STATISTICAL ANALYSIS PLAN

Protocol Title	A Randomized, Double-Blind, Placebo-Controlled Study to Evaluate the Efficacy and Safety of Molindone Hydrochloride Extended-Release Tablets for the Treatment of Impulsive Aggression in Pediatric Patients with Attention Deficit/Hyperactivity Disorder (ADHD) in Conjunction with Standard ADHD Treatment
Protocol Number	810P301
Protocol Version	V9.0
Protocol Date:	01 Nov 2018
Study Drug	Molindone Hydrochloride Extended-Release Tablets (SPN-810)
Study Phase	III
IND Number	106,515
Sponsor	Supernus Pharmaceuticals, Inc. [redacted]
SAP Version	Final v3.0
Date (v2.0) Final (v3.0)	14 Mar 2019 11 June 2019

SIGNATURE PAGE

Prepared by [redacted]

Reviewed by [redacted]

Approved by [redacted]

Approved by [redacted]

Table of Contents

LIST OF ABBREVIATIONS

AE	Adverse event
BID	Bis in die, twice daily
CP	Conditional Power
CRO	Contract Research Organization
CRF	Case report form
ECG	Electrocardiogram
FDA	Food and Drug Administration
GCP	Good Clinical Practice
IA	Impulsive Aggression
ICF	Informed Consent Form
ICH	International Conference on Harmonisation
IND	Investigational New Drug Application
IV	Intravenous
IWRS	Interactive web-based randomization system
kg	kilogram
mcg	Microgram
MedDRA	Medical Dictionary for Regulatory Activities
mg	Milligram
msec	Millisecond
PK	Pharmacokinetic(s)
PCH	Percent Change in the frequency of IA behaviors per 7 days
PO	Oral
PP	Per Protocol
PT	Preferred term in MedDRA and WHODD
SAE	Serious adverse event
SAP	Statistical analysis plan
SM	Study Medication
SOC	System Organ Class in MedDRA
TEAE	Treatment-emergent adverse event
TESAE	Treatment-emergent serious adverse event
WHODD	WHO Drug Dictionary

Summary of Changes to the Version 1.0 SAP Submitted to FDA

Section	Page	Description of Change	Rationale
Signature Page		Updated signatory information	Administrative
1	9	a. The protocol version and date are updated b. The CRO name for MCIC determination is updated c. A sentence is added to state that this SAP does not cover the population PK analysis method	Protocol amendment V8.0 Administrative
2.2	9	"Oppositional defiant disorder" is added to the SNAP-IV rating scale	This subscale was omitted in the last version of SAP
3.1	10-13	a. Age range changed from "6-17" to "6-12". This update and removal of "13-17"are found in multiple sections. This describption will not be repeated below b. Treatment schedule diagrams are updated to reflect the change/ correction in the study day ranges for the screening and baseline periods c. A brief paragraph is added to the end of the section to describe the planned interim analysis	Protocol amendment V8.0
3.2	11	A new section titled "Schedule of Visits and Procedures" is added	Administrative
3.5	17	A paragraph is added to describe the changes in the sample size and randomization scheme after the planned interim analysis and the new blinded SD estimate.	Protocol amendment V9.0 Note the effect of sample size increase on the type-I error is controlled by the method of inverse normal p-value combination test (MINP) (Section 6.5.7)
4	18	Added "Baseline Period", "Titration Period", "Maintenance Period", and "Treatment Period" definitions	Clarification
5.1	19	A clarification of the visit range is added for the treatment period	Clarification
5.2	20-21	a. Added "at Visit 6" to clarify caregiver's timing for CGI-I	Clarification

		b. Added "Oppositional defiant disorder" to SNAP-IV ADHD scores c. Protocol Appendix numbers are updated in this section and various sections thereafter. This same description will not repeated Descriptions for CHQ-PF28 is updated to include reference for the derivation of scores for individual items, multi-item scales, and summaries	Clarification and correction is made to the CHQ-PF28 scoring based on the published manual
5.3	22	Previous Section "5.3 Pharmacokinetic Variables" is removed	PK analysis is not the scope of this SAP
5.4.7	23	Visit range for C-SSRS is corrected	Administrative
5.4.8	24	"by the site coordinator" is added to clarify the subjects who administer the caregivers's assessment of patient's IA daily activities	Clarification
5.4.9	24	Inserted sentences describing the ranges of the scale scores	Clarification
6.1	25	Respecified the number of decimal places for some summary statistics in the table presentation	Administrative
6.2	26	Clarified how unscheduled data points will be used in the analysis	Clarification
6.3	26	Updated the reference for the statistical method and added two more covariates for the adjustment of covariates: pooled site and baseline IA adjusted to 7 days	Administrative
6.4	28	Expanded text to include methods of deriving weekly IA frequency in the treatment and maintenance period when subject discontinued the study or date of last dose is missing; Replaced LOCF analysis with MMRM analysis for secondary endpoints	To provide detail of how primary endpoint is calculated when diary and dosing data are incomplete replace LOCF method by MMRM with respect to secondary endpoints
6.8	37	To add section number for where to find the multiplicity adjustment	Clarification
6.9	35	Deleted age subgroup of "13-17"	Protocol amendment
7.1.2	37	Added a sentence "Subjects will be analyzed according to the treatment they were randomized to" to the definition of "ITT Population"	Clarification

7.1.3	38	a. Changed diary completion compliance requirement from “80%” to “70%” for the per protocol population b. Added words “within the treatment period” in the diary compliance formula	To allow more patients in the PP analysis Clarification
7.2	38-39	a. Corrected a typographic error b. Clarified what populations will be used for the disposition summary	Administrative
7.3	39	Added information for protocol deviation summary	Clarification
7.5	39	Specified analysis population for the baseline summary	Clarification
7.6.1	40	Added word “treatment” to read “treatment compliance” to differentiate from “diary compliance”	Clarification
7.7.2.1	42-43	Expanded the section to provide further details for the imputation procedure	Clarification for missing data imputation
7.7.3	44	Analysis of the maintenance period is added	Administrative
7.7.4	44	Renumbered from previous Section 7.7.2.3	Administrative
7.7.5	44	a. Renumbered from previous Section 7.7.2.4 b. Added 2 new covariates c. Updated the statistical reference	Administrative
7.7.6	44	Defined CGI-I from caregiver and PI as key secondary efficacy variable	Clarification
7.7.7	44-47	a. Clarifies the visit numbers for the by-visit analysis and change from baseline analysis b. MMRM is added for the CGI-I and CGI-S analysis with detailed specifications and sample SAS code c. Clarified other secondary endpoints will be analyzed by ANCOVA d. Added a paragraph to clarify the multiplicity adjustment for the secondary endpoints	Clarification
7.8 (old)		Removed Pharmacokinetic Analysis section	out of the scope for this SAP but will be defined in separate document
7.8.1	48	Added a few clarifications for AE summaries	Clarification

7.8.2	49	Added a clarifications for lab summaries	Clarification
Reference	51	a. Included HealthActCHQ as the reference for CHQ-PF28 scoring b. Reference for the covariate adjustment is updated from “Zink and Koch, 2012” to “Dmitrienko and Koch, 2017”	Administrative
Appendix 2	54	Various tables and listings numbers and titles are updated as a result of the SAP text change	Administrative

1 INTRODUCTION

The purpose of this statistical analysis plan (SAP) is to ensure the credibility of the study findings by specifying the statistical approaches to the analysis of the data prior to database lock. This SAP covers the planned analyses of all data collected on the eCRFs and diary pages, and will identify handling of data issues. The statistical analysis plan presented in this document will supersede the statistical analysis methods described in the clinical protocol. Any deviations/changes from the planned analyses described in this SAP will be identified, with justification, in the appropriate section of the clinical study report. This SAP is based on the clinical study protocol 810P301, version 9.0, dated 01 November 2018 and its associated electronic case report forms (eCRF).

This SAP does not cover the analytic methods for the determination of Minimal Clinically Important Change (MCIC) in clinical efficacy. The assessment of MCIC will be performed by [redacted], which is a designated CRO to perform this analysis. This SAP also does not cover the methods for the population pharmacokinetic modeling and analysis.

2 STUDY OBJECTIVES

2.1 *Primary Objective*

The primary objective is to assess the efficacy and safety of SPN-810 in reducing the frequency of IA behaviors in pediatric patients with ADHD when taken in conjunction with standard ADHD treatment.

2.2 *Secondary Objectives*

The secondary objectives of the study are to assess the following:

- The effect of SPN-810 on the Clinical Global Impression – Improvement Scale (CGI-I)
- The effect of SPN-810 on the Clinical Global Impression – Severity Scale (CGI-S)
- The effect of SPN-810 on the child's overall health as measured by the Child Health Questionnaire Parent Form 28-item (CHQ-PF28)
- The effect of treating the child with SPN-810 on the parent-child relationship as measured by the Parenting Stress Index – Short Form version 4 (PSI-4-SF)
- The effect of SPN-810 on the caregiver-completed CGI-I
- The effect of SPN-810 on inattention, hyperactivity-impulsivity and oppositional defiant disorder measured by the SNAP-IV Rating Scale

2.3 *Tertiary Objectives*

[redacted]

3 STUDY DESIGN

3.1 *General Description*

Protocol 810P301 is a randomized, placebo-controlled, double blind, multicenter, parallel group, fixed dose study to demonstrate the efficacy, safety, and tolerability of SPN-810 in the treatment of IA in patients aged 6-12 years with ADHD in conjunction with standard ADHD treatment. The study design is presented schematically in Figure 1 and Figure 2. The study is divided into three phases: Pre-Treatment, Treatment, and Conversion/Taper.

Figure 1: Treatment Schedule (Conversion)

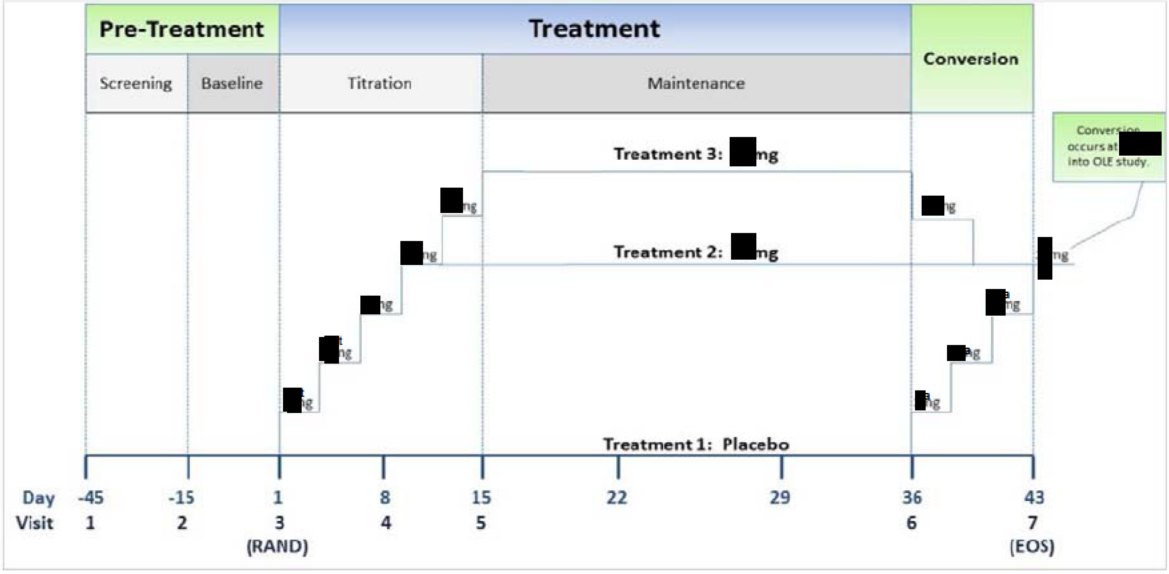
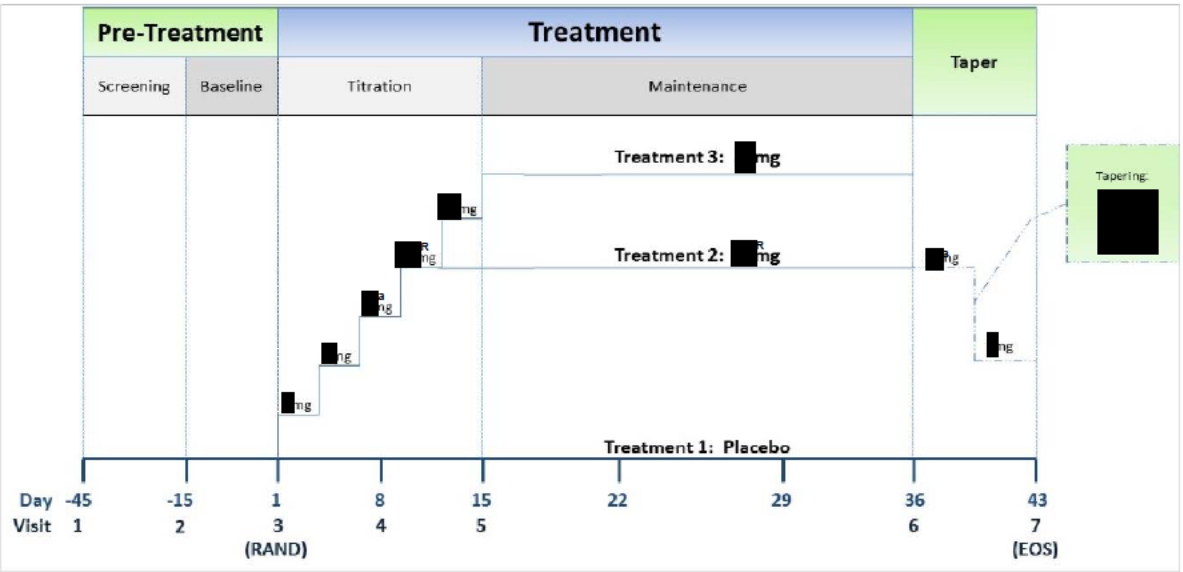


Figure 2: Treatment Schedule (Taper)



Following screening, eligible subjects will enter a flexible 15-day baseline period, at which time the IA diary will be issued to the subject's primary caregiver. At the end of the baseline period, eligible subjects whose primary caregiver has maintained at least 80% compliance with the IA diary will be randomized 1:1:1 to mg/day SPN-810, mg/day SPN-810, or placebo treatment group. A dose titration schedule will be followed, with dosing in the active treatment groups initiated at mg/day and increased approximately every days until the target dose is reached. After completing the two-week titration period and the three-week maintenance period, subjects will enter the conversion or tapering phase prior to discontinuing Study Medication (SM), at which time subjects will have the option to enter an open label extension study. This extension study will be conducted under a separate protocol. Subjects who choose to participate in the open label extension will enter that study at a dose of mg/day SPN-810.

An interim analysis is planned when approximately 50% of patients have completed or discontinued the study. This interim analysis is conducted by an independent unblinded statistician for Stage 2 sample size adjustment, if necessary.

3.2 *Schedule of Visits and Procedures*

All subjects who are randomized and take any study medication (SM) will be followed according to the protocol regardless of the number of doses of SM taken, unless consent for follow-up is withdrawn. Table 1 below presents the schedule of visits and procedures for the study.

Table 1: Schedule of Visits and Procedures

Phase	Pre-treatment		Treatment				Conversion/ Taper
Period	Screening	Baseline	Titration		Maintenance		
VISIT NUMBER	1	2	3	4	5	6	7
DAY	-45	-15	1	8	15	36	43
WINDOW (DAYS)	≤45d prior to Visit 3	≤15d prior to Visit 3	0[c]	7d±2d from Visit 3	14d±2d from Visit 3	21d±3d from Visit 5	7d±1d from Visit 6
Informed Consent/Assent [a]	X [b]						
R-MOAS, K-SADS-PL 2013 & Vitiello Aggression Scale	X						
Medical History	X		X [d]				
Demographics	X						
Physical Examination	X					X	
ECG (12-lead)	X			X		X	
Inclusion/Exclusion Criteria	X						
Randomization			X [b,e]				
Urine Drug Screen	X		X				
Urine Pregnancy Test [f]	X		X				X
Diary Training & Distribution or Evaluation		X	X [e]	X	X	X	
Vital Signs[g]	X						X
Weight, height, BMI	X		X	X	X	X	X
Hematology/chemistry/Urinalysis	X		X			X	X
PK Blood Sampling				X [h]	X [h]		
Columbia Suicide Severity Rating Scale (CSSRS)	X		X	X	X	X	X
Investigator CGI-S	X		X	X	X	X	
Caregiver and investigator CGI-I				X	X	X	
Efficacy scales (SNAP-IV, CHQ-PF28, PSI-4-SF)			X			X	
Safety Scales (Simpson-Angus, Barnes, AIMS)			X	X	X	X	X
Infrequent Behaviors Checklist			X	X	X	X	
Adverse Events			X	X	X	X	X
Concomitant Medications	X	X	X	X	X	X	X
Drug Dispensation			X [b]		X	X	

Phase	Pre-treatment		Treatment				Conversion/ Taper
Period	Screening	Baseline	Titration		Maintenance		
VISIT NUMBER	1	2	3	4	5	6	7
DAY	-45	-15	1	8	15	36	43
WINDOW (DAYS)	≤45d prior to Visit 3	≤15d prior to Visit 3	0[c]	7d±2d from Visit 3	14d±2d from Visit 3	21d±3d from Visit 5	7d±1d from Visit 6
Drug Return and Compliance					X	X	X
Diary Return							X

a Written consent must be obtained prior to performing any study-related procedure.

b Access IWRS.

c Visit 3 will occur at least 15 days following Visit 2

d Assess for any clinically significant change in Medical History since screening.

e Diary compliance must be at least 80% (minimum of 12 days out of 15) to qualify for randomization.

f To be performed for female subjects of childbearing potential prior to administration of first dose of SM and will have to be tested as negative for the subject to continue in the study.

g Heart Rate (HR), blood pressure (BP), temperature, and respiratory rate will be measured.

h Total of 5 PK blood samples will be obtained over one or two visits (Visit 4 and/or Visit 5)

3.3 *Study Treatment*

3.3.1 *Treatment, Dose, and Mode of Administration*

SPN-810 (molindone hydrochloride extended-release tablet) or matching placebo tablets will be administered orally (PO) twice daily (BID). Subjects will be force-titrated over a period of 2 weeks to their final randomized dose.

- Treatment 1: placebo
- Treatment 2: [redacted]mg total daily dose (TDD)
- Treatment 3: [redacted]mg TDD

3.3.2 *Duration of Treatment and Study Duration*

Total subject duration on study is approximately 13 weeks.

- Pre-treatment phase: 45 days
  - Screening period: up to 30 days
  - Baseline period: at least 15 days
- Treatment phase: 5 weeks
  - Titration period: 2 weeks
  - Maintenance period: 3 weeks
- Conversion/Tapering phase: 1 week

3.3.3 *Methods of Assigning Subjects to Treatment Group*

Allocation of study drug will be completed centrally through the use of an interactive web response system (IWRS) that will determine which kit to assign to the subject. Separate schedules for subject randomization and drug list will be created. The randomization scheme assigns treatment to each randomization number in a 1:1:1 allocation ratio.

Upon admission to the study, subjects will be assigned site ([redacted]) numbers and subject numbers [redacted]), in the sequence that they are entered. Subjects who complete the Baseline Period and continue to meet all eligibility criteria will be assigned kit numbers, according to the randomization schedule, by using the IWRS.

3.3.4 *Blinding*

Subjects, all site personnel, the Sponsor and its designated Contract Research Organizations (CROs) are blinded with the following exceptions. The CRO and Sponsor personnel who are

involved in the randomization generation, study drug supply, and PK concentration assays are unblinded. However, they are not involved in the routine study monitoring, data cleaning and review, the design of this SAP or data analysis.

3.4 *Hypotheses*

The null and alternative hypotheses are:

- H_{01}: There is no difference between SPN-810 [redacted]mg and Placebo in the treatment of IA in subjects with ADHD in conjunction with standard ADHD treatment.
- H_{a1}: There is a difference between SPN-810 [redacted] mg and Placebo in the treatment of IA in subjects with ADHD in conjunction with standard ADHD treatment.

- H_{02}: There is no difference between SPN-810 [redacted]mg and Placebo in the treatment of IA in subjects with ADHD in conjunction with standard ADHD treatment.
- H_{a2}: There is a difference between SPN-810 [redacted] mg and Placebo in the treatment of IA in subjects with ADHD in conjunction with standard ADHD treatment.

3.5 *Determination of Sample Size*

Based on results from the Phase 2 study, it is assumed that the average treatment difference between SPN-810 dose groups and placebo is 15 with a common standard deviation of 27.3 in the primary efficacy endpoint. A sample size of 77 subjects per arm (231 subjects for 3 arms) will yield 90% power to detect a non-zero difference between the median of active drug group [redacted] mg or [redacted] mg dose) and the placebo group using the Wilcoxon rank-sum test with a 2-sided significance level α=0.05.

It is assumed that approximately 20% subjects will dropout before the completion of the study and hence, an adjusted total of 291 subjects will be randomized in a 1:1:1 ratio to obtain 231 subjects in the ITT population at the completion of the study.

To estimate the variability for designing Study 810P503, the standard deviation (SD) of the primary endpoint for Study 810P301 was reviewed blindly. This review indicates that the SD for interim data was actually 34.83 instead of the protocol-assumed 27.3 (see above). This increase in the SD resulted in the increase in the sample size to 122 subjects per treatment. Allowing for 20% dropout, the final sample size for randomization is 153 subjects per treatment.

The sample size was calculated using the nQuery Advisor Software, Version 7.

3.6 *Changes to Analyses Planned in the Protocol*

This SAP will be updated as necessary if any changes in the protocol occur due to the expected protocol amendment impacting the SAP.

4 DEFINITIONS AND DERIVED DATASETS

Table 1. Terminology and Definition

Terminology	Definition
Study Medication	SPN-810 or the matching placebo.
Study Day	The first dose date (Day 1) is defined as the date on which a subject took the first dose of double-blind study medication (SM), Other study days are defined relative to the Study Day 1. For visits prior to the first dose of the SM, Study Day is calculated as Visit Date – Day 1. For visits after the first dose, Study Day is calculated as Visit Date – Day 1 +1.
Enrolled	Subject has provided signed informed consent.
Randomized	Subject is assigned with the kit number for the SM.
Completer for the Study	Subject who completes the maintenance period (Visit 6).
Baseline	Unless specified otherwise, baseline is defined as the non-missing value collected most recent to and before the time of the very first dose of the SM.
Baseline Period	For subject daily IA diaries, on and prior to Visit 3 (inclusive, ≤Day 1).
Prior Medication	Medication collected on the Prior/Concomitant Medication CRF, with the start date prior to Study Day 1.
Concomitant Medication	Medication collected on the Prior/Concomitant Medication CRF, with end date on/after Study Day 1.
Titration Period	For subject daily IA diaries from Visit 3 through Visit 5 (v3< to ≤ V5)
Maintenance Period	For subject daily IA diaries from Visit 5 through Visit 6 (V5< to ≤ V6)
Treatment Period	Titration and Maintenance Periods combined
Derived Dataset	Derived dataset (analysis dataset) is any permanently stored collection of data containing raw and/or derived variables created to support the production of statistical summary tables. Derived variables will be created from the “raw” data. Analysis datasets to be created will include ADEFF (efficacy analysis dataset), ADSL (subject-level analysis dataset), ADAE (Adverse event analysis dataset) and other as appropriate. In deriving ADEFF, the source data include both the episodic diary and evening diary data files. Specifications for derived datasets will be developed which will include the names and definitions of derived variables in the derived SAS datasets.

5 EFFICACY AND SAFETY VARIABLES

5.1 *Primary Efficacy Variable*

The primary efficacy endpoint is percent change (PCH_T) in the frequency (unweighted score) of IA behaviors per 7 days in the treatment (titration and maintenance) period relative to the Baseline period calculated over the number of days with non-missing IA diary data.

PCH_T will be calculated as 100 x (T – B)/B, where T and B are IA behavior frequencies per 7 days during the treatment period (from Day 2 through Visit 6, inclusive) and baseline period (Day ≤1), respectively. The IA behavior frequency per 7 days is defined as (SUM/DAY) x 7, where SUM is the total of the IA behaviors reported in the subject IA diary, and DAY is the number of days with non-missing IA score in the subject IA diary during the specified study period.

This primary efficacy variable will be based on a checklist of 15 IA behaviors collected in an electronic IA diary. The IA diary comprises two parts: 1) An episodic diary that will be used by the primary caregiver (or alternate) to enter events as soon as possible after they are observed; and 2) an evening diary that will prompt the caregiver to review events for the day and to enter any events that were not previously captured. Each event will be characterized by a checklist of 15 observed behaviors: Yelling, Screaming, Threatening, Scratching, Throwing, Slamming, Hitting Self, Arguing, Cursing, Name Calling, Shoving, Hair Pulling, Fighting, Hitting Others, and Kicking Others. The checklist will indicate whether each behavior was observed (coded 1) or was not observed (coded 0) during the incidence of an event. Each day can have multiple events. A day can have no event, as can be attested in the evening diary. In this case, if no event is reported during a day, and the evening diary confirms this, the daily event score for that subject will be 0 for each of the 15 IA variables. Behaviors not on this list will not be captured.

5.2 *Secondary Efficacy Variables*

1. Actual CGI-I (Investigator-rated Clinical Global Impression – Improvement Scale) score at Visit 6
2. Change from Visit 3 to Visit 6 in CGI-S (Investigator-rated Clinical Global Impression – Severity) score
3. CHQ-PF28 at Visit 6 (Overall Health Measured by Child Health Questionnaire Parent Form 28-item)
4. PSI-4-SF at Visit 6 (Parenting Stress Index – Short Form)
5. Caregiver-completed CGI-I at Visit 6
6. SNAP-IV ADHD scores at Visit 6 (Swanson, Nolan and Pelham-IV Rating Scale for Inattention, hyperactivity-impulsivity and oppositional defiant disorder measured) in
    - Inattention

- Hyperactivity/Impulsivity
- Oppositional Defiant Disorder
- Combined subscale

CGI-I (Protocol Appendix 10.3), relative to the condition at Visit 3, will be evaluated by the caregiver and by the Investigator at Visits 4 to 6 on a 7-point scale with 1=Very much improved, 2=Much improved, 3=Minimally improved, 4=No change, 5=Minimally worse, 6=Much worse, and 7=Very much worse.

CGI-S (Protocol Appendix 10.3) will be evaluated by the Investigator at Visits 1, 3 - 6 on a 7-point scale with 1=Normal, 2=Borderline ill, 3=Mildly ill, 4=Moderately ill, 5=Markedly ill, 6=Severely ill, and 7=Extremely ill.

The Child Health Questionnaire Parent Form 28-item (**CHQ-PF28**) (Protocol Appendix 10.11) is evaluated by the primary caregiver at Visit 3 and Visit 6 for the subject's health status and health related quality of life. CHQ-PF28 items have four, five, or six response options, divided over eight multi-item scales (Physical Functioning [PF, Q2.1a–2.1c], Behavior [BE, Q5.1a-5.1c, Q5.2], Mental Health [MH, Q6.1a–6.1c], Self-Esteem [SE, Q7.1a-7.1c], General Health Perceptions [GH, Q1.1, Q8.1a-8.1c], Parental Impact - Emotional [PE, Q9.1a-9.1b], Parental Impact - Time [PT, Q9.2a-9.2b], and Family Activities [FA, Q9.3a-9.3b]) and five single item concepts (Global Health [GGH, Q1.1], Role/Social Limitations - Emotional/Behavior [REB, 3.1], Role/Social Limitations - Physical [RP, Q3.2], Bodily Pain/Discomfort [BP, Q4.1], Global Behavior [GBE, Q5.2], Family Cohesion [FC, Q9.4], and Change in Health [CH, Q8.2]). In addition, the individual scale scores will be aggregated to derive 2 summary component scores: the physical functioning and psychosocial health summary scores. Range on subscales and the overall scale is 0–100, where 0 is the worst possible health state and 100 the best possible health state. Details for item and multi-item scale scoring, calculation of standardized scores and Aggregate Physical and Psychological Summary Scores can be found in the Child Health Questionnaire (CHQ) Scoring and Interpretation Manual (HealthActCHQ).

The Parenting Stress Index – Short Form (**PSI-4-SF**) (Protocol Appendix 1.10) is a 36-item self-report measure of parenting stress at Visit 3 and Visit 6. Three subscales (Parental Distress [PD: Q1-12], Parent-Child Dysfunctional Interaction [P-CDI: Q13-24], and Difficult Child [DC: Q25-36]) consist of 12 items each. Parents use a 5-point scale to indicate the degree to which they agree with each statement. The subscale score is the sum of the respective item scores. "Total Stress" score is the sum of the 3 subscale scores. The subscale will be set to missing if 2 or more (>1) item scores in the subscale are missing. Total Stress score is set to missing if 1 or more subscale score is missing.

The **SNAP-IV** (Protocol Appendix 10.7) rating scale includes 18 ADHD and 8 oppositional defiant disorder (ODD) symptoms as specified in the DSM-IV-TR and International Statistical Classification of Diseases and Health Related Problems 10th Revision (ICD-10) Classification of Mental and Behavioral Disorders. The symptoms are scored by assigning a severity estimate for each symptom on a 4-point scale. The SNAP-IV rating is performed by the parent or legal representative at each visit. The ratings from the SNAP-IV scale are grouped into the following 4 subscales:

- ADHD-Inattention (items #1-9),
- ADHD-Hyperactivity/Impulsivity (items #10-18)
- ODD (items #19-26)
- ADHD-Combined subscale: (items #1 – 18)

Each subscale score is the average of the available item scores in the subscale. The SNAP-IV rating scale will be administered at Visit 3 and Visit 6.

5.3 *Pharmacokinetic Variables*

Analysis of pharmacokinetic variables will not be covered in this SAP.

5.4 *Safety Variables*

5.4.1 *Medical History and Adverse Events (AEs)*

AEs that occur on/after signing ICF will be recorded on the AE CRF. Medical history and AEs will be coded for preferred term (PT) and system organ class (SOC) using Medical Dictionary for Regulatory Activities (MedDRA) version 18.1. AEs are evaluated by the investigators for seriousness, relationship, and severity.

5.4.2 *Prior, Concomitant, and Prohibited Medications*

Prior medications and concomitant medications will be extracted from the Prior/Concomitant Medication CRF. Medications taken prior to Study Day 1 will be considered as prior medications and medications taken on or after Study Day 1 will be considered as concomitant medications. Prior and concomitant medications will be coded using WHO Drug Dictionary (WHODD) version 1 June, 2015 for PT and Anatomical Therapeutic Chemistry (ATC) classification.

Prohibited medications include, but are not limited to,

- α_2- adrenergic agonists (e.g. clonidine and guanfacine) used for any other reason except for monotherapy treatment for ADHD

- Anti-psychotics including aripiprazole, risperidone, quetiapine, and ziprasidone
- Anticonvulsants including carbamazepine and valproic acid, antidepressants, mood stabilizers including lithium, benzodiazepines, cholinesterase inhibitors or any drug known to inhibit CYP2D6 activity
- Herbal supplements

5.4.3 *Clinical Laboratory Tests (Hematology, Chemistry, and Urinalysis)*

Laboratory tests will be converted to follow the International System (SI) units. Laboratory tests will be collected at Visits 1, 3, 6, and 7. The tests include hematology, serum chemistry and urinalysis panels. Laboratory values will be categorized according to the appropriate laboratory reference ranges.

5.4.4 *Vital Signs, Height and Weight*

Vital signs include body temperature, heart rate, respiratory rate, systolic blood pressure, and diastolic blood pressure. They will be collected at Visits 1 and 7. Subject height, weight, and BMI will be collected at Visits 1 and 3 – 7.

5.4.5 *12-Lead ECG*

The following ECG variables will be collected at Visits 1, 4, and 6: HR (heart rate), PR (respiratory rate), QRS interval, QT interval, and QTcF (QT interval corrected using the Fridericia's method).

5.4.6 *Physical Examinations*

Physical examinations will be conducted at Visit 1 and Visit 6. Clinically significant findings will be reported as adverse events.

5.4.7 *Columbia Suicide Severity Rating Scale (C-SSRS)*

The C-SSRS (Protocol Appendix 10.6) is a questionnaire assessing suicidal ideation and behavior at baseline and since last visit. This assessment will be conducted at Visits 1, and 3 – 7. The following outcomes are C-SSRS categories and have binary responses (yes/no).

Category 1 – Wish to be dead
Category 2 – Non-specific Active Suicidal Thoughts
Category 3 – Active Suicidal Ideation with Any Methods (Not Plan) without Intent to Act
Category 4 – Active Suicidal Ideation with Some Intent to Act, without Specific Plan
Category 5 – Active Suicidal Ideation with Specific Plan and Intent
Category 6 – Preparatory Acts or Behavior

Category 7 – Aborted Attempt
Category 8 – Interrupted Attempt
Category 9 – Actual Attempt (non-fatal)
Category 10 – Completed Suicide

Three composite binary endpoints and one severity score are derived from the above categories.

Suicidal ideation: A "yes" answer at any time during treatment to any one of the five suicidal ideation questions (Categories 1-5) on the C-SSRS.

Suicidal behavior: A "yes" answer at any time during treatment to any one of the five suicidal behavior questions (Categories 6-10) on the C-SSRS.

Suicidal ideation or behavior: A "yes" answer at any time during treatment to any one of the ten suicidal ideation and behavior questions (Categories 1-10) on the C-SSRS.

Suicidal Ideation score: The maximum suicidal ideation category (1-5 on the C-SSRS) present at the assessment. Assign a score of 0 if no ideation is present.

5.4.8 *Infrequent Behaviors Checklist*

The infrequent behaviors checklist is a checklist of 15 behaviors that (along with the 15 IA Diary behaviors) were qualitatively linked to IA during the development of the IA diary. These behaviors include teasing, spitting, biting, weapons, ripping, breaking, vandalizing, destroying, fire setting, hitting animal, kicking self, kicking animal, severe injury self, severe injury others, severe injury animal. Caregivers will be asked which, if any, of these behaviors have been observed since the patient's last visit. This assessment will be administered at Visits 3 – 6 by the site coordinator.

5.4.9 *Extrapyramidal symptoms (EPS) scales*

The EPS symptom and neuroleptic malignant syndrome will be evaluated using the Simpson-Angus scale (Protocol Appendix 10.8), the Barnes Akathisia scale (Protocol Appendix 10.9), and the Abnormal Involuntary Movement scale (AIMS) (Protocol Appendix 10.10). The Simpson-Angus scale is a 10-item rating scale, consisting of 1 item measuring gait, 6 items measuring rigidity, and 3 items measuring glabella tap, tremor and salivation. The Barnes Akathisia scale is a rating scale for drug-induced akathisia. It includes Objective Akathisia, Subjective Awareness of Restlessness, Subjective Distress Related to Restlessness, and Global Clinical Assessment of Akathisia subscales. Additionally, total score will be calculated by summing up the Objective and two Subjective scale scores (Barnes 1989). The objective and two subjective symptoms are rated on a 4-point scale ranging from 0 to 3. The global clinical assessment is rated on a 5-point scale ranging from 0 to 4. The AIMS test is a rating scale used to measure tardive dyskinesia (Munetz

1988). It includes 12 items to rate involuntary movements of various areas of the subject's body. These assessments will be administered at Visit 3 and all subsequent visits.

6 STATISTICAL METHODS

6.1 *General Methodology*

All statistical analyses will be conducted using SAS Version 9.4 or later on the Microsoft® Windows Operating System.

All tabulations of analysis results will include summaries for the three treatment groups: SPN810 [redacted]mg, SPN810 [redacted] mg and placebo.

Continuous data will be summarized using n (number of subjects with non-missing observations), mean, median, standard deviation (SD), minimum value, and maximum value. Categorical data will be summarized using the frequency count and percentage (n, %) of subjects in each category. Number of subjects with non-missing values or number of subjects with missing values (e.g., Not Done) will be presented, where appropriate. Subjects with missing values will not contribute to the denominator of percentage calculations. Counts of zero in any category will be presented without percentage.

Treatment differences will be evaluated as SPN-810 dose minus placebo. Unless otherwise specified, all statistical hypothesis tests will be 2-sided with a significance level of α=0.05. All confidence interval will be calculated 2-sided from the unadjusted 5% significance level. P-values will be reported to 4 decimal places. P-values <0.0001 will be presented as "<0.0001". Similarly, p-values >0.9999 will be presented as ">0.9999"

The precision rules for the presentation of summary statistics will be:

- Sample size (n, N) and number of missing responses (if displayed) – Integer
- Mean, median, percentiles, and confidence interval – One additional decimal place than reported/collected
- Standard deviation – Two additional decimal places than reported/collected
- Minimum, maximum – Same number of decimal places as reported/collected
- Ratios – two decimal places
- Percentages – one decimal place generally, or two decimal places for <0.1%

The data summaries will be accompanied by individual subject data listings. All data available from diary, eCRF, and external transfer (labs) will be listed and will include relevant and pertinent subject information, e.g., treatment, center, age/gender/race, and study day. The listings will be

sorted in the order of treatment group, unique subject ID, and assessment date/time and assessment name. The unique subject ID consists of 3-digit study number followed by [redacted]-digit site number and [redacted]-digit subject number separated by hyphen (e.g., [redacted]).

For each summary table, a supporting listing will be identified in the footnote.

Dates will be presented in the ISO-8601 format YYYY-MM-DD. Times will be displayed in 24-hour clock format. Numbering for tables, figures and listings will follow ICH E3 Guidelines.

Alternative methods of analysis of the data may be considered prior to un-blinding when some of the assumptions underlying the proposed analyses not be met. However, the reason for any departure from the planned approach and methods will be fully documented as an amendment to the SAP or in the clinical study report.

6.2 *Visit Window and Unscheduled Assessments*

There is no visit window for the by visit analysis. Data will be analyzed according to the visit they are associated with. Data collected at unscheduled time points will not be summarized at the unscheduled time points. However, the unscheduled data will be used for analysis if data are missing at the next scheduled visit due to early termination.

6.3 *Adjustment for Covariates*

An exploratory analysis of the primary efficacy endpoint will be performed using a randomization-based approach (Dmitrienko and Koch 2017). This analysis will include age (categorical and continuous), gender, pooled site and baseline IA adjusted to 7 days as covariates.

6.4 *Handling of Dropouts, Missing Data, and Data Discrepancies*

For the primary analysis of the primary efficacy endpoint, the frequency of IA behaviors during the treatment (titration + maintenance) period will be calculated over the number of days with non-missing IA diary data in the treatment period. No explicit imputation of missing data will be used, but this approach is implicitly equivalent to using the frequency of IA behaviors during the days with non-missing IA diary data to impute the frequency for days after study discontinuation or days with missing IA diary data. In addition to this implicit imputation, multiple imputations under the assumption of missing at random (MAR) and missing not at random (MNAR) will be used for handling missing data for sensitivity analysis. Details of imputation methods are specified in Section 7.7.2.

For analysis of frequency of IA behaviors during the treatment period (titration or maintenance), the following methods will be followed.

1) If subjects discontinued the study during the maintenance period, their maintenance period will be counted from Visit 5 (exclusive) through the date of last dose as documented on the End of Study eCRF page (inclusive). If the date of last dose is not collected, it will be imputed by the subjects' last visit for the study.

2) If subjects have ≥7 daily diaries in the maintenance period, their maintenance IA frequencies will be based on maintenance daily diaries.

3) If subjects have <7 daily diaries in the maintenance period, their maintenance IA frequencies will be calculated pooling the titration period diaries with the maintenance period diaries.

4) If subjects do not have any maintenance diaries, their maintenance IA frequencies will be entirely derived from the titration period diaries.

Missing data for secondary efficacy endpoints will be handled by using the Mixed Model for Repeated Measures method, which is implemented via SAS® PROC MIXED procedure. MMRM assumes data are "Missing At Random (MAR)". That is, the probability that an observation is missing may depend on the observed values but not the missing values in the previously scheduled visits.

AEs with incomplete information for start or stop dates (i.e., either day or month is missing) will be considered treatment-emergent (TEAE) unless the partial start date or the stop date confirms the AE started or ended prior to Study Day 1 (e.g., the day of the AE start date is unknown but the month and year indicate that the AE starts prior to Study Day 1). AEs with missing relationship to study drug will be included in the "Related" category for the summary tables (Section 7.8.1). AEs with missing severity will not be included in the summary table but will be footnoted.

When there is incomplete information regarding dosing dates for prior and concomitant medication, the medication will be considered as a concomitant medication unless it contradicts with the stop date. For example, a medication will be considered a prior medication if the month and year of the end date indicates a date before Study Day 1 even though the start date could be missing.

6.5 *Interim Analysis and Adaptive Design*

The sample size estimation described in Section 3.5 is based on the effect size assumptions from the Phase 2 data in which the primary endpoint was the change from baseline to endpoint in R-MOAS ratings. However, the primary endpoint for this study is percent change in the frequency of IA incidence per 7 days in the treatment period (PCH_T) relative to the Baseline period in the ITT population. Hence, there is uncertainty as to what magnitude of alternative hypothesis

treatment difference and between-subject variability is appropriate for the sample size calculation. To this end, a single interim analysis is planned after approximately 50% of subjects complete the study for the purpose of the Stage 2 sample size adjustment.

6.5.1 *Objective*

The primary objective of this Interim Analysis is to re-evaluate the pre-specified power and sample size to consider sample size adjustment using adaptive statistical methods.

6.5.2 *Responsible Party*

This interim analysis will be performed by an independent unblinded statistician who is not affiliated to the trial sponsor, trial management, data cleaning, and analysis.

The Senior Director of Biostatistics and Executive VP and Chief Medical Officer for Clinical Developmen at Supernus (sponsor) will receive recommendation letter from the independent unblinded statistician. They will make the final decision for the sample size adjustment.

6.5.3 *Confidentiality*

The independent statistician will sign the confidentiality agreement. He/she will not reveal any unblinded data/information to any persons within his/her own organization or to the sponsor, with the exception of sending recommendation as described in Section 6.5.8 below.

6.5.4 *Scope of Data Being Analyzed*

Primary efficacy endpoint – percent change in the frequency of impulsive aggression behaviors per 7 days in the treatment (PCH_T) period relative to Baseline period in the ITT population. The primary efficacy analysis will be performed using the Wilcoxon rank-sum test to compare the medians of each of the two doses of SPN-810 ([redacted]mg and [redacted]mg) with the median of the Placebo.

6.5.5 *Procedures for Interim Analysis*

a. Cutoff dates for collection of diary data for efficacy analysis is established based on an estimated target date of 50% of the randomized subjects completing/discontinuing the study.
b. All data received by the cutoff date will be validated, queries generated and resolved or pending queries documented.
c. A snapshot of the diary data will be taken and used for the interim analysis.
d. From the data obtained at the time of the interim analysis, the independent statistician will calculate

- The Conditional Power (CP_L) associated with SPN-810 [redacted] mg/day
- The Conditional Power (CP_H) associated with SPN-810 [redacted] mg/day

Using the formula derived from Chang (2008a), Hintze (2013), (Chen, 2004), and Jennison and Turnbull (2005),

$$CP\left(n_2, \alpha_{2,}, p_1 \middle| \delta, \sigma\right) = 1 - \Phi\left(\frac{1}{\sqrt{w_2}}\left[Z_{1-\alpha_2} - \sqrt{w_1} Z_{1-p_1}\right] - \frac{\delta}{\sigma}\sqrt{\frac{n_2}{2}}\right)$$

$$1 - \Phi\left(\frac{1}{\sqrt{w_2}}\left[Z_{1-\alpha_2} - \sqrt{w_1} Z_{1-p_1}\right] - Z_{1-p_1}\sqrt{\frac{n_2}{n_1}}\right),$$

where δ/σ is the expected effect size (standardized treatment difference from placebo), $w_1 = \frac{n_1}{n_1+n_2}, w_2 = \frac{n_2}{n_1+n_2}$ are the weights for the average of the Stage 1 and Stage 2 test statistic, n_1 is the sample size at stage 1, n_2 is the planned sample size per group at Stage 2 (increment from interim to final), $\alpha_2 = 0.025$ is the one-sided Type-I error for the final analysis, and Z_{1-p_1} is the Wilcoxon Rank Sum Z statistic with continuity correction at the interim analysis (Stage 1). When the interim analysis is conducted with 50% subjects have completed the study, $w_1 = w_2 = 1/2$. For the sample size re-estimation, Stage 2 data are assumed to follow the trend observed in Stage 1. Hence, the effect size is calculated as $\frac{\delta}{\sigma} = Z_{1-p_1}\sqrt{\frac{2}{n_1}}$. (Chang 2008b).

6.5.6 *Adaptation Rules*

Based on the conditional power calculated in Section 6.5.5, the independent statistician will use the following adaptation rules for sample size adjustment and make recommendation to the sponsor.

Table 2. Sample Size Adjustment Decision Rules

Case	Dose 1 ([redacted] mg)	Dose 2 ([redacted] mg)	Decision for Stage 2 Sample Size Adjustment	Verbiage for Communication to Sponsor (Section 6.5.8)
1	CP_L<10%	CP_H<10%	Terminate the trial.	CASE 1. Terminate the trial.
2	CP_L<10%	10%≤CP_H≤36% or CP_H>80%	Subjects to be randomized to Dose 1 group in Stage 2 will be allocated to Dose 2 and Placebo groups; However, subjects who received dose 1 prior to the interim analysis will continue to receive dose 1 in Stage 2.	CASE 2. Stop enrollment for [redacted]mg group, allocate Stage 2 planned [redacted] mg sample size to [redacted]mg and Placebo.
3	10%≤CP_L≤36% or CP_L>80%	CP_H<10%	Subjects to be randomized to Dose 2 group in Stage 2 will be allocated to Dose 1 and Placebo groups; However, subjects who received Dose 2 prior to the interim analysis will continue to receive Dose 2 in Stage 2.	CASE 3. Stop enrollment for [redacted]mg group, allocate Stage 2 planned [redacted] mg sample size to [redacted]mg and Placebo.
4	CP_L<10%	36%<CP_H≤80%	Enrollment for Dose 1 group will be terminated. Sample size in Dose 2 and Placebo groups will be increased to achieve the conditional power of at least 90% (before the multiplicity adjustment). This increase will include subjects allocated to Dose 2 and Placebo from Dose 1 termination The maximum increase will not result in the doubling of the total sample size. In other words, suppose n_1 and n_2 are the sample sizes per arm at the interim analysis and n_{2adj} is the calculated adjusted sample size. If $(n_1+ n_{2adj}) \le 2*(n_1+ n_2)$ then the adjusted sample size for Stage 2 will be n_{2adj} as calculated. If $((n_1+n_{2adj})> 2*(n_1+n_2))$ then $n_{2adj} = 2*(n_1+n_2) - n_1$.	CASE 4. Stop enrollment for [redacted]mg group, increase Stage 2 sample sizes for [redacted]mg and Placebo to xxx/group.

Case	Dose 1 ([redacted] mg)	Dose 2 [redacted] mg)	Decision for Stage 2 Sample Size Adjustment	Verbiage for Communication to Sponsor (Section 6.5.8)
5	36%<CP_L≤80%	Cp_H<10%	Enrollment for Dose 2 group will be terminated. Sample size in Dose 1 and Placebo groups will be increased to achieve the conditional power of at least 90% (before the multiplicity adjustment). This increase will include subjects allocated to Dose 1 and Placebo from Dose 2 termination. The maximum increase will not result in the doubling of the total sample size.	CASE 5. Stop enrollment for [redacted]mg group, increase Stage 2 sample sizes for [redacted]mg and Placebo to xxx/group.
6	10%≤CP_L≤36% or CP_L>80%	10%≤CP_H≤36% or CP_H>80%	Retain the original sample sizes of each treatment group.	CASE 6. No change in sample sizes.
7	36%<CP_L≤80%	36%<CP_H≤80%	Increase the sample size in all treatment groups to achieve the conditional power of at least 90% (before the multiplicity adjustment) in both treatment groups. The maximum increase will not result in the doubling of the total sample size.	CASE 7. Increase Stage 2 sample sizes for [redacted]mg to xxx [redacted]mg to xxx, and Placebo to xxx.
8	10%≤CP_L≤36% or CP_L>80%	36%<CP_H≤80%	Retain the original sample size in Dose 1 group and increase the sample size in Dose 2 and Placebo groups to achieve the conditional power of at least 90% (before the multiplicity adjustment). The maximum increase will not result in the doubling of the total sample size.	CASE 8. Increase Stage 2 sample sizes for [redacted]mg and Placebo to xxx/group.

Case	Dose 1 (mg)	Dose 2 mg)	Decision for Stage 2 Sample Size Adjustment	Verbiage for Communication to Sponsor (Section 6.5.8)
9	36%<CP_L≤80%	10%≤CP_H≤36% or CP_H>80%	Retain the original sample size in Dose 2 group and increase the sample size in Dose 1 and Placebo groups to achieve the conditional power of at least 90% (before the multiplicity adjustment). The maximum increase will not result in the doubling of the total sample size.	CASE 9. Increase Stage 2 sample sizes for and Placebo to xxx/group.

Note the sample size to achieve the conditional power of CP=90% is calculated as follows (Chang 2008a and 2008b, Hintze 2013, and Jennison and Turnbull 2000),

$$n_2 = \frac{2\sigma^2}{\delta^2}\left\{\frac{1}{\sqrt{w_2}}\left(Z_{1-\alpha_2}-\sqrt{w_1}Z_{1-p_1}\right)-\Phi^{-1}(1-CP)\right\}^2,$$

$$= \frac{n_1}{Z_{1-p_1}^2}\left\{\frac{1}{\sqrt{w_2}}\left(Z_{1-\alpha_2}-\sqrt{w_1}Z_{1-p_1}\right)-\Phi^{-1}(1-CP)\right\}^2.$$

The following rules will be applied to compute the sample sizes across the treatment arms in Stage 2:

- If a decision to discontinue subject enrollment is made for a particular dosing arm at Interim Analysis, subjects who were planned to be enrolled in that arm after Interim Analysis will be re-distributed to the remaining treatment arms. For example, if Dose 1 is dropped and Dose 2 is retained, the sample sizes for the three treatment arms in Stage 2 will be defined as follows:
  - Dose 1: 0
  - Dose 2: $n_{22} + n_{21}/2$
  - Placebo: $n_{23} + n_{21}/2$

  where n_{ij} is the sample size allocated to Stage i for Treatment Group j.

- If a decision is made to increase the sample size in at least one dosing arm, the modified sample size in the placebo arm will be equal to the larger modified sample size in the two dosing arms. For example, if the modified sample sizes in the two dosing arms in Stage 2 are n_{31} and n_{32}, respectively, the modified sample size in the placebo arm in Stage 2 will be set to $\max(n_{31}, n_{32})$. However, if the sample size is increased in one dosing arm, the sample size in the other dosing arm will not be automatically increased.

6.5.7 *Adaptive Design*

This section defines the mathematical approach used in the proposed adaptive design with two stages and data-driven decision rules (e.g., sample size modification at Interim Analysis). The derivation of the adaptive design is based on the combination function approach applied in conjunction with the closure principle, see, for example, Wang et al. (2001), Brannath et al. (2002) and Hommel (2005). This general approach is commonly used to construct adaptive designs in confirmatory trials and guarantees that the overall Type I error rate is protected in the trial at the nominal level (two-sided α= 0.05) regardless of the type of design modifications at an Interim Analysis, including decisions to drop a treatment arm or adjust the sample size in one or more arms.

Let H_1 and H_2 denote the null hypotheses of no treatment effect for the two dose-placebo comparisons (Dose 1 versus placebo and Dose 2 versus placebo). According to the combination function approach, the amount of evidence to reject these null hypotheses will be evaluated separately in each individual stage of the trial and the inferences at the final analysis will be performed by combining the evidence across the two stages. Further, the rejection rules for H_1 and H_2 will be chosen to achieve consistency with the Hochberg procedure (Dmitrienko and D'Agostino, 2013) that was pre-defined as multiplicity adjustment tool in this trial. To achieve this consistency, the Hochberg procedure will be applied within each stage (Hochberg, 1988).

Let p_i and q_i denote the p-values for testing the null hypothesis H_i, i = 1, 2, based on the data in Stages 1 and 2, respectively. The Stage 2 p-values are computed from the samples obtained after Interim Analysis, i.e., the sample size in a dosing arm may be adjusted as explained above. Further, $p_{(1)} < p_{(2)}$ will denote the ordered treatment effect p-values in Stage 1 and similarly $q_{(1)} < q_{(2)}$ will denote the ordered treatment effect p-values in Stage 2. These p-values will be used to construct a closure-based representation of the Hochberg procedure in each of the two stages. To define the closure-based representation, a closed family of intersection hypotheses that includes all possible non-empty intersections of the two null hypotheses will be set up. Table 3 and Table 4 display the intersection hypotheses in the closed family and associated intersection p-values that define the Hochberg procedure in Stage 1 and Stage 2 respectively.

Table 3: Intersection hypothesis and Hochberg intersection p-values in Stage 1

Intersection hypothesis	Hochberg intersection p-value
$H_1 \cap H_2$	$p_{12} = \min(2\,p_{(1)}, p_{(2)})$
H_1	p_1
H_2	p_2

Table 4: Intersection hypothesis and Hochberg intersection p-values in Stage 2

Intersection hypothesis	Hochberg intersection p-value
$H_1 \cap H_2$	$q_{12} = \min(2\,q_{(1)}, q_{(2)})$
H_1	q_1
H_2	q_2

Note that, if a certain dose is dropped due to futility at Interim Analysis, the corresponding p-values in Stage 2 is automatically set to 1 to remove the dosing arm from consideration. This guarantees that a null hypothesis cannot be rejected at Final Analysis if the corresponding dose was dropped due to futility at Interim Analysis. Further, the p-values computed at Interim Analysis (Stage 1 p-values) are not used for inferential purposes and therefore no α will be spent at Interim Analysis to stop the trial early due to overwhelming efficacy.

As an illustration, if Dose 1 is dropped at Interim Analysis, the corresponding treatment effect p-value in Stage 2 (q_1) will be set to 1. As a result, the intersection p-values will be defined in Stage 2 as shown in Table 5 (note that the smaller ordered p-value will be equal to q_2 and the larger ordered p-value will be equal to 1).

Table 5: Intersection hypotheses and Hochberg intersection p-values in Stage 2 if dose 1 is dropped at Interim Analysis

Intersection hypothesis	Hochberg intersection p-value
$H_1 \cap H_2$	$q_{12} = min\ (2\ q_2\ ,\ 1)$
H_1	1
H_2	q_2

The inferences for the two null hypotheses will be performed at final analysis using the composite treatment effect p-values that will be defined as follows. Let w1 and w2 define the stage weights that are prospectively defined based on the expected amount of information at Interim Analysis. Since Interim Analysis is planned to be taken after 50% of the subjects have completed the trial, n_1 (total sample size in Stage 1) and n_2 (total sample size in Stage 2 before sample size adjustment), and thus the pre-specified stage weights are given by

$$w_1 = \frac{n_1}{n_1 + n_2} = 0.5, w_2 = \frac{n_2}{n_1 + n_2} = 0.5.$$

Using the stage weights, the combination function to be used at Final Analysis will be defined as follows:

$$c(x_1, x_2) = 1 - \emptyset\left[\sqrt{w_1}\emptyset^{-1}(1 - x_1) + \sqrt{w_2}\emptyset^{-1}(1 - x_2)\right],$$

where Φ(x) is the cumulative distribution function of the standard normal distribution.

This combination function will be applied to compute composite intersection p-values for each intersection hypothesis in the closed family. The resulting composite intersection p-values are denoted by r_{12}, r_1, r_2, and are defined in Table 6.

Table 6: Intersection hypothesis and composite intersection p-values at final analysis

Intersection hypothesis	Composite intersection p-value
$H_1 \cap H_2$	$r_{12} = c(p_{12}, q_{12})$
H_1	$r_1 = c(p_1, q_1)$
H_2	$r_2 = c(p_2, q_2)$

After that, the null hypothesis H_1 will be rejected at Final Analysis (which implies a statistically significant treatment effect at Dose 1) if the composite p-values for all intersection hypotheses in the closed family that contain H_1 are significant at $\alpha = 0.05$ (two-sided), i.e.,

$$r_{12} \leq \alpha \text{ and } r_1 \leq \alpha.$$

Similarly, the null hypothesis H_2 will be rejected at Final Analysis (and a statistically significant treatment effect will be established at Dose 2) if the composite p-values for all intersection hypotheses in the closed family that contain H_2 are significant at $\alpha = 0.05$, i.e.,

$$r_{12} \leq \alpha \text{ and } r_2 \leq \alpha.$$

As indicated above, the resulting approach accounts for two sources of multiplicity in this clinical trial (multiple data looks and multiple dose-placebo comparisons) and guarantees Type I error rate control at $\alpha = 0.05$. The α control is retained even if the sample size is adjusted in one or more treatment arms at Interim Analysis. In addition, it needs to be emphasized that a null hypothesis cannot be rejected at Final Analysis if the corresponding dose was dropped due to futility at Interim Analysis.

6.5.8 *Communication of the Interim Analysis Results to the Trial Sponsor*

The recommendation will be made to the sponsor Executive VP and Chief Medical Officer for Clinical Development and Senior Director of Biostatistics at Supernus in a formal password protected letter. The verbiage for the letter will be:

Based on the conditional power calculation for the diary data at the cutoff date (DDMMMYYYY) and the pre-specified adaptation rules as documented in the Interim Analysis Plan, a recommendation is made for study 810P301 to the effect that:

'terminate the trial' or *'stop enrollment for [redacted] mg group or allocate Stage 2 planned [redacted] mg sample size to [redacted] mg and Placebo'* or *'stop enrollment for [redacted] mg group, allocate Stage 2 planned [redacted] mg sample size to [redacted] mg and Placebo' or 'stop enrollment for [redacted] mg group, increase Stage 2 sample sizes for [redacted] mg and Placebo to xxx/group' or 'stop enrollment for [redacted] mg group, increase Stage 2 sample sizes for [redacted] mg and Placebo to xxx/group' or 'no change in sample sizes'* or *'increase Stage 2 sample sizes for [redacted] mg to xxx [redacted] mg to xxx, and Placebo to xxx'* or *'increase Stage 2 sample sizes for [redacted] mg and Placebo to xxx/group'* or *'increase Stage 2 sample sizes for [redacted] mg and Placebo to xxx/group.'*

To protect the trial integrity, no other information will be provided in the letter or in any other communication.

6.6 *Timing of Final Analyses*

Data summary and statistical analysis will commence after all subjects have completed study Visit 7 (or discontinued prior to Visit 7) and the study database is cleaned and locked.

6.7 *Multicenter Study*

This is a multicenter study. The primary analysis will have all sites pooled. Where applicable, analysis adjusting for site effect or analysis evaluating homogeneity across sites may be carried out. If conducted, small sites (defined as those with sample size <9 ITT subjects/site) will be pooled with larger sites based on the geographic proximity.

6.8 *Multiple Comparisons/Multiplicity Adjustment*

With respect to the primary endpoint, the details of this multiplicity adjustment in the context of the adaptive design employed in this trial are defined in Sections 6.5.7. Multiplicity adjustment for secondary endpoints is described in Section 7.7.7.

6.9 *Examination of Subgroups*

In addition to the analysis adjusted for the confounding effects by the covariates (see Section 6.3), primary and key secondary efficacy variables will be summarized by subgroup to explore the heterogeneity of the treatment effect across subgroups. Subgroups are defined as follows:

- Gender (Male, Female)
- Age (6-9, 10-12)
- Race (White, Black/African American, Other)

7 STATISTICAL ANALYSIS

7.1 *Analysis Populations*

The population of “all enrolled subjects” consists of all those screened subjects who meet the requirements for study participation and are entered in the Baseline period (i.e., distributed IA diary) of the study. The population of “all randomized subjects” consists of all those enrolled subjects who complete the Baseline Period, meet the inclusion/exclusion criteria and are randomized.

7.1.1 *Safety Population*

The safety population will include all randomized subjects who received at least 1 dose of study drug and have at least one post-randomization safety measurements.

7.1.2 *Intent-to-treat (ITT) Population*

The intent-to-treat population will include all subjects who received at least 1 dose of study drug and have a baseline and at least 1 valid post-randomization assessment of frequency of IA behaviors based on IA diary entry. Subjects will be analyzed according to the treatment to which they were randomized.

7.1.3 *Per Protocol (PP) Population*

The per protocol population will include all of the subjects in the ITT population who completed the treatment period (Completed Maintenance Visit 6) with 70% diary completion compliance and who did not have major protocol deviations. Diary compliance is calculated as

$$\text{diary compliance} = \frac{\text{number of days diary completed within the treatment period}}{\text{Number of treatment days}} \times 100\%$$

Detailed specification of the PP population will be provided prior to the database lock for the final analysis.

7.2 *Disposition of Subjects*

Subject disposition and reasons for screening failure will be summarized for screen-failed subjects (Table 14.1.1).

Subject randomization schedule will be listed. Subjects whose randomization codes are unblinded during the conduct of the study will be listed along with the reason for un-blinding (Listing 16.1.7).

The following summaries and listings will be based on all randomized subjects.

Subjects (n and %) who completed and discontinued from the study will be tabulated by the treatment group for (1) Randomized, (2) ITT, (3) Safety, and (4) PP populations (Table 14.1.2a – Table 14.1.2d). The primary reasons for study discontinuation will be included in the tabulation. The primary reasons may include any of the following:

- Subject withdrew consent
- Lost to follow-up
- Administrative reason
- Adverse event
- Investigator decision
- Failure to follow required study procedures
- Other

In addition, subject disposition will be summarized by treatment period (combined titration and maintenance, titration, maintenance, and tapering) for (1) Randomized, (2) ITT, (3) Safety, and (4) PP populations (Table 14.1.3a – Table 14.1.3d).

In addition, subject's enrollment will be summarized by study site among ITT population (Table 14.1.4).

A data listing will be presented for subjects who do not complete the treatment along with the reasons for discontinuation (Listing 16.2.1.1).

7.3 *Protocol Deviations*

Protocol deviations will include, but are not limited to

- Non-compliance with any scheduled study visit
- Non-compliance with study treatment
- Prohibited medications
- Non-compliance with study inclusion or exclusion criteria
- Non-compliance with study assessment procedures

Subjects with major protocol deviations will be listed based on all randomized subjects. The listing will include a brief description of the deviation, deviation category, and if applicable, study day when deviation occurred along with other pertinent information (Listing 16.2.1.2).

If warranted, subjects (n and %) with major protocol deviations will be tabulated by treatment group and deviation category (Table 14.1.5a).

Additionally, subjects (n and %) in each analysis population with major protocol violations will be tabulated by treatment group (Table 14.1.5b). Subject's inclusion/exclusion from each analysis population will be listed along with the applicable reasons for exclusion (Listing 16.2.1.3).

Summary (n and %) of the analysis population will also be presented by treatment groups for the randomized population (Table 14.1.6).

7.4 *Demographics and Baseline Subject Characteristics*

Subject age, sex, ethnicity, race, height at screening, weight at screening and baseline will be summarized based on the ITT, Safety and PP populations by treatment group. Age will be summarized as a continuous variable as well as a categorical variable (6 to 9 [i.e., <10], and 10 to 12 [i.e., ≥10 and ≤ 12] years, inclusive) (Table 14.1.7a – Table 14.1.7c).

All demographic and baseline characteristics will be included in the subject data listing (Listing 16.2.1.4).

7.5 *Baseline Disease Characteristics*

Summaries of baseline disease characteristics will be based on the ITT population, with the exception of medical history which will be presented for the Safety population.

Subject incidence (n and %) of medical history will be tabulated by MedDRA System Organ Class (SOC) and Preferred Term (PT) for each treatment group as per Section 7.8.1 (Table 14.1.8).

R-MOAS and Vitiello Aggression Scale will be summarized by treatment group as per Section 6.1 (Tables 14.1.9 - 14.1.10).

All baseline disease characteristics will be listed (Listing 16.2.1.5 – Listing 16.2.1.7).

7.6 *Treatments*

7.6.1 *Treatment Compliance and Extent of Exposure*

Summaries of treatment compliance and exposure will be based on the ITT population.

Duration of exposure is defined as the total number of days a subject is exposed to any study treatment. This will be calculated for each subject by taking the difference between the date of last dose minus the date of the first dose, plus 1 (date of last dose – date of first dose +1).

Duration of Treatment exposure will be summarized using descriptive statistics (n, mean, SD, median, minimum, and maximum).

Percent of SM compliance is defined as

$$\text{Treatment Compliance} = \frac{\text{Number of Tablets Dispensed} - \text{Number of Tablets Returned}}{4\text{x}(\text{Date of Last Dose} - \text{Date of First Dose} + 1)} \cdot 100\%.$$

Each subject is expected to take 4 tablets per day as in the following. Subjects randomized to placebo group will take 2 placebo tablets in the morning and 2 placebo tablets in the evening. Subjects randomized to SPN-810 [redacted] mg will take one [redacted] mg active tablet and 1 placebo tablet in the morning and one [redacted] mg active tablet and 1 placebo tablet in the evening. Subjects randomized to SPN-810 [redacted] mg will take two [redacted] mg active tablets in the morning and two [redacted] mg active tablets in the evening. For each treatment, SM compliance will be summarized with number of subjects (n and %) by compliance category (<80%, 80-120%, and >120%). SM compliance will also be summarized as a continuous variable using descriptive statistics (n, mean, standard deviation, median, minimum, and maximum) for each treatment.

Summaries of treatment compliance and exposure will be presented separately for the Titration Period, Maintenance Period, and combined Titration and Maintenance Periods (Table 14.1.11 and Listing 16.2.1.8).

7.6.2 *Prior and Concomitant Medications*

Prior and concomitant medications will be tabulated separately based on the Safety Population. The number (n and %) of subjects taking at least one medication and the number (%) of subjects taking each medication at the WHODD preferred term level will be tabulated by treatment group, Anatomical-Therapeutic-Chemical classification 3 (ATC3), and PT. Subjects taking the same PT medication twice will only be counted once (Table 14.1.12.1 and Table 14.1.12.2).

A subject data listing will be provided to include the reported medication name, the WHODD PT, ATC4, study day and pertinent subject information (Listing 16.2.1.9.1 and Listing 16.2.1.9.2).

A separate data listing will include subjects who received prohibited medications (Listing 16.2.1.9.3).

7.7 *Efficacy Analyses*

7.7.1 *Primary Efficacy Analysis*

The primary analysis is based on the ITT population and the primary efficacy endpoint is the percent change (PCH_T) in the frequency (unweighted score) of IA behaviors per 7 days in the treatment period relative to the Baseline period calculated over the number of days with non-missing IA diary data.

The primary efficacy analysis will be performed using the Wilcoxon rank-sum test to compare the medians of each of the two doses of SPN-810 ([redacted] mg and [redacted] mg) with the median of the Placebo. The Hodges-Lehmann estimate of the difference (SPN-810 dose minus placebo) and the associated 95% confidence interval (CI) around the difference will be calculated. To protect the overall Type-I error rate from inflation due to sample size re-estimation and multi-dose to placebo comparisons, p-values will be adjusted using the method described in Section 6.8. The statistical analysis will be presented in the table along with the summary of the change from baseline frequency for Stage 1 and Stage 2 data (Table 14.2.1.1). The daily IA behavior diary will be listed. The daily frequency, the derived 7-day frequency for each week, and the derived 7-day frequency for the treatment period will be included (Listing 16.2.2.1.1 – Listing 16.2.2.1.2).

Box plots of Wilcoxon scores will be created for [redacted] mg dose SPN-810 and Placebo, and 36 mg dose SPN-810 and Placebo, separately, using SAS PROC NPAR1WAY.

7.7.2 *Sensitivity Analyses*

7.7.2.1 Multiple Imputation Assuming Missing at Random

This sensitivity analysis assumes that the diary data are missing at random (MAR), that is, the probability that an observation is missing may depend on the observed values but not the missing values (e.g., the probability of missing Week 5 diary is related to the observed Week 4 or Week 3 value but not the unobserved Week 5 value). In this sensitivity analysis, missing data is imputed multiple times, thus creating the multiply-imputed diary data. The final analysis incorporates the between-imputation variability with the within-imputation variability.

The SAS MI procedure is implemented using the following four steps.

1. Weekly IA total score is derived from the available Daily IA total scores (see Section 5.1). Weekly IA total score is set to missing if all Daily IA total scores are missing for the week.
2. SAS PROC MI is applied to the input dataset containing all weekly IA total scores during the baseline, titration, and maintenance period. The Markov Chain Monte Carlo (MCMC) method will be used for the multiple imputations. One hundred (100) multiply-imputed datasets will be created.
3. For each of the multiply-imputed data sets, PCH_T will be computed. The primary analysis based on the Wilcoxon rank-sum test will be conducted and the Hodges-Lehmann asymptotic 95% confidence intervals will be constructed.
4. To produce a single confidence interval for each dose and placebo comparison [redacted] mg versus placebo and [redacted] mg versus placebo), PROC MIANALYZE will be used and the Rubin's combination rules will be applied to the treatment effect estimates and associated asymptotic standard errors from the MI data sets. The treatment effect estimates will be defined as the midpoints of the asymptotic confidence intervals and the standard errors will be defined as the asymptotic standard errors (based on the width of the associated 95% confidence intervals) from the Hodges-Lehmann estimate of the individual datasets (Rubin 1987).

A sample SAS code is as follows.

```
PROC MI DATA=adqs NIMPUTE=100 SEED=123456789  MINIMUM=0 OUT=adqs_imp ;
     BY trtp ;
     VAR age sex bwk  twk1 twk2 twk3  twk4  twk5 ;
     MCMC CHAIN=multiple IMPUTE=full INITIAL=em PRIOR=jeffreys ;
RUN ;
```

where trtp is treatment group, bwk, twk1, twk2, twk3, twk4, and twk5 are variables of weekly IA scores during the baseline, and treatment (titration + maintenance) periods. These are derived as (SUM/DAY)*7, where SUM is the total of the IA scores reported on the IA diary in that analysis

week, and DAY is the number of days with non-missing IA score in the IA diary in that analysis week. Analysis week is defined as: 1st week is the 1st 7 days after visit 3 (exclusive) during titration, 2nd week is the rest of the days up to Visit 5 (inclusive) in titration period, 3rd week is the 1st 7 days after Visit 5 (exclusive), 4th week is the following 7 days after 3rd week, the 5th week is the rest of the days in maintenance period up to Visit 6 (inclusive). Adqs is the input data set with missing values, and adqs_imp is the complete imputed data set after 100 imputations. For the weekly IA score calculation, if the week has <3 daily diaries, the weekly IA score for this week will be set to missing and imputed using the above procedure.

From the above multiply-imputed IA data, the Hodges-Lehmann confidence interval will be calculated for each imputed data set. The point estimate for the between group difference and the associated standard error (SEM) will be derived from the confidence interval. These point estimates and the SEMs will be summarized by the SAS MIANALYZE procedure. A sample SAS code is as follows.

```
PROC MIANALYZE DATA=PARMEST ;
        MODELEFFECTS estim ;
        STDERR sem ;
        ODS OUTPUT PARAMETERESTIMATES=parmest2(KEEP= paramcd label estimate
                stderr lclmean uclmean probt) ;
RUN ;
```

where estim and sem are the point estimate and SEM from each imputed set, estimate, stderr, lclmen, uclmean, and probt are the final summarized point estimate, standard error, lower and upper confidence limits, and the p-value for the between group comparison (Table 14.2.1.2.1).

7.7.2.2 Multiple Imputation Assuming Missing Not at Random

This sensitivity analysis assumes that the diary data are missing not at random (MNAR), that is, the probability that an observation is missing may depend on its underlying unobserved value.

This approach may be considered “worst-case” sensitivity analyses as it assumes that after discontinuation subjects from the dosing arms would adopt the outcome model estimated from the placebo arm. To generate missing values from this “placebo-based” imputation model, PROC MI with the MNAR statement will be used (Table 14.2.1.2.2). A sample SAS code is as follows.

```
PROC MI DATA=adqs NIMPUTE=100 SEED=123456789 MINIMUM=0 OUT=adqs_imp ;
        CLASS trtp ;
        VAR age sex  bwk twk1 twk2 twk3 twk4 twk5;
        FCS reg ;
        MNAR MODEL(bwk twk1 twk2 twk3 twk4 twk5 / MODELOBS=(trtp='Placebo')) ;
RUN ;
```

where multivariate imputation is carried out by the fully conditional specification (fcs) method.

7.7.3 *Analysis of PCH Based on Data from Maintenance Period*

For this sensitivity analysis, the primary analysis will be repeated using data restricted to the maintenance period only (Table 14.2.1.2.3).

7.7.4 *Supplementary Analysis*

A supplementary analysis will be performed based on the Per Protocol population and will be presented in Table 14.2.1.2.4.

7.7.5 *Analysis Based on Adjustment of Covariates*

An exploratory analysis of the primary efficacy endpoint will be performed using a randomization-based approach (Dmitrienko and Koch 2017). Separate models will be produced for the following key prognostic baseline characteristics (age [categorical and continuous] and gender [categorical]) will be used in this analysis. In addition, pooled site and baseline IA adjusted to 7 days will be used as covariates (Table 14.2.1.4).

7.7.6 *Examination of Subgroups*

The primary and key secondary efficacy (Caregiver CGI-I and Investigator CGI-I) variables will be summarized by subgroup to explore the heterogeneity of the treatment effect across subgroups. This analysis will be done for the following subgroups: gender (male, female), age (6-9 and 10-12), and race (White, Black/African American, Other) (Table 14.2.1.5a - Table 14.2.1.5c).

7.7.7 *Secondary Efficacy Analyses*

All secondary analyses are based on the ITT population to compare the [redacted] mg dose with the Placebo. The secondary analysis variables are:

1. Actual Caregiver CGI-I score at Visits 4, 5, and 6 (Table 14.2.2 and Listing 16.2.2.2)
2. Actual Investigator CGI-I score at Visits 4, 5, and 6 (Table 14.2.3 and Listing 16.2.2.2)
3. Change from baseline (Visit 3) to Visits 4, 5 and 6 in Investigator CGI-S (Table 14.2.4 and Listing 16.2.2.3)
4. CHQ-PF28 Aggregate Physical and Psychological Scores will be summarized at Visit 6 (Table 14.2.5 and Listing 16.2.2.4a – Listing 16.2.2.4c)
5. PSI-4-SFwill be summarized at Visit 6 (Table 14.2.6, Listing 16.2.2.5a, and 16.2.2.5b)
    a. Parental Distress
    b. Parent-Child Dysfunctional Interaction

c. Difficult Child

6. SNAP-IV ADHD scores will be summarized at Visit 6 in (Listing 16.2.2.6a and Listing 16.2.2.6b):

a. Inattention ratings (Table 14.2.7.1)
b. Hyperactivity/Impulsivity ratings (Table 14.2.7.2)
c. Oppositional Defiant Disorder (Table 14.2.7.3)
d. Combined Scale ratings (Table 14.2.7.4)

Scores of CGI-I and CGI-S will be analyzed using a Mixed-Effects Model for Repeated Measure (MMRM). The model includes treatment, visit, and interaction between treatment and visit as fixed factors, baseline CGI-S as covariate for both CGI-I and CGI-S. The model parameters will be estimated using restricted maximum likelihood method with unstructured variance-covariance matrix and Kenward-Roger approximation to estimate denominator degrees of freedom. In case there is a convergence problem in the MMRM model with the unstructured variance covariance matrix, the first (co)variance structure which does not have convergence problem will be used for the analysis from the following ordered list: 1) Toeplitz 2) Autoregressive of order 1 3) Compound symmetry. The least squares mean of each treatment group, the difference in the least squares mean ([redacted]mg dose minus placebo and [redacted] mg dose minus placebo), and the 2-sided 95% CI for the difference will be obtained.

A sample SAS code is provided as follows.

```
PROC MIXED DATA=XXXX ;
    CLASS trtp(ref="P")  usubjid visit;
    MODEL Response = base trtp visit trtp*visit / DDFM=kr ;
    REPEATED visit / SUBJECT=usubjid(trtp) TYPE=UN ;
    Lsmeans trtp*visit/Pdiff alpha=0.05;"
RUN ;
```

Note: usubjid is subject ID, trtp ("P" denotes placebo treatment) is treatment variable, visit is the nominal study visit, and Response is respective endpoints.

Scores for CHQ-PF28, PSI-4-SF, and SNAP-IV will be analyzed using analysis of variance (ANOVA) at Visit 6.

To preserve the overall type I error rate at 0.05 for the secondary endpoints, a sequential testing procedure will be used with the following features. First only dose or doses that are significantly different from placebo for the primary endpoint will be tested for secondary endpoints. If the primary endpoint analysis indicates only one dose is significant, no multiplicity adjustment is required. Otherwise, multiplicity adjustment will be performed as in the following. The first of the secondary endpoints will be compared to placebo using the Hochberg step up procedure but only using those doses retained as a result of testing the primary endpoint. The second secondary endpoint will be tested in the same manner but only using those doses that were retained from the primary and the first secondary endpoint and so forth. As the endpoints are gone through in

the pre-defined order doses will only be retained if significant for all endpoints tested so far and at the given stage the Hochberg step-up procedure will be applied.

If the null hypothesis for the primary analysis is not rejected, then no multiplicity adjustment will be done for the secondary endpoints. Otherwise, a sequential testing procedure to preserve the type I error rate at 0.05 will be conducted as described below.

The ordering of the secondary endpoints follows the order in the list above.

7.8 *Safety Analyses*

Subject safety is assessed by AEs, labs, vital signs, 12-lead ECGs, and findings from the physical examinations. The occurrence of neurological side effects will be assessed by looking at any worsening in scores for each of the Simpson-Angus scale, Barnes Akathisia scale, and AIMS. Suicidal ideation and suicidal behavior will be measured by C-SSRS. All summaries related to safety analyses will use the safety population and will be presented by treatment groups.

7.8.1 *Adverse Events*

AEs will be summarized using subject incidence table by treatment period (titration, maintenance, combined titration and maintenance, and tapering). An overview of TEAE will be presented. This table will include numbers and percentages of subjects experiencing at least 1 TEAE, study drug-related TEAE, AE leading to study drug discontinuation, severe TEAE, serious TEAE, related TESAE, AESI, and TEAE with outcome of death (Table 14.3.1.1).

Additionally, the number and percent (%) are calculated based on the number of unique subjects within each category (e.g., preferred term) by treatment group. A subject reporting multiple events of the same category will be counted only once for that category. For summary purpose, AE relationship to the study drug will be grouped into “Unrelated” for “unrelated” or “unlikely related” and “Related” for “possibly”, “probably”, or “definitely related”. For subjects with more than one event coded to the same PT, the subjects will be counted for the categories with the strongest relationship and the greatest severity. The following subject incidence tables will be presented.

- TEAEs by PT (Preferred Term) sorted by the decreasing order of subject incidence in the combined active dose group (Table 14.3.1.2.1 – Table 14.3.1.2.3)
- TEAEs by SOC (System Organ Class) and PT sorted alphabetically (Table 14.3.1.3.1 – Table 14.3.1.3.3)
- Study Drug-Related TEAEs by SOC and PT (Table 14.3.1.4.1 – Table 14.3.1.4.3)
- Severity of TEAEs by SOC and PT (Table 14.3.1.5.1 – Table 14.3.1.5.3)
- TESAEs by SOC and PT when applicable (Table 14.3.1.6)

- Number and percent of patients reporting common AEs (≥5% in any group) by PT (Table 14.3.1.7.1 – Table 14.3.1.7.3)
- TEAEs leading to study drug discontinuation by SOC and PT (Table 14.3.1.8)
- Treatment Emergent AESI by SOC and PT (Table 14.3.1.9)

Where warranted by the subject incidence, these same tables may be presented by treatment period (Titration, Maintenance and Conversion/Tapering).

A subject data listing will be provided for all adverse events (Listing 16.2.3.1.1). Included in the listing are the reported term, PT, SOC, TEAE flag, study day when AE starts, duration, relationship, severity, action taken, outcome, and seriousness category.

Data listings will be provided for subjects who die on study, experience SAEs, have TEAEs leading to study drug discontinuation, or AEs of special interest (Listing 16.2.3.1.2 – Listing 16.2.3.1.5).

7.8.2 *Laboratory Test*

Clinical laboratory values will be summarized by treatment group and visit using descriptive statistics for hematology, biochemistry, and urinalysis. For quantitative laboratory parameters, both actual values and change from baseline values will be summarized (Table 14.3.2.1, Table 14.3.3.1, and Table 14.3.4.1).

Laboratory test result will be assigned a low, normal, high (LNH) classification according to whether the value is below (L), within (N), or above (H) the reference range provided by the laboratory. Within-treatment comparisons will be based on three by three tables (shift tables), that, for a particular laboratory test, compare the LNH classification at baseline to the LNH classification at visit. By subject-listings of all laboratory values will be provided for subjects who have L or H classification in any of their laboratory tests (Table 14.3.2.2, Table 14.3.3.2, and Table 14.3.4.2).

Note the summary and shift tables will only use the planned/scheduled tests. However, both scheduled and repeat tests will be in the data listing.

A complete lab data listing, including hematology, biochemistry, and urinalysis will be provided for all subjects. Investigator assessment of clinical significance is based on the local laboratory reference ranges (Listing 16.2.3.2.1 – Listing 16.2.3.2.3).

7.8.3 *Vital Signs, Height, Weight, and BMI*

Vital signs (body temperature, heart rate, respiration rate, systolic blood pressure, and diastolic blood pressure), and their changes from baseline will be summarized by treatment group and

visit (Table 14.3.5, Listing 16.2.3.3). Subject body size (weight, height, and BMI) will be summarized similarly (Table 14.3.6, Listing 16.2.3.4).

7.8.4 *12-Lead ECG*

By-visit tabular summaries of the quantitative ECG parameters and the overall ECG findings (normal, abnormal not clinically significant, or abnormal clinically significant) will be presented. The QT interval will be corrected using Fridericia's method.

ECG results will be summarized by visit by treatment group using descriptive statistics (for quantitative ECG parameters) and frequency tables (for qualitative ECG parameters, including the overall ECG finding). For quantitative ECG parameters, both actual values and change from screening values will be summarized (Table 14.3.7.1 and Table 14.3.7.2).

Data listings will be presented for ECG parameters by visit. QTcF meeting the criteria of >450 but ≤480, >480 but ≤500 or >500 msec for the absolute value, or >30 msec or >60 msec increase from baseline will be identified (Listing 16.2.3.5.1 and Listing 16.2.3.5.2).

7.8.5 *Columbia Suicide Severity Rating Scale (C-SSRS)*

C-SSRS outcomes will be summarized using number and percent of subjects by categories for suicidal ideation only, suicidal behavior only and suicidality (ideation and behavior combined). The summary will be presented by treatment groups. The proportion of subjects in each treatment group will be compared with the proportion of subjects in the placebo group using the Fisher's exact test (Table 14.3.8 and Listing 16.2.3.6).

7.8.6 *Infrequent Behaviors Checklist*

Infrequent behaviors will be listed for each subject by treatment group (Listing 16.2.3.7).

7.8.7 *Other Special Tests*

The occurrence of neurological side effects will be assessed by looking at the changes in scores from baseline to post-baseline visits for each of the Simpson-Angus scale (Listing 16.2.3.8), Barnes Akathisia scale (Listing 16.2.3.9), and AIMS (Listing 16.2.3.10). For each item on each of these scales, the number (and percentage) of subjects with a worse score at any post-baseline visit, compared to baseline, will be presented. A listing of these subjects will also be provided.

8 VALIDATION

The Clinical Operations, Clinical Research, and Statistics groups at Supernus will work with [redacted] and [redacted] to ensure that the data collected for the study are of the highest quality possible. The

study monitor will be responsible for reviewing and verifying the accuracy of the data recorded on the electronic case report forms (eCRFs) direct from source documents at the investigative site. [redacted] will be responsible for performing edit checks and reviewing all data entered into the electronic database to identify discrepant and/or inconsistent values and to send queries to the clinical sites. The Investigator will be responsible for answering queries about discrepant data and providing electronic signatures to confirm data integrity.

The programming of Tables, Listings and Figures (TLFs) based on the clinical data is outsourced to [redacted]. Supernus seeks to ensure the quality of the reports provided by [redacted] in the form of TLFs must pass a rigorous validation process involving the following processes.

- Derived datasets must be independently reprogrammed by a second programmer. The separate datasets produced by the 2 programmers must match 100%.
- Tables must be independently reprogrammed by a second programmer for and the results from both programs must match.
- Figures must be checked for consistency against corresponding tables and listings, or independently reprogrammed if there are no corresponding tables or listings.
- Listings must be checked for consistency against corresponding tables, figures, and derived datasets.

The entire set of TLF must be checked for completeness and consistency prior to its delivery to Supernus.

The above validation process must be repeated any time TLFs are redelivered using different data. Execution of this validation process must be documented through the study Table of Programs that will be provided to Supernus at study conclusion.

9 REFERENCES

Brannath W, M Posch, P Bauer. Recursive combination tests. Journal of the American Statistical Association. 2002; 97: 236-244.

Chang, M. Adaptive Design Theory and Implementation Using SAS and R. Chapman & Hall/CRC, Boca Raton, Florida. 2008a.

Change, M. Classical and Adaptive Clinical Trial Designs Using ExpDesign Studio. Wiley, New Jersey. 2008b.

Chen, YH, DL DeMets, and KKG Lan. “Increasing the sample size when the unblinded interim result is promising.” Statistics in Medicine, 2004. 23: 1023-1038.

HealthActCHQ. The CHQ Scoring and Interpretation Manual. Boston, MA: HealthActCHQ, 2013. 2013.

Dmitrienko, A, RB D’Agostino. Tutorial in Biostatistics: Traditional multiplicity adjustment methods in clinical trials. Statistics in Medicine. 2013; 32:5172-5218.

Dmitrienko A, AC Tamhane, and F bretz. Multiple Testing Problems in Pharmaceutical Statistics. Chapman & Hall/CRC, Boca Raton, Florida. 2010.

Dmitrienko, A and GG Koch. Analysis of Clinical Trials Using SAS: A Practical Guide. 2nd Ed. Cary, NC: SAS Institute Inc. 2017.

Hintze, J. PASS 12, NCSS, LLC. Kaysville, Utah, USA. 2013.

Hommel, G. Adaptive modifications of hypotheses after an interim analysis. Biometrical Journal. 2005; 5: 581-589.

Jennison C and Turnbull BW. Group Sequential Methods with Applications to Clinical Trials. Chapman & Hall/CRC, Boca Raton, Florida. 2000.

Mehta CR and SJ Pocock. “Adaptive increase in sample size when interim results are promising: a practical guide with examples.” Statistics in Medicine, 2000. 00:1-6.

Wang, SJ, HMJ Hung, Y Tsong, L Cui. Group sequential test strategies for superiority and non-inferiority hypotheses in active controlled clinical trials. Statistics in Medicine. 2001; 20: 1903-1912.

Appendix 1. Exclusion Criteria for the Per Protocol Populations

To be determined

Appendix 2. List of Tables, Listings, and Figures

Table No.	Title	Population
14.1.1	Subject Demographics Characteristics, and Reason for Screening Failure	Screening Failure Subjects
14.1.2a	Subject Disposition and Reason for Discontinuation	Randomized Population
14.1.2b	Subject Disposition and Reason for Discontinuation	ITT Population
14.1.2c	Subject Disposition and Reason for Discontinuation	Safety Population
14.1.2d	Subject Disposition and Reason for Discontinuation	PP Population
14.1.3a	Subject Disposition and Reason for Discontinuation by treatment period (combined titration and maintenance, titration, maintenance, and tapering)	Randomized Population
14.1.3b	Subject Disposition and Reason for Discontinuation by treatment period (combined titration and maintenance, titration, maintenance, and tapering)	ITT Population
14.1.3c	Subject Disposition and Reason for Discontinuation by treatment period (combined titration and maintenance, titration, maintenance, and tapering)	Safety Population
14.1.3d	Subject Disposition and Reason for Discontinuation by treatment period (combined titration and maintenance, titration, maintenance, and tapering)	PP Population
14.1.4	Subject Enrollment by Study Site	ITT Population
14.1.5a	Summary of Major Protocol Deviations During the Study, by Treatment Group	ITT Population
14.1.5b	Summary of Major Protocol Violations During the Study, by Treatment Group and Deviation Category	Randomized Population
14.1.6	Summary of Analysis Population	Randomized Population
14.1.7a	Demographics and Baseline Characteristics	ITT Population
14.1.7b	Demographics and Baseline Characteristics	Safety Population
14.1.7c	Demographics and Baseline Characteristics	PP Population

14.1.8	Medical History by System Organ Class and Preferred Term	Safety Population
14.1.9a	Retrospective Modified Overt Aggression Scale at Baseline – Item Scores	ITT Population
14.1.9b	Retrospective Modified Overt Aggression Scale at Baseline – Domain and Total Scores	ITT Population
14.1.10a	Vitiello Aggression Scale at Baseline – Item Scores	ITT Population
14.1.10b	Vitiello Aggression Scale at Baseline – Domain and Total Scores	ITT Population
14.1.11	Exposure to Study Medication and Compliance	ITT Population
14.1.12.1	Prior Medications by ATC Code and Preferred Term	Safety Population
14.1.12.2	Concomitant Medications by ATC Code and Preferred Term	Safety Population
14.2.1.1	Analyses of PCH in the Frequency of IA Behaviors During the Treatment Period - Primary Analysis	ITT Population
14.2.1.2.1	Analyses of PCH in the Frequency of IA Behaviors During the Treatment Period - Sensitivity Analysis, Multiple Imputation with MAR Assumption	ITT Population
14.2.1.2.2	Analyses of PCH in the Frequency of IA Behaviors During the Treatment Period - Sensitivity Analysis, Multiple Imputation with MNAR Assumption	ITT Population
14.2.1.2.3	Analyses of PCH in the Frequency of IA Behaviors During the Maintenance Period - Sensitivity Analysis	ITT Population
14.2.1.2.4	Analyses of PCH in the Frequency of IA Behaviors During the Treatment Period - Supplementary Analysis	PP Population
14.2.1.3	Descriptive Statistics for Frequency of IA Behaviors Per 7 Days and Change from Baseline by Study Period	ITT Population
14.2.1.4	Percent Change from Baseline to Treatment Period in Frequency of IA Behaviors: Adjustment for Covariates using Randomization-Based Approach of Dmitrienko and Koch	ITT Population
14.2.1.5a	Analyses of PCH in the Frequency of IA Behaviors During the Treatment Period by Gender - Primary Analysis	ITT Population
14.2.1.5b	Analyses of PCH in the Frequency of IA Behaviors During the Treatment Period by Age - Primary Analysis	ITT Population

14.2.1.5c	Analyses of PCH in the Frequency of IA Behaviors During the Treatment Period by Race - Primary Analysis	ITT Population
14.2.2	Investigator Clinical Global Impression – Improvement (CGI-I) Score by Visit	ITT Population
14.2.3	Caregiver Clinical Global Impression – Improvement (CGI-I) Score by Visit	ITT Population
14.2.4	Investigator Clinical Global Impression – Severity (CGI-S) Score and Change from Baseline by Visit	ITT Population
14.2.5.1	Child Health Questionnaire Parent Form 28 (CHQ-PF28) - Item Scores and Change from Baseline to Visit 6	ITT Population
14.2.5.2	Child Health Questionnaire Parent Form 28 (CHQ-PF28)-Component Scores and Change from Baseline to Visit 6	ITT Population
14.2.6.1	Parenting Stress Index – Short Form (PSI-4-SF) Total Score, Change from Baseline to Visit 6	ITT Population
14.2.6.2	Parenting Stress Index – Short Form (PSI-4-SF) Parental Distress Subscale Scores, Change from Baseline to Visit 6	ITT Population
14.2.6.3	Parenting Stress Index – Short Form (PSI-4-SF) Parent-Child Dysfunctional Interaction Subscale Scores, Change from Baseline to Visit 6	ITT Population
14.2.6.4	Parenting Stress Index – Short Form (PSI-4-SF) Difficult Child Subscale Scores, Change from Baseline to Visit 6	ITT Population
14.2.7.1	SNAP-IV ADHD Inattention Subscale Score and Change from Baseline to Visit 6	ITT Population
14.2.7.2	SNAP-IV ADHD Hyperactivity/Impulsivity Subscale Score and Change from Baseline to Visit 6	ITT Population
14.2.7.3	SNAP-IV ADHD oppositional defiant disorder Score and Change from Baseline to Visit 6	ITT Population
14.2.7.4	SNAP-IV ADHD Combined Scale Rating Score and Change from Baseline to Visit 6	ITT Population
14.2.8.1a	Subgroup Analysis of Investigator Clinical Global Impression – Improvement (CGI-I) Score by Visit (Age)	ITT Population
14.2.8.1b	Subgroup Analysis of Investigator Clinical Global Impression – Improvement (CGI-I) Score by Visit (Gender)	ITT Population

14.2.8.1c	Subgroup Analysis of Investigator Clinical Global Impression – Improvement (CGI-I) Score by Visit (Race)	ITT Population
14.2.8.2a	Subgroup Analysis of Caregiver Clinical Global Impression – Improvement (CGI-I) Score by Visit (Age)	ITT Population
14.2.8.2b	Subgroup Analysis of Caregiver Clinical Global Impression – Improvement (CGI-I) Score by Visit (Gender)	ITT Population
14.2.8.2c	Subgroup Analysis of Caregiver Clinical Global Impression – Improvement (CGI-I) Score by Visit (Race)	ITT Population
14.3.1.1	Overall Incidence of Treatment-Emergent Adverse Events by Study Period	Safety Population
14.3.1.2.1	Incidence of Treatment-Emergent Adverse Events by Preferred Term – Titration and Maintenance Period	Safety Population
14.3.1.2.2	Incidence of Treatment-Emergent Adverse Events by Preferred Term – Titration Period	Safety Population
14.3.1.2.3	Incidence of Treatment-Emergent Adverse Events by Preferred Term – Maintenance Period	Safety Population
14.3.1.3.1	Incidence of Treatment-Emergent Adverse Events by System Organ Class and Preferred Term – Titration and Maintenance Period	Safety Population
14.3.1.3.2	Incidence of Treatment-Emergent Adverse Events by System Organ Class and Preferred Term – Titration Period	Safety Population
14.3.1.3.3	Incidence of Treatment-Emergent Adverse Events by System Organ Class and Preferred Term – Maintenance Period	Safety Population
14.3.1.4.1	Incidence of Treatment-Emergent Adverse Events by System Organ Class, Preferred Term, Treatment-Relatedness, and Treatment Group – Titration and Maintenance Period	Safety Population
14.3.1.4.2	Incidence of Treatment-Emergent Adverse Events by System Organ Class, Preferred Term, Treatment-Relatedness, and Treatment Group – Titration Period	Safety Population
14.3.1.4.3	Incidence of Treatment-Emergent Adverse Events by System Organ Class, Preferred Term, Treatment-Relatedness, and Treatment Group – Maintenance Period	Safety Population
14.3.1.5.1	Incidence of Treatment-Emergent Adverse Events by System Organ Class, Preferred Term, Severity, and Treatment Group – Titration and Maintenance Period	Safety Population

14.3.1.5.2	Incidence of Treatment-Emergent Adverse Events by System Organ Class, Preferred Term, Severity, and Treatment Group – Titration Period	Safety Population
14.3.1.5.3	Incidence of Treatment-Emergent Adverse Events by System Organ Class, Preferred Term, Severity, and Treatment Group – Maintenance Period	Safety Population
14.3.1.6	Incidence of Treatment-Emergent Serious Adverse Events by System Organ Class and Preferred Term	Safety Population
14.3.1.7.1	Number (%) of Patients Reporting Common AEs (Greater or Equal to 5% in Any Group) by Preferred Term – Titration and Maintenance Period	Safety Population
14.3.1.7.2	Number (%) of Patients Reporting Common AEs (Greater or Equal to 5% in Any Group) by Preferred Term – Titration Period	Safety Population
14.3.1.7.3	Number (%) of Patients Reporting Common AEs (Greater or Equal to 5% in Any Group) by Preferred Term –Maintenance Period	Safety Population
14.3.1.8.1	TEAEs leading to study drug discontinuation by SOC and Preferred Term – Titration and Maintenance Period	Safety Population
14.3.1.8.2	TEAEs leading to study drug discontinuation by SOC and Preferred Term – Titration Period	Safety Population
14.3.1.8.3	TEAEs leading to study drug discontinuation by SOC and Preferred Term –Maintenance Period	Safety Population
14.3.1.9.1	Treatment Emergent AESI by SOC and Preferred Term – Titration and Maintenance Period	Safety Population
14.3.1.9.2	Treatment Emergent AESI by SOC and Preferred Term – Titration Period	Safety Population
14.3.1.9.3	Treatment Emergent AESI by SOC and Preferred Term –Maintenance Period	Safety Population
14.3.2.1	Hematology: Change from Baseline by Visit	Safety Population
14.3.2.2	Hematology: Shift from Baseline Classification by Visit	Safety Population
14.3.3.1	Blood Chemistry: Change from Baseline by Visit	Safety Population
14.3.3.2	Blood Chemistry: Shift from Baseline Classification by Visit	Safety Population
14.3.4.1	Urinalysis: Change from Baseline by Visit	Safety Population

14.3.4.2	Urinalysis: Shift from Baseline Classification by Visit	Safety Population
14.3.5	Vital Signs: Change from Baseline at End of Study	Safety Population
14.3.6	Subject Body Size: Change from Baseline by Visit	Safety Population
14.3.7.1	12-Lead ECG Intervals: Change from Baseline by Visit	Safety Population
14.3.7.2	12-Lead ECG Overall Findings: Shift from Baseline Classification by Visit	Safety Population
14.3.8	Columbia Suicidal Severity Rating Scale by Visit	Safety Population

Listing No.	**Title**	**Population**
16.1.7	Subject Randomization Schedule	Randomized Population
16.2.1.1	Subject Disposition and Reason for Disposition	Randomized Population
16.2.1.2	Subjects Who had Major Protocol Deviations	Randomized Population
16.2.1.3	Subjects Inclusion and the Reason for Exclusion from Analysis Populations	Randomized Population
16.2.1.4	Demographics and Baseline Characteristics	Randomized Population
16.2.1.5	Medical History	Randomized Population
16.2.1.6	R-MOAS Scale at Baseline	Randomized Population
16.2.1.7	Viteillo Aggression Scale at Baseline	Randomized Population
16.2.1.8	Study Drug Accountability and Compliance	
16.2.1.9.1	Prior Medications	Randomized Population
16.2.1.9.2	Concomitant Medications	Randomized Population
16.2.1.9.3	Prohibited Medications	Randomized Population
16.2.2.1.1	Impulsive Aggression Behavior Daily Diary	ITT Population
16.2.2.1.2	Derived Primary Endpoint	ITT Population
16.2.2.2	Investigator and Caregiver Clinical Global Impression - Improvement (CGI-I)	ITT Population
16.2.2.3	Investigator Clinical Global Impression – Severity (CGI-S)	ITT Population

16.2.2.4a	Child Health Questionnaire Parent Form 28 (CHQ-PF28)– Question Text	ITT Population
16.2.2.4b	Child Health Questionnaire Parent Form 28 (CHQ-PF28) – Individual Questions	ITT Population
16.2.2.4c	Child Health Questionnaire Parent Form 28 (CHQ-PF28) – Items, Components and General Health Scores	ITT Population
16.2.2.5a	Parenting Stress Index – Short Form (PSI-4-SF) – Individual Questions	ITT Population
16.2.2.5b	Parenting Stress Index – Short Form (PSI-4-SF) – Domain and Total Scores	ITT Population
16.2.2.6a	SNAP-IV Teacher and Parent Rating Scale – Individual Questions	ITT Population
16.2.2.6b	SNAP-IV Teacher and Parent Rating Scale – Subscales	ITT Population
16.2.3.1.1	All Adverse Events	Safety Population
16.2.3.1.2	Serious Adverse Events	Randomized Population
16.2.3.1.3	Adverse Events with Outcome of Death	Randomized Population
16.2.3.1.4	Adverse Events Leading to Study Drug Discontinuation	Randomized Population
16.2.3.1.5	Adverse Events of Special Interest	Randomized Population
16.2.3.2.1	Clinical Laboratory Tests-Hematology	Randomized Population
16.2.3.2.2	Clinical Laboratory Tests - Blood Chemistry	Randomized Population
16.2.3.2.3	Clinical Laboratory Tests - Urinalysis	Randomized Population
16.2.3.3	Vital Signs	Randomized Population
16.2.3.4	Subject Body Size	Randomized Population
16.2.3.5.1	12-Lead ECG Intervals	Randomized Population
16.2.3.5.2	12-Lead ECG Overall Interpretation	Randomized Population
16.2.3.6	Columbia Suicidal Severity Rating Scale	Randomized Population
16.2.3.7	Infrequent Behaviors	Randomized Population
16.2.3.8	Simpson-Angus Rating Scale	Randomized Population
16.2.3.9	Barness Akathisia Rating Scale	Randomized Population

16.2.3.10	Abnormal Involuntary Movement Scale	Randomized Population

Appendix 3. Table and Listing Shells