STATISTICAL ANALYSIS PLAN

Protocol No.:	ACP-103-070
Protocol Title:	A 52-Week Open-Label Extension Study of Pimavanserin in Children and Adolescents with Irritability Associated with Autism Spectrum Disorder (ASD)
Drug:	Pimavanserin
Sponsor:	Acadia Pharmaceuticals Inc.
Version No. and Date	Version 1.0, February 19, 2025

SIGNATURE/APPROVAL PAGE

AUTHOR

See appended electronic signature page

_____ _____

[redacted]
Director, Biostatistics
Acadia Pharmaceuticals Inc.

Date

APPROVERS

See appended electronic signature page

_____ _____

[redacted]
VP Clinical Development
Acadia Pharmaceuticals Inc.

Date

See appended electronic signature page

_____ _____

[redacted]
Executive Director, Biostatistics
Acadia Pharmaceuticals Inc.

Date

TABLE OF CONTENTS

TABLE OF CONTENTS 3
LIST OF TABLES 6
LIST OF FIGURES 6
LIST OF ABBREVIATIONS 7
1 INTRODUCTION 9
2 OBJECTIVES 9
2.1 Primary Objective 9
2.2 Secondary Objective 9
2.3 Exploratory Objectives 9
3 STUDY DESIGN 9
3.1 General Study Design 9
3.2 Schedule of Assessments 12
3.3 Randomization 12
3.4 Blinding 12
3.5 Determination of Sample Size 12
3.6 Coronavirus Disease 2019 12
4 ANALYSIS SETS 12
5 GENERAL CONSIDERATIONS FOR DATA ANALYSES 12
5.1 Data Presentation Conventions 12
5.1.1 Change from Baseline Summaries 13
5.2 Derived Variables 13
5.2.1 Aberrant Behavior Checklist 14
5.2.2 CGI-S 14
5.2.3 CGI-I 14
5.2.4 Repetitive Behavior Scale-Revised (RBS-R) 15
5.2.5 Vineland Adaptive Behavior Scales–Socialization (VABS-Socialization) 15
5.2.6 Caregiver Strain Questionnaire 16
5.2.7 Columbia-Suicide Severity Rating Scale (C-SSRS) 16
5.2.8 Tanner Staging 17
5.2.9 Extrapyramidal Symptom Rating Scale–Abbreviated (ESRS-A) 18
5.3 Analysis Visits 19
5.3.1 Study Day 19

5.3.2 Analysis Visit Windows 19
5.3.3 Unscheduled Assessments 19
5.3.4 Multiple Measurements within Visit Windows 19
5.4 Missing Data 20
5.4.1 Missing or Incomplete Date for Last Dose of Study Drug 20
5.4.2 Missing or Incomplete Dates for Prior or Concomitant Medications 21
5.4.3 Missing or Incomplete Dates for Adverse Events 21
5.4.4 Missing Severity Assessment for Adverse Events 21
5.4.5 Missing Relationship to Study Drug for Adverse Events 21
5.5 Character Values of Clinical Laboratory Variables 22
6 SUBJECT DISPOSITION 22
7 PROTOCOL DEVIATIONS 22
8 DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS 22
8.1 Demographics and Other Baseline Characteristics 22
8.2 Baseline Disease Characteristics 23
9 MEDICAL HISTORY 23
10 EXTENT OF EXPOSURE AND TREATMENT COMPLIANCE 23
10.1 Exposure to Study Drug 23
10.1.1 Exposure to Open-label Pimavanserin 23
10.1.2 Total Exposure to Pimavanserin Across the Double-blind and Open-label Periods 24
10.2 Measurement of Treatment Compliance 24
11 CONCOMITANT AND POST-TREATMENT MEDICATION 25
12 EFFICACY ANALYSIS CONSIDERATIONS 25
12.1 Efficacy Variables 25
12.2 Adjustment for Covariates 26
12.3 Handling of Missing Data 26
12.4 Multiple Comparisons / Multiplicity 26
12.5 Examination of Subgroups 26
13 METHODS OF EFFICACY ANALYSES 27
13.1 Analysis of Continuous Efficacy Endpoints 27
13.2 Categorical Analyses 27
14 SAFETY ANALYSES 27

14.1 Adverse Events 27
14.2 Clinical Laboratory Variables 28
14.2.1 Analysis of Clinical Laboratory Variables 30
14.3 Vital Signs, Height, Weight and BMI 33
14.4 Electrocardiogram (ECG) 34
14.5 Physical Examination and Tanner Staging 35
14.5.1 Physical Examination 36
14.5.2 Tanner Staging 36
14.5.3 Syncope and somnolence 36
14.6 Other Safety Variables 36
14.6.1 Columbia-Suicide Severity Rating Scale (C-SSRS) 36
14.6.2 Extrapyramidal Symptom Rating Scale–Abbreviated (ESRS-A) 36
15 CLINICAL PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES 36
16 INTERIM ANALYSIS 37
17 DATA MONITORING/REVIEW COMMITTEE 37
18 COMPUTER METHODS 37
19 CHANGES TO ANALYSES SPECIFIED IN PROTOCOL 37
20 REFERENCES 38
21 APPENDICES 39
21.1 Schedule of Events and Assessments 40
21.2 Summary of Version Changes 44

LIST OF TABLES

Table 1 Tanner Staging Classification 17
Table 2 Analysis Visit Windows 19
Table 3 Criteria for Potentially Clinically Important Laboratory Values – Hematology and Chemistry 32
Table 4 Criteria for Potentially Clinically Important Laboratory Values - Urinalysis..... 33
Table 5 Criteria for Potentially Clinically Important Vital Signs 34
Table 6 Criteria for Potentially Clinically Important ECG Values 35

LIST OF FIGURES

Figure 1 Schematic of Study Design.....11

LIST OF ABBREVIATIONS

Terms	Definitions
ABC	Aberrant Behavior Checklist
ABC-I	Aberrant Behavior Checklist-Irritability
ABC-L	Aberrant Behavior Checklist-Lethargy (lethargy and social withdrawal)
AE	adverse event
ASD	autism spectrum disorder
CGI-I	Clinical Global Impression–Improvement
CGI-S	Clinical Global Impression–Severity
CGSQ	Caregiver Strain Questionnaire
COVID-19	coronavirus disease 2019
C-SSRS	Columbia–Suicide Severity Rating Scale
DSMB	data and safety monitoring board
ECG	electrocardiogram
eCRF	electronic case report form
EDC	electronic data capture
eGFR	estimated glomerular filtration rate
ESRS-A	Extrapyramidal Symptom Rating Scale–Abbreviated
ET	early termination
EOT	end of treatment
GSD	Guidance for Site Documentation and Data Management Querying of Data Impacted by COVID-19
HbA_{1c}	glycosylated hemoglobin
IQ	intelligence quotient
MedDRA	Medical Dictionary for Regulatory Activities
Msec	milliseconds
PCI	potentially clinically important
QRS	QRS interval of ECG
QT	QT interval for heart rate of ECG
QTc	corrected QT interval of ECG for heart rate
QTcB	corrected QT interval using Bazett's correction method
QTcF	QT interval corrected for heart rate using Fridericia's formula
RBS-R	Repetitive Behavior Scale–Revised
SAE	serious adverse event
SAP	statistical analysis plan

Terms	Definitions
SD	standard deviation
SE	standard error
SOC	system organ class
TEAE	treatment-emergent adverse event
ULN	upper limit of normal
US	United States
VABS	Vineland Adaptive Behavior Scales

1 INTRODUCTION

This statistical analysis plan (SAP) provides a technical and detailed elaboration of the statistical analyses of efficacy and safety data based on the study protocol Amendment 3.0 dated 21 December 2022. Specifications for tables, figures, and listings will be provided in a separate document.

For Italy, a country-specific protocol amendment (Amendment 3-IT finalized 12 July 2023) added an exclusion criterion is to exclude subjects with a hypersensitivity to pimavanserin, any of its metabolites, or any compounds listed as being present in this medication and the details of the Schwartz equation to exclude patients whose estimated glomerular filtration rate (eGFR) is <30 mL/min as calculated by the bedside Schwartz equation at the last visit of the antecedent study ACP-103-069.

2 OBJECTIVES

2.1 Primary Objective

To evaluate the long-term safety and tolerability of pimavanserin after 52 weeks of treatment in children and adolescents with autism spectrum disorder (ASD).

2.2 Secondary Objective

To evaluate the continued response to long-term pimavanserin treatment in children and adolescents with ASD.

2.3 Exploratory Objectives

To evaluate the continued benefit of long-term pimavanserin treatment on the symptoms of ASD, in children and adolescents with ASD in the following clinical areas:

- Aberrant Behavior Checklist (ABC) symptoms domains
- Maintenance of benefit
- Response to treatment

3 STUDY DESIGN

3.1 General Study Design

This study will be conducted as a 52-week, open-label extension study of the antecedent double-blind study (ACP-103-069) to determine the long-term safety and tolerability of pimavanserin for the treatment of irritability associated with ASD in children and adolescents (5 through 17

years old at the time of enrolling into the antecedent double-blind study). Approximately 60 global sites and approximately eight countries will participate in this study.

The study will have two periods:

- Open-label treatment period (52 weeks)
- Safety follow-up period (at least 30 days)

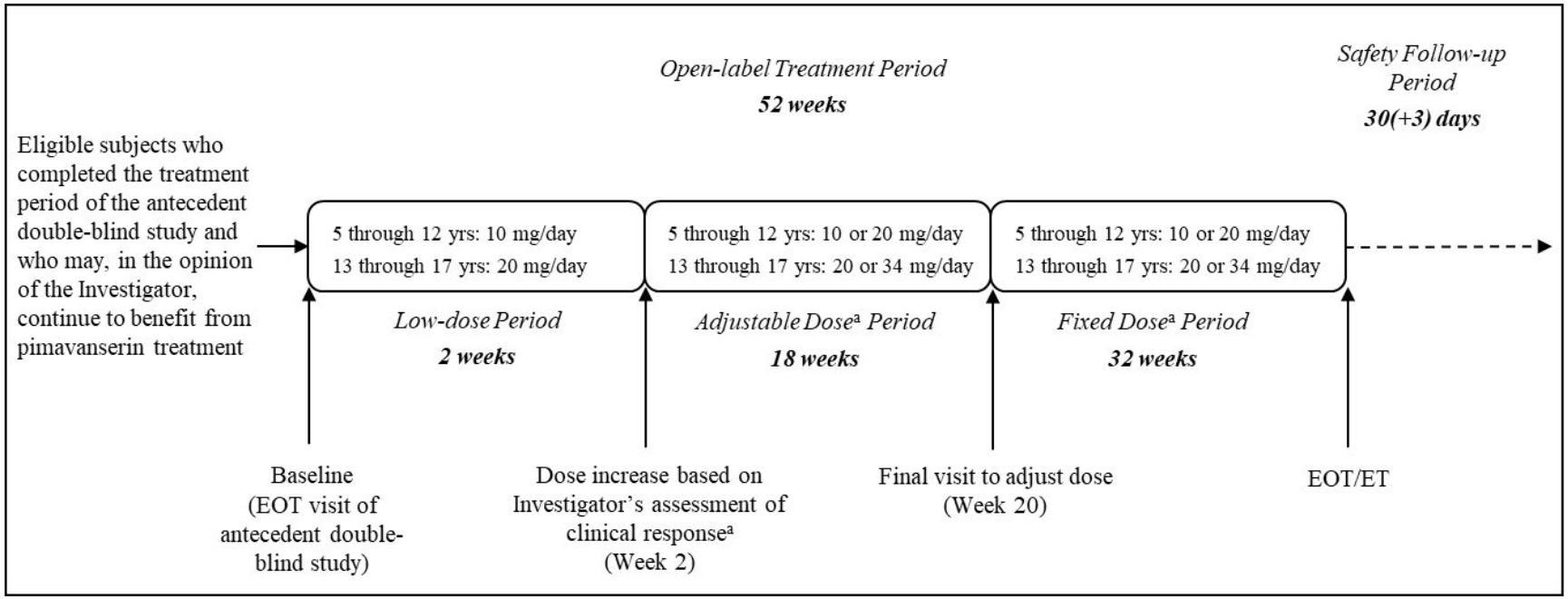
Subjects who have completed the antecedent double-blind study and who have shown no significant worsening of symptoms at the end of the study as evidenced by the CGI-I of irritability will be included in this long-term extension study.

All subjects will receive the “low dose” for their age group (10 mg/day for 5 through 12 years old, and 20 mg/day for 13 through 17 years old) for the first two weeks of the study. At the Week 2 visit, the dose may be increased to the “high dose” (20 mg/day for 5 through 12 years old, and 34 mg/day for 13 through 17 years old) based on the Investigator’s assessment of clinical response. After the Week 2 visit and up to and including the Week 20 visit, dose adjustments, either increases or decreases, may be made at any clinic visit (scheduled or unscheduled), based on the Investigator’s assessment of clinical response and tolerability.

During the treatment period, clinic visits will be conducted at Baseline, and Weeks 2, 6, 12, 20, 28, 36, 44, and 52, or upon early termination (ET) from the study. In addition to the visit performed at time of study completion or premature discontinuation, a follow-up safety assessment will be conducted by telephone call at approximately 4 weeks after the last dose of study drug.

The study design schematic is provided in Figure 1.

Figure 1 Schematic of Study Design



Abbreviations: EOT=end of treatment; ET=early termination; yrs=years

[a] A dose increase, based on the Investigator's assessment of clinical response, may be made at the Week 2 visit. After the Week 2 visit and up to and including the Week 20 visit, dose adjustments, either increases or decreases, may be made at any clinic visit (scheduled or unscheduled), based on the Investigator's assessment of clinical response and tolerability. No further dose adjustments are allowed after the Week 20 visit.

3.2 Schedule of Assessments

The schedule of events and assessments for the study is presented in Appendix 21.1.

3.3 Randomization

Not applicable.

3.4 Blinding

Not applicable.

3.5 Determination of Sample Size

The planned sample size for this study is not based on statistical power but will depend on the number of subjects who complete the antecedent double-blind study, and who rollover into this open-label extension study.

Assuming a drop-out rate of 15% in the antecedent study, it is estimated that approximately 193 subjects will be eligible to enroll in this OLE. Assuming all 193 subjects do enroll, then according to the Rule of Three, the 95% confidence interval for the occurrence rate of an unobserved adverse event is estimated as (0, 1.6%).

3.6 Coronavirus Disease 2019

In the context of the global COVID-19 pandemic or other natural disaster, the mechanisms to record information on the potential impact of the COVID-19 pandemic on data itself, as well as data collection and integrity, will be implemented. The impact of COVID-19 on the statistical analysis is discussed in each of the relevant sections of this SAP.

4 ANALYSIS SETS

Safety Analysis Set

The Safety Analysis Set will include all subjects who received at least one dose of open-label study drug. The Safety Analysis Set will be used for all analyses.

5 GENERAL CONSIDERATIONS FOR DATA ANALYSES

5.1 Data Presentation Conventions

For continuous variables, the following summary statistics will be provided: number of subjects, mean, standard error of the mean (SE), standard deviation (SD), minimum, maximum, and median. Unless specified otherwise, means, medians, and confidence intervals (CIs) will be presented to one more decimal place than the raw data, and the SDs and SEs will be presented to

2 more decimal places than the raw data. Unless otherwise specified, the maximum number of decimal places is 4 and values will be truncated to 4 decimal places in situations where there are more than 4 decimal places. Wherever possible data will be decimal aligned.

For categorical variables, summaries will include the number and percentage of subjects in each category. For demographic and baseline characteristics, the number of subjects and the percentage of subjects with missing data will be summarized (if applicable), and the denominator for percentages will be the total number of subjects in the given cohort. Categories with zero counts will not have zero percentages displayed. Percentages will be presented to 1 decimal place.

No hypothesis testing is planned. Descriptive summaries of all safety and efficacy endpoints will be provided.

Values that are collected with "<" or ">" signs will generally be analysed as the numerical value without the sign in tables and figures. In listings, these data will be reported as collected with the sign.

Duration in years is calculated as number of days/365.25 and will be rounded to 1 digit after the decimal point (tenths) for presentation purposes.

Duration in months is calculated as (number of days/365.25)×12 and will be rounded to 1 digit after the decimal point (tenths) for presentation purposes.

All data captured in the Electronic Data Capture (EDC) system will be presented in the listings, unless otherwise specified.

5.1.1 Change from Baseline Summaries

Change from Baseline results for each continuous measure in the safety and efficacy analyses will be presented in two ways:

1. Main analysis: using the Baseline from ACP-103-070 (study -070) and reporting the changes across study -070 timepoints
2. Exploratory analysis: using the Baseline from the antecedent study (ACP-103-069) and reporting the changes across the timepoints of both the antecedent study and the open-label study -070.

5.2 Derived Variables

Assessment scale total scores and subscores may be derived within the analysis datasets. In the event that total scores and/or subscores are collected on the eCRF as well as derived, the derived values will be used for all analyses. Both the raw and derived scores will be presented in the listings.

5.2.1 Aberrant Behavior Checklist

The Aberrant Behavior Checklist (ABC) is a parent/caregiver-rated scale comprised of five empirically-derived subscales encompassing 58 items that describe various behavior problems (Aman et al. 1985; Kaat et al. 2014). The subscales have been labeled as follows:

I. Irritability (irritability, agitation, and crying): Items 2, 4, 8, 10, 14, 19, 25, 29, 34, 36, 41, 47, 50, 52, and 57 (15 items)

II. Lethargy (lethargy and social withdrawal): Items 3, 5, 12, 16, 20, 23, 26, 30, 32, 37, 40, 42, 43, 53, 55, and 58 (16 items)

III. Stereotypic Behavior: Items 6, 11, 17, 27, 35, 45, and 49 (7 items)

IV. Hyperactivity (hyperactivity and noncompliance): Items 1, 7, 13, 15, 18, 21, 24, 28, 31, 38, 39, 44, 48, 51, 54, and 56 (16 items)

V. Inappropriate Speech: Items 9, 22, 33, and 46 (4 items)

The ABC will be administered at Baseline and at all post-baseline visits from (Week 2 through Week 52). Items are rated on a 4-point Likert scale ranging from 0 (not at all a problem) to 3 (the problem is severe), with higher scores indicating more severe problems. Subscale scores are calculated by summing the items within that subscale.

Missing ABC item scores will not be imputed. If any item is missing within a subscale, then that particular subscale score will be missing.

5.2.2 CGI-S

The CGI-S is a clinician-rated, 7-point scale that is designed to rate the severity of the subject's illness, in this study irritability associated with ASD, at the time of assessment, making use of the clinician's judgment and past experience with subjects who have the same disorder (Guy 1976). The CGI-S of irritability will be administered Baseline and at all post-baseline visits from (Week 2 through Week 52). The possible scores are 1=normal, not at all ill, 2=minimally ill, 3=mildly ill, 4=moderately ill, 5=markedly ill, 6=severely ill, 7=among the most extremely ill.

Missing CGI-S of irritability scores will not be imputed.

5.2.3 CGI-I

The CGI-I is a clinician-rated, 7-point scale that is designed to assess how much the subject's illness, in this study irritability associated with ASD, has improved or worsened relative to a baseline state at the beginning of the intervention (Guy 1976). The CGI-I of irritability will be administered at Baseline and at all post-baseline visits from (Week 2 through Week 52). The

possible scores are 1=very much improved, 2=much improved, 3=minimally improved, 4=no change, 5=minimally worse, 6=much worse, 7=very much worse.

Missing CGI-I of irritability scores will not be imputed.

5.2.4 Repetitive Behavior Scale-Revised (RBS-R)

The RBS-R (Lam and Aman 2007) is a 43-item parent/caregiver-facing questionnaire. Items are conceptually grouped into six subscales:

- stereotyped behavior (movements with no obvious purpose that are repeated in a similar manner): 6 items
- self-injurious behavior (actions that cause or have the potential to cause redness, bruising, or other injury to the body): 8 items
- compulsive behavior (behavior that is repeated and performed according to a rule or involves things being done "just so"): 8 items
- ritualistic behavior (performing activities of daily living in a similar manner): 6 items
- sameness behavior (resistance to change, insisting that things stay the same): 11 items
- restricted behavior (limited range of focus, interest, or activity): 4 items

Items are rated on a 4-point Likert scale ranging from (0) "behavior does not occur" to (3) "behavior occurs and is a severe problem", and raters are asked to refer to the previous month when completing the scale. The total RBS-R score is calculated as the sum of all subscale scores.

Missing RBS-R item scores will not be imputed.

The RBS-R will be administered at Baseline, and at the Week 12, Week 28, Week 36, Week 44, and Week 52 visits.

5.2.5 Vineland Adaptive Behavior Scales–Socialization (VABS-Socialization)

The VABS is a parent/caregiver-facing measure of adaptive behavior, organized into three domains: communication, daily living skills, and socialization (Sparrow et al. 2016). Only the socialization domain will be used in this study.

The socialization domain is a 112 item questionnaire with 3 subdomains: interpersonal relationships, play and leisure and coping skills. Items are rated in a 0, 1, 2 scale. 2 (behavior is usually performed), 1 (behavior is sometimes performed), 0 (behavior is never performed).

Missing VABS-Socialization item scores will not be imputed.

The VABS–Socialization will be administered at Baseline, and at the Week 12, Week 28, Week 36, Week 44, and Week 52 visits.

5.2.6 Caregiver Strain Questionnaire

The Caregiver Strain Questionnaire (CGSQ) (Brannan et al. 1997) is a 21-item parent/caregiver-facing questionnaire of self-reported strain experienced since last visit by parents/caregivers and families of children and adolescents with serious emotional and behavioral disorders. Responses are on a 5-point Likert scale ranging from (1) “not at all a problem” to (5) “very much a problem”.

The following areas of caregiver strain are included:

- disruption of family life and relationships
- demands on time
- negative mental and physical health effects for any member
- financial strain
- sacrifice
- disruption of social/community life
- worry and guilt
- fatigue and strain
- embarrassment
- child-caregiver relationship

The total score is sum of the following three subscale scores:

- objective strain (mean of items 1-11),
- subjective internalized strain (mean of items 12, 16-18, 20-21), and
- subjective externalized strain (mean of items 13, 14 reversed, 15, 19).

Missing CGSQ item scores will not be imputed.

The CGSQ will be administered at Baseline, and at the Week 6, Week 20, Week 28, Week 36, Week 44, and Week 52 visits.

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

The C-SSRS will be administered at Baseline and at all visits from Week 1 through Week 52, and at unscheduled visits. The C-SSRS monitors changes in suicidal thinking and behavior over time,

in order to determine risk (Posner et al. 2011). Four constructs are measured: the severity of ideation, the intensity of ideation, behavior, and lethality.

There are 5 questions about suicidal ideation, representing 5 types of suicidal ideation: wish to be dead; non-specific active suicidal thoughts; active suicidal ideation with any methods (not plan) without intent to act; active suicidal ideation with some intent to act, without specific plan; active suicidal ideation with specific plan and intent. If a subject/caregiver answers "yes" to any of these 5 questions, this subject will be counted as having suicidal ideation.

There are 5 questions about suicidal behavior, representing 5 types of suicidal behavior: preparatory acts or behavior; aborted attempt; interrupted attempt; actual attempt; suicide. If a subject answers "yes" to any of these 5 questions, this subject/caregiver will be counted as having suicidal behavior.

Missing C-SSRS item scores will not be imputed.

5.2.8 Tanner Staging

Tanner staging is a scale of physical development in children, adolescents, and adults. The scale defines physical measurements of development based on external primary and secondary sex characteristics, such as the size of the breasts, genitals, testicular volume, and development of pubic hair.

Tanner staging will be reported (by either the subject or the subject's caregiver) at Baseline, Week 12, Week 28, Week 44, and Week 52 visits. The Tanner stages are defined as follows in Table 1.

Table 1 Tanner Staging Classification

Sexual Characteristic	Tanner Stage	Female	Male
Breast development (female) or stage of genitalia (male)	Stage 1	Elevation of papilla only	Testes <2.5 cm
	Stage 2	Breast bud under the areola	Testes 2.5 to 3.2 cm
	Stage 3	Breast tissue grows but has no contour or separation	Testes 3.3 to 4.0 cm
	Stage 4	Projection of areola and papilla	Testes 4.1 to 4.5 cm
	Stage 5	Adult-type contour	Testes >4.5 cm

		Female/Male
Pubic hair development	Stage 1	Villus hair only
	Stage 2	Sparse hair
	Stage 3	Coarse hair
	Stage 4	Adult coarse hair, curly
	Stage 5	Adult hair, spreads to medial thigh

5.2.9 Extrapyramidal Symptom Rating Scale–Abbreviated (ESRS-A)

The ESRS (Chouinard and Margolese 2005) was developed to assess drug induced movement disorders such as parkinsonism, akathisia, dystonia, and tardive dyskinesia with established reliability, validity, and sensitivity. It consists of a questionnaire of parkinsonian symptoms, physician examination of parkinsonism, dyskinetic movements, and global impression of tardive dyskinesia. The ESRS-A, an accepted modified form of the original ESRS, will be used during the study to monitor for any worsening in extrapyramidal symptoms or signs at scheduled and unscheduled visits.

ESRS-A consists of 4 subscales and 4 clinical global impression movement severity scales of Parkinsonism, dyskinesia, dystonia, and akathisia. The Parkinsonism subscale consists of 10 items, the dyskinesia subscale consists of 6 items, the dystonia subscale consists of 6 items, and the akathisia subscale consists of 2 items. Each item is scored on a 6-point scale (0=absent, 1=minimal, 2=mild, 3=moderate, 4=severe, 5=extreme). The ESRS-A total score is the sum of the 24 item scores with a possible range of 0 to 120. Higher scores denote more severe symptoms of a movement disorder.

Missing ESRS-A item scores of the 4 subscales will not be imputed. The ESRS-A total score will be missing if any item score of the 4 subscales is missing.

Each clinical global impression movement severity scale of Parkinsonism, dyskinesia, dystonia, and akathisia is a single-item 6-point scale (0=absent, 1=minimal, 2=mild, 3=moderate, 4=severe, 5=extreme). The score ranges from 0 to 5, with higher scores indicating more severe movement disorders. Missing clinical global impression movement severity score will not be imputed.

The ESRS-A will be administered at Baseline and at all post-Baseline visits (Week 1 through Week 52), and at unscheduled visits.

5.3 Analysis Visits

5.3.1 Study Day

Study Day 1 is defined as the first dose date of open-label study drug (pimavanserin). If the date of assessment occurs on or after the first dose date, then study day will be calculated as (date of assessment – date of first dose) + 1. If the date of assessment occurs prior to the first dose date, then study day will be calculated as (date of assessment – date of first dose).

5.3.2 Analysis Visit Windows

The Baseline value is defined as the most recent non-missing value collected prior to the administration of the first dose of open-label study drug.

Efficacy and safety assessments will be summarized by analysis visits as presented in Table 2 below.

Table 2 Analysis Visit Windows

Open-label (OL) Analysis Visit Name	Target Study Day[1]	Study Day Interval
OL Baseline	1	≤ 1
OL Week 2	15	2 to 28
OL Week 6	43	29 to 63
OL Week 12	85	64 to 112
OL Week 20	141	113 to 168
OL Week 28	197	169 to 224
OL Week 36	253	225 to 280
OL Week 44	309	281 to 336
OL Week 52	365	≥337

[1] If the assessment date ≥ first OL dose date, study day = assessment date - first OL dose date + 1, otherwise study day = assessment date – first OL dose. Study day 1 is the first OL dose date.

5.3.3 Unscheduled Assessments

Both scheduled and unscheduled assessments, including early termination visits, will be considered for planned timepoint analyses based on the above analysis visit windowing rules. All assessments will be presented in data listings.

5.3.4 Multiple Measurements within Visit Windows

If multiple non-missing continuous measurements exist in an analysis window, the measurements will be selected based on the following rules if a single value is needed:

- For Baseline visit, the assessment closest to and prior to the time of the first dose of open-label study drug, whether scheduled or unscheduled, will be used as the baseline value. If multiple measurements occur on the same day at the same time or the time is not available, the one with later nominal visit (scheduled visits take precedence over unscheduled visit) will be selected.
- For post-baseline visits, the assessment closest to the target study day will be selected for the by-visit analyses. In these analyses, if two assessments are equidistant from the target day then the chronologically last assessment will be used. If multiple assessments have the same date/time, the one with later nominal visit (scheduled visits take precedence over unscheduled visit) will be selected.

If multiple non-missing categorical measurements exist in an analysis window, the measurements will be selected based on the following rules if a single value is needed:

- For Baseline visit, the assessment closest to and prior to the time of the first dose of open-label study drug, whether scheduled or unscheduled, will be used as the baseline value. If multiple measurements occur on the same day at the same time or the time is not available, the value with the lowest severity will be selected.
- For post-baseline visits, the assessment closest to the target study day will be selected for the by-visit analyses. In these analyses, if two assessments are equidistant from the target day then the chronologically last assessment will be used. Exceptions may be made for incomplete assessments, in which case, more complete assessments may be given priority. If multiple measurements occur on the same day, the value with the highest severity will be selected.

For safety analysis where the most extreme values should be selected (e.g. overall post-baseline minimum, overall post-baseline maximum, and potentially clinically important values), all non-missing post-baseline values should be considered, regardless of whether the value is selected for the by-visit summaries. All results will be presented in data listings.

When ECG is collected in triplicate, the average of the triplicate will be considered as one assessment for the analyses.

5.4 Missing Data

In general, missing data will not be imputed unless otherwise specified in this document.

5.4.1 Missing or Incomplete Date for Last Dose of Study Drug

The missing last dose date of study drug for a subject who completed or early terminated from the study will be imputed using the last expected dosing date, which is defined as the minimum

of the non-missing drug return date of the last dispensed drug kits, the last drug dispensed date plus the number of days that the dosing would continue per protocol or the end-of-treatment/early termination visit (EOT/ET) date, whichever occurs earlier. For the incomplete last dose date of the study drug, the imputation will be compared against the last expected dosing date defined as above. Details will be provided in a separate specifications document. The missing or incomplete dates will be displayed in the data listings as reported on the eCRF rather than the imputed dates.

For data summarization before final database lock, if a subject is still ongoing at the time of analysis, then this subject's last dose date will be imputed using the database extract date.

5.4.2 Missing or Incomplete Dates for Prior or Concomitant Medications

Missing or incomplete medication start or end dates will be imputed for the purpose of determining whether the medications are taken concomitantly (see Section 11 for definition). When the chronological order of medication use relative to the study drug treatment period is unclear due to missing or incomplete date(s), the medication will be considered as concomitant. The imputation algorithms will be provided in a separate specifications document. The missing or incomplete dates will be displayed in the data listings as reported on the eCRF rather than the imputed dates.

5.4.3 Missing or Incomplete Dates for Adverse Events

Missing or incomplete adverse event (AE) start dates will be imputed for the purpose of determining whether the AEs are treatment-emergent (see Section 14.1 for definition). When the chronological order of an AE onset relative to the study drug treatment period is unclear due to missing or incomplete date(s), the AE will be considered as treatment-emergent. The imputation algorithms will be provided in a separate specifications document. The missing or incomplete dates will be displayed in the data listings as reported on the eCRF rather than the imputed dates.

5.4.4 Missing Severity Assessment for Adverse Events

If the severity is missing for an AE starting on or after the date of the first dose of study drug, then a severity of "Severe" will be assigned. The imputed values for severity assessment will be used for incidence summaries, and the actual values will be presented in data listings.

5.4.5 Missing Relationship to Study Drug for Adverse Events

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

5.5 Character Values of Clinical Laboratory Variables

If the reported value of a clinical laboratory variable cannot be used in a statistical analysis due to, for example, a character string reported for a numeric variable, an appropriately determined coded value will be used in the statistical analysis. The coding algorithms will be provided in a separate specifications document. The actual values as reported in the database will be presented in data listings.

6 SUBJECT DISPOSITION

The number of sites that enrolled at least one subject and number of subjects enrolled will be summarized by region (United States [US] and rest of world) and overall. Enrolled subjects are defined as those subjects who have informed consent for study -070 and are not recorded as a rollover failure.

For enrolled subjects, the number and percentage of subjects in the Safety Analysis Set will be summarized. A listing will be provided displaying all subjects excluded from the Safety Analysis Set (if any), and will include reason(s) for exclusion.

The number and percentage of subjects who completed the study or discontinued (all discontinued and by primary discontinuation reason including the reason due to COVID-19) will also be summarized. Summaries by region will also be presented. A listing will be provided on the reasons for premature study drug/study discontinuation.

7 PROTOCOL DEVIATIONS

Protocol deviations will be reviewed regularly over the course of the study. The review process, definition of the deviation categories, and the classification of a deviation as major or minor are detailed in the Protocol Deviation Management Plan. Protocol deviations will also be assessed with respect to relationship to COVID-19.

A summary of the number and percentage of subjects with major protocol deviations for each deviation category will be presented in three ways: all protocol deviations, COVID-19 related protocol deviations, and non COVID-19 related protocol deviations.

Two listings of protocol deviations will be provided: all deviations and COVID-19 related protocol deviations.

8 DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS

8.1 Demographics and Other Baseline Characteristics

Demographics and open-label Baseline characteristics will be summarized for the Safety Analysis Set using descriptive statistics. Variables include age, age category (5-12 years and 13-

17 years), sex, primary race (subjects of multi-racial background can only identify/select one primary race on eCRF, or choose "other" and specify), ethnicity, height, weight, body mass index (BMI), intelligence quotient (IQ), IQ categories (<70 and ≥70), region, and caregiver relationship.

The reported age reflects a subject's age at the open-label Baseline visit date. Age will be presented both as continuous and categorical (5 through 12 years old or 13 through 17 years old). If there are discrepancies in the values between the IRT system and the clinical database, categories based on the IRT system and clinical database will each be presented separately.

8.2 Baseline Disease Characteristics

Disease characteristics at Baseline will be summarized for the Safety Analysis Set using descriptive statistics. Baseline ABC subscale scores, CGI-S of irritability score, RBS-R score, VABS-Socialization score, CGSQ score, and ESRS-A score will be presented as continuous variables. CGI-S of irritability score will also be presented as categorical variables using discrete categories.

9 MEDICAL HISTORY

Medical and surgical history data will be coded using Medical Dictionary for Regulatory activities (MedDRA) version 24.0 or higher. The subject incidence will be summarized for each system organ class (SOC) and preferred term by cohort and overall, for Safety Analysis Set. A subject will be counted only once per SOC or per preferred term for the summary.

A listing of the SOC, preferred term, body system, verbatim term for the medical history condition/event, start and stop dates (when available), and an indicator for whether or not the condition is ongoing will be provided.

10 EXTENT OF EXPOSURE AND TREATMENT COMPLIANCE

Summaries of exposure and compliance to study drug will be provided for the Safety Analysis Set.

10.1 Exposure to Study Drug

The pimavanserin dose levels are expressed as free base.

10.1.1 Exposure to Open-label Pimavanserin

For each subject, the duration of exposure to open-label pimavanserin (open-label last dose date – open-label first dose date + 1), cumulative dose (first multiply the number of doses taken by the dose level for each kit utilized during the open-label study period, then sum the results from

all kits), and average daily dose (cumulative dose in mg divided by duration of open-label exposure in days) will be calculated and summarized.

Duration of OL pimavanserin exposure will be summarized as both continuous and categorical variables. For categorical presentation, the number and percentage of subjects in each of the following categories will be presented: <2 weeks (1 to 13 days), 2 to <6 weeks (14 to 41 days), 6 to <12 weeks (42 to 83 days), 12 to <20 weeks (84 to 139 days), 20 to <28 weeks (140 to 195 days), 28 to <36 weeks (196 to 251 days), 36 to <44 weeks (252 to 307 days), 44 to <52 weeks (308 to 363 days), and ≥52 weeks (364 days or longer). Kaplan-Meier curves of duration on study drug will also be presented. In addition, summaries of whether subjects had any dose change (yes vs. no), their highest dose level (20 mg or 34 mg), their lowest dose level (10 mg or 20 mg), and their last dose level (10 mg, 20 mg, or 34 mg) will also be provided.

10.1.2 Total Exposure to Pimavanserin Across the Double-blind and Open-label Periods

For subjects who received pimavanserin in the antecedent study, the total duration of exposure (in days) to pimavanserin will be calculated as the sum of the exposure durations in the double-blind and open-label periods:

[(*date of last double-blind dose*) – (*date of first double-blind dose*) + 1] + [(*date of last open-label dose*) – (*date of first open-label dose*) + 1].

For subjects who received placebo in the antecedent studies, the total duration of exposure to pimavanserin across the double-blind and open-label periods will be the same as the values calculated for the open-label study period alone.

Duration of total exposure to pimavanserin will be summarized as both continuous and categorical variables. For categorical presentation, the number and percentage of subjects in each of the following categories will be presented: <2 weeks (1 to 13 days), 2 to <6 weeks (14 to 41 days), 6 to <12 weeks (42 to 83 days), 12 to <20 weeks (84 to 139 days), 20 to <28 weeks (140 to 195 days), 28 to <36 weeks (196 to 251 days), 36 to <44 weeks (252 to 307 days), 44 to <52 weeks (308 to 363 days), 52 to <60 weeks (364 to 419 days), 60 to <68 weeks (420 to 475 days), 68 to <78 weeks (476 to 545 days), and ≥78 weeks (546 days or longer).

10.2 Measurement of Treatment Compliance

Study drug dosing compliance (in percentage) during the open-label treatment period for a subject is defined as the total number of capsules actually taken divided by the number of capsules expected to be taken and then multiplied by 100. The total number of capsules actually taken is calculated by the total number of capsules dispensed minus the total number of capsules returned. The number of capsules expected to be taken is calculated as the duration of exposure (days) × 1 (number of capsules taken per day).

Compliance will be summarized as both continuous and categorical variables. For categorical presentation, the number and percentage of subjects in each of the following categories will be presented: <80%, 80 to 120%, and >120%.

Listings of study drug accountability will be provided separately.

11 CONCOMITANT AND POST-TREATMENT MEDICATION

Concomitant medication is defined as any medication with a start date prior to the date of the open-label first dose and continuing after the open-label first dose date or with a start date between the open-label first dose date and open-label last dose date, inclusive. Any medication with a start date after the date of the last dose of open-label study drug will be considered as post-treatment medication. Concomitant and post-treatment medications will be summarized separately.

Medications will be coded using WHO Drug Dictionary 2021 March or newer version. The number and percentage of subjects taking each drug class (ATC Level 3) and medication preferred term will be tabulated for the Safety Analysis Set. Multiple medication usage by a subject in the same category will be counted only once.

Relationship to COVID-19 will be assessed for selected medications as detailed in the GSD and provided in a listing. Additional summaries of relationship to COVID-19 may be provided.

Listings of the concomitant and post-treatment medications will be provided.

12 EFFICACY ANALYSIS CONSIDERATIONS

All efficacy analyses will be performed using the Safety Analysis Set. No hypothesis testing is planned. Descriptive summaries of efficacy endpoints will be provided.

12.1 Efficacy Variables

Secondary Efficacy Endpoint

The secondary efficacy endpoint is defined as follows:

- Proportion of subjects who have at least 25% reduction from the antecedent study Baseline in the ABC-I subscale score and a CGI-I of irritability score of 1 (very much improved) or 2 (much improved) compared to the antecedent study baseline status at Week 52.

Exploratory Efficacy Endpoints

- Change from Baseline at Week 52 in the caregiver-rated ABC subscale scores
    - Irritability
    - Stereotypic behavior

  - Lethargy
  - Hyperactivity
  - Inappropriate speech

- Change from Baseline at Week 52 in the CGI-S of irritability score
- CGI-I of irritability score from the antecedent study Baseline at Week 52
- Proportion of subjects who have at least 25% reduction from the antecedent study Baseline in the ABC-I subscale score (ABC-I responders) at Week 52
- Proportion of subjects who have CGI-I of irritability score of 1 (very much improved) or 2 (much improved) compared to the antecedent study baseline status (CGI-I of irritability responders) at Week 52

12.2 Adjustment for Covariates

Not applicable.

12.3 Handling of Missing Data

Any derived scores (i.e., total, domain, or subscale scores) that are missing after the imputation of individual missing items (if any) will not be imputed.

12.4 Multiple Comparisons / Multiplicity

No hypothesis testing is planned.

12.5 Examination of Subgroups

Summaries of the change from Baseline in the ABC-I subscale scores will be presented by subgroups below:

- age group (5 through 12 years old or 13 through 17 years old) as stratified in antecedent study
- sex (male or female)
- primary race (white or non-white)
- IQ from antecedent study (<70 or ≥70)
- region (US. or rest of world) as stratified in antecedent study
- severity of ABC-I at antecedent study entry (<25 or ≥25)
- severity of ABC-Lethargy at antecedent study entry (ABC-L) (<20 or ≥20)

13 METHODS OF EFFICACY ANALYSES

13.1 Analysis of Continuous Efficacy Endpoints

Descriptive statistics for all secondary and exploratory efficacy endpoints listed in Section 12.1 will be tabulated by cohorts based on the double-blind treatment group at scheduled timepoints.

13.2 Categorical Analyses

At any given visit, the proportion of responders will be calculated using observed cases, i.e., number of responders divided by the number of subjects with non-missing assessment.

14 SAFETY ANALYSES

The safety analysis will be performed on the Safety Analysis Set using descriptive statistics. No formal statistical testing will be performed for any of the safety endpoints.

14.1 Adverse Events

All adverse events (AEs) will be coded using MedDRA version 24.0 or higher.

An AE (classified by preferred term) will be considered a treatment-emergent AE (TEAE) if it starts after the first study dose administration and no later than last study dose date + 30 days. AEs reported on Day 1 based on Baseline (pre-dose) findings (e.g. clinically significantly abnormal vital signs, laboratory test results, or electrocardiogram parameters) will not be considered as TEAEs.

The event counts and the number and percentage of subjects reporting TEAEs will be tabulated by SOC and preferred term; and, by SOC, preferred term, and maximum severity. If more than one AE occurs with the same preferred term for the same subject, the subject will be counted only once for that preferred term using the most severe and more related occurrence for the summarization by severity and by relationship to study drug, respectively.

The event counts and the number and percentage of subjects with any TEAEs will also be tabulated by preferred term without SOCs.

The incidence of most frequently reported (preferred terms reported by ≥ 5% of subjects) TEAEs, TEAEs leading to discontinuation, TEAEs related to study drug, TEAEs by maximum severity, fatal TEAEs (i.e., events that cause death), serious TEAEs, and serious TEAEs related to study drug, will be summarized by SOC and preferred term.

The relationship of selected AEs to COVID-19 will be assessed as detailed in the GSD and provided in a listing.

An AE listing by subject will display all events, including those which are not treatment-emergent, and will include the verbatim term in addition to the MedDRA SOC and preferred term. Separate listings will be provided for TEAEs leading to discontinuation, TEAEs related to study drug, SAEs and fatal AEs.

14.2 Clinical Laboratory Variables

Clinical laboratory tests are performed at Baseline, Week 2, Week 6, Week 12, week 20, Week 28, Week 36, Week 44 and Week 52 visits.

Due to COVID-19 disruptions it is possible that some test results may be collected from a local laboratory. Local laboratory results and the associated normal ranges will be converted to SI units; the local laboratory results, in SI units, will then be normalized to central lab ranges to be included in summary data analysis together with the central laboratory results. The normalization will be performed using the following scale transformation equation:

$$s = L_s + (x - L_x)\frac{U_s - L_s}{U_x - L_x}$$

where s is the normalized individual laboratory value to be used for summary; x is the original value from the local lab; L_x and U_x are the lower and upper limits from the local lab; L_s and U_s are the lower and upper limits from the central lab.

For labs with only a single upper (or lower) limit, the following scale transformation equation will be used:

$$s = x\frac{U_s}{U_x}$$

where s is the normalized individual laboratory value to be used for summary; x is the original value from the local lab; U_s is the upper (or lower) limit from the central lab; U_x is the upper (or lower) limit from the local laboratory. Local laboratory results and normalized results will be included in data listings. Only central lab and normalized local lab results will be used for summary of change from baseline, shift, and potentially clinically important (PCI) analyses.

Clinical laboratory tests include the following:

- Clinical chemistry serum tests
    - Sodium (Na), potassium (K), carbon dioxide (CO_2), chloride (Cl), phosphorus (P), calcium (Ca), blood urea nitrogen (BUN), creatinine (CR), glucose, albumin (ALB), total protein

- Alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma-glutamyl transpeptidase (GGT), alkaline phosphatase (ALP), total bilirubin (TBIL), lactate dehydrogenase (LDH)
- Lipid panel
  - Total cholesterol, high-density lipoprotein (HDL)-cholesterol, triglycerides, low-density lipoprotein (LDL)-cholesterol, cholesterol/HDL ratio, non-HDL cholesterol
- HbA_{1c}
- Prolactin
  - Prolactin results should be blinded to the Investigator and the Sponsor at Baseline. Results will be monitored by an independent Medical Monitor from the Clinical Research Organization (CRO).
- Creatine kinase (CK)/creatine phosphokinase (CPK)

- Pregnancy test
  - A urine pregnancy test should be performed at all designated visits for women of child-bearing potential. If positive, the result will be confirmed with a serum pregnancy test.
    - If urine cannot be obtained in women of childbearing potential, a serum pregnancy test should be done in its place

- Hematology tests
  - Complete blood count (CBC) including:
    - White blood cell (WBC) count
    - Complete differential (relative and absolute)
    - Hematocrit, hemoglobin, red blood cells (RBC), platelets
    - Reticulocyte count

- Urinalysis
  - Blood, RBCs, WBCs, protein, glucose, ketones, specific gravity, pH

- Urine toxicology screen

- Urine toxicology screen will test for controlled substances. The following controlled substances may be tested with a urine toxicology screen: amphetamine, barbiturates, benzodiazepines, cocaine, methadone, morphine/opiates, methamphetamine, marijuana (tetrahydrocannabinol), phencyclidine, ecstasy (3,4-methylenedioxymethamphetamine).
- Subjects who test positive and have a valid prescription for a controlled substance may be retested if they agree to abstain from the medication for the length of their participation in the study.

14.2.1 Analysis of Clinical Laboratory Variables

14.2.1.1 Observed Values and Change from Baseline

All laboratory test results will be provided in the listings. The listings will include date and study day of collection. All units will be displayed in Système International [SI] units. Out of range values will be flagged in the data listings (e.g. 'L' or 'H').

Clinical laboratory values for hematology, chemistry and urinalysis will be summarized using descriptive statistics at Baseline, Week 3 and Week 6. The change from Baseline values will also be summarized at Week 3 and Week 6. The overall minimum and maximum post-baseline observed and change from Baseline values will also be summarized. For hemoglobin, hematocrit and uric acid, the above summaries will be presented for each gender as well as for both genders combined. For urinalysis with categorical results (blood, protein, glucose, and ketones), the number and percentage of subjects will be tabulated by category at Baseline, Week 3 and Week 6, with the denominator being the number of subjects with non-missing values for the given parameter, visit and cohort.

14.2.1.2 Shift Tables

Laboratory values will be summarized in shift tables to determine the number and percentage of subjects with values classified as below (low), within (normal), and above (high) normal ranges at scheduled post-baseline minimum relative to the same classification at the Baseline visit. The shifts from Baseline to overall post-Baseline minimum and overall post-Baseline maximum will also be presented, only for the open-label treatment period. For the by-visit shift summary, the denominator is the number of subjects with non-missing values at Baseline and the given visit for the given parameter and cohort. For the shift to the overall post-baseline minimum or maximum, the denominator is the number of subjects with non-missing Baseline and at least one post-baseline value for the given parameter and cohort. For hemoglobin, hematocrit and uric acid, the shift summaries will be presented for each gender as well as for both genders combined.

14.2.1.3 Potentially Clinically Important (PCI) Laboratory Values

The number and percentage of subjects with PCI laboratory values at scheduled post-Baseline visits and overall post-Baseline (for the open-label treatment period) will be summarized by Baseline status (all and within normal range) for selected parameters. PCI criteria are listed in Table 3 and Table 4. For the overall post-baseline summaries of PCI values, all post-baseline values will be considered, including unscheduled and out of window values. Subjects with multiple PCI values for a given parameter will be counted only once for that parameter. For the by-visit summary, the numerator for the percentage is the number of subjects with a post-baseline PCI value for the given parameter, visit and cohort, and the denominator is the number of subjects with non-missing values for the given parameter, visit and cohort. For the overall post-baseline summary, the numerator for the percentage is the number of subjects with at least one post-baseline PCI laboratory value for the given parameter and cohort, and the denominator is the number of subjects with at least one post-baseline laboratory value for the given parameter and cohort. For hemoglobin, hematocrit and uric acid, the count and percentage of subjects with PCI values will be presented for each gender as well as for both genders combined. A listing of all PCI values in study -070 will be provided. This listing will include all observations from study -070 for those subjects and parameters for which at least one PCI value (including Baseline) was observed.

Table 3 Criteria for Potentially Clinically Important Laboratory Values – Hematology and Chemistry

Analyte	Conventional Unit	Low PCI Criteria	High PCI Criteria	SI Unit	Low PCI Criteria	High PCI Criteria
Hematology (whole blood)						
Hemoglobin (male)	g/dL	<11	>17	g/L	<110	>170
Hemoglobin (female)	g/dL	<10	>17	g/L	<100	>170
Hematocrit	%	<30	>50	L/L	<0.3	>0.50
Leukocyte (White Blood Cell Count)	x 10^3/uL	≤2.8	≥15	x 10^9/L	≤2.8	≥15
Neutrophils	x 10^3/uL	≤1.5	No upper limit	x 10^9/L	≤1.5	No upper limit
Eosinophils	%	No lower limit	≥10	L/L	No lower limit	≥.10
Platelet Count	x 10^3/uL	≤75	≥700	10^9/L	≤75	≥700
Chemistry (serum or plasma)						
ALT (SGPT)	U/L	No lower limit	≥3 X ULN	U/L	No lower limit	≥3 X ULN
AST (SGOT)	U/L	No lower limit	≥3 X ULN	U/L	No lower limit	≥3 X ULN
Total Bilirubin	mg/dL	No lower limit	≥1.5 ULN	umol/L	No lower limit	≥1.5 ULN
BUN	mg/dL	No lower limit	≥30.0	mmol/L	No lower limit	≥10.71
Creatine Kinase (CPK)	U/L	No lower limit	≥3 ULN	U/L	No lower limit	≥3 ULN
Sodium	mEq/L	≤125	≥155	mmol/L	≤125	≥155
Potassium	mEq/L	≤3.0	≥5.5	mmol/L	≤3.0	≥5.5
Calcium, total	mg/dL	<8.0	>11.0	mmol/L	<2.0	>2.75
Lactate Dehydrogenase (LDH)	U/L	No lower limit	≥3 X ULN	U/L	No lower limit	≥3 X ULN
Alkaline Phosphatase	U/L	No lower limit	≥3 X ULN	U/L	No lower limit	≥3 X ULN
Uric acid	mg/dL	No lower limit	≥8.5	umol/L	No lower limit	≥505.75
Albumin	g/dL	≤2.6	≥6.0	g/L	≤26	≥60
Total Protein	g/dL	≤5.0	≥10.0	g/L	≤50	≥100
Chloride	mEq/L	≤85	≥120	mmol/L	≤85	≥120
Glucose (random)	mg/dL	≤45.1	≥115.0	mmol/L	≤2.48	≥6.4
Serum Creatinine	mg/dL	Not Applicable	>1.5 ULN	umol/L	Not Applicable	>1.5 ULN
Triglycerides	mg/dL	Not Applicable	>300	mmol/L	Not Applicable	>3.39
Gamma-Glutamyl Transferase (GGT)	U/L	Not Applicable	≥3 ULN	U/L	Not Applicable	≥3 ULN
Phosphorus, inorganic	mg/dL	≤1,0	≥7	mmol/L	≤0.32	≥2.26
Magnesium	mEq/L	≤0.7	≥5.0	mmol/L	≤0.35	≥2.5

Table 3 Criteria for Potentially Clinically Important Laboratory Values – Hematology and Chemistry (Continued)

Total Cholesterol, Fasting	mg/dL	No lower limit	≥240	mmol/L	No lower limit	≥6,21
Hb1Ac	%	No lower limit	≥7	L/L	No lower limit	≥0.07

Table 4 Criteria for Potentially Clinically Important Laboratory Values - Urinalysis

Urinalysis (qualitative dipstick)	Low PCI Criteria	High PCI Criteria
Blood (occult blood)	Not Applicable	≥Moderate or +2 (Q2)
Protein	Not Applicable	≥100 mg/dL (ICON) or +2 (Q2)
Glucose	Not Applicable	≥500 mg/dL (ICON) or +2 (Q2)

The pregnancy results (positive or negative) for female subjects of childbearing potential and the urine drug screen results will be presented in a listing.

14.3 Vital Signs, Height, Weight and BMI

Vital signs, height (as measured by a stadiometer) and weight are assessed at Baseline, at all scheduled visits from Week 2 through Week 52, and at unscheduled visits. BMI will be derived as Weight (kg)/ [Height (m)]^2.

Vital signs, height, weight and BMI will be summarized using descriptive statistics at Baseline and all scheduled post-baseline visits. The change from Baseline values will also be summarized by cohort at the scheduled post-baseline visits.

Vital sign and weight values will be considered PCI if they meet both the observed value criteria and the change from baseline criteria listed in Table 5. The number and percentage of subjects with post-baseline PCI values will be summarized by cohort at each post-baseline visit and for overall post-baseline. For the overall post-baseline summaries, all post-baseline values will be considered, including unscheduled and out of window values. Subjects with multiple PCI values for a given parameter will be counted only once for that parameter. For the by-visit summary, the numerator for the percentage is the number of subjects with a post-baseline PCI value for the given parameter, visit and cohort, and the denominator is the number of subjects with non-missing values for the given parameter, visit and cohort. For the overall post-baseline summary, the numerator for the percentage is the number of subjects with at least one post-baseline PCI value for the given parameter and cohort, and the denominator is the number of subjects with at least one post-baseline value for the given parameter and cohort. A listing of all subjects with any PCI values will be provided.

Table 5 Criteria for Potentially Clinically Important Vital Signs

Vital Sign Parameter	Unit	Criteria[a]		
		Observed Value	And/Or	Change Relative to Baseline
Systolic blood pressure (supine or sitting)	mmHg	≥140	And	Increase of ≥20
		≤70	And	Decrease of ≥20
Diastolic blood pressure (supine or sitting)	mmHg	≥90	And	Increase of ≥15
		≤50	And	Decrease of ≥15
Pulse (supine or sitting)	bpm	≥120	And	Increase of ≥15
		≤50	And	Decrease of ≥15
Weight	kg	Not Applicable		Increase of ≥7%
				Decrease of ≥7%

[a] A post-baseline value is considered as a PCI value if it meets both criteria for observed value and change from baseline.

14.4 Electrocardiogram (ECG)

12-lead electrocardiograms are performed at Baseline, Week 2, Week 6, Week 12, Week 20, Week 28, Week 36, Week 44, and Week 52 visits.

All tracings will be evaluated by a central reading laboratory. At the Baseline visit, the machine-read results will also be recorded. All data, including the machine-read Baseline results, will be listed. ECG data summaries will be performed using the centrally evaluated data.

Observed values of ECG variables (heart rate, PR interval, RR interval, QRS interval, QT interval, and QTc intervals) and the changes from baseline at each assessment timepoint will be summarized . QTc intervals include QTcB (Bazett's formula) and QTcF (Fridericia's formula).

QTcF will also be categorized into the following categories (msec) and the number and percentage of subjects in each category will be summarized at each visit and for the overall post-baseline maximum:

- Observed: ≤450, 451 to 480, 481 to 500, and >500
- Change from Baseline: ≤10, 11 to 30, 31 to 60, and >60

For cardiologist's interpretations, the number and percentage of subjects with ECG results that are determined as normal or abnormal will be summarized at scheduled visits. The overall post-baseline worst interpretation will also be summarized (i.e. if a subject has one or more post-baseline ECG results that is/are considered as abnormal, this subject will be counted in the abnormal category). Cardiologist's interpretations will also be summarized in a shift table to determine the number and percentage of subjects with ECG results classified as normal or abnormal at scheduled post-baseline visits relative to the same classification at the Baseline visit. The shifts from Baseline to overall post-baseline worst interpretation will also be presented. For the by-visit shift summary, the denominator is the number of subjects with non-missing cardiologist's interpretation at Baseline and the given visit for the given cohort. For the summaries of shift from Baseline to the overall post-baseline worst interpretation, the denominator is the number of subjects with non-missing Baseline and at least one post-baseline cardiologist's interpretation for the given cohort.

Electrocardiogram variable values will be considered as PCI values if they meet or exceed the upper limit values listed in Table 6. The number and percentage of subjects with post-baseline PCI values will be summarized at each post-baseline visit and for overall post-baseline. For the by-visit summary, the numerator for the percentage is the number of subjects with a post-baseline PCI value for the given parameter, visit and cohort, and the denominator is the number of subjects with non-missing values for the given parameter, visit and cohort. For the overall post-baseline summary, the numerator for the percentage is the number of subjects with at least one post-baseline PCI value for the given parameter and cohort, and the denominator is the number of subjects with at least one post-baseline value for the given parameter and cohort. A listing of all subjects with any PCI values will be provided.

Table 6 Criteria for Potentially Clinically Important ECG Values

ECG Parameter	Unit	High PCI Criteria
QRS Interval	msec	≥120
PR Interval	msec	≥220
QTcB or QTcF	msec	>500
QTcB or QTcF: change from baseline	msec	>60

14.5 Physical Examination and Tanner Staging

Physical examination and Tanner Staging is performed at Baseline, Week 12, Week 28, Week 44, and Week 52 visits.

14.5.1 Physical Examination

Physical examination results (normal, abnormal, and not done) at Baseline, Week 12, Week 28, Week 44, and Week 52 will be summarized in a frequency table by cohort, body system and visit.

A listing of physical examination data will be provided.

14.5.2 Tanner Staging

The Tanner staging at Baseline, Week 12, Week 28, Week 44, and Week 52 will be summarized in a frequency table by gender, category (Breast development [female] or Genitalia [male], and Pubic Hair Development) and visit.

14.5.3 Syncope and somnolence

Syncope and somnolence are assessed at all scheduled visits from Week 2 through Week 52, and at unscheduled visits. Occurrences syncope and somnolence should be reported as adverse events.

Assessment of syncope and somnolence at each visit will be listed. Incidences of syncope and somnolence during the study will be summarized as TEAEs.

14.6 Other Safety Variables

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

The C-SSRS will be assessed at all scheduled and unscheduled visits. The number and percentage of subjects with suicidal ideation or behavior during the study (post-Baseline) will be tabulated.

For purposes of calculating percentages, the denominator will be the number of subjects within each cohort with at least one post-Baseline C-SRRS.

14.6.2 Extrapyramidal Symptom Rating Scale–Abbreviated (ESRS-A)

The ESRS-A will be assessed at Baseline, at all visits from Week 2 through Week 52, and at unscheduled visits. ESRS-A total scores and the four individual global clinical global impression movement severity scores will be summarized using descriptive statistics at Baseline and each scheduled post-baseline visit. The change from Baseline values will also be summarized at the post-baseline visits.

15 CLINICAL PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES

Not applicable.

16 INTERIM ANALYSIS

No formal interim analyses are planned for this study.

17 DATA MONITORING/REVIEW COMMITTEE

An independent Data and Safety Monitoring Board (DSMB) will review ongoing safety data including data on AEs, SAEs and safety laboratory data. The DSMB will be independent of the Sponsor and the Investigators and will be empowered to recommend stopping the study due to safety concerns. The roles and responsibilities of DSMB members and planned frequency of meetings will be described separately in the DSMB Charter.

18 COMPUTER METHODS

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

Validation and quality control of the tables, listings and figures containing the results of the statistical analyses will follow appropriate standard operating procedures.

19 CHANGES TO ANALYSES SPECIFIED IN PROTOCOL

No changes have been made to the analyses as specified in the protocol.

20 REFERENCES

Aman MG, Singh NN, Stewart AW, Field CJ. Psychometric characteristics of the Aberrant Behavior Checklist. *Am J Ment Defic*. 1985;89(5):492-502.

Brannan AM, Heflinger CA, Bickman L. The Caregiver Strain Questionnaire: measuring the impact on the family of living with a child with serious emotional disturbance. *J Emot Behav Disord*. 1997;5(4):212-222.

Chouinard G, Margolese HC. Manual for the Extrapyramidal Symptom Rating Scale (ESRS). *Schizophr Res.* 2005;76(2-3):247-265.

FDA Guidance on Conduct of Clinical Trials of Medical Products during COVID-19 Pandemic | Guidance for Industry, Investigators, and Institutional Review Boards (March 2020).

Guy W. Clinical Global Impressions. In: *ECDEU Assessment Manual for Psychopharmacology —Revised* (DHEW publication number ADM 76–338). Rockville, MD, U.S. Department of Health, Education, and Welfare, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, NIMH Psychopharmacology Research Branch, Division of Extramural Research Programs, 1976, 218-222.

Kaat AJ, Lecavalier L, Aman MG. Validity of the Aberrant Behavior Checklist in children with autism spectrum disorder. *J Autism Dev Disord*. 2014;44(5):1103-1116.

Lam KSL, Aman MG. The Repetitive Behavior Scale-Revised: independent validation in individuals with autism spectrum disorders. *J Autism Dev Disord*. 2007:37(5):855-866.

Posner K, Brown GK, Stanley B, et al. The Columbia–Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. *Am J Psychiatry*. 2011;168(12):1266-1277.

Sparrow SS, Cicchetti DV, Saulnier CA. *Vineland Adaptive Behavior Scales, Third Edition: Vineland-3*. Pearson; 2016.

21 APPENDICES

21.1 Schedule of Events and Assessments

Period		**Open-label treatment period**									**Follow-up**
	Baseline[w]								**EOT/ET[n]**	**Unscheduled[a]**	**Safety follow-up[b]**
Visit Week	**0**	**2**	**6**	**12**	**20**	**28**	**36**	**44**	**52**		**56**
Visit number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**		**10**
Visit window (days)		**±3**	**±3**	**±3**	**±3**	**±3**	**±3**	**±3**	**±3**		**+3**
Type of visit[c]	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Telephone
Informed consent/assent[d]	X										
Inclusion/exclusion criteria	X[e]										
Physical examination and self- or caregiver-reported Tanner staging[q]	X			X		X		X	X		
Vital signs	X	X	X	X	X	X	X	X	X	X	
Height[s], weight, and BMI	X	X	X	X	X	X	X	X	X	X	
12-lead ECG[f, g, m]	X	X	X	X	X	X	X	X	X		
Clinical laboratory tests[g, m, r, v, x]	X			X		X		X	X		
Pregnancy test[g, h]	X	X	X	X	X	X	X	X	X	X	
Urine toxicology (drug) screen (UDS)[o]	X								X		
ABC[q]	X	X	X	X	X	X	X	X	X		
CGI-S of irritability	X	X	X	X	X	X	X	X	X		
CGI-I of irritability[i]	X	X	X	X	X	X	X	X	X		
RBS-R[q]	X			X		X	X	X	X		
VABS-Socialization[q]	X			X		X	X	X	X		

Table continues on next page

Schedule of Events and Assessments (Continued)

Period		**Open-label treatment period**									**Follow-up**
	Baseline[w]								**EOT/ET[n]**	**Unscheduled[a]**	**Safety follow-up[b]**
Visit Week	**0**	**2**	**6**	**12**	**20**	**28**	**36**	**44**	**52**		**56**
Visit number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**		**10**
Visit window (days)		**±3**	**±3**	**±3**	**±3**	**±3**	**±3**	**±3**	**±3**		**+3**
Type of visit[c]	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Clinic	Telephone
CGSQ[q]	X		X		X	X	X	X	X		
C-SSRS[q]	X	X	X	X	X	X	X	X	X	X	
ESRS-A[q]	X	X	X	X	X	X	X	X	X	X	
Assessment of concomitant medications[p]	X	X	X	X	X	X	X	X	X	X	X
Assessment of adverse events	X	X	X	X	X	X	X	X	X	X	X
Assessment of syncope occurrence[t]	X	X	X	X	X	X	X	X	X	X	X
Assessment of somnolence occurrence[u]	X	X	X	X	X	X	X	X	X	X	X
Study drug dispensation	X[j]	X	X	X	X	X	X	X		X[k]	
Study drug return and accountability[l]		X	X	X	X	X	X	X	X	X	

Abbreviations: ABC=Aberrant Behavior Checklist; ASR=accurate symptom reporting; BMI=body mass index; CGI-I=Clinical Global Impression–Improvement; CGI-S=Clinical Global Impression–Severity; CGSQ=Caregiver Strain Questionnaire; C-SSRS=Columbia–Suicide Severity Rating Scale; ECG=electrocardiogram; eGFR=estimated glomerular filtration rate; EOT=end of treatment; ESRS-A=Extrapyramidal Symptom Rating Scale–Abbreviated; ET=early termination; LAR=legally acceptable representative; RBS-R= Repetitive Behavior Scale–Revised; UDS=urine toxicology (drug) screen; VABS=Vineland Adaptive Behavior Scales

[a] At a minimum the safety assessments indicated should be completed at unscheduled visits. Other assessments may be completed at unscheduled visits at the discretion of the Investigator.

[b] This visit is a safety follow-up telephone call visit for subjects who complete the study or discontinue treatment prematurely from the study. This visit will occur 30(+3) days after the last dose of study drug. The safety follow-up visit will not be done if the subject withdraws consent to participate in all parts of the study.

[c] Circumstances may arise (e.g., pandemic, natural disaster, political upheaval, or to minimize subject and caregiver burden) when on-site assessments of efficacy and/or safety are not possible. In those cases, assessments may be performed offsite by raters either in person, or via video technology or telephone

where possible. For all visits that are conducted remotely, the Investigator **must** contact the Medical Monitor for approval of the plan. Sites must keep a log to identify details of all visits that are administered remotely. Provided that the subject is physically in the clinic, and accompanied by a relative, all caregiver-rated assessments may be provided remotely.

[d] Consent from the subject, parent/LAR, and caregiver (if different from the parent/LAR), for the present study must be obtained for entry into the present study prior to the final procedures being performed at the end of treatment (EOT) visit in the antecedent double-blind study. Data from the EOT/ET visit procedures of the antecedent double-blind study will be carried over as baseline information in the present study, as applicable

[e] All assessments must be completed and subject must meet required eligibility criteria before being enrolled.

[f] 12-lead ECGs should be performed in sequential triplicate. Electrocardiograms should be performed before blood sampling or at least 30 minutes after blood sampling. The subject must rest in a sitting or supine position for 5 minutes before the ECG is obtained. One repeat set of triplicate ECGs is allowed at Baseline.

[g] Mild sedation is allowed exceptionally for ECGs and blood draws during the study (e.g., alprazolam at a pediatric-appropriate dose per age, and the *lowest dose* deemed necessary by the Investigator) just in cases when the subject's agitation/anxiety does not allow a safe and accurate measurement and the Investigator, with agreement from the caregiver, considers it safe and appropriate for the subject.

[h] For female subjects of childbearing potential, a urine pregnancy dipstick test will be completed at all scheduled and unscheduled visits. If positive, the result will be confirmed with a serum pregnancy test.

[i] Relative to baseline status in the antecedent double-blind study.

[j] The subject and their parent/LAR will be provided instructions for the subject's first dose of study drug on the day after the Baseline visit.

[k] Study drug may be dispensed to the subject at unscheduled visits if needed.

[l] If visit is remote, accountability will be assessed verbally with the caregiver, and verified at the next clinic visit.

[m] The involvement of experienced personnel in the conduction of routine procedures such as blood drawing and ECG recording in this population is strongly recommended.

[n] The ET visit will not be done if the subject (and/or parent/LAR) withdraws consent to participate in all parts of the study and withdrawal of consent happens before that timepoint.

[o] A urine toxicology dipstick should be used in addition to the urine toxicology screen at the Baseline visit (to confirm eligibility).

[p] Including coronavirus disease 2019 (COVID-19) vaccination

[q] For scales that require caregiver input, the caregiver should be the parent/LAR or designee. A designee should be a family member, adult and responsible, living with or in very frequent contact with the subject participating in the study, who is committed to providing responses for the caregiver-reported scales for the duration of Studies ACP-103-069 and ACP-103-070. Caregivers providing input for the ABC, RBS-R, and CGSQ scales will be trained in accurate symptom reporting (ASR) prior to completing the scales. ASR training should be done at Screening in the antecedent study (Study ACP-103-069) before the caregiver completes any scales, and repeated whenever there is a change in caregiver or if the site feels a caregiver requires retraining.

[r] Prolactin results should be blinded to the Investigator and the Sponsor at Baseline. Results will be monitored by an independent Medical Monitor from the Contract Research Organization (CRO).

[s] As measured by stadiometer.

[t] If the caregiver reports an occurrence of syncope, the investigator should ask the "syncope adverse event questions" in Appendix G.

[u] If the caregiver reports an occurrence of somnolence, the investigator should ask the "somnolence adverse event questions" in Appendix H.

[v] Circumstances may arise (e.g., pandemic, natural disaster, political upheaval, or technical issues) when on-site clinical laboratory tests are not possible. In those cases, clinical laboratory tests may be performed at the subject's place of residence by study staff or at a local laboratory. The Investigator must contact the Medical Monitor for approval with the plan. Sites must keep a log to identify details of all visits that are administered remotely.

[w] Procedures performed at the EOT visit for Study ACP-103-069 will be carried over as baseline information, if applicable.

[x] The creatinine value should also be used to calculate estimated glomerular filtration rate (eGFR) according to the bedside Schwartz equation eGFR=0.413 × (height/serum creatinine), where height is in cm, and creatinine is in mg/dL. Subjects are excluded if eGFR is <30 ml/min.

.

21.2 Summary of Version Changes

Version No:	Document History Description of Update	Author(s)	Version Date
1.0	Original version	I-Yuan Liu	19 FEB 2025

Signature Page for ACP-103-070 Statistical Analysis PlanV1.0_19FEB2025

Approval w eSig Task	[redacted]ics 21-Feb-2025 19:51:17 GMT+0000

Approval w eSig Task	[redacted] 21-Feb-2025 19:51:24 GMT+0000

Approval w eSig Task	[redacted]s 21-Feb-2025 22:19:00 GMT+0000

Signature Page for ACP-103-070 Statistical Analysis PlanV1.0_19FEB2025