STATISTICAL ANALYSIS PLAN ADDENDUM

Protocol No.:	ACP-103-069
Protocol Title:	A Phase 2, Randomized, Double-blind, Placebo-controlled Study to Evaluate the Efficacy and Safety of Pimavanserin for the Treatment of Irritability Associated With Autism Spectrum Disorder
Drug:	Pimavanserin
Sponsor:	Acadia Pharmaceuticals Inc.
Version Date	October 12, 2024

SIGNATURE/APPROVAL PAGE

AUTHOR

See appended electronic signature page

PPD
Director, Biostatistics
Acadia Pharmaceuticals Inc.

Date

APPROVERS

See appended electronic signature page

PPDPPDPPDPPD
VP Clinical Development
Acadia Pharmaceuticals GmbH

Date

See appended electronic signature page

PPDPPDPPD
Executive Director of Biostatistics, Biometrics
Acadia Pharmaceuticals Inc.

Date

TABLE OF CONTENTS

1 INTRODUCTION

This statistical analysis plan (SAP) addendum provides further details per the Food and Drug Administration (FDA) comments on the SAP (Version 3) of Study ACP-103-069 dated 29 July 2024 for IND 157863.

2 EFFICACY ANALYSES

Section 12.4 Multiple Comparisons / Multiplicity (Page 27):

In response to the FDA's comments, the hierarchical testing procedure has been reverted to the testing procedures outlined in the previous version of the SAP, dated December 22, 2022, which is consistent with the protocol.

A hierarchical testing procedure will be used to control the Type I error rate across the two treatment comparisons (pimavanserin high dose vs. placebo and pimavanserin low dose vs. placebo) for the primary endpoint. The hypotheses testing will be conducted in sequential order: 1. pimavanserin high dose versus placebo, 2. pimavanserin low dose versus placebo. If the first comparison fails to reach statistical significance at the 2-sided 0.05 level, the subsequent comparison will be declared as not statistically significant regardless of the associated nominal p-value. The study-wise Type I error rate will be maintained at the significance level of 0.05.

Section 13.1.2.2 Tipping Point Analyses (Page 32):

In response to the FDA's comments, the tipping point analyses have been updated to remove the limitation on the delta value, allowing it to exceed 100% until the conclusion of the primary analysis is overturned.

This sensitivity analysis of the tipping point analysis will be implemented for Full Analysis Set using delta adjustment imputation. This MNAR (missing not at random) sensitivity analysis is to departure from MAR (missing at random) assumption by progressively increasing the delta until conclusion from the primary analysis is overturned. The delta value is 0% to 100%, 200%,.... in 5% increments of the observed treatment difference between a pimavanserin high dose and placebo, pimavanserin low dose and placebo, from the primary analysis of MMRM (mixed model for repeated measures) model until conclusion of the primary analysis is overturned.

Signature Page for ACP-103-069 Statistical Analysis Plan Addendum

Approval w eSig Task	PPDPPDPPD 14-Oct-2024 17:46:04 GMT+0000

Approval w eSig Task	PPD ics 14-Oct-2024 17:46:43 GMT+0000

Approval w eSig Task	PPD s 14-Oct-2024 18:47:45 GMT+0000

Signature Page for ACP-103-069 Statistical Analysis Plan Addendum

STATISTICAL ANALYSIS PLAN

Protocol No.:	ACP-103-069
Protocol Title:	A Phase 2, Randomized, Double-blind, Placebo-controlled Study to Evaluate the Efficacy and Safety of Pimavanserin for the Treatment of Irritability Associated With Autism Spectrum Disorder
Drug:	Pimavanserin
Sponsor:	Acadia Pharmaceuticals Inc.
Version No. and Date	Version 3.0, July 29, 2024

SIGNATURE/APPROVAL PAGE

AUTHOR

See appended electronic signature page

PPD
Director, Biostatistics
Acadia Pharmaceuticals Inc.

Date

APPROVERS

See appended electronic signature page

PPDPPDPPDPPD
VP Clinical Development
Acadia Pharmaceuticals GmbH

Date

See appended electronic signature page

PPDPPDPPD
Executive Director of Biostatistics, Biometrics
Acadia Pharmaceuticals Inc.

Date

Signature Page for ACP-103-069 Statistical Analysis Plan v3.0

Approve	PPD 30-Jul-2024 19:40:49 GMT+0000

Approve	PPD 30-Jul-2024 19:53:46 GMT+0000

Approve	PPDPPDPPD 30-Jul-2024 19:54:59 GMT+0000

Signature Page for ACP-103-069 Statistical Analysis Plan v3.0

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

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)
AC-279	*N*-desmethyl-pimavanserin, major metabolite of pimavanserin
ADI-R	Autism Diagnostic Interview–Revised
AE	adverse event
ASD	autism spectrum disorder
ASR	Accurate symptom reporting
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
FAS	Full Analysis Set
FDA	Food and Drug Administration
GSD	Guidance for Site Documentation and Data Management Querying of Data Impacted by COVID-19
HbA_{1c}	glycosylated hemoglobin
HIV	human immunodeficiency virus
ICE	intercurrent event
IQ	intelligence quotient
MAR	missing at random
MedDRA	Medical Dictionary for Regulatory Activities
MI	multiple imputation

Terms	Definitions
MINI-KID	Mini International Neuropsychiatric Interview for Children and Adolescents
MMRM	mixed model for repeated measures
MNAR	missing not at random
Msec	milliseconds
PCI	potentially clinically important
PD	pharmacodynamic
PK	pharmacokinetic
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
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 4.0 dated 21 December 2022, and the electronic case report form (eCRF). The SAP will be finalized and approved before database lock of the final analysis.

For Italy, a country-specific protocol amendment (Amendment 4-IT finalized 12 July 2023) added the details of the Schwartz equation and modified the exclusion criterion #23 to also exclude patients whose estimated glomerular filtration rate (eGFR) is <30 mL/min as calculated by the bedside Schwartz equation.

Specifications for tables, figures, and listings will be provided in a separate document. Statistical analyses for population pharmacokinetic (PK) and PK/pharmacodynamics (PD) modeling will be presented in a separate report and therefore will not be included in this SAP.

2 OBJECTIVES

2.1 Primary Objective

To evaluate the efficacy of pimavanserin compared with placebo in the treatment of irritability associated with Autism Spectrum Disorder (ASD) in children and adolescents.

2.2 Secondary Objective

To evaluate the efficacy of pimavanserin compared with placebo, in the treatment of non-irritability symptoms, the magnitude of improvement in irritability, and the response rate, in children and adolescents with ASD.

3 STUDY DESIGN

3.1 General Study Design

This study will be conducted as a 6-week, randomized, double-blind, fixed-dose, placebo-controlled, parallel group study in children and adolescents (5 through 17 years of age) with irritability, agitation, or self-injurious behaviors associated to ASD. Approximately 60 global sites and eight countries will participate in this study.

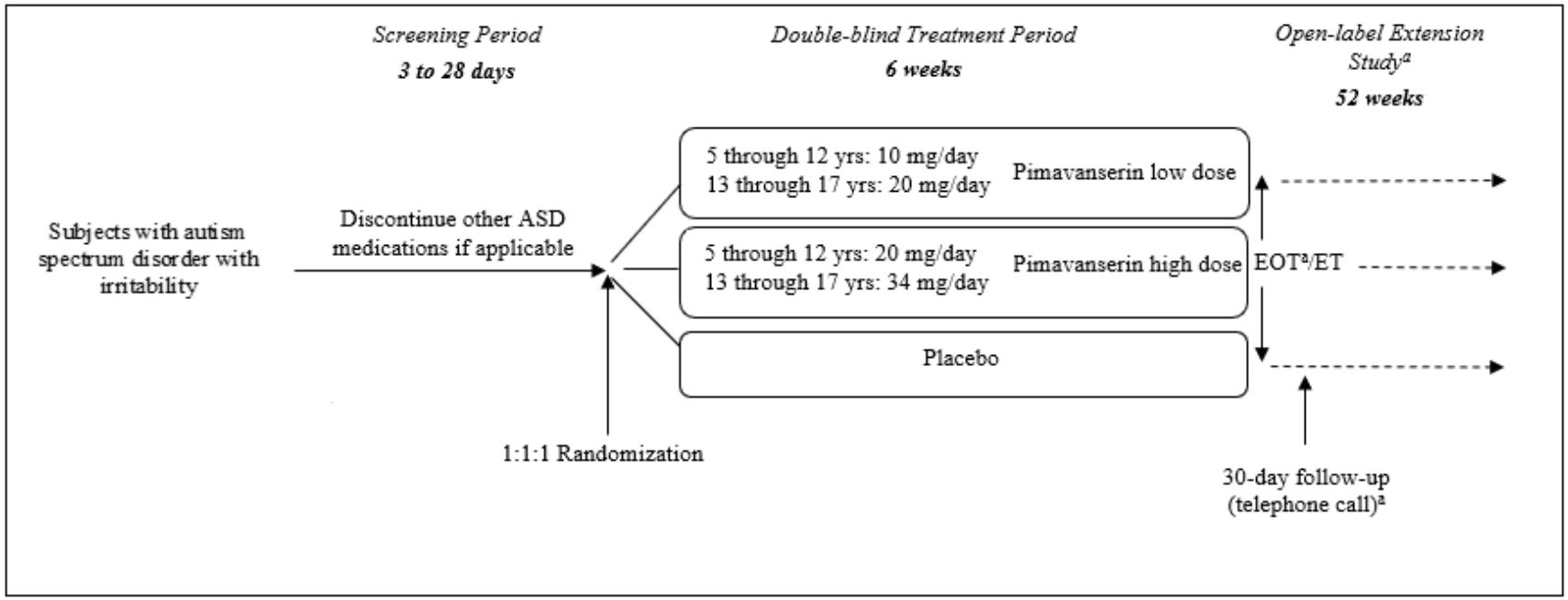
The study will have three periods:

- Screening period: 3-28 days
- Double-blind treatment period: 6 weeks

- Safety follow-up period: 30 (+3) days for those subjects who discontinue prematurely from the study or who do not enroll in the 52-week, open-label extension study (Study ACP-103-070)

The study design schematic is provided in Figure 1.

Figure 1 Schematic of Study Design



Abbreviations: ASD=autism spectrum disorder; EOT=end of treatment; ET=early termination; yrs=years

[a] Subjects who complete the 6-week Treatment Period may be eligible to enroll in a 52-week, open-label extension study (Study ACP-103-070). Subjects entering ACP-103-070 will not complete a follow-up telephone call as they will be immediately enrolled in ACP-103-070.

3.2 Schedule of Assessments

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

3.3 Randomization

Eligible subjects will be randomized in a 1:1:1 ratio to pimavanserin low dose, pimavanserin high dose, or placebo using an interactive response technology (IRT) system. The randomization will be stratified by age group (5- through 12-year-olds or 13- through 17-year-olds) and region (U.S. or rest of world). For the 5- through 12-year-olds, the subjects who are randomized to the low dose group will receive 10 mg/day pimavanserin and the subjects who are randomized to the high dose group will receive 20 mg/day pimavanserin. For the 13- through 17-year-olds, the subjects who are randomized to the low dose group will receive 20 mg/day pimavanserin and the subjects who are randomized to the high dose group will receive 34 mg/day pimavanserin. The assignments will be based on a pre-generated permuted-block randomization schedule.

3.4 Blinding

This is a double-blind study. Blinding will be assured by restricting access of Investigators and Sponsor personnel and/or designee to the treatment codes and providing identical capsules and packaging for the adjunctive pimavanserin and adjunctive placebo treatments. Neither the subjects nor the study personnel at the clinical sites will know which treatment is administered to each subject.

3.5 Determination of Sample Size

The planned sample size is a total of 228 subjects for both age groups combined (76 subjects randomized to each of the three treatment groups of equal sample sizes: pimavanserin high dose group, pimavanserin low dose group, and the placebo group). The total study sample size will be comprised of at least 35% (80) subjects in the younger age group (5- through 12-year-olds).

Assuming the true difference in the mean change in the ABC-Irritability (ABC-I) subscale score from Baseline at Week 6 is 5 points between each pimavanserin dose group and the placebo group, and assuming the common standard deviation is 10 points, 64 evaluable subjects per treatment group will provide 80% power to detect the difference between either pimavanserin dose group and the placebo group at a significance level of 0.05, using a 2-sided t-test. Adjusting for a potential discontinuation rate of up to 15%, approximately 76 subjects per treatment group, or 228 total subjects, will be randomized.

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

Randomized Analysis Set

The Randomized Analysis Set will include all unique subjects who were randomized. Subjects will be analyzed based on their randomized treatment.

Safety Analysis Set

The Safety Analysis Set will include all randomized subjects who received at least one dose of study drug (pimavanserin or placebo). Subjects will be analyzed based on the treatment that they actually received. If a subject received wrong kit, the actual treatment group will be determined based on the majority of treatment the subject received. The Safety Analysis Set will be used for all safety analyses.

Full Analysis Set

The Full Analysis Set will include all randomized subjects who received at least one dose of study drug and who have both a baseline value and at least one post-baseline value for the ABC-I subscale score. Subjects will be analyzed based on their randomized treatment. The Full Analysis Set will be used for the analysis of all efficacy endpoints.

Pharmacokinetics Analysis Set

The Pharmacokinetics Analysis Set will include subjects in the Safety Analysis Set with at least one measurable plasma concentration. The Pharmacokinetics Analysis Set will be used for pimavanserin and AC-279 plasma concentration summaries.

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 treatment group, unless otherwise specified. Categories with zero counts will not have zero percentages displayed. Percentages will be presented to 1 decimal place.

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

Unless specified otherwise, all statistical tests will be 2-sided hypothesis tests performed at the significance level of 0.05 for main effects and all confidence intervals (CIs) will be 2-sided 95% CIs. P-values will generally be presented to 4 decimal places; values less than 0.0001 will be presented as <0.0001.

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.

5.2 Derived Variables

In general, the assessment scale total scores and subscores will be derived within the analysis datasets, if those scores are not automatically generated in EDC. In the event that total scores and/or subscores are also collected on the eCRF, 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:

I. Irritability (irritability, agitation, and crying) (15 items)

II. Lethargy (lethargy and social withdrawal) (16 items)

III. Stereotypic Behavior (7 items)

IV. Hyperactivity (hyperactivity and noncompliance) (16 items)

V. Inappropriate Speech (4 items)

The ABC will be administered at Screening and Baseline, and at all post-baseline visits from Week 1 through Week 6. 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 at Screening and Baseline, and at all visits from Week 1 through Week 6. The possible scores are 1=normal, not 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, i.e., status at Baseline (Guy 1976). The CGI-I of irritability will be administered at all visits from Week 1 through Week 6. 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 Screening and Baseline, and at the Week 3 and Week 6 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 3 and Week 6 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 in the past 6 weeks (Screening visit), or since last visit (all other visits) 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 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).

The CGSQ will be administered at Screening and Baseline, and at the Week 2, Week 4, and Week 6 visits.

Missing CGSQ item scores will not be imputed.

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

The C-SSRS will be administered at Screening and Baseline, at all visits from Week 1 through Week 6, 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. The Baseline/Screening version (Lifetime and past 6 months) will be administered at Screening, and the Since Last Visit version will be administered at subsequent visits including Baseline visit.

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 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 Screening, Baseline, Week 1, Week 2, Week 3, Week 4, Week 5, and Week 6 visits; and at unscheduled visits. The ESRS-A will only be conducted at clinic visits and not at remote visits.

5.3 Analysis Visits

5.3.1 Study Day

Study Day 1 is defined as the first dose date of study drug (pimavanserin or placebo). 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.

Efficacy, safety and PK assessments will be summarized by analysis visits as presented in Table 1 below.

Table 1 Analysis Visit Windows

Analysis Visit Name	Target Study Day	Study Day Interval
Baseline	1	≤ 1
Week 1	8	2 to 11
Week 2	15	12 to 18
Week 3	22	19 to 25
Week 4	29	26 to 32
Week 5	36	33 to 39
Week 6	43	≥40

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, 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 two 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, 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 specified in this document. Handling of missing data in the efficacy analyses are described separately in Section 12.3.

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

For subjects who participate in the screening phase but are not randomized (screen failures), their demographics information (including age, sex, and primary race), screen failure reasons (the

specific inclusion/exclusion criterion (or criteria) not met or other reason including the reason due to coronavirus disease 2019 [COVID-19]) and protocol version will be provided in a listing. If a subject is re-screened, then the re-screening subject ID and the final enrollment status (whether eventually enrolled) will also be displayed in this listing. In addition, the frequency that the screen failure reasons are reported will also be summarized. Note that one subject may be deemed ineligible for multiple inclusion/exclusion criteria. A single rescreening of individuals who fail screening is permitted with the approval of the Sponsor's Medical Monitor.

The number of sites that screened at least one subject, number of sites that randomized at least one subject, number of subjects screened, number of unique subjects screened, and number of randomized subjects will be summarized.

For randomized subjects, number and percentage of subjects in Safety Analysis Set, Full Analysis Set, and PK Analysis Set will be summarized by treatment group and overall. A listing will be provided displaying all subjects excluded from the Safety or Full Analysis Sets, and will include reason(s) for exclusion.

In addition, the number of subjects randomized at each site will also be tabulated by Analysis Set and by treatment group and overall.

For Randomized Analysis Set and Full Analysis Set, 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 by treatment group and overall. 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 by treatment group and overall 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 baseline characteristics will be summarized by treatment group and overall, for Randomized Analysis Set and Full Analysis Sets using descriptive statistics. Variables include age, age category (5-12 years and 13-17 years), mental age, 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 informed consent 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 Randomized Analysis Set and Full Analysis Set using descriptive statistics. Baseline ABC subscale scores, CGI-S of irritability score, RBS-R score, VABS-Socialization score, CGSQ score, ADI-R, MINI-KID, and ESRS-A score will be presented as continuous variables. CGI-S of irritability score will also be presented as a categorical variable using discrete categories. Subject history of syncope, family history of syncope, subject history of somnolence will be summarized using yes/no counts.

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 treatment group 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

For each subject, the duration of exposure to study drug (last dose date – first dose date + 1), will be calculated and summarized by treatment group. Duration of exposure will also be summarized with number and percentage of subjects in each of the following categories: <1 week (1 to 7 days), 1 to <2 weeks (8 to 14 days), 2 to <3 weeks (15 to 21 days), 3 to <4 weeks (22 to 28 days), 4 to <5 weeks (29 to 35 days), 5 to <6 weeks (36 to 42 days), and ≥6 weeks (43 days or longer).

The pimavanserin dose levels are expressed as free base.

10.2 Measurement of Treatment Compliance

Study drug dosing compliance (in percentage) 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 tablets taken per day). Additional details for handling missing number of capsules returned will be provided in a separate specification document.

Compliance will be summarized as both continuous and categorical variables by treatment groups. 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 PRIOR, CONCOMITANT AND POST-TREATMENT MEDICATION

For a subject, prior medication is defined as any medication with stop date prior to the date of the first dose of study drug. Concomitant medication is defined as any medication that is ongoing at the first dose of study drug or with a start date between the dates of the first and last doses of study drug, inclusive. Any medication with a start date after the date of the last dose of study drug will be considered as post-treatment medication. Prior, concomitant, or 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 by treatment group and overall, for 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 Guidance for Site Documentation and Data Management Querying of Data Impacted by COVID-19 (GSD)

and provided in a listing. Listings of the prior, concomitant, or post-treatment medications will be provided.

12 EFFICACY ANALYSIS CONSIDERATIONS

All efficacy analyses will be performed using the planned treatment assignments based on the randomization schedule for the Full Analysis Set.

12.1 Efficacy Variables

Primary Efficacy Endpoint

The primary efficacy endpoint is the change from Baseline at Week 6 in the ABC-I score.

Secondary Efficacy Endpoints

The secondary efficacy endpoints are the following:

- Change from Baseline at Week 6 in the caregiver-rated ABC subscale scores
  - Stereotypic behavior
  - Lethargy
  - Hyperactivity
  - Inappropriate speech
- Change from Baseline at Week 6 in the Clinical Global Impression–Severity (CGI-S) of irritability score
- Clinical Global Impression–Improvement (CGI-I) of irritability score at Week 6
- Change from Baseline at Week 6 in the Repetitive Behavior Scale–Revised (RBS-R) scores
- Change from Baseline at Week 6 in the Vineland Adaptive Behavior Scales (VABS)–Socialization subscale score
- Change from Baseline at Week 6 in the Caregiver Strain Questionnaire (CGSQ) Since last visit scores
- Proportion of subjects who have at least 25% reduction from Baseline in the ABC Irritability subscale score at Week 6
- Proportion of subjects who have CGI-I of irritability score of 1 (very much improved) or 2 (much improved) at Week 6 (CGI-I of irritability responders)
- Proportion of subjects who have at least 25% reduction from Baseline in the ABC Irritability subscale score AND a CGI-I of irritability score of 1 (very much improved) or 2 (much improved) at Week 6

12.2 Adjustment for Covariates

For continuous variables analyzed using the mixed model for repeated measures (MMRM), the stratification factors (age group and region) and the baseline value of the endpoint being analyzed will be included as covariates as described in Section 13. For CGI-I of irritability score, the Baseline CGI-S of irritability score will be included in the MMRM analysis.

12.3 Handling of Missing Data

The primary analysis of the primary efficacy variable will be performed assuming missing at random (MAR) using an MMRM method. ABC-I scores that are missing (based on approaches described in Section 5.2.1) will not be imputed. The MMRM method is unbiased in the estimation of treatment effect under the MAR assumption and can be thought of as aiming to estimate the treatment effect that could have been observed if all subjects had continued on treatment for the protocol planned treatment duration (EMA, 2009). Sensitivity analyses of the primary efficacy variable will be performed as described in Section 13.

12.4 Multiple Comparisons / Multiplicity

A hierarchical testing procedure will be used to control the Type I error rate across the two treatment comparisons (pimavanserin high dose vs. placebo and pimavanserin low dose vs. placebo) for the primary endpoint and an important secondary endpoint CGI-S. The hypotheses testing will be conducted in sequential order:

1. pimavanserin high dose versus placebo on change from Baseline to Week 6 ABC-I score
2. pimavanserin low dose versus placebo on change from Baseline to Week 6 ABC-I score
3. pimavanserin high dose versus placebo on change from Baseline to Week 6 CGI-S score
4. pimavanserin low dose versus placebo on change from Baseline to Week 6 CGI-S score

If the proceeding comparison fails to reach statistical significance at the 2-sided 0.05 level, the subsequent comparison(s) will be declared as not statistically significant regardless of the associated nominal p-value. The study-wise Type I error rate will be maintained at the significance level of 0.05.

No statistical inference will be made for any other secondary endpoints.

12.5 Examination of Subgroups

Treatment comparisons will be made with respect to the primary efficacy variable using the MMRM analysis described in Section 13.1.1 separately for each subgroup by:

- age group (5 through 12 years old or 13 through 17 years old)

  - Only the region stratification factor (U.S. or rest of world) will be included in the MMRM model for this analysis

- sex (male or female)
- primary race (white or non-white)
- IQ (<70 or ≥70)
- region (U.S.A. or rest of world)
  - Only the age group stratification factor (5 through 12 years old or 13 through 17 years old) will be included in the MMRM model for this analysis
- severity of ABC-I at entry (<25 or ≥25)
- severity of ABC-Lethargy at entry (ABC-L) (<20 or ≥20)
- stimulants (defined as ATC level 3 =' PSYCHOSTIMULANTS, AGENTS USED FOR ADHD AND NOOTROPICS') use during treatment period (yes or no)

The LS mean differences with corresponding 95% CIs from the subgroups will also be graphically presented in forest plots for the primary efficacy variable.

13 METHODS OF EFFICACY ANALYSES

13.1 Primary Efficacy Endpoints

13.1.1 Primary Analysis

Estimand

The primary clinical question of interest for the primary objective is: what is the difference in the mean changes from Baseline at Week 6 in the ABC-I score comparing pimavanserin high dose vs. placebo, and pimavanserin low dose versus placebo, in subjects with ASD with irritability, agitation, or self-injurious behaviors, assuming all subjects complete 6 weeks of treatment without extended use of rescue medication or use of prohibited medications?

The primary estimand defining the treatment effect of interest in this study uses the hypothetical strategy specified in the ICH E9 (R1) Addendum. The estimand, or target of estimation, following the hypothetical strategy is the pharmacological effect seen, had no study treatment discontinuation occurred. This hypothetical estimand is justifiable in this case, since the focus is on the pharmacological effect of the drug additional to non-specific effects. Subjects who discontinue from a study treatment either could have lost their treatment effect or could have had their treatment effect confounded by other medication taken after the discontinuation of study

treatment. This means that any observations taken after subjects stop study drug will most likely not contribute relevant information about the pharmacological effect of the study drug.

By this strategy, the last collected efficacy assessment after premature treatment discontinuation will be done only at the early termination (ET) visit. Every effort will be made to complete the ET evaluations prior to administering any additional medications for the treatment of irritability associated with Autism Spectrum Disorder (ASD) or other prohibited medications. In the case of lost to follow-up events or death, no ET evaluations are expected, and only scheduled assessments performed before such an event are expected.

In this hypothetical strategy, the event of premature discontinuation of study medication is considered missing at random (MAR), and the primary endpoint of the study could be considered as a combination of the responses of on-treatment completers at Week 6 and the imputation of the endpoint to Week 6 following the trend in each treatment group using the mixed-effect model repeated measures (MMRM) method for subjects who discontinuation study drug during the study. All data collected in study will be used for statistical analysis. Under the MAR assumption, MMRM provides an unbiased estimate of treatment effect for the treatment period.

The estimand is described by the following attributes:

Target Population: subjects diagnosed with ASD with irritability, agitation, or self-injurious behaviors as defined by the inclusion/exclusion criteria of the study.

Variable (primary endpoint): change from Baseline at Week 6 in the ABC-I score.

Treatment condition: pimavanserin high dose, pimavanserin low dose, and placebo, daily for 6 weeks, without extended use of rescue medications or use of prohibited medications.

Intercurrent events (ICE) and strategies:

Note that in the following discussion of ICEs, the primary reason for treatment discontinuation is used to classify ICEs rather than secondary reasons for discontinuation (if any are recorded).

- The two intercurrent events (ICEs) “premature treatment discontinuation due to COVID-19”, “premature treatment discontinuation not due to COVID-19” (i.e. early dropout prior to Week 6) will be addressed by the hypothetical strategy, i.e., assuming that subjects with these ICEs evolve in the same way after ICE occurrence as the subjects in the same treatment group who complete the treatment.
- If “use of rescue medications beyond maximum allowed period” and “use of prohibited medications” (as listed in the Protocol Appendices A and B) resulted in treatment discontinuation, then the event will be considered as an ICE of treatment discontinuation and will be handled in the analyses according to the strategy for treatment discontinuation ICEs.

- The intercurrent event "remote assessments" will be addressed by the treatment policy strategy, i.e., utilizing measurements of the primary efficacy endpoint regardless of the occurrence of this ICE.

Alternative approaches to handling intercurrent events will be addressed in the sensitivity analyses.

Population-level summary: difference (pimavanserin high dose vs. placebo, pimavanserin low dose vs. placebo) in the mean changes from Baseline at Week 6 in the ABC-I score.

Hypotheses

Let Δ be the difference in the mean change from Baseline at Week 6 in the ABC-I subscale score between either pimavanserin dose group and the placebo group:

The null hypothesis for the primary efficacy endpoint is: $\Delta = 0$

The alternative hypothesis for the primary efficacy endpoint is: $\Delta \neq 0$

The hierarchical testing procedure for the hypothesis testing is described in Section 12.4.

Primary Estimator

The hypothesis testing will be performed for Full Analysis Set using the direct likelihood MMRM method assuming all missing data are missing at random (MAR). The dependent variable will be the change from Baseline in the ABC-I score. The independent variables in the model will include the following: age group (5 through 12 years old or 13 through 17 years old), region (U.S. or rest of world), treatment group (pimavanserin high dose, pimavanserin low dose, or placebo), visit (Week 1, Week 2, Week 3, Week 4, Week 5, or Week 6), treatment-by-visit interaction, and the Baseline ABC-I score. An unstructured covariance matrix will be used, and the Kenward-Roger approximation will be used to adjust the denominator degrees of freedom. Observed cases will be analyzed in the model and the least squares (LS) means will be estimated using observed margins. Sample SAS® code for MMRM analysis is provided in Appendix 21.2.

In the event that the model fails to converge using the unstructured covariance matrix, the following alternative covariance structures will be modeled in the order given: heterogeneous Toeplitz, heterogeneous compound symmetry, heterogeneous autoregressive (1), Toeplitz, compound symmetry, autoregressive (1), variance components. The first alternative covariance structure that allows for convergence will be selected for the final model, and the sandwich estimator (Diggle 1994) will be used to estimate the standard errors of the fixed effects parameters.

At each visit, the effect size (Cohen's d) for the change from Baseline between the treatment groups will be calculated using the following formula:

$$Effect\ size = \frac{LS\ mean\ difference}{\sqrt{variance}}$$

The variance at a given visit will be obtained from the covariance matrix estimated for the MMRM model. The sign (+ or -) of the effect size will be chosen so that a positive value favors pimavanserin.

Summary statistics for the ABC-I score (observed and change from baseline) will be presented for all visits from Baseline through Week 6. For change from Baseline values at each post-baseline visit, LS means and standard errors (SE), the between-group difference in LS means with the corresponding 95% confidence interval, p-value (2-sided) and effect size will also be presented. In addition, LS mean ± SE over time for the change from Baseline values by treatment group will also be presented in line plots.

The hypothesis testing will be performed based on the difference in LS means at Week 6 and will be tested at the significance level of 0.05, following the hierarchical testing procedure described in Section 12.4. The treatment-group comparisons at each of the other timepoints (Weeks 1, 2, 3, 4, or 5) using the same MMRM model will be considered exploratory.

The LS mean differences with corresponding 95% CIs from the primary analysis and sensitivity analyses below will also be graphically presented in a forest plot.

13.1.2 Sensitivity Analysis

The following sensitivity analyses of the primary efficacy endpoint are planned to account for intercurrent events.

13.1.2.1 Pattern-Mixture Models Assuming Missing Not At Random (MNAR) Using Multiple Imputation (MI)

MMRM assumes data are MAR, which is a reasonable assumption in longitudinal clinical trials. However, the possibility of missing not at random (MNAR) data can never be ruled out. In order to further evaluate robustness of the primary results to deviations from MAR assumptions, additional sensitivity analyses may be conducted.

Placebo Based Imputation Methods

Similar to “Standard” multiple imputations, except parameters for imputation model obtained from only the placebo (control) group. Missing data for both placebo and pimavanserin high and low dose groups are imputed based on the imputation model derived from placebo data within each stage. If pimavanserin high and/or low doses improved outcomes prior to dropout, this benefit is carried into subsequent imputed values, but will diminish over time in accordance with the correlation structure.

13.1.2.2 Tipping Point Analyses

Sensitivity analyses based on pattern-mixture model (PMM) using MI will be performed in order to explore data missing mechanisms of MNAR and investigate the response profile of dropout reasons for the following scenarios:

1) Dropout reason due to Adverse Event (AE) as MNAR
2) Dropout reasons due to AE, lack of efficacy (LOE), COVID-19, use of prohibited medication as MNAR
3) All dropouts not due to COVID-19 as MNAR
4) All dropouts as MNAR

This sensitivity analysis will be implemented for Full Analysis Set using delta adjustment imputation. This MNAR sensitivity analysis is to departure from MAR assumption by progressively increasing the delta until conclusion from the primary analysis is overturned. The delta is 0%, 5%, 10%, 15%, .., 100% of the observed treatment difference between a pimavanserin high dose and placebo, pimavanserin low dose and placebo, from the primary analysis of MMRM model until conclusion of the primary analysis is overturned. When delta = 0 the missing data are assumed to be MAR. When delta > 0, the missing data are assumed to be MNAR.

1) Using Monte Carlo Markov Chain (MCMC) methodology from PROC MI to impute the intermittent missing data to a monotone missing pattern
2) Using a standard MAR-based multiple imputation approach from PROC MI to impute the monotone missingness data
3) For subjects in the pimavanserin group with a dropout reason due to AE, LOE, subject withdraw consent, COVID-19, use of prohibited medication, a delta will be added for all the values after the dropout time
4) Using MMRM model in the primary analysis to analyze the completed data using PROC MIXED on the multiple imputed data
5) Obtaining the overall results using PROC MIANALYZE.

13.1.2.3 Sensitivity Analysis Using Randomized Analysis Set

This sensitivity analysis using MI assuming MAR will be performed using the Randomized Analysis Set.

13.1.2.4 Remote Assessments

To assess the impact of the intercurrent events of remote assessments on the primary analysis, an indicator variable will be assigned to flag subjects that had assessments done remotely for the primary efficacy endpoint at any visit including baseline (Yes/No). The remote visit flag indicator variable and the interaction between treatment group and remote visit flag indicator variable will be included as factors in the MMRM mode.

13.2 Secondary and Exploratory Efficacy Analyses

The change from Baseline at each post-baseline timepoint in the ABC subscale scores (lethargy, stereotypic behavior, hyperactivity, inappropriate speech), the CGI-S of irritability score and the CGI-I of irritability score will be analyzed using an MMRM model similar to the primary analysis described above for the primary endpoint, except that the Baseline covariate of the endpoint being analyzed will be included in the model instead of the Baseline ABC-I subscale score. For CGI-I of irritability score, the Baseline covariate include in the MMRM model will be the Baseline CGI-S of irritability score

The change from Baseline at each post-baseline timepoint in the RBS-R scores, the VABS-Socialization subscale score, and the CGSQ scores will be analyzed using an MMRM model similar to the primary analysis described above for the primary endpoint, except that the Baseline value of the endpoint being analyzed will be included in the model instead of the Baseline ABC-I subscale score, also the visit variable included in the model will reflect the visits scheduled for the respective endpoints.

The proportion of subjects with ≥25% reduction in the ABC-I subscale score, the proportion of subjects with a CGI-I of irritability score of 1 or 2, and the proportion of subjects with ≥25% reduction in the ABC-I subscale score and a CGI-I of irritability score of 1 or 2 will be summarized at each post-Baseline timepoint. The proportions will be compared between each pimavanserin dose group and the placebo group using Cochran-Mantel-Haenszel test stratified by age group and region; Newcombe's 95% CI stratified by age group and region will be presented. Subjects with missing score at a given visit are considered as non-responders for that visit.

In addition to the treatment comparisons for the Week 6 timepoint, the treatment groups will also be compared at each of the other timepoints (Weeks 1, 2, 3, 4, or 5). These comparisons will be considered exploratory.

14 SAFETY ANALYSES

The safety analysis will be performed on the Safety Analysis Set based on the actual treatment received. Safety results will be summarized by treatment group using descriptive statistics. No formal statistical testing will be performed for any of the safety endpoints.

Circumstances may arise (e.g., pandemic, natural disaster, political upheaval, or technical issues) when on-site assessments of safety measures may not be possible. In those cases, assessments may be performed at the subject's place of residence either in person or via video technology or telephone, where possible. If a subject is unable to come to the site for lab draws and the site is unable to travel to the subject's place of residence, the subject may visit a local lab to obtain all safety labs. 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.

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 in each treatment group, and the two pimavanserin groups combined 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 display in these tables will be sorted alphabetically by SOC, and then by frequency of the two pimavanserin groups combined within each SOC.

The event counts and the number and percentage of subjects with any TEAEs will also be tabulated by preferred term without SOCs within each treatment group and sorted by frequency of the two pimavanserin groups combined.

The incidence of most frequently reported (preferred terms reported by ≥ 5% of subjects in any treatment group) 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, preferred term, and treatment group. In addition, TEAEs by SOC, preferred term, and stimulant use will be summarized.

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

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. This listing will also include all relevant eCRF data associated with the event: e.g. date of onset, date resolved, date of last dose, severity, frequency, outcome, relationship to study drug, action taken with study drug, and required therapy. When a date is presented, the study day associated with the date will also be displayed. 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 Screening, Baseline, Week 3 and Week 6/ET 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:

$$s = L_{\cdot} + (\quad - \quad _{x}) \frac{s \; - \quad s}{x \; - \quad x}$$

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:

$$= \quad \frac{s}{x}$$

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}
- Magnesium (Mg), Vitamin B12 (only at Visit 1 [Screening])
- Thyroid function tests (only at Visit 1 [Screening])
  - Thyroid stimulating hormone (TSH) and reflex free T4
- Prolactin
  - Prolactin results should be blinded to the Investigator and the Sponsor after 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 serum pregnancy test should only be performed at Visit 1 for women of childbearing potential
  - A urine pregnancy test should be performed at all designated visits after Visit 1 for women of child-bearing potential
    - 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 (Hct), hemoglobin, red blood cells (RBC), platelets

    - Reticulocyte count

- Urinalysis
    - Blood, RBCs, WBCs, protein, glucose, ketones, specific gravity, pH
- Urine toxicology (drug) screen
    - Urine toxicology (drug) 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 [THC]), phencyclidine (PCP), ecstasy (3,4-methylenedioxymethamphetamine [MDMA]).
    - A urine toxicology dipstick should be used at the Baseline and Week 3 visits. At the Baseline visit, the urine toxicology dipstick should be performed before randomization.
    - 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 by treatment group using descriptive statistics at Baseline, Week 3 and Week 6. The change from Baseline values will also be summarized by treatment group 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 treatment group.

14.2.1.2 Shift Tables

Laboratory values (except the ones that were only assessed at screening and urine drug screen) will also be summarized in shift tables by treatment group, to determine the number and

percentage of subjects with values classified as below (low), within (normal), and above (high) normal ranges at Week 3, 6, overall post-baseline minimum and overall post-baseline maximum, relative to the same classification at the Baseline visit. 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 treatment group. 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 treatment group. 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

Number and percentage of subjects with PCI laboratory values at Week 3, 6 and overall post-baseline will be summarized by treatment group for selected parameters. PCI criteria are listed in Table 2 and Table 3. 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 treatment group, and the denominator is the number of subjects with non-missing values for the given parameter, visit and treatment group. 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 treatment group, and the denominator is the number of subjects with at least one post-baseline laboratory value for the given parameter and treatment group. 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 subjects with any PCI values will be provided.

Table 2 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

Table 2 Criteria for Potentially Clinically Important Laboratory Values – Hematology and Chemistry (Continued)

Analyte	Conventional Unit	Low PCI Criteria	High PCI Criteria	SI Unit	Low PCI Criteria	High PCI Criteria
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 3 Criteria for Potentially Clinically Important Laboratory Values - Urinalysis

Urinalysis	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 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 Screening and Baseline, at all visits from Week 1 through Week 6, and at unscheduled visits. BMI will be derived as Weight (kg)/ [Height (m)]^2. Height and weight measurements are optional at the Week 2 and Week 5 visits if these are conducted remotely.

Vital signs, height, weight, and BMI will be summarized by treatment group using descriptive statistics at Baseline and all post-baseline visits. The change from Baseline values will also be summarized by treatment group at the 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 4. The number and percentage of subjects with post-baseline PCI values will be summarized by treatment group 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 treatment group, and the denominator is the number of subjects with non-missing values for the given parameter, visit and treatment group. 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 treatment group, and the

denominator is the number of subjects with at least one post-baseline value for the given parameter and treatment group. A listing of all subjects with any PCI values will be provided.

Table 4 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)

Electrocardiogram is performed at Screening, Baseline, Week 3, and Week 6 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 by treatment group. 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 by treatment group 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 treatment group. 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 treatment group.

Electrocardiogram variable values will be considered as PCI values if they meet or exceed the upper limit values listed in Table 5. The number and percentage of subjects with post-baseline PCI values will be summarized by treatment group 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 treatment group, and the denominator is the number of subjects with non-missing values for the given parameter, visit and treatment group. 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 treatment group, and the denominator is the number of subjects with at least one post-baseline value for the given parameter and treatment group. A listing of all subjects with any PCI values will be provided.

Table 5 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	>60 msec	

14.5 Physical Examination

Physical examination is performed at Screening, Baseline, Week 3 and Week 6 visits.

Physical examination results (normal, abnormal, and not done) at Baseline and all scheduled post-baseline visits will be summarized in a frequency table by treatment group, body system and visit.

A listing of physical examination data will be provided.

14.6 Syncope and somnolence

Syncope and somnolence are assessed at Screening and Baseline, at all visits from Week 1 through Week 6, and at unscheduled visits. Occurrences syncope and somnolence should be reported as adverse events.

Subject history of syncope, family history of syncope, subject history of somnolence will be summarized as Baseline Disease Characteristics. Incidences of syncope and somnolence during the study will be summarized as TEAEs.

Assessment of syncope and somnolence at each visit will be listed.

14.7 Other Safety Variables

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

For the C-SSRS, the number and percentage of subjects with suicidal ideation or behavior during the study will be tabulated.

For calculating the percentages, the denominator will be the number of subjects with at least one post-baseline C-SSRS.

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

ESRS-A total scores and the four individual global clinical global impression movement severity scores will be summarized by treatment group using descriptive statistics at Baseline and each post-baseline visit. The change from Baseline values will also be summarized by treatment group at the post-baseline visits.

15 CLINICAL PHARMACOKINETIC AND PHARMACODYNAMIC ANALYSES

Plasma concentration data for pimavanserin and its metabolite (AC-279) will be listed and summarized using standard summary statistics.

If data allow, population PK and PK/PD analyses will be performed to further characterize the PK profile and exposure response relationship of pimavanserin and its metabolite using measures of safety, and efficacy parameters. The results of population PK and PK/PD modeling will be presented in a separate report. Pimavanserin plasma concentration data will remain blinded to the

Investigators and Sponsor study team until the unblinding of the clinical database at the end of the study.

16 INTERIM ANALYSIS

No interim analyses are planned for this study.

17 DATA MONITORING/REVIEW COMMITTEE

An independent Data and Safety Monitoring Board (DSMB) will review interim 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 detailed in the DSMB Charter.

An independent statistical group not affiliated with the Sponsor will produce unblinded statistical outputs and provide these outputs to DSMB members using a secure method. The Sponsor and the Investigators will remain blinded until the official unblinding of the database at the end of the study.

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 are made to the analyses 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.

Diggle, P.J., K.-Y. Liang, and S.L. Zeger. 1994. *Analysis of Longitudinal Data*. New York: Oxford University Press.

EMA (2009). *Guideline on Missing Data in Confirmatory Clinical Trials, European Medicines Agency*, London, UK.

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.

Ratitch, B. and O'Kelly, M. (2011). Implementation of Pattern-Mixture Models Using Standard SAS/STAT Procedures, in Proceedings of PharmaSUG 2011 (Pharmaceutical Industry SAS Users Group), SP04, Nashville.

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

Yuan, Y. Sensitivity analysis in multiple imputation for missing data, in Proceedings of SAS Global Forum 2014.

21 APPENDICES

21.1 Schedule of Events and Assessments

Period	**Screening**		**Double-blind treatment period**							**Follow-up**
		Baseline						**EOT/ET[r]**	**Unscheduled[a]**	**Safety follow-up[b]**
Visit Week	**-4 to 0**	**0**	**1**	**2**	**3**	**4**	**5**	**6**		**10**
Visit number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**		**9**
Visit window (days)	**+7[c]**		**±3[d]**	**±3[d]**	**±3[d]**	**±3[d]**	**±3[d]**	**±3[d]**		**+3[d]**
Type of visit[e]	Clinic	Clinic	Clinic	Clinic/ Remote	Clinic	Clinic	Clinic/ Remote	Clinic	Clinic	Telephone
Informed consent/assent	X									
Inclusion/exclusion criteria	X	X[f]								
Medical history and demographics	X									
Autism disease history	X									
HBV, HCV and HIV history	X									
COVID-19 history[t]	X									
ADI-R[u, bb]	X									
MINI-KID[u]	X									
Capsule swallowing test[g]	X	X								
Medication history	X	X								
Physical examination	X	X			X			X		
Vital signs	X	X	X	X[v]	X	X	X[v]	X	X	
Height [w], weight, and BMI	X	X	X	X[v]	X	X	X[v]	X	X	
12-lead ECG[h, i, q]	X	X			X			X		
Pharmacokinetic sample collection[i, j, q]		X[k]	X		X			X	X[l]	
Clinical laboratory tests [i, q, s, aa, cc]	X	X			X			X		
Thyroid function tests [i]	X									
Pregnancy test [i, m]	X	X			X			X		

21.1 Schedule of Events and Assessments (Continued)

Period	Screening		Double-blind treatment Period							Follow-up
		Baseline						EOT/ET[r]	Unscheduled[a]	Safety follow-up[b]
Visit Week	-4 to 0	0	1	2	3	4	5	6		10
Visit number	1	2	3	4	5	6	7	8		9
Visit window (days)	+7[c]		±3[d]	±3[d]	±3[d]	±3[d]	±3[d]	±3[d]		+3[d]
Type of visit[e]	Clinic	Clinic	Clinic	Clinic/ Remote	Clinic	Clinic	Clinic/ Remote	Clinic	Clinic	Telephone
Urine toxicology (drug) screen (UDS)[n]	X	X			X			X		
ABC[u]	X	X	X	X	X	X	X	X		
CGI-S of irritability	X	X	X	X	X	X	X	X		
CGI-I of irritability			X	X	X	X	X	X		
RBS-R[u]	X	X			X			X		
VABS-Socialization[u]		X			X			X		
CGSQ[u]	X	X		X		X		X		
C-SSRS[u]	X	X	X	X	X	X	X	X	X	
ESRS-A[u, x]	X	X	X	X[x]	X	X	X[x]	X	X	
Accurate Symptom Reporting (ASR) Training[u]	X									
Assessment of concomitant medications[t]	X	X	X	X	X	X	X	X	X	X
Assessment of adverse events	X	X	X	X	X	X	X	X	X	X
Assessment of syncope occurrence[y]	X	X	X	X	X	X	X	X	X	X
Assessment of somnolence occurrence[z]	X	X	X	X	X	X	X	X	X	X
Randomization		X								
Study drug dispensation		X	X		X	X			X[o]	
Study drug return and accountability[p]			X	X	X	X	X	X	X	

Abbreviations: ABC=Aberrant Behavior Checklist; ADI-R=Autism Diagnostic Interview–Revised; AE=adverse event; BMI=body mass index; 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; ECG=electrocardiogram; eGFR=estimated glomerular filtration rate; EOT=end of treatment; ESRS-A=Extrapyramidal Symptom Rating Scale–Abbreviated; ET=early termination; MINI-KID=Mini International Neuropsychiatric Interview for Children and Adolescents; PK=pharmacokinetic; RBS-R=Repetitive Behavior Scale–Revised; SAE=serious adverse event; 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 discontinue treatment prematurely from the study or who do not participate in the long-term extension 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] The Screening Period can be extended up to an additional 7 days (i.e., for a total of 5 weeks) before the Baseline visit, after discussion with the Medical Monitor. Discussion between the Investigator and Medical Monitor about eligibility is encouraged, to ensure that scientifically informed eligibility decisions are being made with all the necessary information, and in an objective and homogenous manner consistent with the protocol. Following the discussion, the Principal Investigator will document and sign off on the final eligibility decision.

[d] Visit timing and windows are relative to the Baseline visit, not relative to the previous visit.

[e] 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. Remote visits are permitted for Week 2 (Visit 4), Week 4 (Visit 6), and Week 5 (Visit 7). For all other 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.

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

[g] Subjects will be assessed for their ability to swallow a test capsule (i.e., placebo). If the subject is unable to swallow a test capsule at the Screening visit, they may take some capsules home to practice, and will be re-assessed at the Baseline visit.

[h] 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 that can occur either at Screening or Baseline.

[i] Mild sedation is allowed exceptionally for ECGs and blood draws during the study (e.g., alprazolam at appropriate doses according to age) 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.

[j] For all PK samples (scheduled and unscheduled, and except at the Baseline visit) the dates and times of administration of the last three doses of the study drug should be recorded.

[k] At the Baseline visit, a PK sample is drawn predose.

[l] Every effort should be made to collect a PK sample at an unscheduled visit because of an SAE, or an AE leading to discontinuation.

[m] For female subjects of childbearing potential, a serum pregnancy test will be completed at the Screening visit; urine pregnancy dipstick tests will be completed at all other scheduled timepoints.

[n] A urine toxicology dipstick should be used in addition to the urine toxicology screen at both the Baseline visit and visit 8 (to confirm eligibility in Study ACP-103-070). A rapid UDS dipstick will be used at Visit 3.

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

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

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

[r] The ET visit will not be done if the subject withdraws consent to participate in all parts of the study and withdrawal of consent happens before that timepoint.

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

[t] Including COVID-19 vaccination.

[u] 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, that 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. The ASR training should be done at Screening before the caregiver completes any scales, and repeated whenever there is a change in caregiver or if the site feels a caregiver requires retraining.

[v] If visit is conducted remotely, vital signs, height, weight, and BMI are optional.

[w] As measured by stadiometer.

[x] The ESRS-A will only be conducted at clinic visits and not at remote visits.

[y] If the caregiver reports an occurrence of syncope, the investigator should ask the “syncope adverse event questions”, as a tool to guide diagnosis, in Appendix G of protocol.

[z] If the caregiver reports an occurrence of somnolence, the investigator should ask the “somnolence adverse event questions”, as a tool to guide diagnosis, in Appendix H of protocol.

[aa] 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.

[bb] If an ADI-R has been completed for the subject by a certified rater (certified by Dr. Olsen and approved by the Sponsor) within 6 months prior to the Screening visit and if the original complete ADI-R response booklet with the patients’ answers is available as source data, then there is no need to conduct the ADI-R at Screening. The ADI-R can be administered remotely.

[cc] 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 SAS Programming Code for MMRM Model

The following sample SAS® code can be implemented for the MMRM analysis (modify as needed):

21.3 SAS Programming Code for Multiple Imputation under MNAR

The following sample SAS® code can be implemented to generate multiple imputed datasets (modify as needed):

run;

run;

%end;

%mend;

21.4 Summary of Version Changes

Version No:	Document History Description of Update	Author(s)	Version Date
1.0	Original version	Di An	30 APR 2021
1.1	Updated due to protocol amendments 1 and 2	Bruce Coate	10 DEC 2021
2.0	Updated due to protocol amendment 3 and FDA request to include SAS code for tipping point analysis	Bruce Coate	12 APR 2022
2.1	Reference Protocol Amendment 4	Bruce Coate	22 DEC 2022
3.0	Updated for changes in protocol amendment 4-IT; Removed the upper limit of Week 6 analysis visit window; Removed subgroup analysis by ADI-R and added subgroup analysis by stimulant use; Added details for some scales; Updated example SAS codes for sensitivity analysis and PCI cutoff for phosphate and urinalysis; Updated responder analysis to set missing score as non-responder; Updated hierarchical testing by adding the CGI-S; Added a section of Syncope and somnolence; Fixed minor typographical errors	I-Yuan Liu	29 JUL 2024