Otsuka Pharmaceutical Development & Commercialization, Inc.

Investigational New Drug

Brexpiprazole (OPC-34712)

A Phase 3, Multicenter, Randomized, Double-blind, Placebo-controlled Trial of Brexpiprazole in Treatment of children and Adolescents with Irritability Associated with Autism Spectrum Disorder

Protocol No. 331-201-00148
IND No. 141257
EudraCT No. 2019-000723-40

Statistical Analysis Plan

Version: Final 4.0

Date: 28 Sep 2022

CONFIDENTIAL - PROPRIETARY INFORMATION

SAP 331-201-00148

Table of Contents

SAP 331-201-00148

SAP 331-201-00148

List of In-text Tables

SAP 331-201-00148

List of In-text Figures

List of Appendices

SAP 331-201-00148

List of Abbreviations and Definitions of Terms

Abbreviation	Definition
ABC	Aberrant Behavior Checklist
ABC-I	Aberrant Behavior Checklist - Irritability
ADI-R	Autism Diagnostic Interview - Revised
AE	Adverse event
AIMS	Abnormal Involuntary Movement Scale
ALT	Alanine aminotransferase
ANCOVA	Analysis of Covariance
ASD	Autism Spectrum Disorder
AST	Aspartate aminotransferase
BARS	Barnes Akathisia Rating Scale
BMI	Body mass index
CDC	Center for Disease Control and Prevention
CGI-S	Clinical Global Impressions - Severity
CMH	Cochran-Mantel-Haenszel
CRF	Case Report Form
C-SSRS	Columbia-Suicide Severity Rating Scale
DBP	Diastolic blood pressure
DSM-5	*Diagnostic and Statistical Manual of Mental Disorders, 5th edition*
ECG	Electrocardiogram
EPS	Extrapyramidal symptoms
ET	Early termination
FDA	(United States) Food and Drug Administration
GEE	Generalized Estimating Equations
HbA1c	Glycosylated hemoglobin
HDL	High density lipoprotein
IAF	Informed assent form
ICF	Informed consent form
ID	Identification
IMP	Investigational medicinal product
IRE	Immediately reportable event
IRT	Interactive research technology
K-SADS-PL	Kiddie-Schedule for Affective Disorders and Schizophrenia - Present and Lifetime version
LDL	Low density lipoprotein
LOCF	Last Observation Carried Forward
LOE	Lack of Efficacy
MAR	Missing at Random
MCMC	Monte Carlo Markov Chain
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple Imputation
MMRM	Mixed-effect Model Repeated Measures
MNAR	Missing Not at Random
OC	Observed Case

OPDC	Otsuka Pharmaceutical Development and Commercialization, Inc.
PARS	Pediatric Anxiety Rating Scale
PE	Physical examination
PedsQL	Pediatric Quality of Life Inventory
PD	Pharmacodynamic
PK	Pharmacokinetic
PWR	Pediatric Written Request
QD	Once Daily
QTcF	QT interval as corrected for heart rate by Fridericia's formula
QTcN	QT interval as corrected for heart rate by the FDA Neuropharm Division
REML	Restricted Maximum Likelihood
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SAS	Simpson Angus Scale
SAS®	Statistical Analysis System®
SBP	Systolic blood pressure
SD	Standard deviation
SOC	System Organ Class
TEAE	Treatment-emergent adverse event
TSH	Thyroid stimulating hormone
ULN	Upper limit of normal
WHO	World Health Organization

SAP 331-201-00148

1 Introduction

This statistical analysis plan (SAP) expands the statistical section of the protocol 331-201-00148 and documents in detail the statistical methodologies and data analysis algorithms and conventions to be applied to the analysis and reporting of the efficacy and safety data collected for the study. All amendments to the protocol have been taken into consideration in developing this SAP. Of note, this SAP does not contain the analysis plan on the collected pharmacodynamic (PD) and pharmacokinetic (PK) data, which will be separately documented.

2 Study Objectives

The study objective is to evaluate the efficacy, safety and tolerability of brexpiprazole monotherapy in reducing irritability in children and adolescents aged 5 to 17 years with a diagnosis of autism spectrum disorder (ASD).

3 Trial Details

3.1 Study Design

This will be a phase 3, multicenter, randomized, double-blind, parallel-group, placebo controlled trial of brexpiprazole monotherapy (body weight based, flexibly-dosed with 1 to 3 mg/day post-titration) compared with placebo on irritability in children and adolescent subjects, aged 5 to 17 years, with a diagnosis of ASD according to the *Diagnostic and Statistical Manual of Mental Disorders, 5th edition* (DSM-5); the subjects must have a score $\geq$ 18 on the Aberrant Behavior Checklist - Irritability (ABC-I) subscale and a score $\geq$ 4 on the Clinical Global Impressions - Severity (CGI-S) scale. Individual subjects' participation into the trial will be up to 15 weeks, including up to 4 weeks of screening, 8 weeks of double-blind treatment, and 3 weeks of follow-up phase. See Figure 3.1-1 for the trial design schematic. The trial will be organized as follows.

Screening Phase: The screening phase (period) will range from 1 to 28 days and begins after the written informed consent/assent has been obtained. All required assessments during the screening period will be performed. Although the screening period will continue up to the administration of the first dose of the investigational medicinal product (IMP), screening procedures should be initiated with a sufficient amount of time allotted in order to obtain laboratory results and electrocardiogram (ECG) results from the central reader prior to randomization.

SAP 331-201-00148

Double-blind Treatment Phase: Following the screening period, subjects who meet all inclusion criteria and none of the exclusion criteria will be randomized in a 1:1 ratio, with stratification by clinical site, to receive either placebo or brexpiprazole for 8 weeks. The IMP will be dispensed to the subject/caregiver, and the subject will begin taking the IMP daily from the assigned blister card. Refer to Section 3.2 for the dosing schedule. Subjects will be evaluated at the following scheduled visits: Baseline (Day 1), Week 1 (Day 7), Week 2 (Day 14), Week 3 (Day 21) and Week 4 (Day 28), Week 6 (Day 42) and Week 8 (Day 56) during the double- blind treatment phase. The end of study assessments or procedures will be performed at the end of Week 8 (Day 56). Except Baseline, visits are to occur within ± 2 days of the target visit date. If a subject discontinues (or is discontinued) from the trial early, assessments or procedures noted for the Week 8 visit must be completed at the early termination (ET) visit. Attempts should be made to complete all evaluations prior to the administration of any new medications or treatment.

The visits at Weeks 1, 3, 4, and 6 may be conducted either virtually or in clinic (in office). When the investigator and caregiver make the decision to conduct a visit virtually, the visits will be conducted by means of telecommunication technology. That is, the caregiver and subject will remain in their own home and complete trial assessments and questionnaires via an online technology. The Day 1, Week 2 and Week 8 visits will be conducted at clinics (in office).

The Schedule of Assessments (Table 3.7-1 of Protocol 331-201-00148[1]) shows all required evaluations or assessments at all visits.

Follow-up Phase: Eligible subjects who complete all trial visits through the 8 weeks treatment phase may have the option to enroll into the open-label safety trial of brexpiprazole (Protocol 331-201-00191). All subjects who do not enroll in the open-label rollover trial will be followed up by a either a telephone or other acceptable means of contact about 21 (± 2) days after the last dose of IMP to assess any new or ongoing adverse events (AEs), suicidality and to record any concomitant medications. The follow-up contact also applies to subjects who are prematurely withdrawn from the trial.

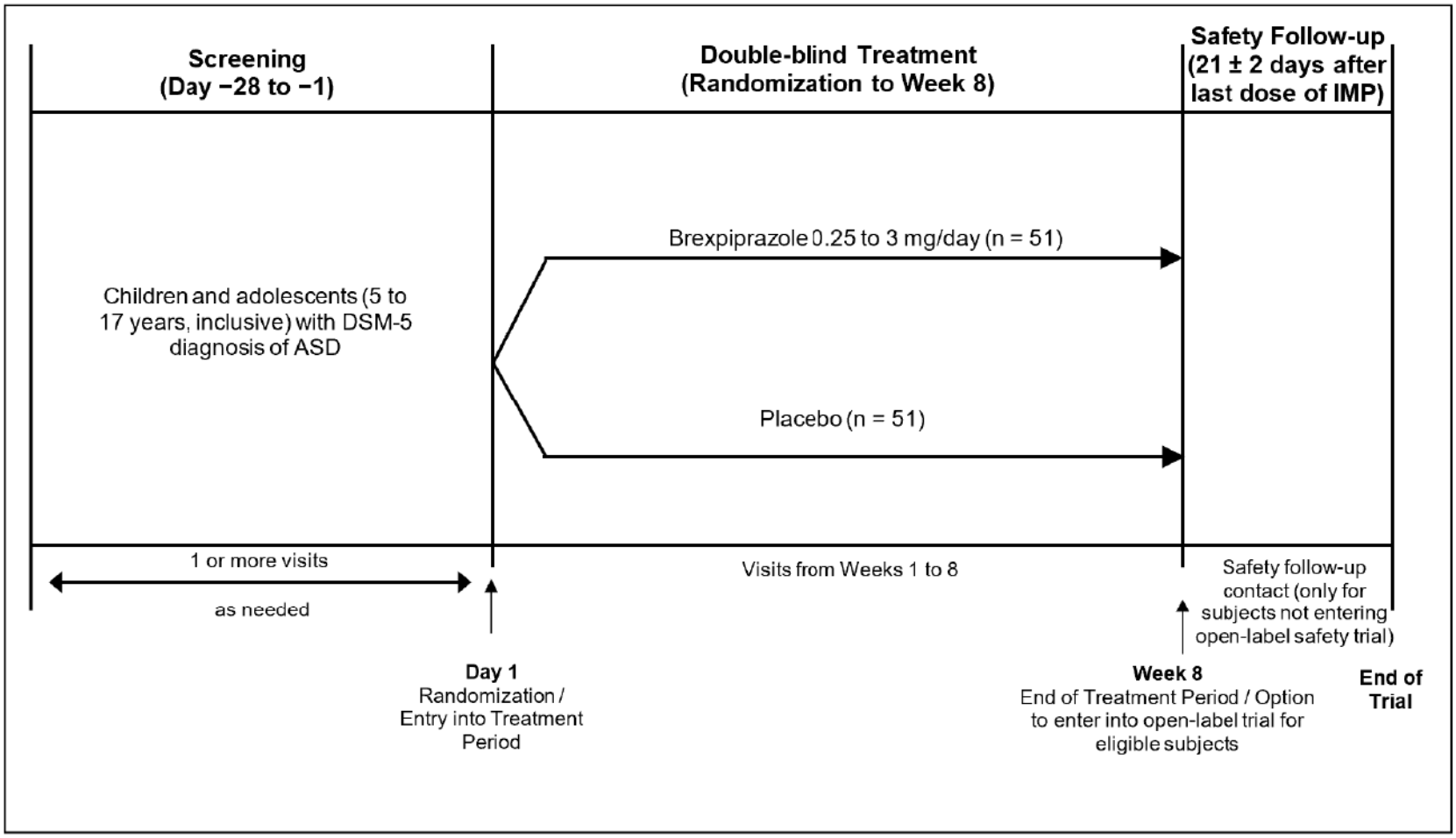


Figure 3.1-1 Trial Design Schematic

3.2 Trial Treatments

During the 8 weeks double-blind treatment phase subjects will receive IMP of either brexpiprazole monotherapy or placebo per the subject's treatment assignment at randomization. Subjects will follow a titration schedule and will have a flexible-dosing which depends upon their body weight stratum (< 50 kg, ≥ 50 kg) at the time of randomization, as shown in Table 3.2-1 below.

Table 3.2-1 Dosing Schedule During the Treatment Period[a]

Trial Day	No. Days on IMP	Brexpiprazole Dose		Placebo
		< 50 kg	≥ 50 kg	
Days 1 to 3	3	0.25 mg QD	0.50 mg QD	placebo tablets QD
Days 4 to 7	4	0.50 mg QD	1.5 mg QD	placebo tablets QD
Days 8 to 14	7	1 mg QD	2 mg QD	placebo tablets QD
Starting at Day 15 (earliest opportunity to increase to maximum dose within target dose range)	Based on investigator discretion to change dose based on therapeutic effect or tolerability	1 mg or 1.5 mg QD	1.5 mg, 2 mg, or 3 mg QD	placebo tablets QD

QD = once daily.

[a]Dose titrations will be performed in conjunction with a trial visit and therefore, trial visit windows are allowed for the titration schedule.

Subjects with body weight < 50 kg will receive a target dose range of 1 to 1.5 mg/day, and those with ≥ 50 kg will receive a target dose range of 1.5 to 3 mg/day. For all subjects, doses can be down-titrated due to tolerability based on investigator judgment. Subjects < 50 kg should receive a dose of 1.5 mg before decreasing the dose to 1 mg. Subjects ≥ 50 kg should receive a dose of 2 mg before decreasing the dose to 1.5 mg and 3 mg before decreasing to 2 mg. If a subject < 50 kg is unable to tolerate the 1 mg dose, the subject will be discontinued. If a subject ≥ 50 kg is unable to tolerate the 1.5 mg dose, the subject will be discontinued. To accommodate maximum flexibility in dosing for the autism subjects, more than 1 dose decrease or increase may be allowed based on the dosing schedule by weight. Any increase or decrease, however, should happen in a stepwise fashion based on investigator discretion to change dose based on therapeutic effect or tolerability according to Table 3.2-1.

SAP 331-201-00148

4 Sample Size and Power Justification

The primary efficacy endpoint is the change from baseline to the end (the Week 8 visit) of the double-blind treatment phase in the ABC-I subscale score. Approximately 102 subjects will be randomized (51 per treatment arm). A sample size of 51 per treatment arm will provide at least 85% power at a nominal 2-sided alpha level of 0.05 to detect a 6.0-point reduction in change from baseline in ABC-I subscale score for brexpiprazole versus placebo, assuming a standard deviation (SD) of 10.

5 Data Sets for Analysis and Missing Data

5.1 Data Sets for Analysis

Three analysis samples are defined for this trial as follows.

Randomized Sample: Comprises all subjects who were randomized into the trial. Subjects are considered randomized when they are assigned a treatment number by the Interactive Response Technology (IRT) at the end of the screening period. Subjects receiving trial treatment outside of the IRT will not be considered randomized, but such subjects will be included into Safety Sample as defined below.

Safety Sample: Comprises all enrolled subjects who received at least 1 dose of the IMP. Subjects will be excluded from this population only if there is documented evidence that the subjects did not take the IMP (ie, number of tablets/capsules dispensed = number of tablets/capsules returned, or no IMP ever dispensed). If a subject is dispensed trial medications and is lost to follow-up, he/she will be considered exposed to the IMP. Of note, a subject is deemed enrolled into the trial if the Informed Consent Form (ICF) / Informed Assent Form (IAF) has been obtained from the subject or his/her legally acceptable representative.

Efficacy Sample: Comprises all randomized subjects who took at least 1 dose of the IMP and had baseline and at least 1 post-baseline assessment of the primary efficacy variable ABC-I subscale score during the double-blind treatment phase.

5.2 Handling of Missing Data

For scoring via aggregating of component items of the efficacy and safety scales and on how to handle the items with unrecorded (or say, missing) ratings for the scoring, refer to Section 11 for details.

The Observed Case (OC) dataset for efficacy analyses will consist of actual observations or assessments of the efficacy parameters (variables) recorded at all scheduled or

unscheduled visits during the double-blind treatment phase. Note that if any observation or assessment is unrecorded at a scheduled visit, the missing value for that scheduled visit will not be imputed. It is expected that subjects who discontinue the double-blind treatment before the Week 8 visit will have an ET visit where efficacy observations or assessments will be collected. Such efficacy data collected at the ET visit will be mapped to a specific visit per the windowing algorithm detailed in Section 10.2. In rare instances where a subject has discontinued the double-blind treatment earlier but somehow does not attend a subsequent ET visit, perhaps due to loss to follow-up, all scheduled assessments after the timepoint of his/her treatment discontinuation will remain missing. The primary mixed-effect model repeated measures (MMRM) analysis under the missing at random (MAR) assumption will be performed based on the OC data only, and all its sensitivity analyses with the employment of the pattern-mixture model and other models under the missing not at random (MNAR) assumption will be based on the OC data as well.

The Last-Observation-Carried-Forward (LOCF) dataset will include all OC data and in addition the imputed data per a simple commonly used imputation rule (the LOCF rule) as follows. If no observation or assessment is recorded at a scheduled visit, the missing value for that visit will be imputed by carrying forward the observed value from the immediately preceding data point (inclusive of the ET visit or any other unscheduled visit). Note that baseline data cannot be carried forward to impute any post-baseline values. The Analysis of Covariance (ANCOVA) analysis will be performed primarily on the LOCF data and, if deemed necessary, on the OC data as well.

6 Study Conduct

6.1 Subject Disposition, Completion Rate and Reasons for Discontinuation

Subject disposition will be summarized for the Randomized Sample by treatment arm and by clinical center. Subject completion rate and reasons for discontinuation will be summarized for the Randomized Sample by treatment arm. Subjects who are evaluated at the last scheduled visit during the treatment period will be defined as trial completers. For purposes of this trial, subjects who complete the Week 8 visit will be defined as trial completers.

6.2 Treatment Compliance

Based on the IMP panel of the Case Report Form (CRF), an individual subject's compliance with taking IMP will be quantified by his/her overall compliance score,

which is calculated by dividing the total number of tablets/capsules (actually) taken by the total number of tablets/capsules scheduled to be taken for the observed treatment period from the first dose date to the last dose date (inclusive), then multiply by 100. Note that premature IMP termination will be taken into account while calculating the denominator (that is, the total number of tablets/capsules scheduled to be taken).

6.3 Protocol Deviation

Protocol deviations including the types of deviations (eg, deviations in entry criteria, dosing, concomitant medications, procedurals, etc.) will be summarized by trial center and treatment group on the Randomized Sample. A listing of protocol deviations will be provided. In addition, protocol deviations affected by COVID-19 will be summarized. Listing of subjects with protocol deviations affected by COVID-19 will also be provided.

7 Baseline Characteristics

7.1 Baseline Definition

Baseline (or say, baseline measurement or value) is defined as the last available measurement or assessment prior to the start (ie, the first dose) of the double-blind IMP.

7.2 Demographic Characteristics

Demographic and baseline characteristics include age, sex, race, ethnicity, height, body weight, waist circumference, and body mass index (BMI). Demographic characteristics will be summarized for the Randomized Sample by treatment group by using descriptive statistics. Refer to Section 10.4 for the conventions on using descriptive statistics. Age, height, body weight, waist circumference and BMI will be treated as continuous variables. Sex, race and ethnicity will be treated as categorical (nominal) variables.

7.3 Medical and Psychiatric History

Medical history and psychiatric history data will be summarized on the Randomized Sample by treatment group by using descriptive statistics.

7.4 Baseline Psychiatric Evaluations

Baseline psychiatric evaluations will be summarized for the following scales or inventories: Aberrant Behavior Checklist (ABC) including all its subscales such as Irritability, Social Withdrawal, Stereotypic behavior, Hyperactivity or noncompliance, and Inappropriate Speech; CGI-S; Pediatric Anxiety Rating Scale (PARS) and Pediatric Quality of Life Inventory (PedsQL). Refer to Section 11 for the explanations of the

efficacy and safety scales and the scoring based on the aggregation of the component items of the scales.

The ABC-I subscale score, all other ABC subscale scores, the CGI-S score, the PARS total score and various scores pertaining to PedsQL will be summarized on the Randomized Sample using descriptive statistics.

8 Efficacy Analysis

All efficacy analyses specified in this section will be conducted on the Efficacy Sample, and subjects will be included in the treatment group to which they were initially randomized. Baseline value is defined as the last available assessment prior to the first dose of the double- blind IMP.

8.1 The Primary Efficacy Endpoint

The primary efficacy endpoint is the change from baseline to the endpoint (ie, Week 8 visit) of the double-blind treatment phase in the ABC-I subscale score. The Week 8 visit is a landmark timepoint for this trial where the last assessment of all the efficacy outcomes will be collected.

8.1.1 The Primary Efficacy Analysis

The primary efficacy analysis will be based on the OC data only. The OC data refer to the actual observations or assessments of the outcomes that are recorded at all visits (including the ET visits) during the double-blind treatment phase, without imputation of any missing observations or assessments.

As the first confirmatory clinical trial of brexpiprazole in autistic children and adolescents, the proposed trial naturally seeks to clarify the efficacy of brexpiprazole in improving the ABC-I subscale in the idealistic scenario of full adherence to the assigned treatment in all subjects. In clinical trial practice, however, instances of non-adherence such as treatment discontinuation are likely to occur before the study endpoint (ie, the Week 8 visit for this trial). Therefore, the hypothetical estimand[2] is the most appropriate to use for the primary efficacy analysis. Specifically, this estimand, or target of estimation, following the hypothetical strategy is the treatment effect seen, assuming no withdrawals occurred. Subjects who withdraw from IMP treatment either could have lost their treatment effect, had the subjects not taken any other treatment after withdrawal, or could have their treatment effect been masked, had the subjects taken other treatment after withdrawal. This means that any observations made after subjects stop IMP will most likely not contribute relevant information about the treatment effect of the drug.

Due to this strategy, the last collected efficacy assessment after premature trial discontinuation will be done only once at the ET Visit. Every effort will be made to complete all of the ET evaluations prior to administering any additional medications for the treatment of ASD or other prohibited medications. In the case of terminal or lost to follow-up events, no ET evaluations would be expected, and only scheduled assessments would be performed before such an event occurred. The estimand is described by four attributes as follows.

- Target Population: Children and adolescent subjects of 5 to 17 years of age with irritability associated with ASD who met the protocol-defined inclusion/exclusion criteria and were qualified for the Efficacy Sample
- Endpoint: Change from Baseline to Week 8 in the ABC-I subscale score.
- Intercurrent Events: Intercurrent events refer to premature treatment discontinuation (ie, early dropout) prior to Week 8 attributable to adverse events, lack of efficacy, withdrawal of consent/assent, or any other causes.
- Measure of Intervention Effect: Difference in endpoint means between the brexpiprazole arm and the placebo arm.

During the course of this trial, the COVID-19 pandemic broke out. The pandemic had a significant impact on many aspects of clinical trials. There are occasionally virtual visits (ie, remote assessments) and possibly early discontinuation of treatment directly or indirectly related to the pandemic. However, in this study subjects were required to attend the screening, Day 1, Week 2, and the end of trial visits in clinic (ie, face-to-face). Note that if above in-clinic visits needed to be replaced by virtual visits, then these virtual visits in the pandemic environment will not be treated as an intercurrent event for the primary analysis. Subjects who drop out with a reason relating to the COVID-19 pandemic will be handled as they would have dropped out for another reason if the pandemic had not happened.

The hypothetical strategy of handling intercurrent events will be used to clarify the efficacy of the brexipiprazole had there be no occurrence of intercurrent events, regardless of being COVID-19 related or not. In other words, the estimand described above uses the hypothetical strategy[2] to address the treatment effect of interest that can be envisioned under the hypothetical setting of no premature treatment discontinuation prior to the end of the planned treatment period.

In this hypothetical strategy, the event of withdrawing IMP is considered MAR, and the primary endpoint of the trial could be considered to be a combination of the responses of on treatment completers at Week 8 and the imputation of the endpoint to Week 8

following the trend in each treatment group, using the MMRM method to impute missing data for subjects who withdraw IMP during the trial. All data collected during the trial treatment period will be used for statistical analysis. For the primary efficacy analysis, the treatment effect will be estimated using the MMRM method described below. Under the MAR assumption, MMRM provides an unbiased estimate of treatment effect for the treatment period.

Mean change from baseline in ABC-I subscale score during the double-blind treatment period will be analyzed using a restricted maximum likelihood (REML)-based repeated measures approach. Specifically, the MMRM analysis will include fixed-class effect terms of treatment, trial center, baseline body weight stratum, visit, and treatment-by-visit interaction, with baseline-by-visit interaction included as a covariate, based on the OC dataset. An unstructured (co)variance structure will be used to model the within-subject errors. If the analysis fails to converge, the following structures will be tested in order of 1) heterogeneous Toeplitz; 2) heterogeneous autoregressive of order 1; and 3) heterogeneous compound symmetry. The first correlation structure to yield convergence will be considered as the primary analysis. The Kenward-Roger approximation will be used to estimate denominator degree of freedom and adjust standard errors. Significance test will be based on least-square means using a two-sided 0.05 level. And statistical significance on the primary endpoint will be claimed if the test is nominally significant. The primary treatment comparison will be the contrast (ie, difference in least-square means) between brexpiprazole and placebo at the Week 8 visit during the double-blind treatment phase. The point estimate and the 95% confidence interval estimate of the mean difference will be reported. Analysis will be implemented using the software Statistical Analysis System® (SAS®, version 9.4 or later) procedure PROC MIXED.

If a structured covariance is to be used, the sandwich estimator of the standard error of the fixed effects parameters will be used in order to deal with possible model misspecification of the covariance matrix.

A small center is defined as a center that does not have at least 1 evaluable subject (with respect to the primary efficacy variable) in each treatment arm. For purpose of analysis, small centers will be pooled to form pseudo-centers so that each treatment group will include at least 1 evaluable subject within a center. Refer to Section 10.3 for the pooling algorithm.

The primary efficacy analysis as elaborated above assumes that the missing data arise from the missing at random (MAR) mechanism, which is often a reasonable assumption in longitudinal clinical trials[3]. However, the possibility of data missing not at random

(MNAR) cannot be ruled out. To assess the robustness of the inferences to departures from model assumptions with respect to missing data mechanism and normality of response variable, sensitivity analyses will be performed. The sensitivity analyses are specified below in Section 8.1.3.

8.1.2 Technical Computation Details for Primary Efficacy Analysis

The SAS code for carrying out the MMRM analysis with an unstructured variance covariance structure is illustrated as follows.

```
proc mixed;
   class treatment center visit stratum subjid;
   model change=treatment center stratum visit treatment*visit baseline*visit  / s cl ddfm=kenwardroger;
   repeated visit /type=un subject=subjid r rcorr;
   lsmeans treament*visit / pdiff cl alpha=0.05 slice=visit;
run;
```

where baseline refers to the baseline ABC-I subscale score and stratum refers to body weight stratum (< 50 kg and ≥ 50 kg) at baseline.

8.1.3 Sensitivity Analyses

8.1.3.1 Sensitivity Analyses for Missing Data Mechanism

Traditionally the dropout mechanisms are divided into three types[4] (Little, 1995): (1) Missing Completely at Random (MCAR), in which the probability of dropout doesn't depend on the observed data and the missing data; (2) Missing at Random (MAR), in which the probability of dropout depends on the observed data, and (3) Missing Not at Random (MNAR), where the probability of dropout depends on the missing data and possibly the observed data.

Most of MNAR methods[5] (Diggle P, Kenward MG, 1994) have treated all observations with dropout as if they fall within the same dropout type. In practice, we would find that different dropout reasons may be related to the outcomes in different ways, for example, detailed dropout reasons for this study are: adverse events (AE), lack of efficacy (LOE), lost to follow-up, protocol deviation, sponsor discontinued study, subject met (protocol specified) withdrawal criteria, subject was withdrawn from participation by the investigator, and subject withdrew consent to participate. Dropout due to an AE and LOE may lead to MNAR dropout. Subject withdrew consent may also lead to MNAR dropout. However, it is debatable whether a dropout caused by subjects withdrew consent is MAR

or MNAR. Except AE, LOE, and subject withdrew consent, all the other dropout reasons may be assumed as either MCAR or MAR dropout. Dropout due to COVID-19 will also be assumed as MAR.

As sensitivity analyses for MAR assumption, analyses for MNAR will be carried out. Pattern Mixture Models (PMM)[4,6,7,8] based on Multiple Imputation (MI) with mixed missing data mechanisms will be used to investigate the response profile of dropout patients by dropout (ie, treatment discontinuation) reason under the MNAR mechanism for the following three scenarios:

1) Dropout reasons due to either AE or LOE as being MNAR
2) Dropout reasons due to either AE or LOE or subject withdrawal of consent/assent as being MNAR
3) All dropouts (due to any known reasons or missing/unknown reason) as being MNAR

Delta Adjustment Imputation Methods

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%, 10%, 20%, 30%, ..., 100% of the expected treatment difference of 6 points and/or the observed mean difference between brexpiprazole and placebo arms from the primary analysis of MMRM model until the conclusion of the primary analysis is overturned. When delta = 0 it is MAR. When delta > 0 it is MNAR.

1) Using Monte Carlo Markov Chain (MCMC) methodology from SAS PROC MI to impute the intermittent missing data to a monotone missing pattern;
2) Using a standard MAR-based multiple imputation approach from SAS PROC MI to impute the monotone missingness data
3) For patients in the treated group and with a dropout reason of AE or LOE or subject withdrew consent, a delta will be added for all the values after the dropout time.
4) Using ANCOVA model in the primary analysis to analyze the completed data using SAS PROC MIXED on the multiple imputed data
5) Obtaining the overall results using SAS PROC MIANALYZE.

Placebo Based Imputation Methods

The placebo based imputations are similar to the standard multiple imputations, except that parameters for imputation model will be obtained from only the placebo (control) group. Missing data for both placebo and drug group are imputed based on the imputation model derived from placebo data. If drug improved outcomes prior to dropout, the benefit

is carried into subsequent imputed values, but will diminish over time in accordance with the correlation structure.

In addition, model based MNAR methods such as the selection model[5], the shared parameter model[9] (Wu and Baily, 1989) and the random coefficient pattern mixture model[7] (Hedeker D, Gibbons RD, 1997) might be performed if deemed useful.

8.1.3.2 Sensitivity Analyses for Violation of Normality Assumption

The MMRM analysis for the primary endpoint is a maximum likelihood method that relies on normality assumption. Residual analyses will be carried out to examine model assumption and outliers.

In the case of gross violations of the normality assumptions, nonparametric Van Elteren test[10] (Van Elteren, 1960) will be performed to compare treatment effect at Week 8 on both LOCF dataset and the Multiple Imputation (MI) data. The Van Elteren test is a generalized Cochran-Mantel-Haenszel (CMH) procedure useful for stratified continuous data in non-normality setting. It belongs to a general family of Mantel-Haenszel mean score tests. The test will be performed via SAS PROC FREQ, by including CMH2 and SCORES = MODRIDIT options in the TABLE statement. The stratification factors will be trial center and baseline body weight stratum.

In addition, other methods which are robust to distributional assumptions will also be performed to provide different views on the primary efficacy result. These robust methods include generalized estimating equations (GEE), weighted GEE (WGEE), and MI-robust regression[11]. For the MI-Van Elteren test and the MI-robust regression, imputation datasets will be generated with SAS MI procedure, each dataset will be analyzed, then an overall estimate will be pooled with SAS MIANALYZE procedure.

8.1.3.3 COVID-19 Pandemic Related Sensitivity Analyses

On March 13, 2020, the national emergence concerning the COVID-19 pandemic was announced in the US. The following analyses will be performed on the Efficacy Sample to evaluate the sensitivity of the primary and key secondary analysis results to the impact of the pandemic. The same model (eg, with the same set of explanatory variables and the response variable) as that for the primary efficacy analysis will be used for these analyses specified below. Of note, the definition of intercurrent events and the strategy for handling intercurrent events are identical to that for the primary efficacy analysis.

1) A MMRM analysis based on the Efficacy Sample excluding the assessments which should be conducted in clinic but actually are conducted remotely.

2) An MMRM analysis using the non-COVID data set based on Efficacy Sample. The non-COVID data set consists of the OC data during the non-COVID treatment period. For each subject, the non-COVID treatment period starts from randomization and ends on the Week 8/ET date or the date before the first COVID-19 related protocol deviation, whichever occurs earlier. The non-COVID treatment period represents the time period when subjects did not have any COVID-19 related protocol deviations during the double-blind treatment period.
3) An MMRM analysis based on the non-COVID Sample. The non-COVID Sample comprises those subjects in Efficacy Sample who did not have any COVID-19 related protocol deviations.
4) To explore the impact of assessments before versus after the COVID breakout, a subgroup analysis for subjects who completed or discontinued from the study before March 13, 2020 and subjects who completed or discontinued from the study on or after March 13, 2020 will be performed. An MMRM analysis based on the Efficacy Sample will be performed to the extent where data allows or the summary statistics will be provided.

8.1.4 Subgroup Analyses

For this study, subgroup analyses will serve as exploratory or hypothesis-generating purpose, given the smaller subgroup sample size. Subgroup analyses for the primary efficacy endpoint will be performed on the OC data for the following subgroups:

- Age (Ages 5 to 12 and Ages 13 to 17)
- Body weight (< 50 kg and ≥ 50 kg)
- Sex (Male and Female, based on biological status)
- Race (White and All Other Races)

All subgroup analyses will be conducted using the same MMRM model as that for the primary efficacy analysis except that the fixed effect term for trial center and baseline body weight stratum will be excluded from the model. Adding a BY statement into the syntax of SAS PROC MIXED will enable separate analyses of observations in subgroups that are defined by a subgrouping variable.

Interaction effects of treatment-by-subgroup will be assessed at Week 8 for the subgroups identified in the previous paragraph. The same MMRM model will be used as for the primary efficacy analysis with the addition of terms for subgroup-by-week and treatment-by-subgroup- by-week. These treatment-by-subgroup interaction analyses will be presented in statistical documentation.

SAP 331-201-00148

8.2 The Key Secondary Endpoint Analysis

The key secondary efficacy endpoint is the change from baseline to the endpoint (the Week 8 visit) of the double-blind treatment phase in the CGI-S score. The mean of which will be analyzed analogously to that of the primary endpoint by the approach of a restricted maximum likelihood (REML)-based MMRM. The MMRM model will be the same as in the primary analysis with a replacement of the interaction term of visit by CGI-S baseline score as a covariate. An unstructured (co)variance structure will be used to model the within-subject errors. If the analysis fails to converge, different covariance structure will be explored in the same order as that listed for the primary analysis (see Section 8.1.1). Small centers, if any, will be pooled per the pooling algorithm detailed in Section 10.3. The MMRM analysis will be based on the OC data only.

In order to control the overall type I error rate for this key secondary efficacy analysis, a hierarchical testing procedure will be used so that the overall family-wise type I error rate is maintained at 0.05. If the primary efficacy analysis for the ABC-I subscale score (as described in Section 8.1.1) yields a statistically significant result at 0.05 (two-sided) for the comparison of the brexpiprazole vs. placebo, then the corresponding comparison for this key secondary efficacy endpoint (CGI-S score) will be tested at an alpha level of 0.05 (two-sided). Should the primary analysis not achieve statistical significance, the key secondary endpoint will be examined as another efficacy endpoint

Serving as supplementary supporting evidence regarding the magnitude of the treatment effect on the CGI-S, the ANCOVA analysis will be used to estimate the least square mean difference between brexpiprazole and placebo groups in change-from-baseline CGI-S score at Week 8. Point estimate and interval (95%) estimate of the mean difference (with placebo being the reference group) and the corresponding two-sided p-value will be provided. The ANCOVA model includes change-from-baseline CGI-S score at Week 8 as the dependent variable, baseline CGI-S value as the covariate, and treatment group, study center, and baseline body weight stratum as main effects. The ANCOVA analysis will also be applied on the OC dataset, where study center and baseline body weight stratum will not be included in the model.

8.3 Analyses of Other Efficacy Endpoints

Other secondary efficacy endpoints are specified as follows. All other efficacy endpoints will be evaluated at a nominal 0.05 level (2-sided) without adjusting for multiplicity.

1) Change from baseline in ABC-I subscale score at each visit during the double-blind treatment phase other than the timepoint of the Week 8 visit.
2) Change from baseline in CGI-S score at each visit during the double-blind treatment period other than the timepoint of the Week 8 visit.
3) Response rate at Week 8 of the double-blind treatment phase, where response is defined as having ≥ 25% improvement (reduction) in ABC-I subscale score and ≥ 1 improvement (reduction) in CGI-S score compared to baseline.
4) Response rate at each visit during the double-blind treatment phase other than the timepoint of the Week 8 visit, where response at each visit is defined as having ≥ 25% improvement in ABC-I subscale score and ≥ 1 improvement in CGI-S score compared to baseline. Note that a timepoint (visit) will constitute a clinical endpoint with respect to the response rate.

Inferences for the above-mentioned endpoints (1) will be based on the same MMRM model as that for the primary efficacy endpoint, and inferences for (2) will be based on the same MMRM model as that for the secondary efficacy endpoint.

The endpoints specified in (3) and (4) will be analyzed by the stratified Cochran-Mantel-Haenszel (CMH) test controlling for trial center and baseline body weight stratum, on LOCF data set. Of note, small centers will be pooled to form pseudo-centers using the algorithm described in Section 10.3. The test for the hypothesis "Row Mean Score Differ" will be chosen for inferences. Because the treatment and response variables form a 2 by 2 table for each trial center and baseline body weight stratum, the tests for "Non-zero Correlation" and for "General Association" will yield the same inferences as that for "Row Mean Score Differ". Corresponding to each clinical endpoint in (3) and (4), the point estimate and the 95% interval estimate of the common relative risk and the corresponding p-value (one-sided) will be presented. The relative risk herein is defined as the proportion of responder in brexpiprazole group over the proportion of responder in placebo group. A higher value (say, greater than 1) of the relative risk estimate is favorable to the brexpiprazole group. OC analyses will also be conducted for (3) and (4) but will not control for trial center and baseline body weight stratum.

SAP 331-201-00148

CCI

CCI

8.6 Interim Analysis

No interim analysis is planned for this trial; however, variability will be monitored in a strictly blinded fashion and sample size could be adjusted upward to achieve the specified target power. Specifically, a sample size reassessment will be conducted based on pooled data (Week 8 LOCF) when more than 75% of subjects having completed the Week 8 visit or prematurely discontinued from the trial. The simple adjustment approach of Gould and Shih[13] (1992) will be used to re-estimate the common within-group standard deviation.

In July 2022, a blinded interim analysis was conducted when about 81% of subjects having completed the Week 8 visit or prematurely discontinued from the trial. The results showed that the current sample size (ie, 102 randomized subjects) will adequately maintain 85% study power. As a result, no sample size increase is needed.

9 Safety Analysis

All safety analyses will be performed on the Safety Sample. Standard safety variables to be analyzed include Adverse Events (AEs), routine clinical laboratory tests (hematology, serum chemistry, and urinalysis), HbA1c, vital signs, body weight, waist circumference, body mass index (BMI), 12-lead ECGs, physical examinations, coagulations, prolactin concentrations, TSH, and free T4. In addition, data from the following safety scales will be evaluated: Abnormal Involuntary Movement Scale (AIMS), Simpson-Angus Scale (SAS), Barnes Akathisia Rating Scale (BARS) and Columbia-Suicide Severity Rating Scale (C-SSRS).

A subject's baseline value in a safety variable is defined as the last observation or assessment of the variable on the subject prior to his/her the first dose of IMP. In fact, the baseline definition for this trial is the same for all analyses (demographics, efficacy and safety).

Prospectively defined criteria will be used to identify potentially clinically relevant abnormal values for clinical laboratory tests, vital signs and ECGs.

9.1 Adverse Events (AEs)

All adverse events will be coded by System Organ Class (SOC) and Preferred Term (PT) according to the latest version (Version 22.0 or newer) of Medical Dictionary for Regulatory Activities (MedDRA). AEs that are sex-specific, eg, ovarian cancer, will have the incidence rates evaluated for the specific sex. Treatment-emergent AEs (TEAEs) are defined as AEs with an onset date on or after the start of double-blind treatment period. In more detail, TEAEs are all adverse events which started after start of the double-blind IMP; or if the event was continuous from baseline and was worsening, serious, study drug related, or resulted in death, discontinuation, interruption or reduction of study therapy. Adverse events occurring up to 30 days after the last dose of IMP will be included in the summary tables. AEs having occurred 30 days after the last dose of IMP, if collected, will be listed only.

The incidence of the following events in the double-blind treatment period will be tabulated by treatment group and overall:

a) TEAEs
b) TEAEs by severity
c) TEAEs potentially causally related to the IMP
d) TEAEs with an outcome of death
e) Serious TEAEs
f) TEAEs leading to discontinuations of the IMP

The above summaries (b), (e), and (f) will also be prepared for TEAEs potentially causally related to the IMP.

In addition, TEAE incidence with rate ≥ 5% in the brexpiprazole arm and greater than that of the placebo arm at any SOC and PT levels will be provided. Incidence of TEAEs by SOC and PT will also be summarized for subgroups defined by sex, race, age and body weight stratum at randomization.

9.2 Clinical Laboratory Tests

Change from baseline at the Week 8 visit in the routine clinical laboratory parameters (ie, hematology, serum chemistry, and urinalysis), glycosylated hemoglobin (HbA1c), serum prolactin concentrations, coagulation parameters, thyroid stimulating hormone (TSH), and free T4 (if the TSH is abnormal) will be summarized on the OC data and last visit by using descriptive statistics for continuous variables. Of note, metabolic parameters (ie, total cholesterol, low density lipoprotein (LDL), high density lipoprotein (HDL) and triglycerides) and serum glucose will be separately summarized for fasting and non-fasting tests, and serum prolactin concentrations will be separately summarized for male and female subjects.

Potentially clinically relevant laboratory test results in the double-blind treatment period will be identified, and the number and percentage of subjects who had such a result will be tabulated by laboratory parameter and treatment group based on the observation from the scheduled (ie, Week 8) and the unscheduled (if any) post-baseline visits.

Potential serious hepatotoxicity is an immediately reportable adverse event of interest. The case of potential serious hepatotoxicity at a post-baseline visit is defined as having (1) aspartate aminotransferase (AST) or alanine aminotransferase (ALT) that is ≥ 2 times the upper limit of normal (ULN), and (2) the total Bilirubin ≥ 1.6 times the ULN. The incidence of potential serious hepatotoxicity will be tabulated by treatment group based on the observation from the scheduled (ie, Week 8) and the unscheduled (if any) post-baseline visits.

All potentially clinically relevant laboratory abnormalities will be listed by treatment group, patient ID, baseline value, age, sex, body weight, visit, and dose at onset of the

event if applicable. Criteria for identifying laboratory values of potential clinical relevance for subjects of 5 to 17 years of age are provided in Appendix 2.

9.3 Vital Signs

Vital sign measurements include: body temperature, systolic blood pressure (SBP), diastolic blood pressure (DBP), and heart rate. Considering two positions (the supine and the standing) being used in blood pressure and heart rate measurements, there will be 7 parameters for the vital signs. For the clinic visit (ie, the Day 1 and the Weeks 2 and 8), equipment at the site will be used for the measurements. For the virtual visits (ie, the Weeks 1, 3, 4, 6), equipment will be shipped to the subject's home with instructions for use.

Change from baseline at each visit during the double-blind treatment period in each of the 7 vital parameters will be summarized by using descriptive statistics on the last visit and the OC data as well.

Potentially clinically relevant vital signs abnormalities in the double-blind treatment period will be identified, and the number and percentage of subjects who had such an abnormality in vital signs will be tabulated based on the observation from the scheduled and the unscheduled (if any) post-baseline visits. And all potentially clinically relevant vital signs abnormalities will be listed by treatment group, patient ID, baseline value, age, sex, body weight, visit number, type of visit (virtual or clinic), and dose at onset of the event if applicable.

Criteria for identifying vital signs of potential clinical relevance for subjects of 5 to 17 years of age are provided in Appendix 1.

9.4 12-Lead Electrocardiogram (ECG)

Standard 12-lead ECG measurements will include QTc, heart rate, PR interval, and QRS complex. For the analysis of QT and QTc, data from three consecutive complexes (representing three consecutive heart beats) will be measured to determine average values. The following QT corrections will be used for reporting purposes:

1) QTcF is the length of the QT interval corrected for heart rate by the Fridericia formula: $QTcF = QT/(RR)^{0.33}$
2) QTcN is the length of the QT interval corrected for heart rate by the FDA Neuropharm Division formula: $QTcN = QT/(RR)^{0.37}$

For each of the quantitative ECG parameters, change from baseline to Week 8 of the treatment period (OC) data and last visit will be summarized by using descriptive statistics for continuous variables.

Criteria for identifying potentially clinically relevant ECG abnormality for subject of 5 to 17 years of age are provided in Appendix 3. All potentially clinically relevant ECG abnormalities during the treatment period will be listed by treatment group, patient number, sex, age, body weight, baseline value, visit and dose at onset of the event if applicable. In addition, the number and percentage of patients who had a potentially clinically relevant ECG abnormality will be tabulated by treatment group based on the observation at the scheduled (ie, Week 8) and the unscheduled (if any) post-baseline visits.

Categorical Change in ECG parameters during the double-blind treatment period will be summarized based on the following criteria:

Table 9.4-1 Categorical Change Criteria in QT/QTc Parameters

Classification	Category	Criteria
QT	New onset (≥ 450 msec for men or ≥ 470 msec for women)	New onset in QT means a subject who attains a cut off value during treatment period but not at baseline.
QTc[a]	New onset (≥ 450 msec for men or ≥ 470 msec for women)	New onset in QTc means a subject who attains a cut-off value during treatment period but not at baseline.
	New onset (≥ 450 msec for men or ≥ 470 msec for women) and > 10% increase	New onset and > 10% increase in QTc means a subject who attains a cut off value and > 10% increase during treatment period but not at baseline
	New onset (> 500 msec)	New onset (> 500 msec) in QTc means a subject who attains a value > 500 msec during treatment period but not at baseline.
	Increase 30 - 60 msec	Increase from baseline value > 30 and ≤ 60 msec in QTc
	Increase > 60 msec	Increase from baseline value > 60 msec in QTc

[a]QTc categorical change criteria apply to QTcF and QTcN.

9.5 Physical Examinations

By-patient listings will be provided for physical examinations (PE).

9.5.1 Body Weight, Waist Circumference and BMI

For parameters such as body weight, waist circumference and body mass index (BMI), the mean change from baseline to Week 8 (OC) and last visit of the double-blind treatment period will be analyzed via an ANCOVA model. The ANCOVA models for both the OC and last visit analyses will include baseline value as the covariate and treatment group as the main effect of interest. Of note, BMI is defined as weight in kilograms divided by the square of height in meters.

Percentage of subjects showing significant weight gain (≥ 7 % increase in body weight relative to baseline), or significant weight loss (≥ 7 % decrease in body weight relative to baseline) at Week 8 (OC) and last visit will be respectively analyzed by using CMH General Association Test (testing the association between treatment and clinically significant weight change).

9.5.2 Z-score

Weight z-score is a variable of interest to be derived due to the natural growth of children and adolescent subjects. It describes how similar a subject is to his/her age and gender peers by determining the number of standard deviations from the expected weight. For each visit, weight z-score is calculated as the deviation of a subject's weight from the mean weight of the reference population divided by the standard deviation for the reference population. Weight z-score and change from baseline in weight z-scores will be summarized on the OC data by specified visit (ie, the Day 1 and Week 8) and last visit by using descriptive statistics for continuous variables.

Height z-score and BMI z-score will be similarly calculated and similarly summarized as that for weight. Refer to Section 9.5.2.1 for the calculation of z-score for body weight, height, and BMI. Furthermore, the number and percentage of subjects with the change of BMI z-score (relative to baseline) ≥ 0.5 or ≤ -0.5 at Week 8 will be tabulated by treatment group, based on the OC data and last visit.

9.5.2.1 Calculation of Z-scores for Body Weight, Height, and BMI

Age and gender adjusted z-scores for body weight, height and BMI will be calculated using the approach of the Center for Disease Control and Prevention (CDC), USA. The CDC provides a reference dataset (CDCref_d.sas7bdat in sas data format or CDCref_d.csv in csv data format) and a sas program (cdc-source-code.sas) along with detailed instructions for the calculations at

https://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm

Z-scores are calculated as Z = [((value / M)**L) - 1] / (S * L) , in which "value" is the child's BMI, weight, height, etc. The L, M, and S values are in the reference dataset and vary according to the child's sex and age. The following rules will be observed:

1) Age (in months) at the day of assessment, which is an input for the calculations, will be calculated as: (assessment date - birth date + 1)/(365.25/12).
2) In principle, BMI and its z-score will be calculated only if weight and height are both taken on the same day.

3) The above calculations are applicable in children and adolescents older than 24 months but younger than 20 years of age.

It is not anticipated that CDC's information (including the reference dataset or the SAS program or the instructions for the calculations) will be updated in the short run.

Nevertheless, the information at the CDC website will be re-checked for any updates within a month prior to the final database lock, and updated information will be used if any.

9.6 Extrapyramidal Symptoms Scales (SAS, AIMS, and BARS)

Extrapyramidal symptoms scales (SAS, AIMS, and BARS) will be assessed at two specified visits (ie, the Week 2 and Week 8) at clinical site. The mean change from baseline in variables SAS total score, AIMS total score, and the BARS global clinical assessment score (see Section 11.5, Section 11.6 and Section 11.7 for the construction of these scores) will be respectively analyzed on the OC data by specified visit by employing an ANCOVA model. In addition, analyses will be performed using the maximum (ie, the worst) value observed during the double-blinded treatment period and the last visit data to determine the change from baseline score. The ANCOVA model for the OC data set will include the baseline measure and the treatment group. The ANCOVA model for change at the last visit and for change to the maximum value will include the baseline measure, study center, baseline body weight stratum, and treatment group. For the AIM individual item score on Item 8, 9 and 10, analyses will be performed analogously to the above.

In addition, incidence of BARS global clinical assessment during the double-blinded treatment period by severity category will be provided using summary statistics on OC data and last visit.

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

Suicidality will be monitored during the trial using the C-SSRS (Child Version). Suicidality data will be collected at all visits during the study, including the follow-up visit.

Variables pertaining to suicidality assessments are categorical in nature. Thus, summaries of C-SSRS data will be provided in the form of tabulation of the number and percentage of subjects who have had certain outcomes for each of the suicidality categories of interest.

The number and percentage of subjects with the presence of suicidality during the entire study will be tabulated. Note that, the presence of suicidality is defined as reporting of at least one occurrence of any types of suicidal ideation or behavior at any visits during the study (including the follow-up period). Note that a subject will be counted only once if he/she had multiple occurrences of suicidal ideation or behavior during the study.

And the following four categories of treatment emergent suicidality will be analogously summarized: (1) Emergence of suicidal ideation, which is defined as reporting of any types of suicidal ideation during the treatment and the follow-up period while there was no suicidal ideation at baseline; (2) Emergence of serious suicidal ideation, which is defined as having observation of suicidal ideation severity rating of 4 or 5 during the treatment and the follow-up period while there was no suicidal ideation at baseline; (3) Worsening of suicidal ideation, which is defined as having a suicidal ideation severity rating that is more severe compared to baseline; (4) Emergence of suicidal behavior, which is defined as reporting of any types of suicidal behavior during the treatment and follow-up period while there was no suicidal behavior at baseline. Specifically, for each of the above categories of treatment emergent suicidality, descriptive statistics in the form of count and percentage of the subjects who experienced such a treatment emergent event at any visits during the study will be provided by treatment group.

There are 4 specific types of suicidal behavior and 5 specific types of suicidal ideation. The number and percentage of subjects who experienced any of those types of behavior or ideation will be tabulated by treatment group.

Wherever applicable, assessments on all individual questionnaires pertaining to suicidal behavior or ideation will be summarized by visit by using the descriptive statistics for categorical variables.

9.8 Concomitant Medications

The number and percentage of patients taking concomitant medications prior to the double-blind treatment period, during the double-blind treatment period, and during the follow-up period will be respectively tabulated by drug classification using the latest version of the World Health Organization (WHO) drug dictionary.

9.9 Extent of Exposure

An individual subject's Total Days of Exposure to IMP is calculated as: last dose date - first dose date + 1, regardless of any gaps in treatment (such as dose interruption or omission). The variable Total Days of Exposure will be summarized on the Safety Sample by descriptive statistics for a continuous variable.

The number and percentage of subjects who received the double-blind IMP will be presented by dosing week and by treatment group. The dosing week will be based on the actual week since Day 1 (ie, Day 1 to 7 as Week 1, Day 8 to 14 as Week 2, etc.).

An individual subject's Average Daily Dosage (unit: mg/day) will be calculated for each dosing week by dividing the subject's total doses in mg taken over the week by the number of his/her exposure days in that week. Note that the number of exposure days in a week is not always equal to 7. It can be less than 7 for the week in which early treatment discontinuation occurs. Note that the number of exposure days in the week will not be adjusted for any gaps in treatment during the week. The variable Average Daily Dosage will be summarized by dosing week and treatment group by using descriptive statistics.

10 Conventions

10.1 Software for Statistical Analysis and Reporting

The Statistical Analysis System® (SAS®, version 9.3 or newer) will be used for all statistical analysis and reporting.

10.2 Visit Window and the Derived Analysis Visit

Analysis visit will be derived and be used for purposes of data summary and analysis for all efficacy parameters and selected safety parameters (including all ABC subscales, CGI-S, PedsQL, PARS, SAS, AIMS and BARS). Observations or assessments collected at all the protocol-specified scheduled visits and the ET visit during the double-blind treatment period will be mapped to an analysis visit using the windowing algorithm given in Table 10.2-1. Note that for Table 10.2-1 the variable Study Day is defined as the number of days since the start of IMP dosing. To formulate it, Study Day = date of observation - IMP dosing start date + 1, if date of observation is on or after IMP dosing start date; Study Day = date of observation - IMP dosing start date, if date of observation is before IMP dosing start date. Of note, the IMP dosing start date is referred to as Day 1.

Table 10.2-1 Mapping of the Analysis Visit

Analysis Visit	Target Study Day	Study Day Interval (endpoints inclusive)
Day 1 Visit	1	1
Week 1 (Day 7) Visit	7	2-10
Week 2 (Day 14) Visit	14	11-17
Week 3 (Day 21) Visit	21	18-24
Week 4 (Day 28) Visit	28	25-35
Week 6 (Day 42) Visit	42	36-49
Week 8 (Day 56) Visit	56	50-63

In the instances where multiple observations fall within one visit window (or say, study day interval), the last observation within the window will be selected for analysis use and will be mapped to an analysis visit. However, the observations that are not selected will be listed or documented. Evaluations or assessments that take place more than 7 days after the last dose of IMP will not be mapped into an analysis visit and will be excluded from data analysis.

For the listing presentation of data, the derived variables Study Day and Analysis Visit will be listed along with the original CRF visit (ie, the visit as recorded on the CRF). Note that the original CRF visit will be used for data summary and analysis for the parameters such as AEs, Concomitant Medication, and C-SSRS.

(Note that, if any primary assessments are made more than 7 days after treatment discontinuation or after the ET visit, the assessments will not be used in the primary analysis but will be reported in listing.)

10.3 Pooling of small centers

A small center is defined as a center that does not have at least one evaluable subject (with respect to the efficacy variable of interest) in each treatment arm in the double-blind treatment period. Note that an evaluable subject is defined as an efficacy subject (ie, subjects in the Efficacy Sample) who has baseline assessment and at least one post-baseline assessment of the efficacy variable of interest during the double-blind treatment phase. For the primary efficacy analysis for the ABC-I subscale score, the set of evaluable subjects is simply the Efficacy Sample. For the secondary efficacy analysis for the CGI-S, the set of evaluable subjects could be a proper subset of the Efficacy Sample if there is any efficacy subject who has an ABC assessment but does not have any post-baseline CGI-S assessment during the double-blind treatment phase.

Because of stratified randomization by trial center, the problem of small center is unlikely to occur. If it does occur, small centers will be pooled to form pseudo-centers for purpose of analysis according to the algorithm as follows. Small centers will be ordered from the largest to the smallest based on the number of evaluable subjects. The process will start by pooling the largest of the small centers with the smallest of the small centers until a non-small center is formed. This process will be repeated using the centers left out of the previous pass. In case of ties in center size, the center with the smallest center code will be selected. If any centers are left out at the end of this process, they will be pooled with the smallest pseudo- centers, or if no pseudo-center exists, they will be pooled with the smallest non-small center.

The pseudo-center will be used for all efficacy analyses where the model is adjusted for study center.

10.4 Descriptive Statistics for Continuous and Categorical Variables

For continuous variables, descriptive statistics refer to N (ie, number of subjects with non- missing value), mean, standard deviation, median, quartile, minimum and maximum. For categorical (eg, nominal or dichotomous) variables, descriptive statistics refer to frequency distribution (eg, count and percentage). Unless stated otherwise, ordinal variables such as assessment scores on all the scales used for this trial will be treated as continuous. Wherever possible, (model-based) inferential statistics should be accompanied by descriptive statistics for presentation. Descriptive statistics can be used alone for data summary purpose, for example, for summarizing baseline demographic characteristics, medical history, etc. For this study, descriptive statistics will be provided by treatment (ie, separately for each treatment group). Unless specified otherwise, the denominator used for calculating percentage will be the number of unique subjects in the analysis sample. And whenever applicable, a row of "Missing/Unknown" will be presented along with rows of recorded outcomes for the tabulation of categorical variables.

11 Scales: Rules for Scoring and Handling of Missing Data

11.1 Aberrant Behavior Checklist (ABC) Subscales

The ABC scale has 58 items, which divide into 5 subscales as follows: (1) Irritability, Agitation; (2) Lethargy, Social Withdrawal; (3) Stereotypic Behavior; (4) Hyperactivity, Noncompliance; and (5) Inappropriate Speech. Each of the 58 ABC items is rated on a 4-point scale (0 = not at all a problem; 1 = the behavior is a problem, but slight in degree; 2 = the problem is moderately serious; 3 = the problem is severe in degree).

The Irritability subscale (ABC-I) total score is the sum of the ratings over 15 ABC items as follows. Item 2: injuries self on purpose; Item 4: verbally or physically aggressive to other children or adults; Item 8: scream inappropriately; Item 10: temper tantrums or outbursts; Item 14: irritable and whiny; Item 19: yells at inappropriate times; Item 25: depressed mood; Item 29: demands must be met immediately; Item 34: cries over minor annoyances and hurts; Item 36: mood changes quickly; Item 41: cries and screams inappropriately; Item 47: stamps feet or bangs objects or slam doors; Item 50: deliberately hurts himself/herself; Item 52: does physical violence to self; Item 57: has

temper outbursts or tantrums when he/she does not get own way. Thus, ABC-I total score ranges from 0 to 45. For a visit, a subject's ABC-I total score will be unevaluable and set to missing if less than 12 of the 15 items are recorded for the visit. If 12, 13 or 14 of the 15 items are recorded, the ABC-I total score will be the mean of the recorded items multiplied by 15 and then rounded to the first decimal place. To state equivalently, if more than 20% of the component items have missing ratings (or say, not recorded), the ABC-I total score will be set to missing; otherwise the total score will be the mean of the recorded items multiplied by 15.

The ABC subscale score on Social Withdrawal, Stereotypic Behavior, Hyperactivity/Noncompliance, and Inappropriate Speech is respectively the sum of ratings over 16, 7, 16 and 4 individual ABC items. For each of these 4 subscales, if more than 20% of the component items have missing ratings (of note, at the subject-visit level), the subscale score will be set to missing; otherwise the subscale score will be the mean of the recorded items multiplied by the total number of component items for the subscale.

11.2 Clinical Global Impressions - Severity (CGI-S)

The severity of illness for subjects with ASD will be rated using the CGI-S with a focus on symptoms of irritability. To perform this assessment, the rater or investigator will answer the following question: "Considering your total clinical experience with this particular population, how ill is the patient at this time with regard to symptoms of irritability?" Response choices include: 0 = not assessed; 1 = normal, not at all ill; 2 = borderline; 3 = mildly ill; 4 = moderately ill; 5 = markedly ill; 6 = severely ill; and 7 = among the most extremely ill patients. The response choice 0 (= not assessed) will be set to missing. The CGI-S is thus a 7-point scale from 1 to 7, and the CGI-S score ranges from 1 to 7.

11.3 Pediatric Quality of Life Inventory (PedsQL)

Quality of life will be assessed for subjects via parent proxy report using the PedsQL™ Generic Core Scales Version 4.0 for the corresponding age group (Teens aged 13 to 18, Children aged 8 to 12, Young Children aged 5 to 7). The "acute" version measuring the past seven days will be used. The 23-item PedsQL Generic Core Scale measures health-related quality of life across 4 dimensions, consisting of an 8-item Physical Function, a 5-item Emotional Function, a 5-item Social Function, and a 5-item School Function. For children and adolescents aged ≥ 8, each item will be rated on a 5-point Likert scale from 0 to 4, with a score of 0 representing "Never" and 4 representing "Almost Always". For young children aged 5 to 7, to facilitate their comprehension, each item will be rated on a

3-point Likert scale, with a score of 0 representing "Not at all", 2 representing "Sometimes", and 4 representing "A lot". Items must be reversed scored and linearly transformed to a 0 to 100 scale as follows: 0 = 100, 1 = 75, 2 = 50, 3 = 25 and 4 = 0. The transformed scale will be used for the scoring of the following variables of interest.

1) The PedsQL Generic Core Scale Average Score is the average of the transformed score over 23 items across the 4 dimensions as aforementioned.
2) The PedsQL Psychosocial Health Average Score is the average of the transformed score over 15 items across 3 dimensions (Emotional Function, Social Function and School Function).
3) The PedsQL Physical Health Average Score is the average of the transformed score over the 8 items from the Physical Function dimension.
4) The PedsQL Emotional Function Average Score is the average of the transformed score over the 5 items from the Emotional Function dimension.
5) The PedsQL Social Function Average Score is the average of the transformed score over the 5 items from the Social Function dimension.
6) The PedsQL School Function Average Score is the average of the transformed score over the 5 items from the School Function dimension.

Quality of life function for the subject's family will be assessed via parent report using the PedsQL 2.0 Family Impact Module measuring the 8 dimensions (domains) as follows: Physical Functioning (6 items), Emotional Functioning (5 items), Social Functioning (4 items), Cognitive Functioning (5 items), Communication (3 items), Worry (5 items), Daily Activities (3 items), and Family Relationships (5 items) that are impacted by the child's disease. The "acute" version measuring the past seven days will be used. Each item will be rated on a 5-point Likert scale, with a score of 0 representing "Never" and 4 representing "Almost Always". Items must be reversed scored and linearly transformed to a 0 to 100 scale as follows: 0 = 100, 1 = 75, 2 = 50, 3 = 25 and 4 = 0. The transformed scale will be used for the scoring of the following variables of interest.

7) Family Impact Module Average Score is the average of the transformed score over 36 items across the 8 dimensions as aforementioned.
8) Parent HRQL Average Score is the average of the transformed score over 20 items across the 4 dimensions (Physical, Emotional, Social and Cognitive functioning).
9) Family Functioning Average Score is the average of the transformed score over 8 items across the 2 dimensions (Daily Activities and Family Relationship).

All the above constructed variables take values (average scores) from 0 to 100. Higher scores indicate better health-related quality of life. Note that, if more than 50% of the

component items in the scoring (ie, constructing the average score of interest as described above) are not rated, the average score will be set to missing.

11.4 Pediatric Anxiety Rating Scale (PARS)

The PARS is a clinician-rated instrument for assessing the severity of anxiety symptoms associated with common *DSM-5* anxiety disorders in children. The PARS contains two components: a symptom checklist and severity items. The checklist has 50 items of symptoms, each item will be rated by the rater (interviewing clinician) as present or absent (yes/no). The PARS has 7 severity items as follows: 1) Overall number of anxiety symptoms; 2) Overall frequency of anxiety symptoms; 3) Overall severity of anxiety feelings; 4) Overall severity of physical symptoms of anxiety; 5) Overall avoidance of anxiety-provoking situations; 6) Interference with family relationship and/or performance at home; 7) Interference with peer and adult relationship and/or performance outside of home. For each of the 7 severity items, severity is rated with a 6-point Likert scale from 0 to 5 (0 for none, and 1 to 5 for minimal to extreme). The PARS total score is calculated as summing 5 (specifically, Item 2, 3, 5, 6 and 7) of the 7 severity items, and thus the total score ranges from 0 to 25. If 3 or more out of the 5 summands have missing values, the PARS total score will be set to missing; otherwise the PARS total score will be the mean of the rated items multiplied by 5.

11.5 Simpson Angus Scale (SAS)

The SAS will be used to evaluate extrapyramidal symptoms (EPS). It consists of a list of 10 symptoms of Parkinsonism (gait, arm dropping, shoulder shaking, elbow rigidity, wrist rigidity, head rotation, glabella tap, tremor, salivation, and akathisia). Each item will be rated on a 5-point Likert scale, with 0 representing absence of symptoms and 4 representing a severe condition. The SAS total score is the sum of ratings over all 10 items, with possible total scores ranging from 0 to 40. The SAS total score will be un-evaluable if less than 8 of the 10 items are recorded. If 8 or 9 of the 10 items are recorded, the total score will be the mean of the recorded items multiplied by 10 and then rounded to the first decimal place.

11.6 Abnormal Involuntary Movement Scale (AIMS)

The AIMS is a 12-item scale. The first 10 items are rated on a Likert 5-point scale from 0 to 4 (0 = best, 4 = worst). An item score of 0, depending on specific item, means either “no abnormal Involuntary movement (AIM)” or “no incapacitation due to AIM” or “no awareness of AIM”. An item score of 4 means either “severe AIM” or “severe incapacitation due to AIM” or “being aware of, and severe distress caused by AIM”.

Items 11 and 12 are related to dental status, taking dichotomous response: 0 = no and 1 = yes. The AIMS total score is the sum of the ratings over the first 7 items, with possible total score ranging from 0 to 28. The AIMS total Score will be un-evaluable if less than 6 of the first 7 items are recorded. If 6 of the items are recorded, then the total score will be the mean of the recorded items multiplied by 7 and then rounded to the first decimal place.

11.7 Barnes Akathisia Rating Scale (BARS)

The BARS consists of 4 items related to akathisia as follows. Item 1: objective observation of akathisia by the investigator; Item 2: subjective feelings of restlessness by the subject; Item 3: subjective distress due to akathisia; and Item 4: global clinical assessment of akathisia. The first 3 items will be rated on a 4-point Likert scale from 0 to 3, with 0 representing absence of symptoms and 3 representing a severe condition. The BARS global clinical assessment score refers to the ratings from the fourth item Global Clinical Assessment of akathisia, which is a 6-point Likert scale from 0 to 5, with 0 representing absence of symptoms and 5 representing severe akathisia. Thus, the BARS global clinical assessment score ranges 0 to 5.

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

Suicidality will be monitored during the trial using the C-SSRS (Child Version). Suicidality data will be collected at all the visits (including all the scheduled visits, the ET visit, and the follow-up visit). This trial will use the “baseline/screening” and “Since Last Visit” versions of the scale. The “baseline/screening” version, which assesses the lifetime experience of the subject with suicide events and suicidal ideation and the occurrence of suicide events and/or ideation within a specified period prior to entry into the trial, will be completed for all subjects at screening to determine eligibility. Any subject with active suicidal ideation within the last 3 months, suicidal behaviors within the last year, or who in the clinical judgment of the investigator presents a serious risk of suicide should be excluded from the trial. The “Since Last Visit” C-SSRS form will be completed at all other visits.

Suicidal ideation has 5 types of ideation with increasing severity from Type 1 to Type 5. A subject can report or can be rated with multiple types of ideation at the visit level. The most severe type of ideation on the subject at the visit level will be used for data summary purpose. The variable Intensity of Suicidal Ideation also has 4 other dimensions, which are the frequency of ideation, duration of ideation, controllability and

deterrents, each of which is rated on a scale of 0 to 5 or 1 to 5, with higher integers representing more worse clinical outcomes.

12 References

1 Clinical Protocol, OPDC Protocol No: 331-201-00148, Version 6.0. A Phase 3, Multicenter, Randomized, Double-Blind, Placebo-controlled Trial of Brexpiprazole in Treatment of Children and Adolescents with Irritability Associated with Autism Spectrum Disorder. 07 Jul 2022.

2 ICH E9 (R1), Estimands and Sensitivity Analysis in Clinical Trials, Step 2 Draft Guideline, 2017.

3 Siddiqui O, Hung JHM, O'Neill R. MMRM vs. LOCF: A comprehensive comparison based on simulation study and 25 NDA datasets. J Biopharmaceutical Stats. 2009; 19(2):227-46.

4 Little RJA. Modeling the drop-out mechanism in repeated measures studies. J Am Stat Assoc. 1995; 90:1112-21.

5 Diggle P, Kenward MG. Informative drop-out in longitudinal data analysis. Applied Statistics. 1994; 43:49-93.

6 Little RJA. Pattern-mixture models for multivariate incomplete data. J Am Stat Assoc. 1993; 88:125-34.

7 Hedeker D, Gibbons RD. Application of random effects pattern-mixture models for missing data in longitudinal studies. Psychological Methods. 1997; 2:64-78.

8 Ali MW, Siddiqui O. Multiple imputation compared with some information dropout procedures in the estimation and comparison of rates of change in longitudinal clinical trials with dropouts. J Biopharmaceutical Stats. 2000;10(2):165-81.

9 Wu MC, Bailey KR. Estimation and comparison of changes in the presence of informative right censoring: Conditional linear model. Biometrics. 1989; 45:939-55.

10 Van Elteren, PH. On the combination of independent two sample tests of Wilcoxon. Bull Int Stat Inst. 1960; 37:351-61.

11 Mehrotra D, Li X, Liu J, Lu K. Analysis of Longitudinal Clinical Trials with Missing Data Using Multiple Imputation in Conjunction with Robust Regression. Biometrics 2012; 68:1250-1259.

12 Cortina, J. M., & Nouri, H. (2000). *Effect Size for ANOVA Designs*. Thousand Oaks, Calif.: Sage Publications.

13 Gould, A. L., & Shih, W.J. Sample-Size Re-estimation without Unblinding for Normally Distributed Data with Unknown Variance. Communication in Statistics - Theory and Methods. 1992; 21(10): 2833-2853.

14 Hughes HK, Kahl LK. The Johns Hopkins Hospital: The Harriet Lane Handbook. 21st edition. 2015.

15 Flynn JT, Kaelber DC, Baker-Smith CM, Blowey D, Carroll AE, Daniels SR, et al. Clinical practice guideline for screening and management of high blood pressure in children and adolescents. Pediatrics. 2017;140(3).

16 Soghier L, Pham K, Rooney S. Reference range values for pediatric care. The American Academy of Pediatrics. 2014.

17 Covance. Central laboratory services manual; Protocol 331-201-00148. Created 28 Sep 2019, revised 11 Dec 2019.

SAP 331-201-00148

[18] American Diabetes Association. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2019. Diabetes Care. 2019;42(Suppl 1): S13-S28.

SAP 331-201-00148

13 Potential Clinical Relevance Criteria from Protocol

Appendix 1 Criteria for Identifying Vital Signs Outside of Normal Range Values and of Potential Clinical Relevance[a]

Variable	Criterion Value	Change Relative to Baseline
Heart Rate[14]	< 60 bpm or > 110 bpm	Increase or decrease of ≥ 15 bpm
Systolic Blood Pressure[15]		
Preschooler (5 y)	< 80 mmHg or > 115 mmHg	Increase or decrease of ≥ 20 mmHg
School-age (6-9 y)	< 85 mmHg or > 115 mmHg	Increase or decrease of ≥ 20 mmHg
Preadolescent (10-12 y)	< 90 mmHg or > 120 mmHg	Increase or decrease of ≥ 15 mmHg
Adolescent (13-17 y)	< 90 mmHg or > 120 mmHg	Increase or decrease of ≥ 15 mmHg
Diastolic Blood Pressure[15]		
Preschooler (5 y)	< 45 mmHg or > 80 mmHg	Increase or decrease of ≥ 15 mmHg
School-age (6-9 y)	< 50 mmHg or > 80 mmHg	Increase or decrease of ≥ 15 mmHg
Preadolescent (10-12 y)	< 60 mmHg or > 80 mmHg	Increase or decrease of ≥ 15 mmHg
Adolescent ≥ 13 y	< 60 mmHg or > 85 mmHg	Increase or decrease of ≥ 15 mmHg

[a]The criterion value and change relative to baseline represented in this table are intended to identify on-treatment values outside of normal changes and that could potentially be clinically relevant. In order to be identified as potentially clinically relevant, the on-treatment value must meet the "Criterion Value" and also represent a change from the subject's baseline value of at least the magnitude shown in the "Change Relative to Baseline" column. When evaluating these values the emotional state of the subject must be taken into consideration (eg, crying, screaming) and documented as applicable.

Adapted information.[14,15]

SAP 331-201-00148

Appendix 2 Criteria for Identifying Laboratory Values of Potential Clinical Relevance[14,16,17,18]

Laboratory Tests	Criteria (Normal Ranges) for Subjects 5 to 17 Years of Age
Chemistry	
AST	≥ 2 × ULN
ALT	≥ 2 × ULN
ALP	≥ 2 × ULN
Bicarbonate	
< 6 y	< 17.0 or > 26.0 mEq/L (17.0-26.0 mEq/L)
6 y to 12 y	< 19.0 or > 27.0 mEq/L (19.0-27.0 mEq/L)
≥ 12 y	< 19.3 or > 29.3 mEq/L (19.3-29.3 mEq/L)
BUN	≥ 24 mg/dL (≤ 4 mg/dL or ≥ 24 mg/dL)
Creatinine ≤ 12 y	≥ 0.7 mg/dL (≤ 0.2 mg/dL or ≥ 0.7 mg/dL)
≥13 y	≥ 1.1 mg/dL (≤ 0.3 mg/dL or ≥ 1.1 mg/dL)
Uric Acid	
≤ 12 y	≥ 6.7 mg/dL (≤ 1.6 mg/dL or ≥ 6.7 mg/dL)
≥ 13 y	≥ 8.2 mg/dL (≤ 2.2 mg/dL or ≥ 8.2 mg/dL)
Bilirubin (total)	≥ 1.6 mg/dL (≤ 0.2 mg/dL or ≥ 1.6 mg/dL)
CPK	≥ 2 × ULN
Prolactin	
≤ 12 y	≥ 21.00 ng/dL (≤ 2.63 ng/dL or ≥ 21.00 ng/dL)
≥ 13 y	≥ 39.00 ng/dL (≤ 2.52 ng/dL or ≥ 39.00 ng/dL)
Hematology	
Hematocrit	
≤ 12 y	≤ 33 % (≤ 33 % or ≥ 44%)
≥ 13 y	≤ 34 % (≤ 34 % or ≥ 54%)
Hemoglobin	
≤ 12 y	≤11.2 g/dL (≤ 11.2 g/dL or ≥ 15.5 g/dL)
≥ 13 y	≤ 11.6 g/dL (≤ 11.6 g/dL or ≥ 18.1 g/dL)
White blood count	≤ 4.35 × 10^3/uL (≤ 4.35 × 10^3/uL or ≥ 13.65 × 10^3/uL)
Eosinophils	
≤ 12 y	≥ 4.8%
≥ 13 y	≥ 4.1%
Neutrophils	≤ 40.5% (≤ 40.5% or ≥ 75.0 %)
Absolute neutrophil count	
≤ 12 y	≤ 1.00 × 10^3/uL or ≥ 9.00 × 10^3/uL
≥ 13 y	≤ 1.35 × 10^3/uL or ≥ 8.15 × 10^3/uL
Platelet count	
≤ 12 y	≤ 130 × 10^3/uL (≤ 130 × 10^3/uL or ≥ 570 × 10^3/uL)
≥ 13 y	≤ 140 × 10^3/uL (≤ 140 × 10^3/uL or ≥ 400 × 10^3/uL)
Urinalysis	
Protein	Change from baseline
Glucose	Presence
Additional Criteria	
Chloride	≤ 94 mEq/L or ≥ 112 mEq/L
HbA1c	≥ 5.7%
ACTH	< 7.2 pg/mL - > 63.3 pg/mL
Cortisol	AM: 6.7 ug/dL - 22.60 ug/dL PM: < 10 ug/dL
Potassium	≤ 3.3 mEq/L or ≥ 5.2 mEq/L
Sodium	≤ 132 mEq/L or ≥ 148 mEq/L

SAP 331-201-00148

Laboratory Tests	Criteria (Normal Ranges) for Subjects 5 to 17 Years of Age
Calcium	≤ 8.3 mg/dL or ≥ 10.9 mg/dL
Glucose	
Fasting	≥ 100 mg/dL (≥ 70 mg/dL and ≤ 100 mg/dL)
Nonfasting	≥ 139 mg/dL (≥ 70 mg/dL and ≤ 139 mg/dL)
Total Cholesterol, Fasting	
≤ 12 y	≥ 217 mg/dL (≤ 97 mg/dL or ≥ 217 mg/dL)
≥ 13 y	≥ 217 mg/dL (≤ 124 mg/dL or ≥ 217 mg/dL)
LDL Cholesterol, Fasting	≥ 130 mg/dL
HDL Cholesterol, Fasting	
≤ 12 y	≤ 34 mg/dL (≤ 34 mg/dL or ≥ 75 mg/dL)
≥ 13 y	≤ 30 mg/dL (≤ 30 mg/dL or ≥ 74 mg/dL)
Triglycerides, Fasting	
≤ 12 y	≥ 131 mg/dL (≤ 30 mg/dL or ≥ 131 mg/dL)
≥ 13 y	≥ 148 mg/dL (≤ 32 mg/dL or ≥ 148 mg/dL)
TSH	
≤ 12 y	≤ 0.34 mIU/mL or ≥ 5.40 mIU/mL
≥ 13 y	≤ 0.34 mIU/mL or ≥ 5.60 mIU/mL
Free T4	
≤ 12 y	≤ 9 pmol/L or ≥ 30 pmol/L
≥ 13 y	≤ 10 pmol/L or ≥ 24 pmol/L
PT	≥ 12.3 sec (≤ 9.7 sec or ≥ 12.3 sec)
aPTT	≥ 29.4 sec (≤ 21.9 sec or ≥ 29.4 sec)
INR	
Not taking anticoagulants	≥ 1.2 (≤ 0.8 or ≥ 1.2)
Taking anticoagulants	≥ 3.0 (≤ 2.0 or ≥ 3.0)

The recommended criteria represented in this table are intended to identify on-treatment outside of normal values that could potentially be clinically relevant. Variations based on local laboratory ranges may need to be considered.

Adapted information.[14,16,17,18]

SAP 331-201-00148

Appendix 3 Criteria for Identifying ECG Measurement of Potential Clinical Relevance

Variable	Criterion Value[a]	Change Relative to Baseline[a]
Rhythm		
Sinus tachycardia[b]	≥ 110 bpm	increase of ≥ 15 bpm
Sinus bradycardia[c]	≤ 60 bpm	decrease of ≥ 15 bpm
Supraventricular premature beat	all	not present → present
Ventricular premature beat	all	not present → present
Supraventricular tachycardia	all	not present → present
Ventricular tachycardia	all	not present → present
Atrial fibrillation	all	not present → present
Atrial flutter	all	not present → present
Conduction		
1° atrioventricular block	PR ≥ 200 msec	increase of ≥ 50 msec
2° atrioventricular block	all	not present → present
3° atrioventricular block	all	not present → present
Left bundle-branch block	all	not present → present
Right bundle-branch block	all	not present → present
Pre-excitation syndrome	all	not present → present
Other intraventricular conduction block[d]	QRS ≥ 120 msec	increase of ≥ 20 msec
Infarction		
Acute or subacute	all	not present → present
Old	all	not present → present ≥ 12 weeks post-trial entry
ST/T Morphological		
Myocardial Ischemia	all	not present → present
Symmetrical T-wave inversion	all	not present → present
Increase in QTc	QTcF ≥ 450 msec for males, ≥ 470 msec for females	increase of 60 msec from baseline

[a]The criterion value and change relative to baseline represented in this table are intended to identify on-treatment values outside of normal changes and that could potentially be clinically relevant. In order to be identified as potentially clinically relevant, the on-treatment value must meet the "Criterion Value" and also represent a change from the subject's baseline value of at least the magnitude shown in the "Change Relative to Baseline" column. When evaluating these values the emotional state of the subject must be taken into consideration (eg, crying, screaming) and documented as applicable.

[b]No current diagnosis of supraventricular tachycardia, ventricular tachycardia, atrial fibrillation, atrial flutter, or other rhythm abnormality.

[c]No current diagnosis of atrial fibrillation, atrial flutter, or other rhythm abnormality.

[d]No current diagnosis of left bundle branch block or right bundle branch block.

SAP 331-201-00148

14 Proposed List of Summary Tables

CT-1.1 Subject Disposition
CT-1.2 Subject Disposition by Center
CT-1.3 Subject Completion Rates by Week (Randomized Sample)
CT-1.4.1 Enrollment by Age Group (Randomized Sample)
CT-1.4.2 Enrollment by Body Weight Group (Randomized Sample)
CT-1.5 Summary of Proportion of the Face-to-Face and Virtual Visits by Study Week (Randomized Sample)
CT-2.1 Reasons for Discontinuation (Randomized Sample)
CT-2.2 Reasons for Discontinuation due to COVID-19 (Randomized Sample)
CT-3.1 Demographic Characteristics (Randomized Sample)
CT-3.2.1 Medical History (Randomized Sample)
CT-3.2.2 Psychiatric History (Randomized Sample)
CT-3.3 Baseline Psychiatric Scale Evaluation (Randomized Sample)
CT-4.1.1 Concomitant Medications: Medications Taken Prior to Start of Study Therapy (Safety Sample)
CT-4.1.2 Concomitant Medications: Medications Taken During Study Therapy (Safety Sample)
CT-4.1.3 Concomitant Medications: Medications Taken Post Study Therapy (Safety Sample)
CT-5.1 Summary of Efficacy Results at Week 8 of the Double-Blind Treatment Period (Efficacy Sample)
CT-5.2.1 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC-Irritability Subscale Score - MMRM, UN (Efficacy Sample)
CT-5.2.2 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC-Irritability Subscale Score - MMRM, TOEPH Empirical (Efficacy Sample)
CT-5.2.3 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC-Irritability Subscale Score - MMRM, ARH1 Empirical (Efficacy Sample)
CT-5.2.4 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC-Irritability Subscale Score - MMRM, CSH Empirical (Efficacy Sample)
CT-5.2.5 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC-Irritability Subscale Score - MMRM, DDFM=SATTERTHWAITE (Efficacy Sample)
CT-5.2.6.1 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC-Irritability Subscale Score - ANCOVA, LOCF (Efficacy Sample)
CT-5.2.6.2 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC-Irritability Subscale Score - ANCOVA, OC (Efficacy Sample)
CT-5.2.7.1 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC-Irritability Subscale Score Excluding Remote Assessments Which Should Be In-clinic per Protocol - MMRM, UN (Efficacy Sample)
CT-5.2.7.2 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC-Irritability Subscale Score Non-COVID Data Set - MMRM, UN (Efficacy Sample)
CT-5.2.7.3 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC-Irritability Subscale Score Non-COVID Sample - MMRM, UN (Efficacy Sample)
CT-5.2.7.4 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC-Irritability Subscale Score in Subjects Who Completed or Discontinued on or After March 13, 2020 - MMRM, UN (Efficacy Sample)
CT-5.2.7.5 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC-Irritability Subscale Score in Subjects Who Completed or Discontinued Before March 13, 2020 - MMRM, UN (Efficacy Sample)
CT-5.3.1 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in Clinical Global Impression - Severity (CGI-S) Score - MMRM, UN (Efficacy Sample)
CT-5.3.2 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in Clinical Global Impression - Severity (CGI-S) Score - ANCOVA, LOCF (Efficacy Sample)
CT-5.3.3 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in Clinical Global Impression - Severity (CGI-S) Score - ANCOVA, OC (Efficacy Sample)

CT-5.3.4.1 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in Clinical Global Impression - Severity (CGI-S) Score Excluding Remote Assessments Which Should Be In-clinic per Protocol - MMRM, UN (Efficacy Sample)
CT-5.3.4.2 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in Clinical Global Impression - Severity (CGI-S) Score Non-COVID Data Set - MMRM, UN (Efficacy Sample)
CT-5.3.4.3 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in Clinical Global Impression - Severity (CGI-S) Score Non-COVID Sample - MMRM, UN (Efficacy Sample)
CT-5.3.4.4 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in Clinical Global Impression - Severity (CGI-S) Score in Subjects Who Completed or Discontinued on or After March 13, 2020 - MMRM, UN (Efficacy Sample)
CT-5.3.4.5 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in Clinical Global Impression - Severity (CGI-S) Score in Subjects Who Completed or Discontinued Before March 13, 2020 - MMRM, UN (Efficacy Sample)
CT-5.4.1 Summary of Proportion of Responders defined as Having at Least 25% Reduction in ABC-Irritability Subscale Score and at Least 1 Point Reduction in CGI-S Score during the Double-Blind Treatment Period by Study Week - LOCF (Efficacy Sample)
CT-5.4.2 Summary of Proportion of Responders defined as Having at Least 25% Reduction in ABC-Irritability Subscale Score and at Least 1 Point Reduction in CGI-S Score during the Double-Blind Treatment Period by Study Week - OC (Efficacy Sample)
CT-5.5 Summary of Mean Change from Baseline to the Endpoint (Week 8) of the Double-Blind Treatment Period in Pediatric Anxiety Rating Scale (PARS) Total Score - ANCOVA, LOCF (Efficacy Sample)
CT-5.6.1 Summary of Mean Change from Baseline to the Endpoint (Week 8) of the Double-Blind Treatment Period in PedsQL Generic Core Scale Average Score - ANCOVA, LOCF (Efficacy Sample)
CT-5.6.2 Summary of Mean Change from Baseline to the Endpoint (Week 8) of the Double-Blind Treatment Period in PedsQL Psychosocial Health Average Score - ANCOVA, LOCF (Efficacy Sample)
CT-5.6.3 Summary of Mean Change from Baseline to the Endpoint (Week 8) of the Double-Blind Treatment Period in PedsQL Physical Health Average Score - ANCOVA, LOCF (Efficacy Sample)
CT-5.6.4 Summary of Mean Change from Baseline to the Endpoint (Week 8) of the Double-Blind Treatment Period in PedsQL Emotional Function Average Score - ANCOVA, LOCF (Efficacy Sample)
CT-5.6.5 Summary of Mean Change from Baseline to the Endpoint (Week 8) of the Double-Blind Treatment Period in PedsQL Social Function Average Score - ANCOVA, LOCF (Efficacy Sample)
CT-5.6.6 Summary of Mean Change from Baseline to the Endpoint (Week 8) of the Double-Blind Treatment Period in PedsQL School Function Average Score - ANCOVA, LOCF (Efficacy Sample)
CT-5.6.7 Summary of Mean Change from Baseline to the Endpoint (Week 8) of the Double-Blind Treatment Period in PedsQL Family Impact Module Average Score - ANCOVA, LOCF (Efficacy Sample)
CT-5.6.8 Summary of Mean Change from Baseline to the Endpoint (Week 8) of the Double-Blind Treatment Period in PedsQL Parent HRQL Average Score - ANCOVA, LOCF (Efficacy Sample)
CT-5.6.9 Summary of Mean Change from Baseline to the Endpoint (Week 8) of the Double-Blind Treatment Period in PedsQL Family Functioning Average Score - ANCOVA, LOCF (Efficacy Sample)
CT-5.7.1 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC- Social Withdrawal Subscale Score - MMRM, UN (Efficacy Sample)
CT-5.7.2 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC- Stereotypic Behavior Subscale Score - MMRM, UN (Efficacy Sample)
CT-5.7.3 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC- Hyperactivity/Noncompliance Subscale Score - MMRM, UN (Efficacy Sample)
CT-5.7.4 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC- Inappropriate Speech Subscale Score - MMRM, UN (Efficacy Sample)

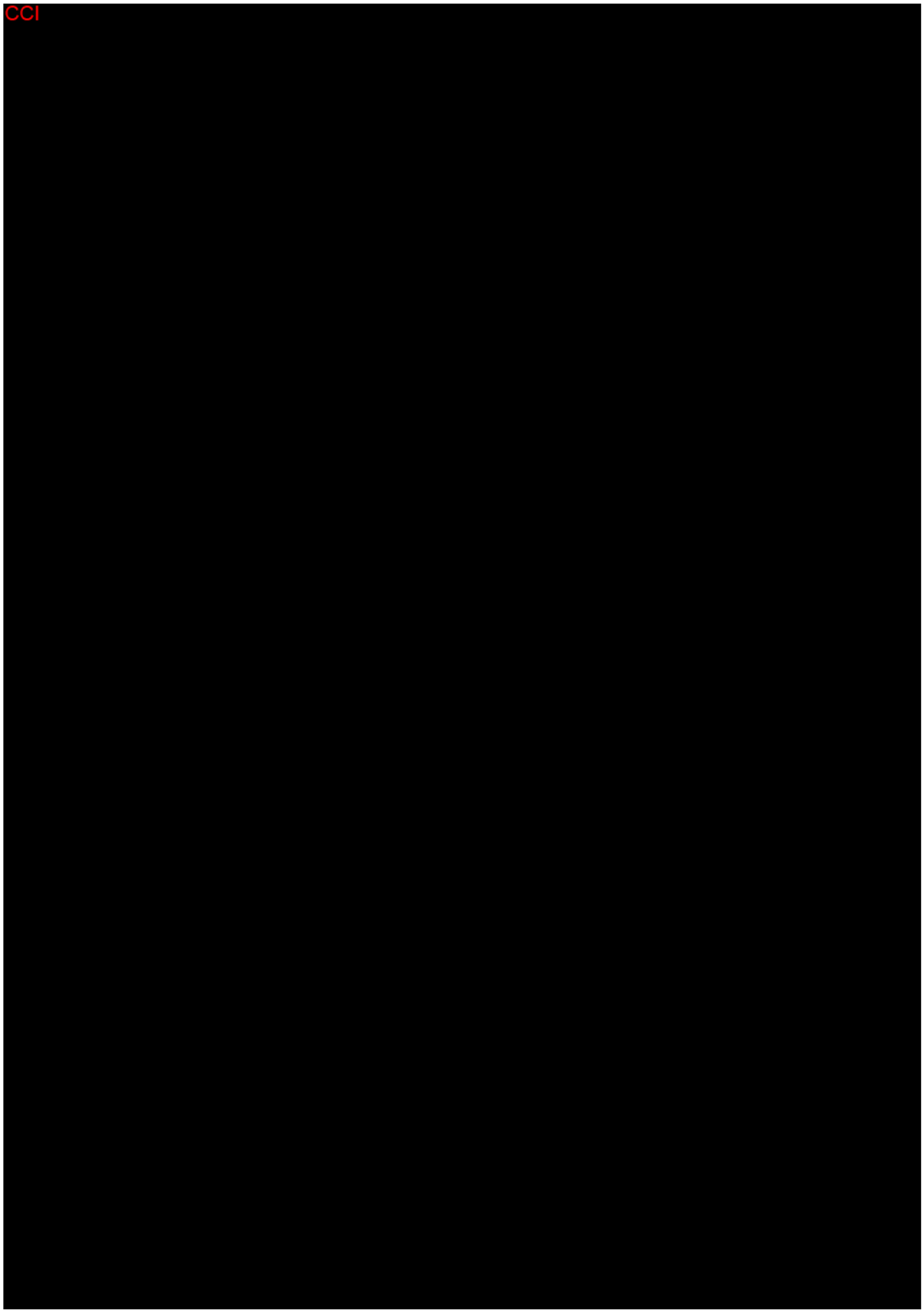
CT-5.8.1 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC-Irritability Subscale Score - MI, VAN ELTEREN TEST (Efficacy Sample)
CT-5.8.2 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC-Irritability Subscale Score - GEE (Efficacy Sample)
CT-5.8.3 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC-Irritability Subscale Score - WGEE (Efficacy Sample)
CT-5.9.1 Sensitivity Analysis of MNAR using Pattern Mixture Model with Multiple Imputation (Delta Method 1) in ABC-Irritability Subscale Score Assume All Dropouts as MNAR (Efficacy Sample)
CT-5.9.2 Sensitivity Analysis of MNAR using Pattern Mixture Model with Multiple Imputation (Delta Method 2) in ABC-Irritability Subscale Score Assume All Dropouts as MNAR (Efficacy Sample)
CT-5.9.3 Sensitivity Analysis of MNAR using Pattern Mixture Model with Multiple Imputation (Delta Method 1) in ABC-Irritability Subscale Score Dropout due to AE or Lack of Efficacy (LOE) or Withdrawal of Consent as MNAR (Efficacy Sample)
CT-5.9.4 Sensitivity Analysis of MNAR using Pattern Mixture Model with Multiple Imputation (Delta Method 2) in ABC-Irritability Subscale Score Dropout due to AE or Lack of Efficacy (LOE) or Withdrawal of Consent as MNAR (Efficacy Sample)
CT-5.9.5 Sensitivity Analysis of MNAR using Pattern Mixture Model with Multiple Imputation (Delta Method 1) in ABC-Irritability Subscale Score Dropouts due to AE or LOE as MNAR (Efficacy Sample)
CT-5.9.6 Sensitivity Analysis of MNAR using Pattern Mixture Model with Multiple Imputation (Delta Method 2) in ABC-Irritability Subscale Score Dropouts due to AE or LOE as MNAR (Efficacy Sample)
CT-5.9.7 Sensitivity Analysis of MNAR using Placebo Based Imputation in ABC-Irritability Total Score (Efficacy Sample)
CT-5.9.8 Sensitivity Analysis of MNAR using Pattern Mixture Model with Multiple Imputation (Delta Method 1) in Clinical Global Impression - Severity (CGI-S) Score Assume All Dropouts as MNAR (Efficacy Sample)
CT-5.9.9 Sensitivity Analysis of MNAR using Pattern Mixture Model with Multiple Imputation (Delta Method 2) in Clinical Global Impression - Severity (CGI-S) Score Assume All Dropouts as MNAR (Efficacy Sample)
CT-5.9.10 Sensitivity Analysis of MNAR using Pattern Mixture Model with Multiple Imputation (Delta Method 1) in Clinical Global Impression - Severity (CGI-S) Score Dropout due to AE or Lack of Efficacy (LOE) or Withdrawal of Consent as MNAR (Efficacy Sample)
CT-5.9.11 Sensitivity Analysis of MNAR using Pattern Mixture Model with Multiple Imputation (Delta Method 2) in Clinical Global Impression - Severity (CGI-S) Score Dropout due to AE or Lack of Efficacy (LOE) or Withdrawal of Consent as MNAR (Efficacy Sample)
CT-5.9.12 Sensitivity Analysis of MNAR using Pattern Mixture Model with Multiple Imputation (Delta Method 1) in Clinical Global Impression - Severity (CGI-S) Score Dropouts due to AE or LOE as MNAR (Efficacy Sample)
CT-5.9.13 Sensitivity Analysis of MNAR using Pattern Mixture Model with Multiple Imputation (Delta Method 2) in Clinical Global Impression - Severity (CGI-S) Score Dropouts due to AE or LOE as MNAR (Efficacy Sample)
CT-5.9.14 Sensitivity Analysis of MNAR using Placebo Based Imputation in Clinical Global Impression - Severity (CGI-S) Score (Efficacy Sample)
CT-5.9.15 Sensitivity Analysis of MNAR Using Model Based Methods: Shared Parameter Model and Random Coefficient Pattern Mixture Model - OC (Efficacy Sample)
CT-6.1.1 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC-Irritability Subscale Score in the Subgroup of Females - MMRM (Efficacy Sample)
CT-6.1.2 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC-Irritability Subscale Score in the Subgroup of Males - MMRM (Efficacy Sample)
CT-6.2.1 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC-Irritability Subscale Score in the Subgroup of Whites - MMRM (Efficacy Sample)
CT-6.2.2 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC-Irritability Subscale Score in the Subgroup of All Other Races - MMRM (Efficacy Sample)

CT-6.3.1 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC-Irritability Subscale Score in Subjects ages 5 to 12 - MMRM (Efficacy Sample)
CT-6.3.2 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC-Irritability Subscale Score in Subjects ages 13 to 17 - MMRM (Efficacy Sample)
CT-6.4.1 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC-Irritability Subscale Score in Subjects with Body Weight < 50 kg - MMRM (Efficacy Sample)
CT-6.4.2 Summary of Mean Change from Baseline during the Double-Blind Treatment Period by Study Week in ABC-Irritability Subscale Score in Subjects with Body Weight ≥ 50 kg - MMRM (Efficacy Sample)
CT-7.1 Extent of Exposure to Study Medication in the Double-Blind Treatment Period (Safety Sample)
CT-7.2 Number and Percentage of Subjects Having Received Study Medication and Mean and Range of Average Daily Dose in the Double-Blind Treatment Period (Safety Sample)
CT-8.1 Adverse Events (All Causalities) (Safety Sample)
CT-8.2.1 Incidence of TEAEs by System Organ Class and MedDRA Preferred Term (Safety Sample)
CT-8.2.2 Incidence of TEAEs by System Organ Class, MedDRA Preferred Term and Severity (Safety Sample)
CT-8.2.3 Incidence of TEAEs with Rate being at least 5% in Brexpiprazole Group and Greater Than Placebo by System Organ Class and MedDRA Preferred Term (Safety Sample)
CT-8.2.4 Incidence of TEAEs by System Organ Class and MedDRA Preferred Term, by Sex (Safety Sample)
CT-8.2.5 Incidence of TEAEs by System Organ Class and MedDRA Preferred Term, by Race (Safety Sample)
CT-8.2.6 Incidence of TEAEs by System Organ Class and MedDRA Preferred Term, by Age Group (Safety Sample)
CT-8.2.7 Incidence of TEAEs by System Organ Class and MedDRA Preferred Term, by Body Weight Group (Safety Sample)
CT-8.2.8 Incidence of Non-Serious TEAE with Rate being at Least 5% in Brexpiprazole Group and Greater Than Placebo by System Organ Class and MedDRA Preferred Term (Safety Sample)
CT-8.2.9 Incidence of TEAEs with Rate being at Least 2% in Brexpiprazole Group and Greater Than Placebo by System Organ Class and MedDRA Preferred Term (Safety Sample)
CT-8.2.10 Occurrence (Number of Events) of Serious TEAEs System Organ Class and MedDRA Preferred Term (Safety Sample)
CT-8.2.11 Occurrence (Number of Events) of Potentially Drug-Related Serious TEAEs by System Organ Class and MedDRA Preferred Term (Safety Sample)
CT-8.2.12 Occurrence (Number of Events) of Non-Serious TEAEs by System Organ Class and MedDRA Preferred Term (Safety Sample)
CT-8.3.1 Incidence of Potentially Drug-Related TEAEs by System Organ Class and MedDRA Preferred Term (Safety Sample)
CT-8.3.2 Incidence of Potentially Drug-Related TEAEs by System Organ Class, MedDRA Preferred Term and Severity (Safety Sample)
CT-8.4 Incidence of Deaths Due to TEAEs by System Organ Class and MedDRA Preferred Term (Safety Sample)
CT-8.5.1 Incidence of Serious TEAEs by System Organ Class and MedDRA Preferred Term (Safety Sample)
CT-8.5.2 Incidence of Serious TEAEs by System Organ Class, MedDRA Preferred Term and Severity (Safety Sample)
CT-8.6.1 Incidence of TEAEs Resulting in Discontinuation from Study Medication by System Organ Class and MedDRA Preferred Term (Safety Sample)
CT-8.6.2 Incidence of TEAEs Resulting in Discontinuation from Study Medication by System Organ Class, MedDRA Preferred Term and Severity (Safety Sample)
CT-8.7 Incidence of TE EPS-related AEs by EPS Category and MedDRA Preferred Term (Safety Sample)
CT-8.8 Incidence of Onset of Akathisia Adverse Event by Week during Double-Blind Treatment Period (Safety Sample)
CT-8.9.1 Listing of TEAEs that Are Related or Probably Related to COVID-19 (Safety Sample)

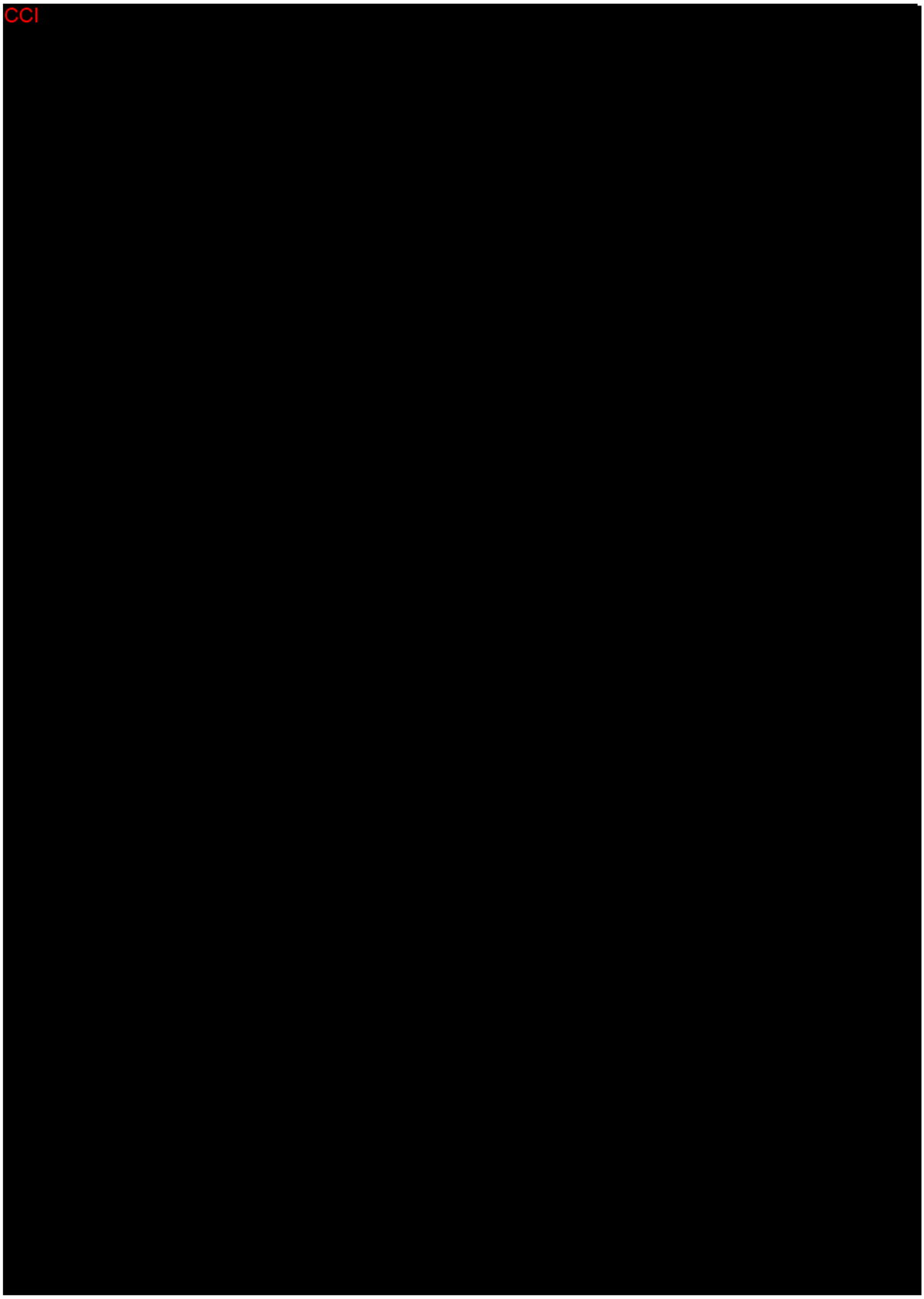
CT-8.9.2 Listing of Serious TEAEs or Deaths that Are Related or Probably Related to COVID-19 (Safety Sample)
CT-9.1 Listing of Deaths (Safety Sample)
CT-9.2 Listing of Serious Adverse Events (Safety Sample)
CT-9.3 Listing of Discontinuations due to Adverse Events (Safety Sample)
CT-10.1 Criteria for Laboratory Test Values with Potentially Clinical Relevance
CT-10.2.1 Listing of Laboratory Test Values with Potentially Clinical Relevance by Subject (Safety Sample)
CT-10.2.2 Listing of Laboratory Test Values with Potentially Clinical Relevance by Test (Safety Sample)
CT-10.2.3 Incidence of Laboratory Test Values with Potentially Clinical Relevance (Safety Sample)
CT-10.3.1 Summary of Change from Baseline in Clinical Laboratory Test Results - Serum Chemistry (Safety Sample)
CT-10.3.2 Summary of Change from Baseline in Clinical Laboratory Test Results - Hematology (Safety Sample)
CT-10.3.3 Summary of Change from Baseline in Clinical Laboratory Test Results - Urinalysis (Safety Sample)
CT-10.3.4 Summary of Change from Baseline in Clinical Laboratory Test Results - Prolactin, by Sex (Safety Sample)
CT-10.3.5 Summary of Change from Baseline in Clinical Laboratory Test Results - Other Tests (Safety Sample)
CT-10.4.1 Incidence of Potential Serious Hepatotoxicity (Safety Sample)
CT-10.4.2 Listing of Potential Serious Hepatotoxicity (Safety Sample)
CT-10.5.1 Incidence of Laboratory Test Values with Potential Clinical Relevance - Serum Prolactin (Safety Sample)
CT-10.5.2 Listing of Laboratory Test Values with Potential Clinical Relevance - Serum Prolactin (Safety Sample)
CT-11.1 Criteria for Potentially Clinically Relevant Abnormalities in Vital Signs
CT-11.2.1 Listing of Potentially Clinically Relevant Abnormalities in Vital Signs (Safety Sample)
CT-11.2.2 Incidence of Potentially Clinically Relevant Abnormalities in Vital Signs (Safety Sample)
CT-11.2.3 Summary of Change from Baseline in Vital Signs by Study Week (Safety Sample)
CT-12.1.1 Criteria for Potentially Clinically Relevant Abnormalities in ECG Evaluations
CT-12.1.2 ECG Diagnosis Mapping for Potentially Clinically Relevant Electrocardiogram Abnormalities (Safety Sample)
CT-12.2.1 Listing of Potentially Clinically Relevant Abnormalities in ECG Evaluations (Safety Sample)
CT-12.2.2 Incidence of Potentially Clinically Relevant Changes in ECG Evaluations (Safety Sample)
CT-12.2.3 Summary of Mean Change from Baseline in Electrocardiogram Results (Safety Sample)
CT-12.3.1 Listing of Categorical Changes in QT/QTc (Safety Sample)
CT-12.3.2 Incidence of Categorical Changes in QT/QTc (Safety Sample)
CT-13.1 Summary of Mean Change from Baseline to the Endpoint (Week 8) of the Double-Blind Treatment Period in Simpson-Angus Scale (SAS) Total Score (Safety Sample)
CT-13.2 Summary of Mean Change from Baseline to the Endpoint (Week 8) of the Double-Blind Treatment Period in Abnormal Involuntary Movement Scale (AIMS) Total Score and Item Score on Item 8, 9 and 10 (Safety Sample)
CT-13.3.1 Summary of Mean Change from Baseline to the Endpoint (Week 8) of the Double-Blind Treatment Period in Barnes Akathisia Rating Scale (BARS) Global Clinical Assessment Score (Safety Sample)
CT-13.3.2 Incidence of Barnes Akathisia Rating Scale (BARS) Global Clinical Assessment of Akathisia During the Double-Blind Treatment Period (Safety Sample)
CT-14.1.1 Columbia-Suicide Severity Rating Scale(C-SSRS) during the Treatment and the Follow-up Period, Suicidality (Safety Sample)
CT-14.1.2 Columbia-Suicide Severity Rating Scale(C-SSRS) during the Treatment and the Follow-up Period, Suicidal Ideation by Type (Safety Sample)
CT-14.1.3 Columbia-Suicide Severity Rating Scale(C-SSRS) during the Treatment and the Follow-up Period, Suicidal Behavior by Type (Safety Sample)

CT-14.1.4 Columbia-Suicide Severity Rating Scale(C-SSRS) during the Treatment and the Follow-up Period, Treatment Emergent Suicidal Ideation and Behavior (Safety Sample)
CT-14.1.5 Columbia-Suicide Severity Rating Scale (C-SSRS) - Listing of Treatment Emergent Suicidal Ideation during the Treatment and the Follow-up Period (Safety Sample)
CT-14.1.6 Columbia-Suicide Severity Rating Scale (C-SSRS) - Listing of Treatment Emergent Suicidal Behavior during the Treatment and the Follow-up Period (Safety Sample)
CT-14.1.7 Columbia-Suicide Severity Rating Scale (C-SSRS) - Listing of Treatment Emergent Serious Suicidal Ideation during the Treatment and the Follow-up Period (Safety Sample)
CT-14.1.8 Columbia-Suicide Severity Rating Scale (C-SSRS) - Listing of Worsening Suicidal Ideation during the Treatment and the Follow-up Period (Safety Sample)
CT-15.1.1 Summary of Mean Change from Baseline to the Endpoint (Week 8) of the Double-Blind Treatment Period in Body Weight (Unit: kg) (Safety Sample)
CT-15.1.2 Summary of Proportion of Patients with Potentially Clinically Significant Weight Gain or Weight Loss during the Double-Blind Treatment Period (Safety Sample)
CT-15.1.3 Incidence of Potentially Clinically Relevant Weight Gain or Loss by Baseline BMI (Safety Sample)
CT-15.2 Summary of Mean Change from Baseline to the Endpoint (Week 8) of the Double-Blind Treatment Period in Waist Circumference (Unit: cm) (Safety Sample)
CT-15.3 Summary of Mean Change from Baseline to the Endpoint (Week 8) of the Double-Blind Treatment Period in BMI (Unit: kg/m2) (Safety Sample)
CT-15.4.1 Summary of Change from Baseline in Age-and-Gender-Adjusted Z-scores for Body Weight, during the Double-Blind Treatment Period by Study Week (Safety Sample)
CT-15.4.2 Summary of Change from Baseline in Age-and-Gender-Adjusted Z-scores for Height during the Double-Blind Treatment Period by Study Week (Safety Sample)
CT-15.4.3 Summary of Change from Baseline in Age-and-Gender-Adjusted Z-scores for BMI during the Double-Blind Treatment Period by Study Week (Safety Sample)
CT-15.4.4 Incidence of Exceeding 0.5 in the change from Baseline Body Weight Z-score for the Double-Blind Treatment Period (Safety Sample)
CT-15.4.5 Number and Percentage of Subjects with Significant Change in BMI Z-score for the Double-Blind Treatment Period (Safety Sample)

SAP 331-201-00148

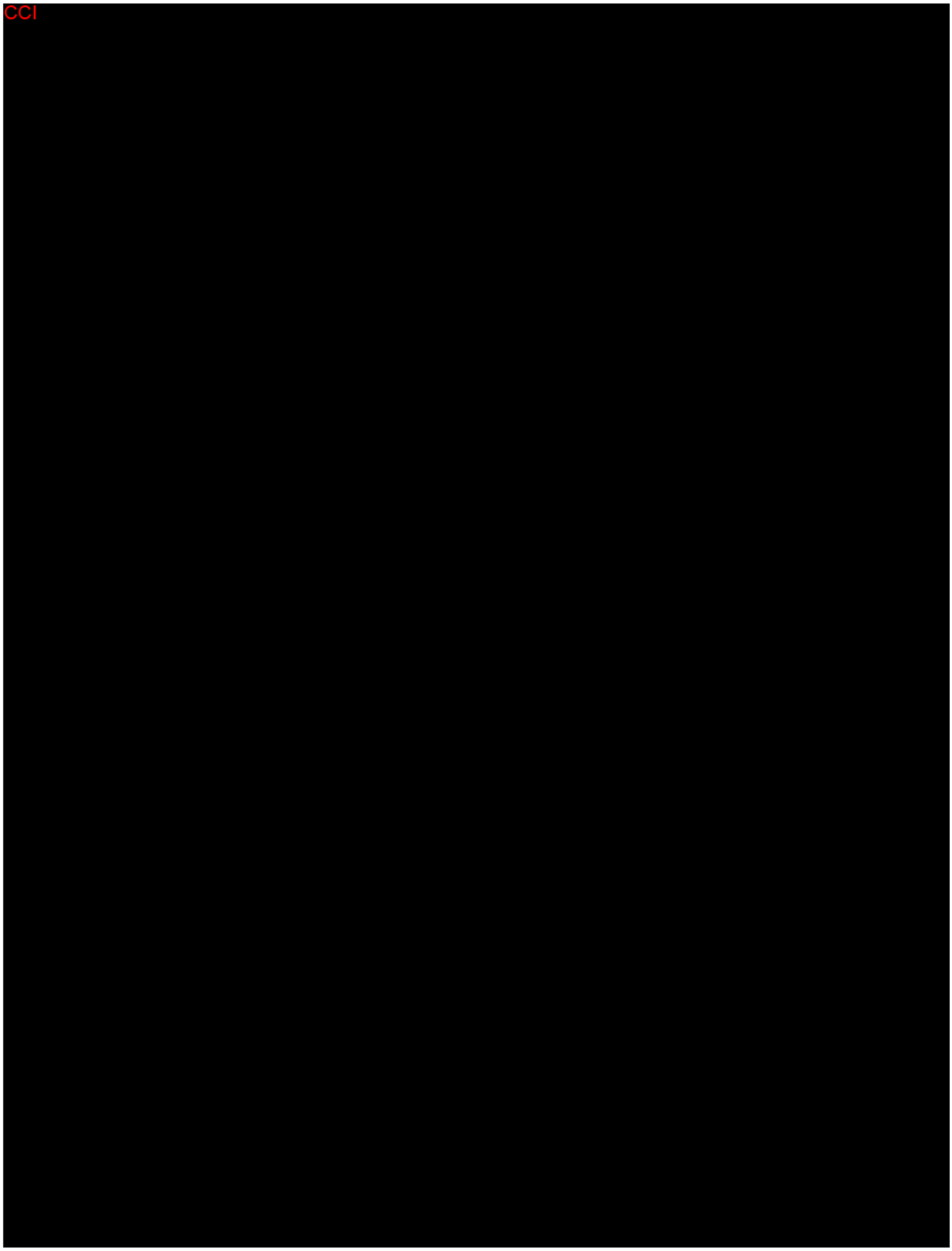


SAP 331-201-00148



SAP 331-201-00148

CCI



This page is a manifestation of an electronically captured signature

SIGNATURE PAGE

Document Name: 331-201-00148_SAP

Document Number: CCI [redacted]

Document Version: 4.0

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy hh:min) - UTC timezone
PPD	Biostatistics Approval	28-Sep-2022 21:22:03
PPD	Clinical Approval	28-Sep-2022 21:13:52