Otsuka Pharmaceutical Development & Commercialization, Inc.

Investigational New Drug

Brexpiprazole (OPC-34712)

A Phase 3, Multicenter, Randomized, Double-blind Trial of Brexpiprazole as Combination Therapy with Sertraline in the Treatment of Adults with Post-traumatic Stress Disorder

Protocol No. 331-201-00072
IND No. 117549

Statistical Analysis Plan

Version: 2.0

Date: 24 May 2023

CONFIDENTIAL – PROPRIETARY INFORMATION

SAP 331-201-00072

Table of Contents

SAP 331-201-00072

SAP 331-201-00072

SAP 331-201-00072

List of In-text Tables

List of In-text Figures

SAP 331-201-00072

List of Appendices

CCI

SAP 331-201-00072

1 Introduction

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 study 331-201-00072. All amendments to the protocol and Addendum to the protocol amendment are taken into consideration in developing this SAP. In addition, if the analyses described in the protocols differ from those in this SAP, the methods of the SAP prevail.

2 Study Objectives

Primary: To evaluate the efficacy of brexpiprazole + sertraline in adult subjects with post-traumatic stress disorder (PTSD).

Secondary: To evaluate the safety and tolerability of brexpiprazole + sertraline in adult subjects with PTSD.

3 Trial Details

3.1 Study Design

This is a phase 3, randomized, double-blind, active-controlled, fixed dose trial to evaluate the efficacy, safety, and tolerability of brexpiprazole (fixed doses 2 or 3 mg/day) as combination therapy with sertraline in adult subjects with post-traumatic stress disorder (PTSD). Efficacy of the combination will be measured against the efficacy of sertraline monotherapy. See Figure 3.1-1 for a schematic of the trial design.

CCI

CCI

CCI

CCI

CCI

CCI

[redacted]

5 Data Sets for Analysis and Missing Data

5.1 Data Sets for Analysis

- Enrolled Sample - all subjects enrolled in placebo run-in period
- Randomized Sample - all subjects randomized into this trial
- Enriched Randomized Sample - all subjects who were randomized satisfying the Enriched Subjects Criteria, where the Enriched Subjects Criteria are defined as CAPS-5 total score of at least 27 at the randomization visit (Week 1), and an improvement (in terms of reduction in CAPS-5 total score) in CAPS-5 total score of less than 50% at end of the placebo run-in phase (from baseline [Day 0] to randomization visit [Week 1]).
- Safety Sample - all subjects in the randomized sample who were administered at least 1 dose of double-blind IMP
- FAS - all subjects in the Randomized Sample who took at least one dose of double-blind IMP and have a baseline value (Week 1) and at least one postbaseline evaluation for the Clinician-Administered PTSD Scale for DSM-5 (CAPS-5) total score
- FAS for Enriched Subjects - all subjects in the Enriched Randomized Sample who received at least 1 dose of double-blind IMP, have a baseline value (Week 1) and at least 1 post baseline efficacy evaluation for CAPS-5 total score.

In general, the baseline value of an efficacy endpoint is defined as the last available measurement before the first dose of double-blind IMP, scheduled at the Week 1 visit. Randomization in this trial will be stratified by site, and whether or not a subject meets the Enriched Subjects criteria.

5.2 Handling of Missing Data

The CAPS-5 is a clinician-rated, structured interview designed to assess PTSD diagnostic status and symptoms severity as defined by the DSM-5. This trial will use the CAPS-5 Past Month (at screening only) and CAPS-5 Past Week (at all other assessment timepoints) versions of the scale. The CAPS-5 measures overall PTSD severity by quantifying subjective distress, interpersonal dysfunction, and difficulty with important life tasks associated with each of the DSM-5 PTSD diagnostic criteria..

The CAPS-5 total symptom severity score is calculated by summing severity scores for the 20 DSM-5 PTSD symptoms (items 1-20) from the following categories:

Category B: Intrusion symptoms (5 items);

Category C: Avoidance symptoms (2 items);

Category D: Cognition and mood symptoms (7 items); and

Category E: Arousal and reactivity symptoms (6 items).

If there are more than 1 item in each of categories B, D and E with a missing score, or any item in Category C with a missing score, the CAPS-5 total score is set to missing. For each category (B, D or E) with a missing score, the imputed subscores for that category is the average of the existing scores in the category multiplied by the total number of questions in the category. The CAPS-5 total score will then be imputed by adding all the subscores from categories B, C, D and E.

In general, missing data will be handled by analysis of mixed-effect model repeated measures (MMRM) methodology based on observed-case (OC) data from protocol-specified visits under the assumption of missing at random.

MMRM assumes data are missing at random (MAR), which is a reasonable assumption in longitudinal clinical trials in PTSD.[1] However, the possibility of "missing not at random" (MNAR) data can never be ruled out. As sensitivity analyses, selection model[2], pattern-mixture model[3,4,5,6], and/or shared parameter model[7] will be used to explore data missing mechanisms of MNAR and investigate the response profile of dropout reason.

SAP 331-201-00072

The Observed Cases (OC) data set will consist of actual observations recorded at each visit during Period B and no missing data will be imputed. MMRM, Wu-Bailey, and pattern-mixture model will be performed on the OC dataset.

The Last Observation Carried Forward (LOCF) data set will include data recorded at a scheduled Period B visit or, if no observation is recorded at that visit, data carried forward from the previous scheduled Period B visit. Baseline (Week 1 visit) data will not be carried forward to impute missing values for the LOCF data set.

For categorical response/remission variables, OC analyses will be performed in addition to LOCF analyses. Study center will not be included in the models for OC analyses.

6 Study Conduct

6.1 Subject Disposition, Completion Rate and Reasons for Discontinuation

Subject disposition will be summarized for the Randomized Sample by treatment group, and by center.

Subject completion rate and reasons for discontinuation will be summarized for the Randomized Sample by treatment group. Subject completion rate for Period B will also be summarized by treatment group and week for the Randomized Sample.

6.2 Treatment Compliance

Based on the Investigational medicinal product (IMP) panel of the CRF, compliance in taking IMP is calculated by dividing the number of tablets/capsules taken by the total number of tablets/capsules the patients were scheduled to take during the study period. Compliance is calculated on double-blind IMP for Period B. For lost-to-follow up patients, last IMP end date record will be used as the treatment end date.

6.3 Protocol Deviation

Protocol deviations will be summarized by center and type of deviation for randomized subjects by treatment group. A listing of protocol deviations will be provided.

7 Baseline Characteristics

7.1 Baseline Definition

For analyses of the double-blind treatment period (Period B) data, baseline measurement is defined as the last available measurement prior to the first dose of double-blind IMP, scheduled at the Week 1 visit except for Brief Inventory of Psychosocial Function (B-

SAP 331-201-00072

IPF) CCI [redacted] which is at the Day 0 visit.

7.2 Demographic Characteristics

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, range and standard deviation will be used to describe continuous variables such as age. Frequency distributions will be tabulated for categorical variables such as race.

7.3 Medical and Psychiatric History and Treatment for PTSD

A summary of medical and psychiatric history will be presented for the Randomized Sample (by treatment group and overall). Summarized data will include the number of years since index traumatic event that led to development of PTSD and the number of years since onset of first symptoms leading to the diagnosis of PTSD. A summary of PTSD treatment based on the Emory Treatment Resistance Interview (E-TRIP) will also be presented for the Randomized Sample (by treatment group and overall).

7.4 Neuropsychiatric Diagnosis

A summary of the MINI International Neuropsychiatric Interview (M.I.N.I.) will be presented for the Randomized Sample (by treatment group and overall). Summarized will be the number and percentage of patients who meet each diagnosis criteria, and number and percentage of patients with each primary diagnosis.

A summary of the Life Events Checklist (LEC-5) extended version will also be presented for the Randomized Sample (by treatment group and overall). Summarized will be the number and percentage of patients with each response to each question in Part 1.

In addition, a summary of the Ohio State University Traumatic Brain Injury Identification (OSU TBI-ID) method will be presented for the Randomized Sample (by treatment group and overall). Summarized will be the number and percentage of patients' answers (yes/no) to each question in Part 1 and the first question in Part 3.

7.5 Baseline PTSD Evaluation

For the Randomized Sample, baseline PTSD scale evaluation will be summarized by treatment group and overall. The mean, median, range and standard deviation will be used to summarize the assessments of: CAPS-5 total score, Clinical Global Impression – Severity (CGI-S) Score, Brief Inventory of Psychosocial Function (B-IPF) score, PTSD Checklist for DSM-5 (PCL-5) total score, Hospital Anxiety and Depression Scale

(HADS) subscale scores (depression and anxiety), and CCI [redacted]

8 Efficacy Analysis

CCI [redacted]

All efficacy analyses will be performed on the FAS for Enriched Subjects and on the FAS unless specified otherwise, with FAS for Enriched Subjects being primary. Statistical comparisons are based on 2-sided, 0.05 significance level.

8.1 Primary Efficacy Endpoint

The primary efficacy endpoint is the change from baseline to the Week 10 Visit in the CAPS-5 total score.

8.1.1 Primary Estimand

The objective of the primary analysis is to evaluate the efficacy of brexpiprazole + sertraline in adult subjects with PTSD.

The primary estimand defining the treatment effect of interest in the trial uses the hypothetical strategy specified in the draft ICH E9 (R1) Addendum. The objective of the primary analysis is to evaluate the efficacy of brexpiprazole as combination treatment with sertraline in adult subjects with PTSD versus sertraline monotherapy. The estimand, or target of estimation, following the hypothetical strategy is the pharmacological effect seen, had no withdrawals occurred. This hypothetical estimand is justifiable in this case since the focus is on the pharmacological effect of the drug additional to nonspecific effects. Subjects who withdraw from a symptomatic IMP treatment either could have lost their treatment effect, had the subjects not taken any other symptomatic medication after withdrawal, or could have their treatment effect been masked, had the subjects taken other symptomatic medication after withdrawal. This means that any observations taken after subjects stop IMP will most likely not contribute relevant information about the pharmacological 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 PTSD or other prohibited medications. Evaluations occurring more than 7 days will be excluded from the analysis. In the case of terminal or lost to follow-up events, no ET evaluations would be expected, and only scheduled assessments performed before such an event has occurred.

SAP 331-201-00072

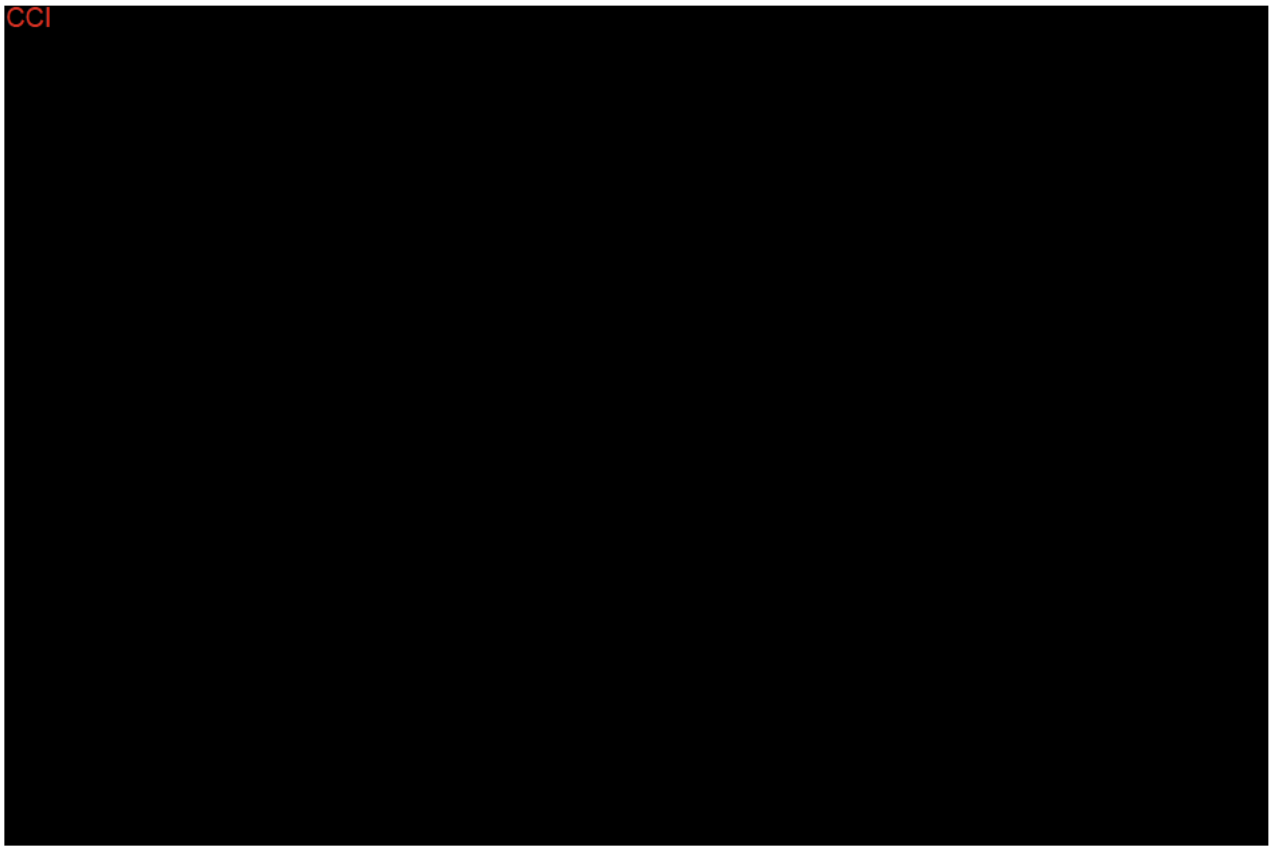
The primary estimand for this trial is defined by the following components:

- Target Population: FAS for Enriched Subjects
- Endpoint: Change from baseline to Week 10 in the CAPS-5 total score
- Intercurrent Events: Premature treatment discontinuation
- Measure of Intervention Effect: Difference in endpoint means between brexpiprazole and sertraline combination and sertraline monotherapy.

In this hypothetical strategy, the event of withdrawing IMP is considered missing at random (MAR), and the primary endpoint of the trial could be considered as a combination of the responses of on-treatment completers at Week 10 and the imputation of the endpoint to Week 10 following the trend in each treatment group using the Mixed Model Repeated Measurements (MMRM) method 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 in Section 8.1.2. Under the MAR assumption, MMRM provides an unbiased estimate of treatment effect for the treatment period. Analyses with missing values imputed by multiple imputation under MNAR, and other methods will be performed as sensitivity analyses.

It is assumed that the placebo effect is reduced in the FAS for Enriched Subjects comparing to the FAS.

CCI



CCI

CCI

where baseline is the CAPS-5 total score at end of Period A (Week 1 visit).

8.1.4 Sensitivity Analyses

8.1.4.1 Sensitivity Analyses for Missing at Random (MAR) Assumption

The possibility of "missing not at random" (MNAR) data will also be considered. As sensitivity analyses, selection model, pattern-mixture model, and/or shared parameter model will be used to explore data missing mechanisms of MNAR and investigate the response profile of dropout reason.

SAP 331-201-00072

Pattern Mixture Models 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 3 scenarios: 1) Dropout reasons due to either AE or lack of efficacy (LOE) as MNAR; 2) Dropout reasons due to either AE or LOE or subject withdrew consent as MNAR; 3) All dropouts as MNAR using both 1) Delta adjustment imputation method which is to departure from MAR assumption by progressively increasing the delta until conclusion from the primary analysis is overturned and 2) Placebo-based imputation methods in which 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, this benefit is carried into subsequent imputed values, but will diminish over time in accordance with the correlation structure. Traditionally the dropout mechanisms are divided into three types (Little, 1995): (1) Missing Completely at Random (MCAR), in which the probability of dropout does not 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[2] 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.

As sensitivity analyses for missing at random (MAR) assumption, analyses for missing not at random (MNAR) will be carried out. Pattern Mixture Models (PMM) based on Multiple Imputation (MI) with mixed missing data mechanisms will be used to investigate the response profile of dropout patients by last dropout reason under MNAR mechanism for the following 3 scenarios:

1) Dropout reasons due to either AE or LOE as MNAR
2) Dropout reasons due to either AE or LOE or subject withdrew consent as MNAR
3) All dropouts as MNAR

SAP 331-201-00072

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 5 points and/or the observed treatment difference between the active arm and placebo from the primary analysis of MMRM model until 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 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 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 MMRM model in the primary analysis to analyze the completed data using PROC MIXED on the multiple imputed data
5) Obtaining the overall results using PROC MIANALYZE

The details of the imputation model under the MI procedure and related SAS codes are provided in Appendix 4.

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 group 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.

In addition, model based MNAR methods such as the shared parameter model[7] and random coefficient pattern mixture model[5] will be also performed.

8.1.4.2 Sensitivity Analyses for Violation of Normality Assumption

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

In the case of gross violations of the normality assumptions, nonparametric van Elteren test[8] will be performed to compare treatment effect at Week 10 on Multiple Imputation (MI) data. The van Elteren test is a generalized Cochran-Mantel-Haenszel (CMH)

SAP 331-201-00072

procedure useful for stratified continuous data in non-normal setting. It belongs to a general family of Mantel-Haenszel mean score tests. The test is performed via SAS procedure PROC FREQ, by including CMH2 and SCORES=MODRIDIT options in the TABLE statement. The stratification factor is trial center and previous pharmacological treatment intervention for PTSD (Yes/No).

In addition, other methods that are robust to distributional assumption will also be performed to provide different views on the primary efficacy result, these include generalized estimating equations (GEE), weighted GEE (WGEE), and MI-robust regression.[9]

For MI-van Elteren test and MI-robust regression, imputation datasets will be generated with SAS MI procedure, each dataset will be analyzed, then an overall estimate is derived with SAS MIANALYZE procedure.

8.1.4.3 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 FAS for Enriched Subjects to evaluate the sensitivity of the primary and key secondary analysis results to the impact of the pandemic. The same model (e.g., 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) An MMRM analysis excluding the virtual assessments based on the FAS for Enriched Subjects.
2) An MMRM analysis using the non-COVID data set based on the FAS for Enriched Subjects. 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 12/ET date, the date before the first virtual assessment 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 or virtual assessments during the double-blind treatment period.
3) An MMRM analysis based on the non-COVID Sample. The non-COVID Sample comprises those subjects in the FAS for Enriched Subjects who did not have any virtual assessments nor COVID-19 related protocol deviations.

In addition, demographics and baseline characteristics by subgroup of subjects with or without any virtual visits will be provided.

SAP 331-201-00072

8.1.5 Subgroup Analyses

CCI

[redacted]

All subgroup analyses will be conducted using the same MMRM analysis as for the primary efficacy analysis except that the fixed class effect term for trial center and previous pharmacological treatment intervention for PTSD (Yes/No) will not be included in the model.

Interaction effects of treatment-by-subgroup will be assessed at Week 10 for the subgroups identified in the previous paragraph. MMRM analyses will be performed by adding 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.

8.2 Key Secondary Efficacy Endpoint

The key secondary efficacy endpoints are the change from baseline to Week 10 in the Clinical Global Impression – Severity (CGI-S) score and the change from baseline (at Day 0) to Week 12 in Brief Inventory of Psychosocial Function (B-IPF) score. Both endpoints will be analyzed using an MMRM model similar to that prespecified for the primary efficacy endpoint, correcting for the relevant values at randomization. The key secondary efficacy endpoints will be tested at the same level as the primary endpoint.

8.3 Control of Experiment-wise Type 1 Error

To control the family-wise type I error when testing for both the primary efficacy endpoint and the key secondary efficacy endpoints, the following testing procedure will be applied.

First, on the primary efficacy endpoint, the family-wise error rate (FWE) of multiple comparisons is controlled by testing the difference between the average effect of 2 mg and 3 mg of brexpiprazole + sertraline combination versus sertraline monotherapy first, at an alpha level of 0.05. If this global test is significant, then comparisons for each group (2 mg of brexpiprazole + sertraline combination and 3 mg of brexpiprazole + sertraline combination) versus sertraline monotherapy will be performed at significant level 0.05,

separately. A primary comparison is significant at the 0.05 FWE significance level if the global test and either individual test of brexpiprazole dose (2mg, 3mg) + sertraline combination versus sertraline monotherapy are significant at 0.05.[10]

Secondly, the analysis of the first key secondary endpoint (CGI-S score) will be conducted if both global test and both individual tests of brexpiprazole dose (2mg, 3mg) + sertraline combination versus sertraline monotherapy are all significant for the primary endpoint. In this case, the alpha level used in the analysis of the first key secondary efficacy endpoint is 0.05. In order to handle multiplicity issue in the analysis of the first key secondary efficacy endpoint, a test procedure similar to the one used in the analysis of the primary efficacy endpoint will be adopted. That is, a global test in the difference between the average effect of 2 mg and 3 mg of brexpiprazole + sertraline combination versus sertraline monotherapy will be conducted first, at an alpha level of 0.05. If this global test is significant, the two individual comparisons, 2 mg of brexpiprazole + sertraline combination versus sertraline monotherapy and 3 mg of brexpiprazole + sertraline combination versus sertraline monotherapy, will be performed in a parallel way, each at significant level 0.05. A comparison is significant at the 0.05 FWE significant level if the global test and either individual test of the comparison are significant at 0.05 level for the first key secondary endpoint.

Thirdly, the analysis of the second key secondary endpoint (B-IPF score) will be conducted if the global test and both individual tests of brexpiprazole dose (2mg, 3mg) + sertraline combination versus sertraline monotherapy are all significant for the primary endpoint and the first key secondary endpoint. In this case, the alpha level used in the analysis of the second key secondary efficacy endpoint is 0.05. In order to handle multiplicity issue in the analysis of the second key secondary efficacy endpoint, a test procedure similar to the one used in the analysis of the primary efficacy endpoint and first key secondary endpoint will be adopted. That is, a global test in the difference between the average effect of 2 mg and 3 mg of brexpiprazole + sertraline combination versus sertraline monotherapy will be conducted first, at an alpha level of 0.05. If this global test is significant, the two individual comparisons, 2 mg of brexpiprazole + sertraline combination versus sertraline monotherapy and 3 mg of brexpiprazole + sertraline combination versus sertraline monotherapy, will be performed in a parallel way, each at significant level 0.05. A comparison is significant at the 0.05 FWE significant level if the global test and either individual test of the comparison are significant at 0.05 level for the secondary key secondary endpoint.

8.4 Other Efficacy Endpoints

Other efficacy endpoints are as follows:

1) Change from baseline to Week 10 in PTSD Checklist for DSM-5 (PCL-5) score during the double-blind randomization period
2) Change from baseline to Week 10 in the Hospital Anxiety and Depression Scale - Anxiety subscale (HADS-A) and Hospital Anxiety and Depression Scale - Depression subscale (HADS-D) score during the double-blind randomization period
3) Response defined by decrease ≥ 30% from baseline to Week 10 in the CAPS-5 total score during the double-blind randomization period
4) Change from baseline to Week 10 in the CAPS-5 subscales/domain scores during the double-blind randomization period

The continuous efficacy endpoints will be analyzed using MMRM model similar to that of pre-specified for change from baseline in CAPS-5 total score, correcting for the relevant values at randomization.

The response in variables will be evaluated by the Cochran-Mantel-Haenszel (CMH) General Association Test controlling, in LOCF analysis, for study center and previous pharmacological treatment intervention for PTSD (Yes/No).

All the other efficacy variables will be evaluated at a nominal 0.05 level (2-sided) without adjusting for multiplicity.

CCI

SAP 331-201-00072

CCI

[redacted]

SAP 331-201-00072

CCI

[redacted]

9 Safety Analysis

Standard safety variables to be analyzed include adverse events (AEs), clinical laboratory tests, vital signs, body weight, waist circumference, BMI, 12-lead electrocardiograms (ECGs), and physical examinations. 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).

CCI

[redacted]

9.1 Adverse Events

All adverse events will be coded by System Organ Class (SOC) and Preferred Term (PT) according to the Medical Dictionary for Regulatory Activities (MedDRA). AEs that are sex-specific, e.g., ovarian cancer, will have their incidence rates evaluated for the specific sex.

CCI

[redacted]

SAP 331-201-00072

CCI

[redacted]

Incidence of TEAEs by SOC and PT will be summarized for sex, race, and age subgroups.

Unless otherwise specified, in general, analysis of safety data will be performed on observed case and for last visit.

9.2 Clinical Laboratory Tests

Summary statistics for mean and mean change from baseline in the routine clinical laboratory measurements and prolactin concentrations will be provided by treatment group and by visit.

CCI

[redacted]

9.2.1 Potentially Liver Injury Related Laboratory Test

Total bilirubin level will be checked for any subject with increased (alanine aminotransferase (ALT) or aspartate aminotransferase (AST) levels ≥ 3 times the upper normal limits (ULN) or baseline.

CCI

[redacted]

9.2.2 Metabolic Change

In addition to mean change from baseline, incidence of treatment-emergent significant changes in fasting lipids and fasting glucose and treatment-emergent significant changes in metabolic syndrome will be summarized by treatment group using the criteria in Table 9.2.2-1 and Table 9.2.2-2, respectively.

Table 9.2.2-1 Criteria for Treatment-Emergent significant Change in Lipids and Glucose		
Lab Parameter	**Baseline[1]**	**Anytime Postbaseline**
LDL Direct, Fasting (mg/dL)	Borderline 100 – < 160	High ≥ 160
	Normal/Borderline < 160	High ≥ 160
	Normal < 100	Borderline/High ≥ 100
	Any Value	Increased ≥ 30
HDL Cholesterol, Fasting (mg/dL)	Normal ≥ 40	Low < 40
	Any Value	Decreased ≥ 20
Triglycerides, Fasting (mg/dL)	Normal < 150	High 200 – < 500
	Borderline 150 – < 200	High 200 – < 500
	Normal/Borderline < 200	High 200 – < 500
	Normal < 150	Borderline/High/Very High ≥ 150
	Any Value	Increased ≥ 50
Glucose Fasting, Serum (mg/dL)	Normal < 100	High ≥ 126
	Impaired 100 – < 126	High ≥ 126
	Normal/Impaired < 126	High ≥ 126
	Any Value	Increased ≥ 10

[a]Baseline is calculated from Week 1.

Table 9.2.2-2 Criteria for Treatment-Emergent Metabolic Syndrome	
Description	**Anytime Postbaseline[a]**
Central Obesity	Waist Circumference ≥ 102 cm (male), ≥ 88 cm (female)
Dyslipidemia	Triglycerides ≥ 150mg/dL
Dyslipidemia	HDL < 40mg/dL (male), < 50mg/ dL (female)
Supine Blood Pressure	Systolic ≥ 130 mm Hg and Diastolic ≥ 85 mm Hg
Glucose Fasting, Serum	≥ 100 mg/dl
Metabolic Syndrome	Met 3 or more of the above criteria at a visit

[a]Baseline is calculated from Week 1.

9.3 Vital Signs

CCI

CCI

[redacted]

9.4 12-Lead ECG

Summary statistics and incidence of potentially clinically relevant changes will be provided for ECG parameters.

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 in the clinical study report:

1) QTcB is the length of the QT interval corrected for heart rate by the Bazett formula: $QTcB=QT/(RR)^{0.5}$ and
2) QTcF is the length of the QT interval corrected for heart rate by the Fridericia formula: $QTcF=QT/(RR)^{0.33}$
3) QTcN is the length of the QT interval corrected for heart rate by the FDA Neuropharm Division formula: $QTcN=QT/(RR)^{0.37}$

CCI

[redacted]

SAP 331-201-00072

CCI

[redacted]

9.5 Physical Examinations

By-patient listings will be provided for physical examination.

9.5.1 Body Weight, Waist Circumference and BMI

CCI

[redacted]

Percentages of patients showing significant weight gain (≥ 7 % increase in weight), as well as percentages of patients showing significant weight loss (≥ 7 % decrease in weight) from end of Period A to Week 12 (OC and LOCF) will be analyzed using CMH General Association Test.

Body mass index is defined as weight in kilograms divided by the square of height in meters.

9.6 SAS, AIMS, and BARS

CCI

[redacted]

CCI

[redacted]

9.9 Extent of Exposure

The start date of double-blind IMP will be the first day of double-blind dosing. The number and percentage of patients who receive double-blind IMP (brexpiprazole or sertraline), will be presented by week and by treatment group. Each dosing week will be based on the actual week, i.e., Day 1-7 in Week 1, Day 8-14 in Week 2. This summary will be performed on the Safety Sample.

SAP 331-201-00072

The mean daily dosage will be summarized by week and treatment group using descriptive statistics. The mean daily dosage per patient per week will be determined for each week of the study. 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 patients receiving double-blind IMP, and the mean and range of the mean daily dose for each week.

All the exposure analyses will be done for brexpiprazole or sertraline, separately.

10 Conventions

10.1 Study Visit Windows

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 (CAPS-5, CGI-S, B-IPF, PCL-5, HADS, SAS, AIMS, and BARS). This derived study window variable will be named as WEEK and will be footnoted. In listings, it will be listed along with the CRF study visit.

CCI

SAP 331-201-00072

10.2 Pooling of small centers

CCI [redacted] 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 among the largest centers, the center with the smallest center code will be selected. In case of ties in center size among the smallest centers, the center with the largest 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 centers exist, they will be pooled with the smallest non-small center.

10.3 Scales: Rules for Scoring and Handling of Missing Data

10.3.1 CAPS-5

The CAPS-5 is a clinician-rated, structured interview designed to assess PTSD diagnostic status and symptoms severity as defined by the DSM-5. This trial will use the CAPS-5 Past Month (at screening only) and CAPS-5 Past Week (at all other assessment timepoints) versions of the scale. The CAPS-5 measures overall PTSD severity by quantifying subjective distress, interpersonal dysfunction, and difficulty with important life tasks associated with each of the DSM-5 PTSD diagnostic criteria.

The CAPS-5 total symptom severity score is calculated by summing severity scores for the 20 DSM-5 PTSD symptoms (items 1-20) from the following categories:

Category B: Intrusion symptoms (5 items);

Category C: Avoidance symptoms (2 items);

Category D: Cognition and mood symptoms (7 items); and

Category E: Arousal and reactivity symptoms (6 items)

SAP 331-201-00072

If there are more than 1 item in each of Categories B, D and E with a missing score, or any item in Category C with a missing score, the CAPS-5 total score is set to missing. For each category (B, D or E) with a missing score, the imputed subscores for that category is the average of the existing scores in the category multiplied by the total number of questions in the category. The CAPS-5 total score will then be imputed by adding all the subscores from Categories B, C, D and E.

10.3.2 CGI-S

The severity of illness for each subject will be rated using the CGI-S. CGI-S items 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; 7 = among the most extremely ill patients. The score 0 (= not assessed) will be set to missing. The CGI-S is therefore a 7-point scale from 1 through 7.

10.3.3 B-IPF

The Brief Inventory of Psychosocial Function (B-IPF) is a short patient-reported questionnaire consisting of 7 questions which measure PTSD-specific psychosocial function on a 7-point Likert scale (0 = not at all to 6 = very much, and a not applicable option) with a recall period of 30 days. The B-IPF measures the concepts of romantic relationships, parenting, family, friendships and socializing, work, education, and self-care.

The B-IPF score is calculated by summing the scored items to create a total score, dividing the total score by the maximum possible score, based on the number of items scored, and multiplying by 100. Thus, the B-IPF score represents an overall index of functioning, with higher scores indicating greater impairment.[12] If there are more than one item with a missing score, the B-IPF score is set to missing.

10.3.4 PCL-5

The PCL-5 is a checklist of problems that people sometimes have in response to a very stressful experience. Subjects need to indicate a number to the right of each problem to indicate how much they have been bothered by that problem in the past month. The scale rates items from 0 (not at all), 1 (a little bit), 2 (moderately), 3 (quite a bit), and 4 (extremely).

The PCL-5 total score is calculated by summing severity scores for scores from all 20 tems. The PCL-5 total score will be un-evaluable if fewer than 16 of the 20 items are recorded. If 16 to 19 of the 20 items are recorded, the PCL-5 total score will be the mean of the recorded items multiplied by 20 and then rounded to the first decimal place.

SAP 331-201-00072

10.3.5 HADS

The HADS is a subject-rated scale designed to screen for anxiety and depressive states in medical subjects. The HADS consists of 2 subscales: The D-scale measures depression and the A-scale measures anxiety. Each subscale contains 7 items, and each item is rated from 0 (absent) to 3 (maximum severity). The score of each subscale is the sum of scores from each item in the subscale. Each subscale ranges from 0 to 21, The subscales will be analyzed separately. If there are more than 1 item with missing score for a give subscale, the score of the subscale is set to be missing. If 6 of the 7 items are recorded, the score of the subscale will be the mean of the recorded items multiplied by 7 and then rounded to the first decimal place.

CCI

10.3.7 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 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.

SAP 331-201-00072

10.3.8 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 seven items. The possible total scores are 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.

10.3.9 BARS

The BARS 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.

10.3.10 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.

11 References

1 Siddiqui O, Hung JH, O'Neill R. MMRM vs. LOCF: A comprehensive comparison based on simulation study and 25 NDA datasets. J Biopharmaceutical Stats. 2009;19(2):227-246.

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

3 Little RJA. Pattern-mixture models for multivariate incomplete data. J Am Stat Assoc. 1993;88(421):125-134.

4 Little RJA. Modeling the drop-out mechanism in repeated measures studies. J Am Stat Assoc. 1995;90(431):1112-1121.

5 Hedeker D, Gibbons RD. Application of random effects pattern-mixture models for missing data in longitudinal studies. Psychological Methods. 1997;2(1):64-78.

6 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):165-1681.

7 Wu MC, Bailey KR. Estimation and comparison of changes in the presence of informative right censoring: Conditional linear model. Biometrics. 1989;45(3):939-955.

8 van Elteren PH. On the combination of independent two sample tests of Wilcoxon. Bull Int Stat Inst. 1960;37:351-361.

9 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(4):1250-1259.

10 Kong L, Koch G, Liu T, Wang H. Performance of some multiple testing procedures to compare three doses of a test drug and placebo. Pharma Stat. 2005;4(1):25-35.

11 Cortina JM, Nouri H. Effect size for ANOVA designs. Sage Publications. 2000;129:4-6.

12 Kleiman SE, Bovin MJ, Black SK, Rodriguez P, Brown LG, Brown ME, et al. Psychometric properties of a brief measure of posttraumatic stress disorder-related impairment: the brief inventory of psychosocial functioning. Psychol Serv. 2018.

CCI

CCI

CCI

CCI

CCI

CCI

This page is a manifestation of an electronically captured signature

SIGNATURE PAGE

Document Name: 331-201-00072_Statistical_Analysis_Plan_V2

Document Number: CCI [redacted]

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy hh:min) - UTC timezone
PPD	Clinical Approval	24-May-2023 22:24:03
	Biostatistics Approval	24-May-2023 22:30:59