Table of Contents

[redacted]

Investigational New Drug
Brexpiprazole (OPC-34712)

Protocol 331-10-234
IND No. 101,871
EudraCT No. 2017-001447-12

A Multicenter, Randomized, Double-blind, Placebo- and Active-controlled Trial to Evaluate the Efficacy of Brexpiprazole Monotherapy for the Treatment in Adolescents (13-17 years old) With Schizophrenia

Statistical Analysis Plan

Version: 2.0
Date: 14 Mar 2023

Protocol Date: 05 Jul 2022

SAP 331-10-234

Table of Contents

SAP 331-10-234

List of In-text Figures

SAP 331-10-234

List of In-text Tables

SAP 331-10-234

List of Appendices

CCI [redacted]

[redacted]

[redacted]	[redacted]
AE	Adverse event
AIM	Abnormal involuntary movement
AIMS	Abnormal Involuntary Movement Scale
ALT	Alanine transaminase
ANCOVA	Analysis of covariance
AST	Aspartate transaminase
BARS	Barnes Akathisia Rating Scale
BMI	Body mass index
CDC	Center for Disease Control and Prevention
CGAS	Children's Global Assessment Scale
CGI-I	Clinical Global Impression Improvement
CGI-S	Clinical Global Impression Severity
CL	Total body clearance of drug from plasma following intravascular administration
CMH	Cochran-Mantel-Haenszel
CRF	Case report form
C-SSRS	Columbia-Suicide Severity Rating Scale
ECG	Electrocardiogram
EPS	Extrapyramidal symptoms
ET	Early Termination
EudraCT	European Clinical Trial Data Base
FDA	Food and Drug Administration
FU	Follow-up
ICF	Informed consent form
IMP	Investigational medicinal product(s)
IND	Investigative new drug
IWRS	Interactive web response system
kg	Kilogram
L	Liter
LOCF	Last observation carried forward
LS	Least squares
MAR	Missing at random
MCMC	Monte Carlo Markov Chain
MedDRA	Medical Dictionary for Regulatory Activities
mg	Milligrams
MI	Multiple Imputation
MMRM	Mixed model repeated measures
NY-AACENT	New York Assessment for Adverse Cognitive Effects of Neuropsychiatric Treatment
OC	Observed case
PANSS	Positive and Negative Syndrome Scale
PMM	Pattern Mixture Models
QD	Dosing once per day

SAP 331-10-234

RR	Relative risk
s	Second
SAP	Statistical analysis plan
SAS	Simpson Angus Scale
SD	Standard deviation
TEAEs	Treatment-emergent AEs
UKU	Udvalg for Kliniske Undersogelser
ULN	Upper limit of normal
US	United States

SAP 331-10-234

1 [redacted]

This statistical analysis plan (SAP) documents the statistical methodology and data analysis algorithms and conventions to be applied for statistical analysis and reporting of efficacy and safety data of trial 331-10-234. All amendments to the protocol are taken into consideration in developing this SAP.

2 [redacted]

The objective of the trial is to evaluate the short-term efficacy and safety of brexpiprazole monotherapy in the treatment of adolescents with schizophrenia.

3 [redacted]

3 [redacted]

This is a multicenter, randomized, double-blind, placebo- and active-controlled trial designed to assess the effect of brexpiprazole compared to placebo in adolescent subjects, ages 13 to 17 years at the time of informed consent/assent and at baseline (Day 1), with a *Diagnostic and Statistical Manual of Mental Disorders - Fifth Edition* (DSM-5) diagnosis of schizophrenia and by utilizing the Kiddie Schedule for Affective Disorders and Schizophrenia-Present and Lifetime version (K-SADS-PL). The trial is planned to be conducted on an outpatient basis. Any hospitalization needed for prospective subjects, such as hospitalization needed to allow subjects to change medications in preparation for this trial requires discussion with the Medical Advisor.

This trial has a 6-week treatment double-blind period.

After a minimum 3-day washout period, subjects who continue to meet all entrance criteria (including Positive and Negative Syndrome Scale [PANSS] Total Score ≥ 80) at the baseline visit (Day 1) will be randomized 1:1:1 to 1 of 3 double-blind treatment arms:

- Brexpiprazole 2 - 4 mg daily flexible dose
- Aripiprazole 10 - 20 mg daily flexible dose
- Placebo

Subjects will be randomly assigned through an interactive web response system (IWRS). Treatment assignments will be based on a computer-generated randomization code provided by the Otsuka Pharmaceutical Development & Commercialization, Inc. Biometrics Department

SAP 331-10-234

Brexpiprazole will be titrated to the minimum target dose of 2 mg in 8 days and aripiprazole will be titrated to the appropriate therapeutic dose consistent with the US package insert (minimum dose of 10 mg). Subjects who are unable to tolerate the minimum target dose will be discontinued. At the end of the titration period, investigators will be able to adjust the dose of double-blind IMP not to exceed 4 mg/day in the brexpiprazole treatment arm or 20 mg/day in the aripiprazole treatment arm. If subjects are not able to tolerate the minimum dose, they will be discontinued from the trial.

During the double-blind treatment phase, mandatory evaluations will take place at Day 1 (baseline), Day 4 (telephone call), and Weeks 1, 2, 3, 4, 5, and 6. However, at the discretion of the treating physician, more frequent evaluations are permitted. Data collected from subjects undergoing ad hoc evaluations may be excluded from efficacy analysis.

Eligible subjects who complete the trial may have the option to enroll into an open-label safety trial of brexpiprazole (Protocol 331-10-236). All subjects who do not enroll in the open label rollover trial (Protocol 331-10-236) will be assessed either via site visit or telephone assessment 21 (± 2) days after the last dose of IMP to assess adverse events (AEs).

A schematic of the trial design is shown in Figure 3.1-1.

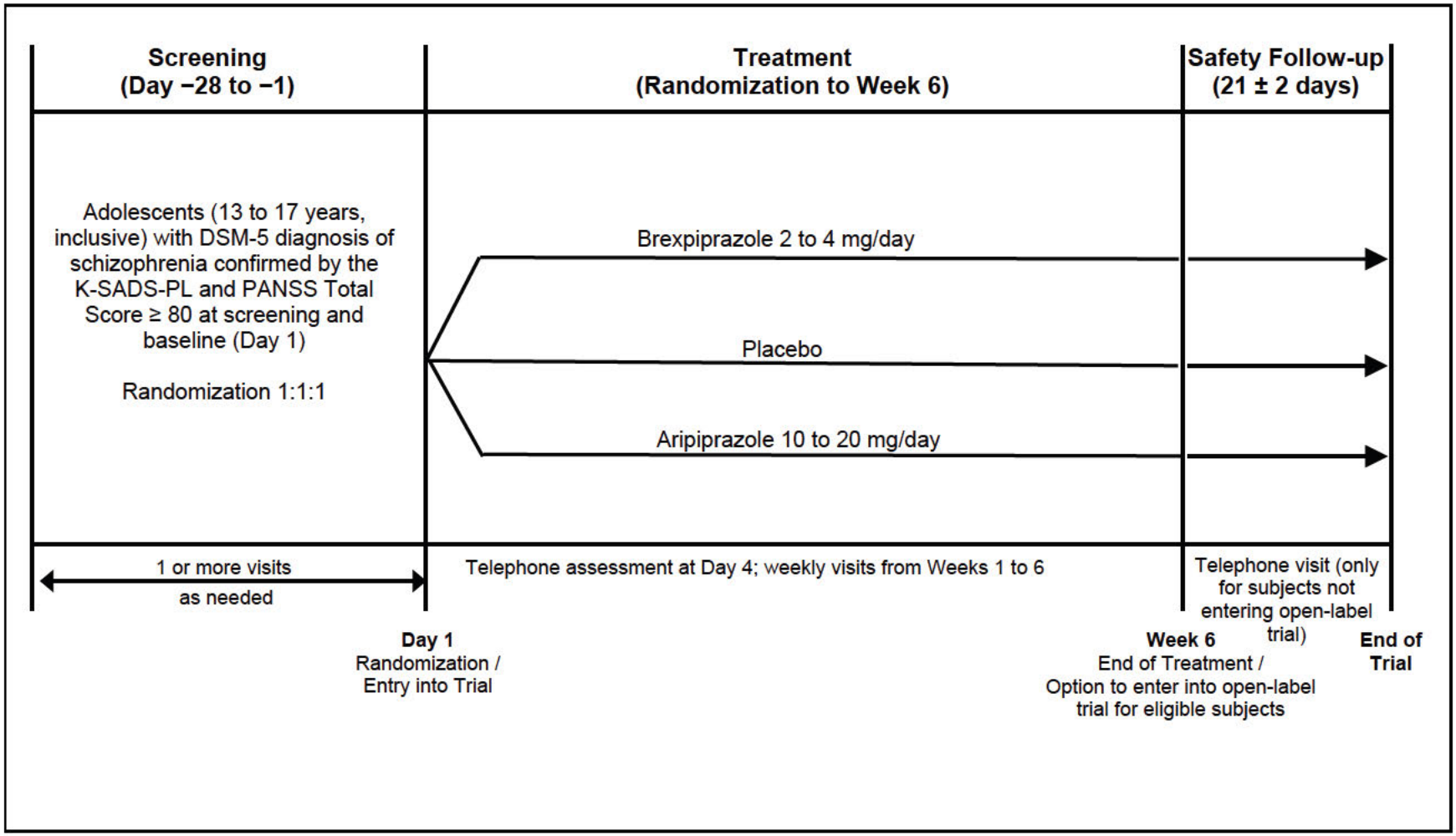

Figure 3.1-1 Trial Design Schematic

SAP 331-10-234

3.2 Trial Treatments

During double blinded treatment period, subjects will receive IMP, consisting of brexpiprazole monotherapy, aripiprazole, or placebo, depending on the subject's treatment assignment.

All doses of double-blind IMP are to be taken orally once daily and can be administered without regard to meals.

For all subjects, the number of tablets taken during the titration period (Day 1 to Day 7) will be identical to ensure the double-blind nature of the trial. As shown in Table 3.2-1, subjects randomized to brexpiprazole will take 0.5 mg the first 4 days as a titration, then take 1 mg per day from Days 5 to 7. From Days 8 to 14, subjects will take the minimum dose of 2 mg. From Days 15 to 21, the dose may be changed from 2 mg to 3 mg, or it may remain at 2 mg. After this titration period, investigators may keep the subject at a maintenance dose, increase the dose by 1 mg to a maximum of 4 mg/day, or decrease the dose by 1 mg. Subjects randomized to aripiprazole will take 2 mg the first 4 days as a titration, then take 5 mg per day from Days 5 to 7, and 10 mg from Days 8 to 14. Beginning on Day 15, the dose may be changed from 10 mg to 15 mg, or it may remain at 10 mg. After Day 21, investigators may keep the subject at a maintenance dose, increase the dose by 5 mg to a maximum of 20 mg, or down-titrate the subject's dose if tolerability is an issue.

Following the first dose increase, if the dose adjustment is not well tolerated, the dose may be reduced back to the minimum dose allowed by protocol (ie, brexpiprazole 2 mg and aripiprazole 10 mg). Subjects who not able to tolerate the decreased dose or minimum dose of 2 mg brexpiprazole or 10 mg aripiprazole will be discontinued from the trial.

Table 3.2-1 Dosing Schedule During the Treatment Period

Trial Day/Week	Brexpiprazole 2 - 4 mg/day	Aripiprazole 10 - 20 mg/day
Days 1 - 4	0.5 mg	2 mg
Days 5 - 7	1 mg	5 mg
Days 8 - 14	2 mg	10 mg
Days 15 - 21	2 - 3 mg with option to maintain 2 mg or increase[a] to 3 mg	10 mg or 15 mg with option to maintain 10 mg or increase[a] to 15 mg
Day 22+ (Weeks 4 - 6[b])	2, 3, or 4 mg with option to remain at the same dose, or increase or decrease dose by 1 mg	10, 15, or 20 mg with option to remain at the same dose, or increase or decrease dose by 5 mg

[a]Dose increases are allowed only at weekly intervals. Dose decreases are allowed at any time for tolerability as long as the minimum required dose is maintained.

[b]Beginning at Day 22, the dose may be increased or decreased if needed for tolerability or efficacy.

SAP 331-10-234

3.3 Trial Population

The trial population will include male and female subjects between 13 and 17 years of age, inclusive, at the time of informed consent/assent and at baseline (Day 1) who continue to have a confirmed DSM-5 diagnosis of schizophrenia and a PANSS Total Score ≥ 80 at screening and at baseline (Day 1). Approximately 645 subjects are anticipated to be screened with the expectation that 315 subjects will be randomized.

3.4 Trial Visit Window

Study visit windows will be used to map visits using study day intervals. This visit window convention applies to tables and listings for all efficacy and safety scales when presented by scheduled visit (PANSS, CGAS, CGI-S, CGI-I, CCI SAS, AIMS, BARS, NY-AACENT).

The scheduled visits for PANSS measurement and other scales assessments are Day 1 and Week 1, 2, 3, 4, 5, and 6, respectively. The first day of double-blind dosing is defined as "Day 1". A window of 2 days is allowed for week 1 to 6 per protocol. Trial Days are derived from the formula: Trial Day = Date of assessment - Date of first dose of IMP +1. Based on the trial days, observations are mapped to the corresponding week in the summary tables as per Table 3.4-1. If more than 1 observation falls within a particular study day interval, then the last observation within that interval is used. Evaluations occurring more than 7 days after the last double-blind dosing date will not be mapped into trial visit windows and will be excluded from the analysis.

Table 3.4-1 Time of Window for Mapping Visit

Week	Target Day	Trial Day Interval
1	7	2 - 10
2	14	11 - 17
3	21	18 - 24
4	28	25 - 31
5	35	32 - 38
6	42	39 - 49[a]

[a]Evaluations occurring more than 7 days after the last dosing date of IMP in the double-blind treatment period will be excluded from the analyses.

4 Sample Size

A sample size of 105 subjects per arm will provide at least 80% power at a nominal 2-sided alpha level of 0.05 to detect a −7.4 point reduction in PANSS Total score change from baseline to Week 6 for brexpiprazole vs placebo assuming an SD of 19. With a

SAP 331-10-234

1:1:1 allocation ratio, the overall sample size of this trial is planned to be 315 subjects. This will provide a sample size comparable to other similar trials.[1,2]

5 Statistical Analysis Datasets

5.1 Enrolled Sample

All subjects who signed an informed consent form (ICF).

5.2 Randomized Sample

All subjects who are randomized into the trial. Subjects are considered randomized when they are assigned a treatment group by IWRS.

5.3 Safety Sample

All subjects who are randomized into the trial and who receive at least 1 dose of IMP. Subjects will only be excluded from this population if there is documented evidence (ie, drug dispensed = drug returned or no IMP dispensed) that the subject did not take IMP. If a subject is dispensed IMP and is lost to follow up, he/she will be considered exposed.

5.4 Efficacy Sample

All subjects who are randomized into the trial who take at least 1 dose of IMP and who have a baseline and at least 1 postbaseline efficacy evaluation for the PANSS Total score.

The efficacy sample will be the core dataset for all efficacy analyses. The observed case (OC) dataset will be used in the primary analysis of efficacy endpoints.

5.5 Handling of Missing Data

The PANSS scale is utilized as the primary efficacy assessment in this trial. The PANSS Total Score is the sum of the rating scores for 7 positive scale items, 7 negative scale items, and 16 general psychopathology scale items from the PANSS panel in the eSource. If less than 24 of the 30 items are recorded, the PANSS Total Score is not evaluable. If 24 to 29 of the 30 items are recorded, the PANSS Total Score is equal to the mean of the recorded items multiplied by 30 and then rounded to the first decimal place.

In general, missing data will be handled by analysis of mixed model repeated measures (MMRM) methodology based on all data from protocol-specified visits in the efficacy sample OC dataset under the assumption of missing at random (MAR).

The OC dataset will consist of actual observations recorded at each visit during double-blind treatment and no missing data will be imputed. Observations at Early Termination

SAP 331-10-234

(ET) will be assigned to Week 1 to Week 6 visits based on their visit windows (see Section 3.4). MMRM (mixed model repeated measurements) will be performed on the OC dataset.

In order to explore the robustness of the primary analysis based on the MAR assumption, sensitivity analyses of the primary efficacy endpoint under MNAR (Missing not at Random) assumption will be conducted using pattern-mixture model[3,4,5]. Details are provided in Section 8.1.2.

In addition, in order to assess sensitivity of results due to missing data, a last observation carried forward (LOCF) analysis will be performed as a sensitivity analysis. In contrast to the OC dataset, the LOCF dataset will include data recorded at a scheduled visit, ie, all OC data, or, if no observation is recorded at that visit, data carried forward from the previously scheduled visit. Baseline data will not be carried forward to impute missing values for the LOCF dataset.

6 Primary and Secondary Outcome Variables:

6.1 Primary Outcome Variables

The primary efficacy endpoint is the change from baseline to Week 6 in PANSS Total Score.

6.1.1 Primary Estimand

The primary clinical question of interest is: What is the intervention difference in change from baseline in PANSS after 6 weeks of treatment where no patient would discontinue treatment due to any reason.

The primary estimand is defined by the following components:

- **Population:** Adolescents (13 - 17 years old) with a primary diagnosis of Schizophrenia who would benefit from pharmacological treatment.
- **Endpoint:** Change from Baseline to Week 6 in the PANSS Total Score
- **Intercurrent Events:** Refers to premature treatment discontinuation (ie, early dropout) prior to Week 6 attributable to adverse events, lack of efficacy, withdrawal of consent/assent, or any other causes The intercurrent event will be handled through the hypothetical strategy.
- **Treatment condition:** one of the randomized treatment groups of the following per protocol: brexpiprazole, aripiprazole or placebo
- **Population-level summary**: Difference in endpoint means between the brexpiprazole arm and the placebo arm.

SAP 331-10-234

Rationale for estimand: hypothetical estimand justifiable to evaluate the pharmacological effect, had 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 had their treatment effect masked had the subjects taken other treatment after withdrawal. This would mean 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 efficacy assessment after premature trial discontinuation will be assessed only once at the Early Termination (ET) Visit. Every effort will be made to complete all of the ET evaluations prior to administering any additional medications for the treatment of schizophrenia 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.

For the primary efficacy analysis, the estimator will be the Mixed Model Repeated Measurements (MMRM) estimate for treatment difference at Week 6, based on all observed case (OC) data until discontinuation from the trial. Under the MAR assumption, MMRM provides an unbiased estimate of treatment effect for the treatment period. Analyses with missing values imputed by MI under MNAR and other methods will be performed as sensitivity analyses.

Of note, the population attribute is slightly different in this SAP compared to the Protocol (dated 05 Jul 2022). Adjustment has been made based on FDA feedback to better reflect the target population expected to use the product.

[redacted]

[redacted]

The secondary efficacy endpoints are as follows:

- Change in the PANSS Positive and Negative Subscale Scores
- Percentage of subjects achieving response. Response is defined as at least 30% improvement from baseline in PANSS Total Score or CGI score of 1 or 2.
- Percentage of subjects achieving remission. Remission is defined as a score of ≤ 3 on each of the following specific PANSS items: delusions (P1), unusual thought content (G9), hallucinatory behavior (P3), conceptual disorganization (P2). mannerisms/posturing (G5), blunted affect (N1), passive/apathetic social withdrawal (N4), and lack of spontaneity and conversation flow (N6).
- Change in the Children's Global Assessment Scale (CGAS) Score
- Change in the Clinical Global Impression Severity (CGI-S) scale
- Clinical Global Impression Improvement (CGI-I) scale

SAP 331-10-234

6.2.2 Secondary Safety Endpoints

Safety will be assessed by the following secondary endpoints:

- The frequency and severity of AEs, serious AEs (clinical and laboratory), and discontinuation from the trial due to AEs
- Weight, height, body mass index, and waist circumference
- Analysis of potential suicide events recorded on the Columbia-Suicide Severity Rating Scale (C-SSRS)
- Clinical laboratory tests and urinalysis results (including serum prolactin), vital signs, physical examinations, and electrocardiogram (ECG) parameters
- Changes on the Simpson Angus Scale (SAS), Abnormal Involuntary Movement Scale (AIMS), and Barnes Akathisia Rating Scale (BARS)
- Comprehensive psychotropic side effects as assessed by the Udvalg for Kliniske Undersogelser (UKU) side effect rating scale
- Cognitive adverse effects as assessed by the New York Assessment for Adverse Cognitive Effects of Neuropsychiatric Treatment (NY-AACENT)

CCI

[redacted]

7 Disposition and Demographic Analysis

7.1 Subject Disposition

Subject disposition (such as “Enrolled”, “Treated”, “Discontinued”, “Completed”, etc.) will be summarized on the Enrolled Sample. Study completion rate and reasons for discontinuation will also be summarized.

For purposes of this trial, subjects who completed the end-of-trial visit (ie, the Week 6 visit) will be defined as trial completers.

Disposition will be summarized by treatment group and by subgroup of sex, race, region and age category (< 15, ≥ 15 years).

Reasons for discontinuation will be summarized for the randomized sample by treatment group and by subgroup of sex, race, age category (< 15, ≥ 15 years) and region (US, Europe [continent], Mexico).

SAP 331-10-234

7.2 Demographic and Baseline Characteristics

Baseline measurements of efficacy and safety variables are defined as their last measurements prior to the first dose of IMP.

Baseline demographic characteristics include age, sex, race, ethnicity, height, weight, waist circumference, and body mass index (BMI). For the Randomized Sample, demographic characteristics will be summarized by treatment group.

Mean, median, SD, minimum and maximum values will be used to describe continuous variables such as age. Frequency distributions will be tabulated for categorical variables such as race and age category (< 15, ≥ 15 years).

Demographics will be summarized by treatment group and by subgroup of sex, race, age category (< 15, ≥ 15 years) and region (US, Europe [continent], Mexico).

7.3 Baseline Disease Evaluation

A summary of medical and psychiatric history will be presented for the Randomized Sample (by treatment group and overall). Prior antipsychotic medication will be summarized in frequency. Medical history will also be summarized by using frequency distribution of subjects across the medical histories.

Baseline values in the following parameters related to psychiatric evaluation will be summarized on the Randomized Sample by using descriptive statistics: PANSS total score, PANSS Positive subscale score, PANSS Negative subscale score, CGI-S score, CGAS score CCI

Baseline characteristics will be summarized by treatment group and by subgroup of sex, race age category (< 15, ≥ 15 years) and region.

7.4 Treatment Compliance

Treatment compliance will be reported, by treatment group, in the Safety Sample, by percentage of subjects taking at least 70%, 80%, and 90% of medications based on IMP Accountability panel of the Case Report Form.

7.5 Prior and Concomitant Medications

Number and proportion of subjects taking concomitant medications prior to IMP, during the double-blind treatment period, and after IMP are tabulated by drug classification using the World Health Organization drug dictionary.

SAP 331-10-234

[redacted]

Protocol deviations data will be summarized by type of deviations by trial site and treatment group for the Randomized Sample. In addition, a subjects listing will be provided describing the deviations for each subject.

8 [redacted]

8 [redacted]

The primary efficacy endpoint is the change from baseline to Week 6 in PANSS Total Score. The primary statistical comparison of interest is brexpiprazole 2 - 4 mg versus placebo based on all available data (OC dataset). All randomized subjects who have both baseline and post-baseline PANSS total score will be included in the primary efficacy analysis.

8 [redacted]

The primary statistical comparison of interest is brexpiprazole 2 - 4 mg versus placebo. With the assumption of MAR, an MMRM analysis with fixed-effect factors of treatment, trial site, visit, treatment by visit interaction, and fixed effect covariates of baseline and baseline by visit interaction will be applied to the change from baseline from Week 1 to Week 6 in PANSS Total Score based on all available data (OC dataset). The data will be modeled using an unstructured variance covariance matrix for the within subject variation. A statistical test of the least squares (LS) mean differences at Week 6 of the MMRM analysis will serve as the analysis of the primary endpoint. The Kenward-Roger approximation will be used to estimate denominator degree of freedom and adjust standard errors.

The contrast (ie, difference in least-square means between brexpiprazole and placebo) at the Week 6 visit will be estimated from the interaction term of treatment by visit week and will serve as the primary treatment comparison. The point estimate and the 95% confidence interval estimate of the contrast at the Week 6 visit will be reported. Significance test will be based on the contrast estimate at the Week 6 visit by using a two-sided 0.05 level.

In case there is a convergence problem with the MMRM model with the unstructured variance covariance matrix, the following structures other than unstructured will be used in order of 1) heterogeneous toeplitz, 2) heterogeneous autoregressive of order 1, and 3) heterogeneous compound symmetry, and the first (co)variance structure converging to the best fit will be used as the primary analysis. If a structured covariance has 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.

8.1.2 Sensitivity Analyses

In order to explore the robustness of the primary analysis based on the MAR assumption, sensitivity analysis of the primary efficacy endpoint under the MNAR assumption will be conducted.

Pattern Mixture Models (PMM) based on Multiple Imputation (MI) with mixed missing data mechanisms will be used to investigate the response profile of dropout subjects by last dropout reason under MNAR mechanism for the following 4 scenarios:

1) Dropout reasons due to AE as MNAR
2) Dropout reasons due to AE or lack of efficacy as MNAR
3) Dropout reasons due to either AE or subject withdrew consent as MNAR
4) All dropouts as 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 -7.4 points and/or the observed treatment difference between brexpiprazole and placebo from the primary analysis of MMRM model until conclusion of the primary analysis is overturned. When delta = 0 the missing data are assumed to be MAR. When delta > 0, the missing data are assumed to be MNAR.

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

CCI [redacted]

SAP 331-10-234

Placebo Based Imputation Methods

Similar to "Standard" multiple imputations, except parameters for imputation model obtained from only the placebo (control) group. Missing data for both placebo and drug groups (brexpiprazole and aripiprazole) are imputed based on the imputation model derived from placebo data. If drug improved outcomes prior to dropout, this benefit is carried into subsequent imputed values, but will diminish over time in accordance with the correlation structure.

Detailed SAS program is provided in Appendix 5.

Additional Sensitivity Analyses

In order to assess sensitivity of results due to missing data, a last observation carried forward (LOCF) analysis will be performed as a sensitivity analysis. In contrast to the OC dataset, the LOCF dataset will include data recorded at a scheduled visit, ie, all OC data, or, if no observation is recorded at that visit, data carried forward from the previously scheduled visit. Baseline data will not be carried forward to impute missing values for the LOCF dataset.

Change from baseline for the PANSS Total Score at Week 6 will be evaluated using ANCOVA with baseline value as covariate and treatment, trial site as main effects, on the LOCF dataset. In addition, ANCOVA with baseline value as covariate and treatment, as main effects will be performed on the OC dataset.

In case of gross violations of the MMRM model assumptions, additional supportive analyses will be provided using the generalized version of the Cochran-Mantel-Haenszel (CMH) procedure, controlling for trial site.

8.1.3 Technical Computational Details for Primary Efficacy Analysis

The following algorithm will be used for the primary analysis.

- PANSS total score is the sum of the rating scores for 7 positive scale items, 7 negative scale items and 16 general psychopathology scale items from the PANSS panel in the eSource. If less than 24 of the 30 items are recorded, the PANSS total score is not evaluable. If 24 to 29 of the 30 items are recorded, the PANSS total is equal to the mean of the recorded items multiplied by 30, and then rounded to the first decimal place.
- All scheduled visits during the 6 weeks after randomization, regardless of compliance to study assessment schedule, will be included in the analysis. If more than 1 observation falls within a particular trial day interval, then only the last observation within that interval is used in the analysis.

SAP 331-10-234

- All early terminated (ET) visits before week 6 will be included according to visit window.
- All ET visits 6 weeks after randomization and beyond Day 49 will not be included in the analysis.
- Evaluations occurring more than 7 days after the last double-blind dosing date will not be mapped into trial visit windows, and will be excluded from the analysis.
- Follow-up (FU) visit will not be included in the analysis.
- The week number for an ET visit will be calculated by (date of ET visit randomization/drug start date + 1) according to visit window.

Visits windows details are provided in Section 3.4.

8.1.3.1 Pooling of small sites

Small trial sites will be defined as sites that do not have at least 1 evaluable subject (evaluable with regard to the primary efficacy variable) in each treatment arm. All small sites will be pooled to form "pseudo sites" for the purpose of analysis according to the following algorithm. Small sites will be ordered from the largest to the smallest based on the number of evaluable subjects (ie, subjects who have a Baseline value and at least one post-randomization value for PANSS Total Score in the double-blind treatment period). The process will start by pooling the largest of the small sites with the smallest of the small sites until a non-small site is formed. This process will be repeated using the site left out of the previous pass. In case of ties in site size, the site with the smallest site code will be selected. If any sites are left out at the end of this process, they will be pooled with the smallest pseudo sites, or if no pseudo site exist, they will be pooled with the smallest non-small site. The pooling of the small sites will be performed within each region (US, Europe [continent], Mexico).

8.1.3.2 SAS code

The MMRM analysis for the primary endpoint will be analyzed using PROC MIXED of SAS (with default REML option). Specifically, the following SAS statements will be used for the primary analysis.

SAP 331-10-234

The test of the contrast of brexpiprazole versus placebo at Week 6 will be performed by using the following estimate statement for MMRM analysis:

CCI

[redacted]

8.1.4 COVID-19 Pandemic Related Sensitivity Analyses

The following analyses will be performed on the Efficacy Sample to evaluate the sensitivity of the primary analysis results to the impact of 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. The date of 13 Mar 2020, date of COVID-19 pandemic announcement in the US, is taken as landmark date for analyses as considered sufficiently close to the date of significance disturbance in the other countries.

1) An MMRM analysis based on the pre-COVID Efficacy Sample. The pre-COVID Efficacy Sample comprises those subjects in Efficacy Sample who had completed or discontinued the trial before 13 Mar 2020.
2) An MMRM analysis based on the Efficacy Sample by using the subset of pre-COVID data. The pre-COVID data consist of the data collected before 13 Mar 2020.
3) An MMRM analysis based on the non-COVID Efficacy Sample. The non-COVID Efficacy Sample comprises those subjects in Efficacy Sample who had no remote assessment of the primary endpoint.
4) An MMRM analysis based on the Efficacy Sample by using the subset of non-COVID data. At the subject level, the non-COVID data refer to the data that exclude his/her first remote assessment of the primary endpoint and all the assessments thereafter.
5) An MMRM analysis based on the Efficacy Sample by using the subset of data that exclude all remote assessments.

The point estimate along with the 95% confidence interval estimate of the treatment contrast and p-values will be presented for each of the above sensitivity analyses.

In addition, the number of face-to-face and virtual visits will be tabulated.

SAP 331-10-234

[redacted]

The Ukraine-Russia conflict started on 24 Feb 2022. This trial enrolled many subjects from Ukraine and Russia. However, the number of subjects from Russia and Ukraine starting treatment in the 9 weeks before the start of the conflict is very limited (N = 6), most completed the treatment and had PANSS assessment. No subject from these countries started treatment after 22 Feb 2022. Therefore, no specific sensitivity analyses will be performed.

[redacted]

The following are the secondary efficacy variables in this trial:

1) Change in the PANSS Positive and Negative Subscale Scores
2) Percentage of subjects achieving response. Response is defined as at least 30% improvement from baseline in PANSS Total Score or CGI score of 1 or 2.
3) Percentage of subjects achieving remission. Remission is defined as a score of ≤ 3 on each of the following specific PANSS items: delusions (P1), unusual thought content (G9), hallucinatory behavior (P3), conceptual disorganization (P2). mannerisms/posturing (G5), blunted affect (N1), passive/apathetic social withdrawal (N4), and lack of spontaneity and conversation flow (N6).
4) Change in the Children's Global Assessment Scale (CGAS) Score
5) Change in the Clinical Global Impression Severity (CGI-S) scale
6) Clinical Global Impression Improvement (CGI-I) scale

For continuous secondary endpoints (change in CGI-S, change in CGAS, PANSS Positive Subscale score, and PANSS Negative Subscale score), an MMRM analysis with factors of treatment, trial site, visit, treatment by visit interaction, and baseline and baseline by visit interaction as covariates with an unstructured variance covariance structure will be applied based on all available data (OC dataset) from the Efficacy Sample. ANCOVA with baseline value as covariate and treatment as main effects will also be performed on the OC dataset. Data will be presented by week.

The CGI-I score will be evaluated by the Cochran-Mantel-Haenszel row mean score differ test controlling for trial site in last-observation-carried-forward (LOCF) and Observed Cases (OC) analyses. Descriptive statistics and differences between arms with 95% confidence intervals will also be presented.

Response and remission endpoints will be evaluated by the CMH General Association Test controlling for trial site, by week, in OC and LOCF analyses, for the Efficacy Sample.

SAP 331-10-234

8.2.1 Technical Computational Details for the Analysis of the Secondary Efficacy Endpoints

1. The same time windows as defined in Table 3.4-1 will be used for mapping the trial weeks are used for secondary efficacy endpoints.

2. Computing of Scale Scores:

(1) The CGI-S and CGI-I of 0s will be set to missing because value 0s of CGI-S and CGI-I mean "not assessed".

(2) PANSS positive subscale score is the sum of the rating scores for the 7 positive scale items from the PANSS panel in CRF. PANSS negative subscale score is the sum of the rating scores for the 7 negative scale items from the PANSS panel in CRF. If less than 6 of the 7 scale items of a PANSS subscale are recorded, the PANSS subscale score is not evaluable. If 6 of the 7 scale items are recorded, PANSS subscale score is equal to the mean of the recorded items multiplied by 7, and then rounded to the first decimal place.

By-visit analyses will also be conducted as previously described for the primary efficacy endpoint to inform the onset of treatment effect on these endpoints.

8.3 Subgroup Analyses

Subgroup analyses of the change from baseline in PANSS Total score at every study week will be conducted by sex, race (White and Other races), age category ($< 15, \geq 15$ years) and region (US, European [continent], Mexico) using MMRM analysis with factors of treatment, visit, treatment by visit interaction, and baseline and baseline by visit interaction as covariate with an unstructured variance covariance structure.

In addition, change in PANSS Positive and Negative subscale Scores, change in CGAS Score and CGI-S scale at Week 6 (MMRM analysis with factors of treatment, visit, treatment by visit interaction, and baseline and baseline by visit interaction as covariates, (with LS Mean and 95% CI), percentage of subjects achieving response and remission (with relative risk and 95% CI), and CGI-I (difference and 95% CI) will also be computed by age category ($< 15, \geq 15$ years).

SAP 331-10-234

9 Safety Analyses

Safety will be assessed by the following secondary endpoints:

- The frequency and severity of AEs, serious AEs (clinical and laboratory), and discontinuation from the trial due to AEs
- Weight, height, body mass index, and waist circumference
- Analysis of potential suicide events recorded on the Columbia-Suicide Severity Rating Scale (C-SSRS)
- Clinical laboratory tests and urinalysis results (including serum prolactin), vital signs, physical examinations, and electrocardiogram (ECG) parameters
- Changes on the Simpson Angus Scale (SAS), Abnormal Involuntary Movement Scale (AIMS), and Barnes Akathisia Rating Scale (BARS)
- Comprehensive psychotropic side effects as assessed by the Udvalg for Kliniske Undersogelser (UKU) side effect rating scale
- Cognitive adverse effects as assessed by the New York Assessment for Adverse Cognitive Effects of Neuropsychiatric Treatment (NY-AACENT)

Safety analysis will be conducted based on the Safety Sample, which is defined in Section 5. In general, baseline of a safety variable is defined as the last observation of the variable before taking the first dose of IMP, unless specified otherwise.

Concomitant medication analyses are described under Section 7.5.

SAP 331-10-234

9 [redacted]

The start date of double-blind IMP with brexpiprazole, aripiprazole or placebo will be the first day of double-blind dosing. The number and percentage of subjects who receive double-blind IMP will be presented by week and by treatment group. Each dosing week will be based on the actual week; ie, Day 1 to 7 in Week 1, Day 8 to 14 in Week 2, etc. This summary will be performed on the Safety Sample.

The number and percentage of completers will be presented by week and by treatment group.

The mean daily dosage will be summarized by week and treatment group using descriptive statistics. The mean daily dosage per subject per week will be determined for each week of the trial. This will be calculated by dividing the sum of individual total doses by the number of days in the week interval. The summary will contain for each treatment group the number of subjects receiving double-blind IMP, and the mean and range of the mean daily dose for each week.

Extent of exposure will also be summarized by sex, race (White and All Other races), age category (< 15, ≥ 15 years) and region (US, Europe [continent], Mexico).

9 [redacted]

All AEs will be coded by system organ class and Medical Dictionary for Regulatory Activities (MedDRA) preferred term. The incidence of the following events will be summarized by treatment group:

1) Treatment-emergent AEs (TEAEs),
2) TEAEs by severity
3) TEAEs potentially causally related to the IMP,
4) TEAEs potentially causally related to the IMP by severity,
5) TEAEs with an outcome of death,
6) Serious TEAEs,
7) TEAEs leading to discontinuations of the IMP.

Treatment-emergent AE is defined as an AE that started after trial drug treatment, or if the event was continuing from baseline and was serious, trial drug related, or resulted in death, discontinuation, interruption or reduction of trial therapy.

Incidence of TEAEs will also be summarized by sex, race (White and All Other races), age category (< 15, ≥ 15 years) and region (US European [continent], Mexico).

SAP 331-10-234

9.3 Clinical Laboratory Data

During the 6 weeks treatment period, clinical laboratory assessments in the fields of hematology, serum chemistry (including prolactin), and urinalysis will be conducted at the following scheduled visits: Day 1 (baseline), Week 4 and Week 6. HbA1c and TSH are only performed at baseline and Week 6.

Summary statistics for the routine clinical laboratory measurements, prolactin concentrations, HbA1c, and TSH at baseline, post-baseline visits, and changes from baseline will be provided. Incidence of potentially clinically significant abnormal lab results will also be summarized by treatment groups. Criteria of potentially clinically significant lab test abnormalities are provided in Appendix 1.

Shift tables will be produced assessing status (low-normal-high) changes from baseline.

Total bilirubin level should be checked for any subject with increased ALT or AST levels ≥ 3 times the upper normal limits.

Liver injury related laboratory test results will be summarized for subjects who met following criteria, based on Hy's law lab criteria. The corresponding listing will be provided as well.

- AST or ALT $\geq 3 \times$ upper limit of normal (ULN) and
- T_Bili $\geq 2 \times$ ULN

Prolactin data will be presented overall and by sex, and its change from baseline will be summarized.

9.4 Vital Signs Data

Descriptive statistics will be provided by treatment group, for both vital signs and change from baseline in vital signs. Vital signs include systolic blood pressure, diastolic blood pressure, heart rate in supine, sitting, and standing position. In addition, body weight, height, and waist circumference are measured. Body mass index (BMI) in kg /m^2 is calculated as weight in kilogram / (height in meter)^2.

Incidence of potentially clinically significant vital sign results will also be summarized by treatment groups. Criteria of potentially clinically significant vital sign abnormalities are provided in Appendix 2.

For a subject with several measures in either vital signs or Lab tests at a visit, the last repeat will be used in the by visit summary. However, for outlier analysis (such as

clinically significant abnormalities), data from all visits, no matter they are from the original visits, repeats, or unscheduled visits, will be included.

9.5 Body Weight, Waist Circumference and Body Mass Index

Analyses of body weight, waist circumference and BMI will be performed for the Safety Sample. Descriptive statistics based on observed cases data will be provided by trial weeks for change from baseline in body weight, as well as incidence of clinically significant changes in body weight, by baseline BMI status. Incidence of potentially clinically relevant weight gain will also be presented by region. Clinically significant changes are described in Appendix 1, Appendix 2 and Appendix 3 respectively. Descriptive statistics will also be summarized by visit for change from baseline in waist circumference and BMI.

9.5.1 Calculations of Z-scores for Body Weight, Height, and BMI

Age and sex adjusted z-scores for body weight 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.

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 interim database lock, and updated information will be used if any.

Mean changes in z-scores for body weight from baseline will be summarized by visit, and by sex, age category (< 15, ≥ 15 years) and region. Mean changes from baseline in z-scores for BMI will be summarized by visit. Furthermore, the number and percentage

of subjects with the change of body weight and BMI z-score (relative to baseline) ≥ 0.5 or ≤ -0.5 through Week 6 will be tabulated by treatment group.

9.6 Physical Examination Data

Physical examination findings will be listed by subject.

9.7 Electrocardiogram Data

ECG data will be summarized by mean change from baseline by treatment group by visit. For the analysis of QT and QTc, data from 3 consecutive complexes (representing 3 consecutive heart beats) will be measured to determine average values. The following QT corrections will be used:

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: $QTcN=QT/(RR)^{0.37}$

Incidence of potentially clinically significant treatment emergent ECG abnormalities will also be summarized by treatment groups. Criteria of potentially clinically significant ECG abnormalities are provided in Appendix 3.

9.8 Extrapyramidal Symptoms Scales

Change from baseline in each of the 3 EPS variables (SAS total score, AIMS total score, and BARS global clinical assessment score) will be summarized on the OC data by visit and on the LOCF data for the last assessment visit using descriptive statistics and ANCOVA model. In addition, BARS global clinical assessment will be treated as categorical variable taking values (categories) of "Absent", "Questionable", "Mild Akathisia", "Marked Akathisia" and "Severe Akathisia", and thus will be summarized on OC data by visit for each severity category, by using count and frequency.

Incidence of Treatment Emergent EPS-related AEs by EPS Category and MedDRA Preferred Term. The most up to date Otsuka list of EPR-related AEs at the time of the lock will be used.

9.9 Columbia-Suicide Severity Rating Scale

Suicidality monitored during the trial using the C-SSRS will be summarized as the number and percentage of subjects reporting any suicidal behavior, ideation, behavior by type (4 types), ideation by type (5 types), and treatment-emergent suicidal behavior and ideation, based on the Safety Sample.

SAP 331-10-234

9.10 Other Safety Data

Incidence of Onset of Akathisia Adverse Event by Week will be presented.

Incidence of sign/symptom items as recorded in the NY-AACENT scale will be summarized. Specifically, incidence (and severity) of the sign/symptom items will be tabulated by study visit, based on the OC data. The summarization will be repeated for the subset of drug-related signs or symptoms and for the subset of function impaired signs or symptoms, respectively. Additionally, change from baseline in the NYAACENT total score will be summarized on the OC data by study visit and on the LOCF data for the last assessment visit.

The UKU, NY-AACENT scale will also be summarized by week using the OC datasets of the Safety Sample.

The summary of Tanner Scale for OC datasets of the Safety Sample will be provided for week 6 and shift from baseline. The collection of Tanner Staging data is required for this trial, and every attempt will be made to collect this information. Tanner Staging will be completed together with the physical examination. When Tanner Staging is not completed at a required visit, it will be collected at the next trial visit. A subject who reaches Stage 5 (both in pubic hair and genitalia) does not need to continue with Tanner Staging.

The incidence of the individual side effects and categories of side effects recorded in the UKU will be summarized by treatment group for the Safety Sample by severity and causal relationship. In addition, the global assessment of side effects and the measures taken on the basis of the side effects will be summarized by incidence in a similar fashion to the above.

10 Scales: Rules for Scoring and Handling of Missing Data

10.1 Positive and Negative Syndrome Scale (PANSS)

The PANSS[4] Total Score is the sum of the rating scores for 7 positive scale items (ie, 7 positive symptom constructs: delusions, conceptual disorganization, hallucinatory behavior, excitement, grandiosity, suspiciousness/persecution, and hostility), 7 negative scale items (ie, 7 negative symptom constructs: blunted affect, emotional withdrawal, poor rapport, passive/apathetic social withdrawal, difficulty in abstract thinking, lack of spontaneity and conversation flow, stereotyped thinking), and 16 general psychopathology scale items (ie, 16 symptom constructs: somatic concern, anxiety, guilt feelings, tension, mannerisms and posturing, depression, motor retardation,

uncooperativeness, unusual thought content, disorientation, poor attention, lack of judgment and insight, disturbance of volition, poor impulse control, preoccupation, and active social avoidance).

If fewer than 24 of the 30 items are recorded, the PANSS Total Score is not evaluable. If 24 to 29 of the 30 items are recorded, the PANSS Total Score is equal to the mean of the recorded items multiplied by 30 and then rounded to the first decimal place.

PANSS positive subscale score is the sum of the rating scores for the 7 positive scale items from the PANSS panel in CRF. PANSS negative subscale score is the sum of the rating scores for the 7 negative scale items from the PANSS panel in CRF. If less than 6 of the 7 scale items of a PANSS subscale are recorded, the PANSS subscale score is not evaluable. If 6 of the 7 scale items are recorded, PANSS subscale score is equal to the mean of the recorded items multiplied by 7, and then rounded to the first decimal place.

10.2 Children's Global Assessment Scale

The CGAS[6] is a 100-point rating scale measuring psychological, social and school functioning for children aged 6-17. The CGAS is a valid and reliable tool for rating a child's general level of functioning on a health-illness continuum. The CGAS was developed by Schaffer and colleagues to provide a global measure of severity of disturbance in children and adolescent. The scale is separated into 10-point sections that are headed with a description of the level of functioning and followed by examples matching the interval. This score represents severity of disturbance with score 1 to10 means need of constant supervision whereas score 91 to 100 means superior functioning in all areas.

10.3 Clinical Global Impression - Severity of Illness scale (CGI-S)

The severity of agitation for each subject will be rated using the CGI-S. To perform this assessment, the investigator (or designee) will answer the following question: Considering your total clinical experience with this particular population, how mentally ill (as related to agitation) is the subject at this time?" Response choices are 0 = not assessed; 1 = normal, not at all ill; 2 = borderline mentally ill; 3 = mildly ill; 4 = moderately ill; 5 = markedly ill; 6 = severely ill; and 7 = among the most extremely ill subjects. The score 0 (= not assessed) will be set to missing. The CGI-S is therefore a 7-point scale from 1 through 7.

10.4 Clinical Global Impression-Improvement Scale (CGI-I)

The efficacy of brexpiprazole in the treatment of agitation will be rated for each subject using the CGI-I. The investigator (or designee) will rate the subject's total improvement

SAP 331-10-234

(as related to agitation) whether or not it is due entirely to drug treatment. Response choices are 0 = not assessed, 1 = very much improved, 2 = much improved, 3 = minimally improved, 4 = no change, 5 = minimally worse, 6 = much worse, and 7 = very much worse. This is also a 7-point scale (1 to 7), with 1 being very much improved and 7 being very much worse. The scale also includes 0: not assessed, which will be set to missing for purposes of analysis. At each visit other than randomization, the global improvement will be judged with respect to subject's condition at randomization.

CCI

10.6 Simpson Angus Scale (SAS)

The SAS[7] 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 scale, with a score of zero representing absence of symptoms, and a score of 4 representing a severe condition. The SAS Total score is the sum of ratings for all 10 items, with possible Total scores 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.

10.7 Abnormal Involuntary Movement Scale (AIMS)

The AIMS is a 12-item scale. The first 10 items are rated from 0 to 4 (0 = best, 4 = worst). An item score of 0, depending on the item, either means: no abnormal involuntary movement (AIM), or no incapacitation due to AIM, or no awareness of AIM. An item score of 4 either means: severe AIM, or severe incapacitation due to AIM, or being aware of, and severe distress caused by AIM. Items 11 and 12, related to dental status, have dichotomous responses, 0 = no and 1 = yes. The AIMS Total Score is the sum of the ratings for the first 7 items. The possible total scores are from 0 to 28. The

SAP 331-10-234

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.

10.8 Barnes Akathisia Rating Scale (BARS)

The BARS[8] consists of 4 items related to akathisia: objective observation of akathisia by the investigator, subjective feelings of restlessness by the subject, subjective distress due to akathisia, and global clinical assessment of akathisia. The first 3 items will be rated on a 4-point scale, with a score of zero representing absence of symptoms and a score of 3 representing a severe condition. The global clinical evaluation will be made on a 6-point scale, with zero representing absence of symptoms and a score of 5 representing severe akathisia.

10.9 Udvalg for Kliniske Undersogelser (UKU)

The UKU is used to assess side effects of subjects being treated with antipsychotic drugs and to determine whether there is a causal relationship. Assessment of the individual symptoms is best accomplished by a semi structured interview with the subject during which the scale is gone through point by point as well as obtaining information from case records and clinical observation.

10.10 New York Assessment for Adverse Cognitive Effects of Neuropsychiatric Treatment (NY AACENT)

The NY-AACENT is intended to be used to detect changes in cognitive function subsequent to pharmacological or similar treatments for neurological or psychiatric problems. It is specifically designed to be used in pediatric populations (ages 12 - 17), but can be utilized with other age groups as appropriate. Each of the 7 items is derived from the 7 domains identified by MATRICS[9].

Each item score is derived as follows: 0 = not present in the past week; 1 = present (during past week) and mild; 2 = present and moderate; 3 = present and severe; and 4 = present and extreme.

The NY-AACENT Total Score is calculated by adding the individual item scores.

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

Suicidality will be monitored during the trial using the C-SSRS. 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 time 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 6 months, suicidal behaviors within the last 2 years, 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 also be completed at all visits after screening.

10.12 Tanner scale

The scale defines physical measurements of development based on external primary and secondary sex characteristics, such as the size of the breasts, genitals, testicular volume and development of pubic hair. Boys and girls are rated on a 5-point scale.

11 Pharmacokinetic Analyses

Pharmacokinetic data will be analyzed separately.

12 Pharmacodynamic Analyses

There is no pharmacodynamic analyses in this trial.

13 Pharmacogenomic Analyses

Pharmacogenomic data will be analyzed separately.

14 Interim Analysis

No unblinded interim analysis is planned for this trial.

15 Changes in the Planned Analyses

None.

16 Revision History

Table 16-1 Document History

Version	Key Changes
Version 2.0, 14 Mar 2023	• Added CCI and CCI describing the SAS programs for multiple imputation delta method and placebo based imputations • Section 3.4: corrected typo for target day on week 5 • Section 8.1.2: indicated that in case of convergence problems ANCOVA will be used and refer to the appendix for programming details • Section 8.1.3.1: corrected the region categories to "(US, Europe [continent], Mexico)"

SAP 331-10-234

Table 16-1 Document History

Version	Key Changes
	• Section 8.2: removed parenthesis "Van Elteren"
Version 1.0. 02 Feb 2023	• Initial version

17 Reference

1 US Food and Drug Administration. Clinicaltrial.gov register. Efficacy and Safety of Cariprazine in the Treatment of Adolescent Participants (13 to 17 Years of Age) With Schizophrenia. Available at: https://www.clinicaltrials.gov/ct2/show/NCT03817502. Assessed on 08 March 2022.

2 European Medicines Agency/59804/2021. Committee for Medicinal Products for Human Use (CHMP). Latuda Assessment Report. 23 July 2020. Procedure No. EMEA/H/C/002713/II/0029.

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

4 Kay SR, Opler LA, Fiszbein A. Positive and Negative Syndrome Scale (PANSS) rating criteria. North Tonawanda, NY: Multi-Health Systems; 1999.

5 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.

6 Shaffer D, Gould MS, Brasic J, Ambrosini P, Fisher P, Bird H, et al. A Children's Global Assessment Scale (CGAS). Arch Gen Psychiatry. 1983;40:1228-31.

7 Simpson GN, Angus JWS. A rating scale for extrapyramidal side effects. Acta Psychiatr Scand. 1970;212(Suppl 44):S11-9.

8 Barnes TR. A rating scale for drug-induced akathisia. Br J Psychiatry. 1989;154:672-6.

9 Green MF, Nuechterlein KH, Gold JM, Barch DM, Cohen J, Essock S, Fenton WS, Frese F, Goldberg TE, Heaton RK, Keefe RS, Kern RS, Kraemer H, Stover E, Weinberger DR, Zalcman S, Marder SR. Approaching a consensus cognitive battery for clinical trials in schizophrenia: the NIMH-MATRICS conference to select cognitive domains and test criteria. Biol Psychiatry. 2004; 56(5):301-7.

SAP 331-10-234

[redacted]

[redacted]	[redacted]
[redacted]	
AST (SGOT)	≥ 3 x upper limit of normal (ULN)
ALT (SGPT)	≥ 3 x ULN
Alkaline phosphatase	≥ 3 x ULN
LDH	≥ 3 x ULN
BUN	≥ 30 mg/dL
Creatinine	≥ 2.0 mg/dL
Uric Acid	
Men	≥ 10.5 mg/dL
Women	≥ 8.5 mg/dL
Bilirubin (total)	≥ 2.0 mg/dL
CPK	≥ 3 x ULN
Prolactin	> ULN
[redacted]	
Hematocrit	
Men	≤ 37 % and decrease of ≥ 3 percentage points from Baseline
Women	≤ 32 % and decrease of ≥ 3 percentage points from Baseline
Hemoglobin	
Men	≤ 11.5 g/dL
Women	≤ 9.5 g/dL
White blood count	≤ 2,800/ mm^3 or ≥ 16,000/ mm^3
Eosinophils	≥ 10%
Neutrophils	≤ 15%
Absolute neutrophil count	≤ 1,000/ mm^3
Platelet count	≤ 75,000/ mm^3 or ≥ 700,000/ mm^3
[redacted]	
Protein	Increase of ≥ 2 units
Glucose	Increase of ≥ 2 units
Casts	Increase of ≥ 2 units
[redacted]	
Chloride	≤ 90 mEq/L or ≥ 118 mEq/L
Potassium	≤ 2.5 mEq/L or ≥ 6.5 mEq/L
Sodium	≤ 126 mEq/L or ≥ 156 mEq/L
Calcium	≤ 8.2 mg/dL or ≥ 12 mg/dL
Glucose	
Fasting	≥ 100 mg/dL
Non-Fasting	≥ 200 mg/dL
Total Cholesterol, Fasting	≥ 240 mg/dL
LDL Cholesterol, Fasting	≥ 160 mg/dL
HDL Cholesterol, Fasting	
Men	< 40 mg/dL
Women	< 50 mg/dL
Triglycerides, Fasting	≥ 150 mg/dL

SAP 331-10-234

[redacted]

[redacted]	[redacted] [a]	[redacted] [a]
Heart Rate[b]	> 120 bpm < 50 bpm	≥ 15 bpm increase ≥ 15 bpm decrease
Systolic Blood Pressure[b]	> 180 mmHg < 90 mmHg	≥ 20 mmHg increase ≥ 20 mmHg decrease
Diastolic Blood Pressure[b]	> 105 mmHg < 50 mmHg	≥ 15 mmHg increase ≥ 15 mmHg decrease
Orthostatic Hypotension	≥ 20 mmHg decrease in systolic blood pressure and a ≥ 25 bpm increase in heart rate from supine to sitting/standing	Not Applicable (baseline status not considered)
Weight	-	≥ 7% increase ≥ 7% decrease

[a]In order to be identified as potentially clinically relevant, an 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.

[b]As defined in "Supplementary Suggestions for Preparing an Integrated Summary of Safety Information in an Original NDA Submission and for Organizing Information in Periodic Safety Updates," FDA Division of Neuropharmacological Drug Products draft (2/27/87).

SAP 331-10-234

		[a]		[a]
R				
	Tachycardia	≥ 120 bpm	increase of ≥ 15 bpm	
	Bradycardia	≤ 50 bpm	decrease of ≥ 15 bpm	
****				
	Sinus tachycardia[b]	≥ 120 bpm	increase of ≥ 15 bpm	
	Sinus bradycardia[c]	≤ 50 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	
****				
	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	
****				
	Acute or subacute	all	not present → present	
	Old	all	not present → present ≥ 12 weeks post study entry	
****				
	Myocardial Ischemia	all	not present → present	
	Symmetrical T-wave inversion	all	not present → present	
	Increase in QTc	QTcF or QTcN ≥ 450 msec for males, ≥ 470 msec for female		

[a] In order to be identified as potentially clinically relevant, an 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.

[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.

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

CCI

This page is a manifestation of an electronically captured signature

SIGNATURE PAGE

Document Name: SAP 331-10-234

Document Number: CCI [redacted]

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy hh:min) - UTC timezone
PPD	Clinical Approval	14-Mar-2023 12:44:19
PPD	Biostatistics Approval	14-Mar-2023 13:45:35

Otsuka Pharmaceutical Development & Commercialization, Inc.

Investigational New Drug
Brexpiprazole (OPC-34712)

Protocol 331-10-234
IND No. 101,871
EudraCT No. 2017-001447-12

A Multicenter, Randomized, Double-blind, Placebo- and Active-controlled Trial to Evaluate the Efficacy of Brexpiprazole Monotherapy for the Treatment in Adolescents (13-17 years old) With Schizophrenia

Statistical Analysis Plan Addendum to v2.0

Version: Final
Date: 21 Aug 2023

Protocol Date: 05 Jul 2022

Table of Contents

List of Abbreviations and Definitions of Terms

Abbreviation	Definition
FDA	Food & Drug Administration
MedDRA	Medical Dictionary for Regulatory Activities
PANSS	Positive and Negative Syndrome Scale
SAP	Statistical analysis plan
TEAE	Treatment emergent adverse event

1 Introduction

This addendum to the statistical analysis plan (SAP) version 2.0 is to include additional analyses by all race categories and by ethnicity, after the approval of the original SAP (version 2.0, issued on 14 Mar 2023). These additional analyses will allow the review of more granular data on race as proposed by the Food and Drug Administration (FDA) Draft Guidance on Diversity Plans.[1] The statistical methodology, data analysis algorithms and conventions remain the same as those in the original SAP version 2.0.

2 Addition of Analyses by Race and Ethnicity

The SAP version 2.0 included analyses by race (White and All Other Races). In addition, the following analyses will be performed by all categories of race separately (White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, Other, Missing Race) and by ethnicity (Hispanic or Latino, Not Hispanic or Latino, Other, Missing Ethnicity):

- Subject Disposition
- Reasons for Discontinuation
- Demographic Characteristics
- Baseline Disease Characteristics
- Summary of Mean Change from Baseline to Double Blind Period by Study Week in Positive and Negative Syndrome Scale (PANSS) Total Score
- Extent of Exposure to Study Medication During Double Blind Period
- Incidence of Treatment Emergent Adverse Events (TEAEs) during Double Blind Period by System Organ Class and MedDRA Preferred Term

3 References

[1] US Food and Drug Administration. Diversity Plans to Improve Enrollment of Participants from Underrepresented Racial and Ethnic Populations in Clinical Trials Guidance for Industry. Draft Guidance, April 2022. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/diversity-plans-improve-enrollment-participants-underrepresented-racial-and-ethnic-populations

This page is a manifestation of an electronically captured signature

SIGNATURE PAGE

Document Name: P331-10-234_SAP v2.0 Addendum

Document Number: CCI [redacted]

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy hh:min) - UTC timezone
PPD	Clinical Approval	28-Aug-2023 13:13:18
PPD	Biostatistics Approval	21-Aug-2023 14:11:06