal allocation to the treatment groups, it is necessary to have 18 completers for each arm. Assuming a 20% dropout rate, a total of 24 subjects for each treatment group are required to be randomized. Therefore, the overall sample size of the study will be sufficient for dose-response analysis

2.6 Interim Analysis and Multiplicity Adjustments for Type I Error Control

The study implements a group sequential design with 1 interim analysis. The interim analysis is proposed at approximately 60% of planned subjects, who either complete or discontinue the study.

The interim analysis will be conducted on the interim analysis dataset, which is defined as all data from 60% of initially planned subjects who have either completed or discontinued the study, together with data for subjects who are in the study up to the interim analysis data cut time point. The efficacy, safety, and sensitivity on missing data mechanisms analyses at the interim analysis will primarily be conducted using this dataset.

2.6.1 Waiving of Interim Analysis

It is estimated that the interim analysis will take up to 8 weeks from data cutoff to its completion. If the recruitment rate is high and the remaining 40% planned subjects are projected to be all or nearly all enrolled within those 8 weeks, the interim analysis will be waived with reasons documented. In this case, the end of study analysis will be conducted at the 2-sided significance level of 0.05 with primary and key secondary endpoints tested sequentially. Furthermore, the dose-response relationship will be evaluated as well.

2.6.2 Interim Analysis

The Lan-DeMets alpha spending function with O'Brien-Fleming boundary will be used for alpha spending between the interim and end of study analysis for the primary efficacy endpoint. Following evaluations in the stage-wise hierarchical setting (Glimm et al. 2010), the Lan-DeMets alpha spending function with Pocock boundary will be applied to analyze the key secondary efficacy endpoint.

If the hypothesis for the primary efficacy endpoint is rejected at the interim analysis, then the study will be stopped for efficacy and all ongoing subjects from the study (SPD489-347) will be immediately rolled over to the open-label extension study (Study SPD489-348). In this case, using the same data cut point that contains the same information fractions as the primary efficacy interim analysis, the key secondary efficacy endpoint will be tested, and all safety summaries that the DMC to be needed for their decision will be generated, which will include disposition, demographic and baseline characteristics, main AE and vital signs information. Additionally, the incremental data collected during the interim analysis process, i.e., the overrun, together with the interim analysis dataset will be utilized in sensitivity analyses for efficacy and safety. Otherwise, if the primary efficacy criterion is not met at the interim analysis, then the hypothesis for the key secondary endpoint will not be tested at this time, and the study will continues to its completion, at which time the primary efficacy endpoint will be tested. If the primary efficacy test is rejected, then the key secondary efficacy endpoint will be tested. This testing procedure strongly protects the overall alpha level for the primary and key secondary efficacy analyses. All other efficacy and safety analyses will be conducted at this time as well.

At the interim or final efficacy analysis, only the primary efficacy result will be used to determine if the study meets the efficacy criterion. The nominal significance level for the interim or final primary and key secondary efficacy endpoints for a 2-sided at a 0.05 significant level are specified below, calculated using East® 6.4. The calculations are based on 60% of

planned sample size at the interim analysis, and the significance values may be updated (in a blinded fashion), based on the actual information fraction at the interim analysis.

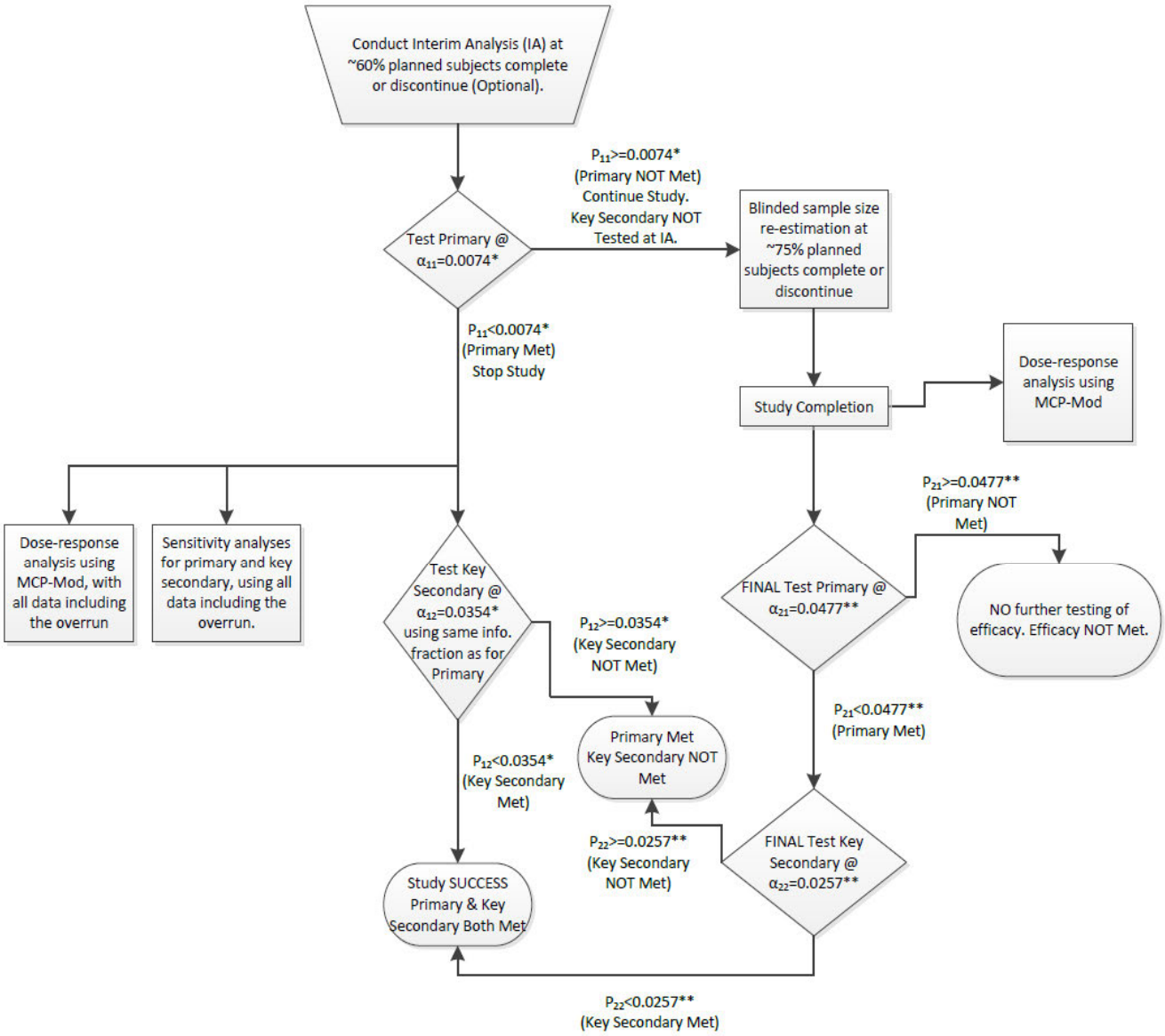
	Efficacy Boundary p-value Scale	
	Interim Analysis	**Final Analysis**
Primary Endpoint	0.0074	0.0477
Key Secondary Endpoint	0.0354	0.0257

In general, the nominal significance level used for the interim analysis will no longer apply for the sensitivity analysis, it should be recalculated. However, if the information fraction increase is limited, for example, if the increase is less than 1%, then the nominal significance levels for the interim analysis can be used as a good approximation (Proschan et al. 2006) for the sensitivity analyses.

If the primary efficacy criterion is not met at the interim analysis, and the blinded sample size re-estimation causes a sample size increase, then the nominal significance levels at end of the study will be recalculated, such that the overall alpha level is not greater than 0.05 for the primary and key secondary analyses together. The calculations will be based on the new increased sample size, the actual information ratio, and the alpha already spent. The new increased sample size extends the information time scale from interval [0, 1] to a wider one. In order to find the critical value for the completion study analysis, the extended time interval will be converted to the original time scale. The information fractions of the interim analysis can be identified within the new time scale. Applying it along with the increased new sample size, the critical values for the completion study analysis can be calculated. This process will be documented prior to database lock and unblinding.

Separately, if the primary efficacy criterion is met at the interim analysis, the interim analysis dataset together with data for the subjects in the study up to the interim analysis data cut time point will be used to evaluate the dose-response relationship over the DRAS. Otherwise, all cumulative data up to completion of the study will be used.

2.7 Decision Tree of Efficacy Analyses



Note: α_{11} and α_{21} are based on Lan-DeMets alpha spending function with O'Brien-Fleming Boundary
α_{12} and α_{22} are based on Lan-DeMets alpha spending function with Pocock Boundary
*: efficacy boundaries (p-value scale) may be updated based on actual information fraction at the IA
**: will be adjusted based on alpha-spending at IA and the actual sample size at EOS

3. OBJECTIVES

3.1 Primary Objective

The primary objective is to evaluate the efficacy of SPD489 compared to placebo in preschool children (4-5 years of age inclusive at the time of consent) with ADHD. The primary measure of efficacy will be the clinician-administered ADHD Rating Scale Preschool Version (ADHD-RS-IV Preschool Version) Total Score.

3.2 Secondary Objectives

- **Key Secondary:** To evaluate the efficacy of SPD489 compared to placebo, using the global clinical measure of improvement, the Clinical Global Impression – Global Improvement (CGI-I) scale.

- **Secondary:**
    - To evaluate dose-response relationship in preschool children with ADHD, using the ADHD-RS Preschool Version Total Score.
    - To evaluate the safety and tolerability of SHP489 based on the occurrence of TEAEs, specific evaluation of vital signs (systolic and diastolic blood pressure and pulse), height, weight, and body mass index (BMI), clinical laboratory evaluations and electrocardiogram (ECG) results, sleep assessment, and the Columbia Suicide Rating Scale (C-SSRS).

4. SUBJECT POPULATION SETS

The following subject sets are applicable to this study.

4.1 Screened Set

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

4.2 Randomized Set

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

4.3 Safety Analysis Set

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

4.4 Full Analysis Set

The Full Analysis Set (FAS) will consist of all subjects in the Safety Analysis Set who have at least 1 post-dose ADHD-RS-IV-Preschool Version Total Score assessment.

4.5 Dose-response Analysis Set

The Dose Response Analysis Set (DRAS) will consist of all subjects in the Safety Analysis Set who have at least 1 valid primary efficacy measurement (ADHD-RS- IV Preschool Version Total Score) on the randomized target dose level of the investigational product.

The above analysis sets will be utilized for both the interim and the end of study analyses.

Subjects are randomized at baseline (Visit 0) in a 5:5:5:5:6 ratio to SPD489 5, 10, 20, 30 mg or placebo group. If not stated otherwise, each dose strength and pooled doses (10, 20, 30 mg pooled together) will be used for reporting the study variables, as below.

Parameter	Placebo	SPD489				
		5mg	10mg	20mg	30mg	Pooled 10, 20, 30 mg
Baseline Characteristics	✓	✓	✓	✓	✓	✓
Primary Efficacy	✓					✓
Key Secondary Efficacy	✓					✓
All Other Efficacy	✓					✓

Parameter	Placebo	SPD489				
		5mg	10mg	20mg	30mg	Pooled 10, 20, 30 mg
Dose-response Relationship	✓	✓	✓	✓	✓	✓
Safety Reporting	✓	✓	✓	✓	✓	✓

5. SUBJECT DISPOSITION

A listing of all Screen Failures (ie, 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 (ie, Screened, Randomized, Safety, FAS and DRAS) will be summarized by treatment group, pooled doses 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 Double-blind Evaluation Period will be presented for each treatment group, pooled doses and overall for the Safety Set. Reasons for premature discontinuation from the Double-blind Evaluation Period, as recorded on the end of study discontinuation page of the electronic case report form, will be summarized (number and percentage) by treatment group, pooled doses and overall for the Safety Set. The subjects who completed the study are those that completed the final scheduled visit at the end of the Double-blind Evaluation Period. When the interim analysis is conducted, for reporting purpose, an ongoing category will be added, and it will be reported by treatment, pooled doses and overall.

All subjects who prematurely discontinued during the Double-blind Evaluation Period will be listed by discontinuation reason for the Randomized Set.

The number of subjects randomized and completed will be tabulated by site. In addition, the duration of enrollment, in days, will be summarized for each site, and overall. Duration of enrollment will be calculated as (last date of contact for any subject at that site - the first date of informed consent for any subject at that site + 1).

Follow-up phone call information will be listed for the Safety Analysis Set.

6. PROTOCOL DEVIATIONS

Protocol deviations will be recorded by the site separately from the clinical database. The contract research organization will classify the protocol deviations per the agreed protocol deviation 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 deviation 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 deviations will be documented in the Protocol Deviation tracker for the study. Protocol deviations will be summarized using the Randomized Set by category and site for each treatment group, pooled doses and overall. Protocol deviations will be listed for the Randomized Set.

7. DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS

7.1 Demographic and Other Baseline Characteristics

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 demographics and baseline data will be summarized by treatment group, pooled doses and overall for the Safety Analysis Set, FAS and DRAS. Listings will be provided as well.

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

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

The Screening height and body weight will be used to calculate BMI using the following formula:

$$\text{BMI} = \frac{\text{Weight [kg]}}{(\text{Height [m]})^2}$$

The BMI should be rounded to 1 decimal place for reporting. BMI categories will be derived using the Centers for Disease Control and Prevention (CDC) BMI percentiles for children: Underweight – BMI less than the 5th percentile; Healthy Weight – BMI 5th percentile up to less than the 85th percentile; Overweight – BMI 85th percentile to less than the 95th percentile; Obese – BMI greater than or equal to the 95th percentile.

The following demographic characteristics will be summarized: age, sex, ethnicity, race, weight, height, BMI, and BMI category. In addition, other baseline characteristics to be summarized include the Peabody Picture Vocabulary Test standard score, Children's Global Assessment Scale scores, Clinical Global Impression – Severity of Illness (CGI-S), and ADHD subtypes. Continuous variables will be summarized by descriptive statistics including number of subjects, mean, SD, median, minimum, and maximum. Categorical variables will be summarized by the number of subjects in each category and the percentage of subjects out of the total in the respective analysis set.

All demographic and baseline characteristics will be listed for the Safety Analysis Set.

7.2 Medical History

Medical history is collected at the Screening Visit (Visit -1). Prior and ongoing psychiatric evaluation will be established with the Screening Visit (Visit -1) interview using the Kiddie-Schedule for Affective Disorders and Schizophrenia for School-age Children-Present and Lifetime Version-Diagnostic Interview (K-SADS-PL). Both of the assessment results will be listed for the Safety Analysis Set.

8. EXTENT OF EXPOSURE AND TREATMENT COMPLIANCE

8.1 Exposure to Investigational Product

Exposure to double-blind investigational product for the Safety Analysis Set will be summarized by treatment, pooled doses. The following statistics will be calculated for the Fixed-dose Titration Period, the Dose Maintenance Period, and the whole Double-blind Evaluation Period for the investigational product:

- Total days of dosing: sum of the days of dosing for each period. Within each period, it is defined as (date of last dose - date of first dose + 1)
- Average daily dose (mg/day) for SPD489: the total dose/total days of dosing
- Maximum dose for SHP489: the highest dose level received

An appropriate statistical summary will be applied to present for each of the above statistical quantity for each of the study period respectively. In addition, person-time (overall total exposure in days) will be derived. It is calculated as total number of days in which the investigational product was taken for each subject and then sum over all subjects for the whole duration of the study.

Dosing information 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 the subjects lost to follow up after the Baseline visit (Visit 0), we will not assume any ingestion of the investigational product unless there is a post-baseline safety assessment. Subjects that do not return the investigational product but return to the site for a following visit will have zero capsule returned entered into the database. In this case, it will be assumed for analysis purposes that all capsules were ingested.

8.2 Measurement of Treatment Compliance

Investigational product dosing compliance, during the Dose Optimization Period, the Dose Maintenance Period, and the whole Double-blind Evaluation Period, respectively, is defined as the total number of capsules taken by a subject during that period divided by the number of capsules expected to be taken during the same period multiplied by 100, that is (number of capsules actually taken × 100 / number of capsules expected to be taken). The number of capsules taken is calculated by the 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 0. The number of capsules expected to be taken is calculated as the number of days the subject is in the particular period multiplied by the number of capsules to be taken per day during that period. If the number of capsules taken or returned is missing, then the compliance will be missing.

For each study period, the compliance data will be reported using number of subjects, mean, SD, median, minimum, and maximum for the Safety Analysis Set by treatment group and pooled

doses. Furthermore, the compliance data will be categorized as <80%, 80-120%, or >120%, and it will be presented using subject counts and percentages for the Safety Analysis Set by treatment group and pooled doses.

Treatment compliance data will also be listed for all subjects in the Safety Analysis Set.

9. PRIOR AND CONCOMITANT MEDICATION

The most updated available version of the World Health Organization drug dictionary will be used to classify prior and concomitant medications by therapeutic class.

Prior medication is defined as any medication with a start date prior to the date of the 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 and last doses of investigational product, inclusive. Any medication with a start date after the date of the last dose of investigational product will not be considered a concomitant medication.

Both prior and concomitant medication usage (including ADHD medication) will be summarized by the number and proportion of subjects within each preferred term for the Safety Analysis Set by treatment group and pooled doses. 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. Both prior and concomitant ADHD medication usage (based on indication), excluding investigational product, will also be summarized by the number and proportion of subjects within each preferred term for the Safety Analysis Set by treatment group and pooled doses.

All prior and concomitant medication will be listed for the Safety Analysis Set. Separate listings will also be provided for prior and concomitant ADHD medication and prior and concomitant ADHD behavioral therapy.

10. EFFICACY ANALYSES

The primary and key secondary analyses, including their sensitivity analyses, will be based on the FAS and will be conducted on either the interim analysis dataset, which is defined in Section 2.6, or all cumulative data up to study completion, whenever it is appropriate. Specifically, these analyses will compare placebo and pooled SPD489 10, 20, 30 mg dose strengths together, excluding the 5 mg arm. For primary and key secondary endpoints, statistical tests are planned to control the study-wise type I error at the 2-sided 5% level, as described in Section 2.6. All other statistical tests will be 2-sided hypothesis tests performed at the 5% level of significance, with no adjustment for multiplicity. All confidence intervals (CIs) will be 2-sided 95% CIs), unless stated otherwise. Additionally, the dose response relationship evaluations will be performed over the DRAS and using appropriate datasets. For this analysis, all randomized investigational product groups, which include 0 (placebo), 5, 10, 20, 30 mg (SPD489) dose strengths will be used.

Baseline for all efficacy analyses is defined as the last valid value for the efficacy assessment prior to the first dose of investigational product. An efficacy measurement that is assessed more than 3 days after the date of the last dose of investigational product will not be included for analysis.

10.1 Primary Efficacy Endpoint and Analysis

The primary efficacy endpoint is defined as the change from baseline in clinician-administered ADHD-RS-IV Preschool Version Total Score at Visit 6 (Week 6).

The statistical hypotheses of the primary efficacy analysis are:

- The null hypothesis is: there is no treatment difference between pooled SPD489 (10, 20 and 30mg) dose group and placebo at Visit 6 (Week 6)
- The alternate hypothesis is: there is a treatment difference between pooled SPD489 (10, 20 and 30mg) dose group and placebo at Visit 6 (Week 6).

The estimate of primary interest is estimate of the average effect attributable to the experimental treatment (pooled SPD489 10, 20 and 30 mg) as compared to placebo at Visit 6 (Week 6), for all subjects in FAS of the study.

The primary efficacy endpoint will be analyzed using the linear mixed-effects model for repeated measures (MMRM). The analysis includes the fixed categorical effects of treatment, visit, and treatment-by-visit interaction, the covariate of baseline ADHD-RS-IV Preschool Version Total Score, and with the baseline ADHD-RS-IV Preschool Version Total Score-by-visit interaction adjusted in the model. The restricted maximum likelihood method will be used, with an unstructured covariance structure (UN) shared across treatment groups used to model the within-patient errors. The Kenward-Roger method is used to estimate denominator degrees of freedom and adjust standard errors (SEs). The primary contrast of interest will be at Visit 6 (Week 6) for pooled SPD489 (10, 20 and 30mg) dose group compared with placebo.

For the above MMRM analysis, contrast coefficients that are based on the actual sample size of placebo or each dose in the pooled SPD489 group will be used to obtain a weighted average for performing the comparison.

If the UN covariance structure fails to converge, as an alternative, the Toeplitz (TOEP) or Autoregressive (1) [AR(1)] covariance matrix will be applied in that order, that is, test the TOEP structure first, if it does not converge then test the AR(1). Both of these covariance matrices can model data with correlations that decline over time. Comparing AR(1) with TOEP, the AR(1) has less constraining conditions, so when the TOEP structure fails to converge, the AR(1) may still converge. Regardless which covariance structure is chosen, missing data will not be imputed.

Descriptive statistics will be displayed for pooled SPD489 (10, 20 and 30mg), and placebo group. To compare pooled SPD489 dose group and placebo, the difference of LS means, the corresponding 95% CI and p-value, and the effect size (Cohen 1988) (defined as difference of the least squares means divided by the estimated SD from the corresponding covariance matrix) will be displayed. The p-value and effect size will only be presented for Visit 6 (Week 6).

Additionally, a corresponding line graph of the LS means of change from baseline and associated 95% CI in ADHD-RS-IV Preschool Version Total Score by pooled SPD489 (10, 20 and 30mg) dose group and placebo, and visit will be presented.

10.2 Sensitivity Analyses for the Primary Efficacy – Missing Not at Random

The primary efficacy analysis (MMRM) relies on the assumption that the missing data mechanism follows the missing at random (MAR) scenario. In this case, it is assumed that the reason for data being missing may depend on observed data, but not on the unobserved missing data. The likelihood-based MMRM analysis is an appropriate method for the statistical analysis under MAR assumption.

The following 2 sensitivity analysis models, which are within the pattern-mixture model framework, will be used to examine the robustness of the primary efficacy analysis results for the missing not at random (MNAR) mechanisms. Under MNAR, it is assumed that the reason for data being missing is related to the unobserved missing data.

Model 1: Placebo Multiple Imputation

Rationale: The imputations are based on the distribution of placebo group responses over time. The underlying assumption is that the missing data for a subject on the active treatment follow the distribution of the placebo responses, i.e., the mean values and intra-subject correlations based on the placebo responses will be applied.

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

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 pooled active treatment group, while the other 100 will be 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 datasets for imputations within a treatment group will be ordered from 1 to 100 and combined between pooled active treatment group and placebo, for a total of 100 completely imputed datasets.

Step 2: Analysis of Complete Datasets

The primary efficacy endpoint will be analyzed for each of the 100 complete datasets with imputed data using an analysis of covariance with treatment group as factors and the baseline value as a covariate.

Step 3: Inference

The LS mean difference estimates will be averaged, and the associated SEs 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.

If the primary efficacy criterion is met at the interim analysis, then the above process will be adjusted. For subjects who are ongoing at the time of study stop, the posteriors from the corresponding treatment group will be applied. For dropouts at the time of the interim analysis, the placebo posteriors will be applied.

Model 2: Multiple Imputations with Penalties Applied to Dropouts

Rationale: The underlying assumption is that subjects who dropout performing 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 1*a* and 1*b* are provided in

Step 1*a*: Imputations

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

Step 1*b*: Application of Penalty

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

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

If the primary efficacy criterion is met at the interim analysis, then the penalties will only apply to dropouts. Multiply imputed values for ongoing subjects at the time of study stop will be used in the analysis without penalties.

10.3 Key Secondary Efficacy Endpoint and Analysis

The key secondary efficacy endpoint CGI-I provide an overall assessment of global symptom improvement. It will be analyzed using the same analysis method (MMRM) as for the primary efficacy endpoint. The baseline CGI-S score will be used as the covariate. The contrast of interest will be at Visit 6 (Week 6) for the pooled SPD489 group compared with placebo.

The CGI-I will be analyzed using the 2 sensitivity analysis models described in Section 10.2 to examine the robustness of the key secondary efficacy analysis (MMRM) results.

10.4 Dose-response Analyses

The primary and key secondary efficacy analyses, and dose-response analysis will be conducted separately, and therefore no multiplicity adjustment will be considered for these objectives together.

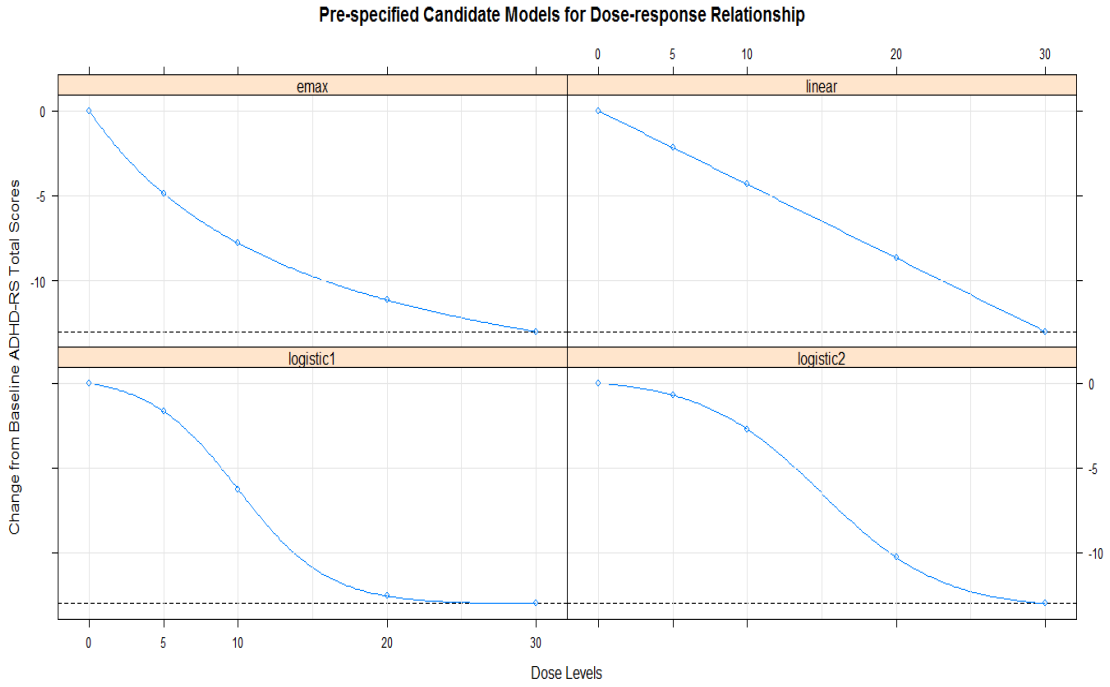
In this section all analyses will be performed on DRAS using R MCPMod package, and treatment arms 0 mg (placebo), 5, 10, 20, 30 mg (SPD489) will not be pooled.

10.4.1 Dose-response Candidate Curves

The MCP-Mod approach was created for dealing with model-uncertainty at the study design stage, by pre-specified set of candidate dose-response curves, combining principles of multiple comparisons with modeling techniques (Bretz et al., 2005). For this study, the candidate dose-response curves were selected based on prior knowledge of SPD489 dose-response, and assumptions on fixed-doses effects (placebo, SPD489 5, 10, 20, and 30 mg). The assumed family of plausible dose-response curves consists of linear, E_{max}, and logistic relationships, with appropriate parameters, covering the dose ranges of 0mg (placebo) to 30mg of investigational product. The candidate models and their figures are presented below.

Family of the Curves	Candidate Dose-response Curves
Linear	$-0.4\times d$
E_{max}	$-19.5\times d/(15.0+d)$
Logistic	$0.5-13.5/\{1+Exp[(10-d)/3]\}$
	$0.3-13.6/\{1+Exp[(15-d)/4]\}$

Where d stands for dose levels, d = 0, 5, 10, 20, 30 mg.



10.4.2 Dose-response Relationship Evaluations

When the study completes, regardless at the interim analysis or at end of the study, using the MCP-Mod methodology (Pinheiro et al, 2006) and all cumulative available data, the dose-response relationship of SPD489 as measured by the change from baseline for ADHD-RS-IV Preschool Version Total Score will be evaluated on the DRAS, fitting with an optimized candidate curve to identify the minimum effect dose (MED).

The dose response modelling in MCP-Mod uses a nonlinear least squares routine to fit the data to each of the candidate curve. Each of the dose response shapes in the candidate set will be tested using an appropriate contrast T-statistic. Using of the multivariate t-distribution to determine a critical value for all of the contrast T-statistics controls the overall Type-I error rate. The maximum contract test statistic will be used to select a dose response curve out of the set of significant curves, if there are any significant curves. If at least 1 of the specified candidate curve is selected, using a 2-sided significance level of 0.05, then for this selection process, dose response for SPD489 is established.

The dose response model will include sex (female vs. male), baseline ADHD subtypes and baseline ADHD-RS total scores as explanatory variables in the dose response analysis. The baseline ADHD-RS score will be a continuous covariate, whereas sex and ADHD subtypes will be factors.

The modeling results of the candidate curves will be tabulated that include T-statistic and corresponding adjusted p-values for each curve. Also parameters estimated for the final dose response relationship curve, and figures showing the dose response relationship will be presented. Fitted significant curves with ADHD-RS total score change from baseline data will be displayed as well.

10.4.3 Minimum Effective Dose

If a significant dose-response relationship can be established using the maximum contrast test statistic, then a clinically relevant threshold,– 5.6 reduction from baseline in ADHD-RS-IV Preschool Version Total Score, which derived from assumed minimum clinical meaningful effect size of 0.40 and common standard deviation of 14, will be used for the MED derivations. The 95% CI of the MED will be bootstrapped, that is, subjects will be sampled with replacement within the randomized treatment group for resampling, and the full MCP-Mod procedure will be applied to the resampled data, producing a bootstrap sample of MED values. The 95% CI for the MED correspond to the 2.5% and 97.5% quantiles of the bootstrap samples.

By default, MCP-Mod reports the MED as missing if any of the significant curves fails to converge or if an estimate cannot be determined from a particular converged model.
If an MED estimate cannot be determined from the selected significant curve, then available MED estimates from other converged models will be used to calculate an MED using the average AIC method.

10.4.4 Additional Analyses

A line-plot of mean change from baseline with 95% CI for each randomized treatment group will be generated.

Each SPD489 arm will be compared with placebo in ADHD-RS-IV Preschool Version Total Score change from baseline using MMRM. The comparisons will be conducted in a sequential order for adjusting for the multiplicity, which will test the highest dose level, 30 mg, first. The model structure will be similar to the primary efficacy analysis. However, this analysis is not powered for the comparisons.

A sigmoid nonlinear E_{max} model (MacDougall, J. 2006) below will be fitted, and the sigmoid E_{max} model curve will be plotted.

$$R_i = E_0 + \frac{D_i^m \times E_{max}}{D_i^m + ED_{50}^m} + \varepsilon_i$$

Where i = the subject indicator; R_i = dose response for subject i that measured as ADHD-RS-IV Preschool Version Total Score change from baseline; E_0 = the dose response when dose strength is 0 mg (placebo); D_i = the dose level for subject i; m = the slope factor determining the steepness of the dose response curve ($m > 0$); E_{max} = the maximum possible effect for the investigational product; ED_{50} = The dose level, which produces half of E_{max}; ε_i = the random error term for subject i.

For the E_{max} model evaluations, initial parameters will be set to $E_0 = 0$; $E_{max} = -19.5$; $ED_{50} = 15.0$ and m = 3.17. Where the m was calculated using the formula

$$ED_p = ED_{50} \left(\frac{p}{1-p}\right)^{1/m}$$

let $p = 0.9$, ED_{90} =30.0.

The MCP-Mod methodology provides a flexibility of modeling for dose response relationship, while preserving the robustness to model misspecification associated with multiple comparison procedures (Bretz et al. 2005). However, if the MCP-Mod fails to converge to any pre-specified candidate curves, the sigmoid nonlinear E_{max} model, as a backup, will provide a dose response relationship. Using the relationship and clinically relevant threshold– 5.6 reduction from baseline in ADHD-RS-IV Preschool Version Total Score together, an approximately estimate of MED will be obtained.

10.5 Supportive Analyses

10.5.1 Subgroup Analysis

The subgroups considered for analyses will be sex (male, female), race (white, non-white), and ethnicity. There is no formal hypothesis testing is planned for the pooled SPD489 (10, 20 and 30 mg) dose group and placebo for the subgroup analysis.

ADHD-RS-IV total score and CGI-I data will be summarized descriptively for each subgroup that is defined above at each visit by pooled SPD489 (10, 20, 30 mg) and placebo group.

10.5.2 Other Analyses

The items in the ADHD-RS-IV Preschool Version will also be grouped into 2 subscales: hyperactivity/impulsivity (even-numbered items 2-18) and inattentiveness (odd-numbered items 1-17). Each subscale will be analyzed using a similar model as the primary efficacy endpoint and summarized descriptively at each visit by pooled SPD489 treatment group and placebo. A corresponding line graph of the LS means of change from baseline and associated 95% CI in ADHD-RS-IV Preschool Version subscale scores by pooled SPD489 treatment, placebo and visit will be presented.

The ADHD-RS-IV total score will be summarized descriptively at each visit by pooled SPD489 treatment and placebo.

The key secondary efficacy measurement, CGI-I, will also be analyzed using the proportion of subjects with an “improved” CGI-I measurement at endpoint. The CGI-I categories will be dichotomized into 2 categories: “very much improved” and “much improved” classified as “improved,” and all other assessed categories grouped together as “not improved.” The key secondary efficacy analysis will be conducted to compare pooled SHP489 treatment and placebo on the FAS for the “improved” rate using a Cochran-Mantel-Haenszel test stratified by CGI-S value at baseline.

Dichotomized CGI-I values will be summarized by visit and pooled SHP489 treatment and placebo. A corresponding bar chart showing the percentage of subjects improved by visit, pooled SHP489 treatment, and placebo group will be presented. CGI-I values and CGI-S values will be listed.

The CGI-I will be summarized descriptively at each visit by pooled SHP489 treatment and placebo.

11. SAFETY ANALYSES

The Safety Analysis Set based on either the interim analysis dataset that for the safety summary at interim analysis, or the interim analysis dataset together with the overrun if the efficacy criterion meets, or all cumulative data up to study completion, whenever it is appropriate, will be used to report the safety data.

Safety data will be summarized by SPD489 dose strengths, pooled doses and placebo. Safety variables include AEs, clinical laboratory variables, vital signs, physical examinations, Children's Sleep Habits Questionnaire (CSHQ), sleep diary, ECG, and C-SSRS variables. For each safety variable, the last value collected before the first dose of investigational product will be used as baseline for all analyses of that safety variable.

11.1 Adverse Events

Adverse events will be coded using Version 18.0 or newer of the Medical Dictionary for Regulatory Activities.

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

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

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

The incidence of TEAEs, serious TEAEs, and TEAEs leading to withdrawal will be summarized by SOC, PT, and treatment group, sorted in decreasing frequency for subjects on SPD489 then for placebo. The incidence of common (≥2% of subjects in any treatment group) will be summarized by PT and treatment group, sorted in decreasing frequency for subjects on SPD489 and then for placebo.

The number and percentage of subjects reporting TEAEs over the Safety Analysis Set will also be presented by week of AE onset, preferred term and treatment group, where week of AE onset is calculated as (AE occurred date – first dose date + 1) / 7, then rounded up to an integer.

All information about AEs collected on the electronic case report form will be listed alongside the dose, PT, and SOC. For serious TEAEs, deaths, AEs related to the investigational product, and TEAEs leading to study discontinuation, a separate listing will also be provided.

11.2 Clinical Laboratory Variables

Descriptive statistics (n, mean, SD, minimum, median, and maximum) for clinical laboratory values (in SI units) and changes from Baseline to Visit 6/early termination (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	Calcium
Aspartate transaminase	Uric acid
Phosphorus	Blood urea nitrogen
Alanine transaminase	Total bilirubin
Sodium	Creatinine
Alkaline phosphatase	Glucose
Potassium	Albumin
Gamma glutamyl transferase	Total protein
Thyroid stimulating hormone	Lactate dehydrogenase
Free thyroxine	

Hematology

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

Urinalysis

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

If urinalysis detects protein and/or blood, then a microscopic examination will be conducted. The microscopic examination will consist of red blood cell, white blood cell, casts, and bacteria. These variables will be listed but not summarized.

Qualitative urinalysis variables will be summarized by the number of subjects in each category and the percentage of subjects out of the total in the Safety Analysis Set with a valid result at both baseline and the given visit.

All clinical laboratory values will be listed for the Safety Analysis Set.

Clinical laboratory test values are potentially clinically important (PCI) if they meet either the low or high PCI criteria listed in Table 2. 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 at least 1 available post-baseline assessment per parameter. 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, site, baseline, and post-baseline values.

Table 2: Criteria for Potentially Clinically Important Laboratory Tests

Parameter	SI Unit	Lower Limit	Higher Limit
Biochemistry			
Albumin	g/dL	<3g/dL	-
Aspartate transaminase (AST)	-	-	≥3 x ULN
Alanine transaminase (ALT)	-	-	≥3 x ULN
Blood urea nitrogen	mg/dl	-	>30 mg/dl or >2.5 x ULN
Calcium	mg/dL	<8mg/dL	>11.5mg/dL
Total cholesterol	mg/dL	-	>300mg/dl
Creatinine	mg/dl	-	>2 mg/dl or >1.5 x ULN
Gamma glutamyl transferase	-	-	≥2.5 x ULN
Glucose	mg/dL	<55mg/dL	>160mg/dL
Lactate dehydrogenase	-	-	>3 x ULN
Phosphorus	mg/dL	<2.5mg/dL	>7mg/dL

Table 2: Criteria for Potentially Clinically Important Laboratory Tests

Parameter	SI Unit	Lower Limit	Higher Limit
Potassium	mmol/l	<3mmol/l	>5.5mmol/l
Thyroid stimulating hormone	-	<LLN	>2 x ULN
Sodium	mEq/L	<130mEq/L	>150mEq/L
Uric acid	mg/dL	-	>10mg/dL
Total bilirubin	mg/dL	-	>2 mg/dL
Total protein	g/dL	<5g/dL	>9g/dL
Hematology			
Bands	-	-	>0.27x10^3/µL or >5%
Basophils	-	-	>10%
Hemoglobin	g/dL	<9g/dl	>16g/dl
Hematocrit	%	<30%	>50%
Platelet count	10^3/µL	<75x10^3/µL	>600x10^3/µL
Red blood cells	10^6/µL	<2.5x10^6/µL	-
White blood cell count	10^3/µL	<3x10^3/µL	>16x10^3/µL
Neutrophils	-	<1x10^3/µL or <30%	-
Eosinophils	%	-	>10%
Lymphocytes	%	<10%	>70%
Monocytes	%	-	>20%
Urinalysis			
Glucose	-	-	Positive Value (excluding trace)
Protein	-	-	Positive Value (excluding trace)
Blood	-	-	Positive Value (excluding trace)
Ketones	-	-	Positive Value (excluding trace)
Bilirubin	-	-	Positive Value (excluding trace)

LLN=lower limit of normal value provided by the laboratory; ULN=upper limit of normal value provided by the laboratory

11.3 Vital Signs

Descriptive statistics (n, mean, SD, minimum, median, and maximum) for vital sign variables (systolic and diastolic blood pressure, pulse rate, body weight, temperature, and respiration rate), and their changes from baseline at each assessment time point will be presented by treatment group.

Body weight is collected at every visit, while height is collected only at Screening and the last scheduled visit/ET. At each visit excluding the last scheduled visit/ET, the Screening height will

be used to derive BMI values. For the last scheduled visit/ET, corresponding height and body weight will be used to calculate the BMI values. The resulting BMI values will be rounded to 1 decimal place for reporting. Descriptive statistics for BMI will be tabulated by visit and treatment group for actual values and change from baseline.

Percentiles of body weight and BMI can be normalized by sex and age using the CDC growth charts (Kuczmarski et al. 2002). Descriptive statistics of percentiles, for both, body weight, and BMI will be generated by visit and treatment group, for actual values and change from baseline.

Furthermore, the body weight percentiles can be categorized as <5th, 5^{th} to $<95^{th}$, and $\geq 95^{th}$ percentiles, and the BMI percentiles can be categorized as: Underweight – BMI less than the 5th percentile; Healthy Weight – BMI 5th percentile up to less than the 85th percentile; Overweight – BMI 85th percentile to less than the 95th percentile; and Obese – BMI greater than or equal to the 95th percentile. Post-baseline shifts in body weight and BMI percentile category from baseline at each visit will be presented by treatment group. The data also will be presented categorically using number of subject and percentages at each visit by treatment group.

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

All vital sign values will be listed for the Safety Analysis Set.

All vital sign values, include an unscheduled visit, will be considered PCI if they meet either the observed value criteria or the change from baseline criteria listed in Table 3. The number and percentage of subjects with PCI post-baseline values will be tabulated by treatment group. 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 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.

Table 3: Criteria for Potentially Clinically Significant Vital Signs

Vital Sign Parameter	Flag	Criteria[a]	
		Observed Value	Change from Baseline
Sitting systolic blood pressure (mmHg)	High	≥120	Increase of >10
	Low	<75	Decrease of >10
Sitting diastolic blood pressure (mmHg)	High	≥85	Increase of >10
	Low	<40	Decrease of >10
Pulse rate (beats per minute)	High	≥130	Increase of >15
	Low	≤55	Decrease of >15
Weight (kg)	High	-	Increase of ≥7%
	Low	-	Decrease of ≥7%

Table 3: Criteria for Potentially Clinically Significant Vital Signs

Vital Sign Parameter	Flag	Criteria[a]	
		Observed Value	Change from Baseline
Temperature (deg. C)	High	>39	-
	Low	<35	-
BMI (kg/m^2)	High	>95th percentile for age and sex	-
	Low	<5th percentile for age and sex	-

BMI=body mass index
[a] A post-baseline value is considered as a PCI value if its meets either the criteria for observed value or the criteria for change from baseline.

11.4 Electrocardiogram (ECG)

Descriptive statistics (n, mean, SD, minimum, median, and maximum) for ECG variables (heart rate, PR interval, RR interval, QRS interval, QT interval, and QTc interval) and their changes from baseline at each assessment time point will be presented by treatment group. QTc interval will be calculated using both Bazett ($QTcB = QT/(RR)^{1/2}$) and Fridericia ($QTcF = QT/(RR)^{1/3}$) corrections; if RR is not available, it will be replaced with 60/hr in the correction formula. ECG interpretation will be summarized by visit (and Visit 6/ET) and treatment group. A shift table from baseline to each visit (and Visit 6/ET) for qualitative ECG results will also be presented.

All ECG values will be listed for the Safety Analysis Set.

Electrocardiogram variable values, include unscheduled visit, will be considered PCI if they meet either the observed value criteria or the change from baseline criteria listed in Table 4. The number and percentage of subjects with post-baseline PCI values will be tabulated by treatment group. 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. A listing of all subjects with post-baseline PCI value will be provided including the subject number, site, baseline, and post-baseline PCI values.

Table 4: Criteria for Potentially Clinically Important ECG Values

ECG Parameter	Unit	Observed Value		Change from Baseline	
		Lower Limit	Higher Limit	Lower Limit	Higher Limit
ECG Result	-	-	Abnormal (core lab) and clinically significant from investigator	-	-
Heart Rate	Beats/minute	<55	>130	Decrease >15	Increase >15

Table 4: Criteria for Potentially Clinically Important ECG Values

PR Interval	msec	-	≥200	-	-
QT Interval	msec	-	≥440	-	≥30 and <60 ≥60
QTcF Interval	msec	-	≥440 and <480 ≥480 and <500 ≥500	-	≥30 and <60 ≥60
QTcB Interval	msec	-	≥440 and <480 ≥480 and <500 ≥500	-	≥30 and <60 ≥60
QRS Interval	msec	-	≥90	-	-
Rhythm	-	-	Any rhythm other than sinus rhythm* **ECG evaluation – abnormal rhythm** • Complete Heart Block • Tachycardia • Bradycardia • Wandering Atrial Pacemaker • Ectopic Atrial Rhythm • Atrial Fibrillation • Atrial Flutter • Multifocal Atrial Tachycardia • Supraventricular Tachycardia • Atrial Bigeminy • Ventricular Bigeminy • Atrial Couplets • Ventricular Couplets • Ventricular Tachycardia • Torsade des Pointes • Ventricular Fibrillation • Junctional Rhythm • Idioventricular Rhythm • Escape Beat • Atrial Pacemaker	-	-

11.5 Other Safety Variables

11.6 Children's Sleep Habits Questionnaire and Sleep Diary

The 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:

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)

Each subscale and each individual item will be summarized using descriptive statistics (n, mean, SD, minimum, median, and maximum) at each visit (and Visit 8/ET) by treatment group. All CSHQ data will be listed for the Safety Analysis Set.

Sleep diary data will be summarized at each visit (and Visit 8/ET) for total daytime napping time, nighttime sleep time and time to fall asleep using descriptive statistics (n, mean, SD, minimum, median, and maximum) by treatment group. Sleep diary data are collected daily. The average daily value per visit and their change from baseline will be summarized by treatment group. Sleep diary data will also be listed for the Safety Analysis Set.

11.7 Columbia-Suicide Severity Rating Scale

Number of subjects and frequency with suicide-related events, based on the C-SSRS data, will be tabulated for the Safety Analysis Set for the following categories:

1. Wish to be dead
2. Non-specific active suicidal thoughts
3. Active suicidal ideation with any methods (not plan) without intent to act
4. Active suicidal ideation with some intent to act, without specific plan
5. Active suicidal ideation with specific plan and intent
6. Preparatory acts or behavior
7. Aborted suicide, attempt
8. Interrupted attempt
9. Actual attempt (non-fatal)
10. Completed suicide

11. Self-Injurious behavior without suicidal intent

A listing of the C-SSRS data will be provided for subjects with a positive response for the Safety Analysis Set.

11.8 Follow-up Phone Call

A listing will be created to show the date of the follow-up phone call.

12. CLINICAL PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES

Not applicable.

13. OTHER ANALYSES

See Section 10.5.2.

14. INTERIM ANALYSIS

See Sections 2.6 and 15.

15. DATA MONITORING/REVIEW COMMITTEE

An external independent Data Monitoring Committee (DMC) was set up to review the data pertaining to safety, tolerability, and benefit/harm of the study therapy for the duration of this Pediatric Written Request program, which includes studies SPD489-211, SPD489-347, and SPD489-348. The same DMC will evaluate the efficacy analysis results and supportive safety summaries to determine if the interim efficacy criteria are met for this study.

Confidentiality of the unblinded DMC analyses is a critical concern, and to address this, an unblinded independent reporting team will be identified within an external contract research organization. The independent reporting team will have no involvement in the conduct of the study. Further details regarding the DMC and the independent reporting team can be found in the DMC charter.

The DMC and the independent reporting team will not be involved in the blinded sample size re-estimation and dose-response relationship evaluations.

16. COMPUTER METHODS

Statistical analyses will be performed using Version 9.4 (or newer) of SAS® on a suitably qualified environment.

17. CHANGES TO ANALYSES SPECIFIED IN PROTOCOL

Not applicable.

18. DATA HANDLING CONVENTIONS

18.1 General Data Reporting Conventions

Continuous variables will be summarized using the following descriptive statistics: n, mean, median, SD, minimum, and maximum. 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 one more decimal place than the raw data. The SD will be presented to 2 more decimal places than the raw data, whenever it is appropriate. Categorical and count variables will be summarized by the number of subjects (n) and the percent of subjects in each category. Percentages will be reported to 1 decimal place, except when the percentage equals exactly 100 where it will be displayed as an integer (100). For zero, only count and no percentage will be displayed.

18.2 Derived Efficacy Endpoints

18.2.1 Primary Efficacy Endpoint

The ADHD-RS-IV Preschool Version 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 Fourth Edition – Text Revision criteria. Each item is scored on a 4-point scale ranging from 0 (never or rarely) to 3 (very often) with total score ranging from 0 to 54. The 18 items may be grouped into 2 subscales: hyperactivity/impulsivity (even-numbered items 2-18) and inattentiveness (odd-numbered items 1-17).

A missing individual item in the ADHD-RS-IV Preschool Version will be imputed as follows:

- 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 will be 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 will be set to missing.

- Both subscales (hyperactivity/impulsivity and inattentiveness) must be non-missing in order to calculate the total score.

18.2.2 Key Secondary Endpoint

The key secondary efficacy endpoint, CGI-I, provides an overall assessment of global symptom improvement. For a supportive analysis, the CGI-I categories will be dichotomized into 2 categories: “very much improved” and “much improved” classified as “improved,” and all other assessed categories grouped together as “not improved.”

18.3 Repeated or Unscheduled Assessments of Safety Parameters

If a subject has repeated assessments before the start of investigational product, then the results from the final assessment made prior to the start of investigational product will be used as baseline. If end-of-study assessments are repeated or unscheduled, the last post-baseline assessment while on investigational product will be used for generating descriptive statistics. However, all post-baseline assessments will be used for PCI value determination and all assessments will be presented in the data listings.

18.4 Missing Date of Investigational Product

When the date of the last dose of investigational product is missing for a subject in the Safety Analysis 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 treatment duration.

18.5 Missing Date Information for Prior or Concomitant Medications

For prior or concomitant medications, including rescue medications, incomplete (ie, 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.

18.6 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.

18.6.1 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 investigational product, then the day and month of the date of the first dose of 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 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 investigational product, then 01 January will be assigned to the missing fields.

18.6.2 Missing Month Only

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

18.6.3 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 investigational product, then the day of the date of the first dose of 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 investigational product or if both years are the same but the month is before the month of the date of the first dose of 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 investigational product or if both years are the same but the month is after the month of the date of the first dose of investigational product, then the first day of the month will be assigned to the missing day.

18.7 Incomplete Stop Date

The following rules will be applied to impute the missing numerical fields. If the date of the last dose of 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.

18.7.1 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 investigational product, then the day and month of the date of the last dose of 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 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 investigational product, then 01 January will be assigned to the missing fields.

18.7.2 Missing Month Only

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

18.7.3 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 investigational product, then the day of the date of the last dose of 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 investigational product or if both years are the same but the month is before the month of the date of the last dose of 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 investigational product or if both years are the same but the month is after the month of the date of the last dose of investigational product, then the first day of the month will be assigned to the missing day.

18.8 Missing Date Information for Adverse Events

For AEs, the default is to only impute incomplete (ie, partially missing) start dates. Incomplete stop dates may also be imputed when calculation of the duration of an AE is required per the protocol. If imputation of an incomplete stop date is required and both the start date and the stop date are incomplete for a subject, impute the start date first.

18.9 Missing Severity Assessment for Adverse Events

If severity is missing for an AE starting prior to the date of the first dose of 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 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.

18.10 Missing Relationship to Investigation 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 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.

18.11 Character Values of Clinical Laboratory Variables

If the reported value of a clinical laboratory variable cannot be used in a statistical analysis because, for example, a character string is reported for a numerical variable, then the appropriately determined coded value will be used in the statistical analysis. However, the actual values as reported in the database will be presented in data listings.

Table 5: Examples for Coding of Special Character Values for Clinical Laboratory Variables

Clinical Laboratory Test	Possible Results (in SI units)	Coded Value for Analysis
Chemistry: ALT	<5	0
Chemistry: AST	<5	0
Chemistry: Total Bilirubin	<2	0
Urinalysis: Glucose	≥55	Positive
	≤0	Negative
Urinalysis: pH	≥8.0	8.0

ALT=alanine transaminase; AST=aspartate transaminase

19. REFERENCES

Bretz, F., Pinheiro, J., and Branson, M. (2005). *Combining multiple comparisons and modeling techniques in dose response studies.* Biometrics 61, 738-748.

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences,* Hillsdale, NJ, Lawrence Earlbaum Associates.

Friede, T. & Kieser, M. 2006. Sample size recalculation in internal pilot study designs: a review. *Biom J,* 48, 537-55.

Glimm, E., Maurer, W. & Bretz, F. 2010. Hierarchical testing of multiple endpoints in group-sequential trials. *Stat Med,* 29, 219-28.

Kuczmarski, R. J., Ogden, C. L., Grummer-Strawn, L. M., Flegal, K. M., Guo, S. S., Wei, R. & Et Al 2002. 2000 CDC growth charts for the United States: Methods and development. *Vital Health Stat,* 11.

MacDougall, J. (2006). Analysis of dose-response studies - Emax model, *in* N. Ting (ed.), *Dose Finding in Drug Development*, Springer, New York, pp. 127–145

Pinheiro J, Bornkamp B, Bretz F 2006. *Design and analysis of dose-finding studies combining multiple comparisons and modeling procedures.* Biopharml Sta; 16:(5): 639-56.

Proschan, M., Lan, K. & Wittes, J. 2006. *Statistical Monitoring of Clinical Trials: A Unified Approach*, Springer Science + Business Media, LLC.

20. APPENDIX

20.1 Appendix 1 Algorithms for Deviation of the Pooled Common Variance

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 detail:

- 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 8.4, 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 and Kieser, 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 and Kieser, 2006).

20.2 Appendix 2 Sample SAS Codes

Mixed Model Repeated Measures (MMRM)

```
proc mixed /* select derived dataset */;
    class trt subjid visit;
    model chg = trt visit 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;
```

Missing Not at Random (MNAR) Sensitivity Analysis

Model 1 - Placebo Multiple Imputation

```
*Step 1;
proc mi data=unimputed out=discard nimpute=200 seed=347 noprint;
  where treatment=1; *where 1 is placebo;
  var baseline chy1-chy7
  mcmc outset=posteriors; *picks up the placebo posteriors;
run;

*100 posteriors for the placebo, rest for active;
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 active;
    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-chy7;
  mcmc inest=posteriors; *use the placebo posteriors;
run;

*Step 2;
data endpoint;
  set imputed;
  endpoint=chy7;
run;
```

```
proc sort data=endpoint;
  by _imputation_;
run;

proc glm data=endpoint;
  by _imputation_;
  class treatment;
  model endpoint = baseline treatment / solution;
  estimate 'Active 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=(chg4=.);
run;

proc sort data=unimputed;
  by treatment;
run;

proc mi data=unimputed out=imputed nimpute=200 seed=347 noprint;
  by treatment;
  var baseline chy1-chy7;
run;

%let SD=10;
* This will be replaced by the estimated SD[chg7] 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, etc) as per SAP;

data imputed;
```

```
  set imputed;
  if missing then chg7=chg7 + &penalty; * + or -, depending on the measure;
run;

*Step 2 (analysis of completed datasets) and Step 3 (inference) are the same as Step 2
and Step 3, respectively, for Model 1;
```