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. 331-201-00071
IND No. 117549

Statistical Analysis Plan

Version: 2.0

Date: 22 May 2023

CONFIDENTIAL – PROPRIETARY INFORMATION

SAP 331-201-00071

Table of Contents

SAP 331-201-00071

List of In-text Tables

List of In-text Figures

List of Appendices

SAP 331-201-00071

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 331-201-00071. 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 trial to evaluate the efficacy, safety, and tolerability of brexpiprazole (flexible dose 2 - 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.

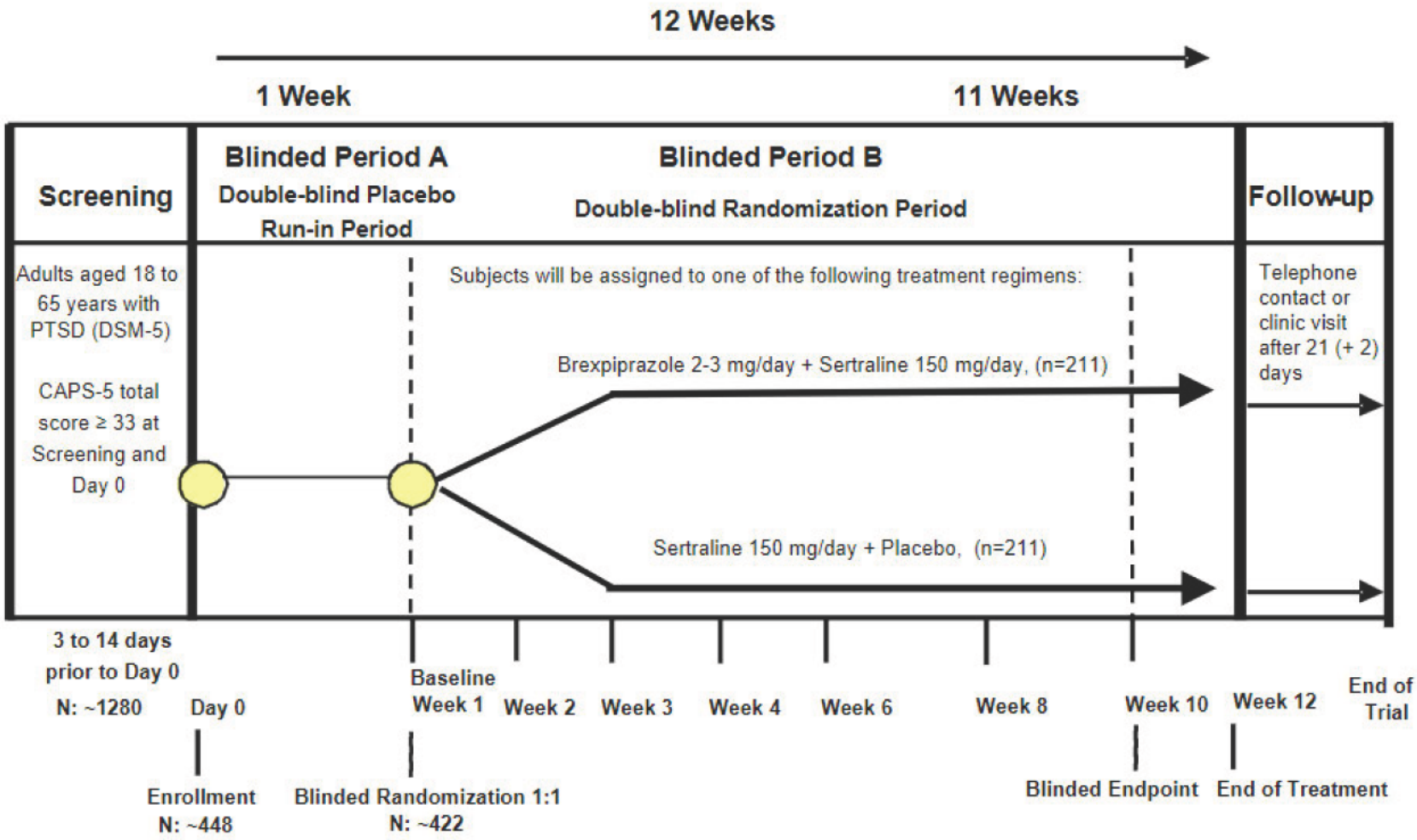

Figure 3.1-1 Trial Design Schematic

SAP 331-201-00071

3.1.1 Screening

Screening will begin when consent has been obtained. The screening visit will take place between Day −14 and Day −1 prior to enrollment and subjects will participate in screening activities for a minimum of 3 days. In the event that a subject cannot be enrolled prior to expiration of the 14-day screening period, an additional 14-day extension of screening may be requested from the medical monitor.

The purpose of the screening period is to assess eligibility criteria at 1 or more visits and to washout prohibited concomitant pharmacotherapy, if applicable. An eSource method will be used to obtain an identification (ID) number for each subject with documented consent.

Subjects will be between 18 and 65 years of age, inclusive, at the time of screening and will have a diagnosis of PTSD as defined by DSM-5 criteria and confirmed by the Mini International Neuropsychiatric Interview (MINI). All subjects must agree to discontinue all prohibited medications during the screening period in order to meet the protocol-specified washout periods.

3.1.2 Blinded Period A - Double-blind Placebo Run-In Period

Subjects who are deemed eligible for trial participation will enter into the treatment period. Subjects will be enrolled into a 1-week, double-blind placebo run-in period at baseline (Day 0).

3.1.3 Blinded Period B - Double-blind Randomization Period

All subjects will be randomized regardless of their response during the placebo run-in period. Placebo-responders will be randomized and will participate in the trial to maintain blinding and to allow for collection of additional safety data. Their efficacy data will not be included in the primary analysis.

At Week 1, subjects will be randomized in a 1:1 ratio to 1 of the following double-blind treatment regimens:

- brexpiprazole + sertraline
- sertraline + placebo

Subjects will receive 11 weeks of treatment with their assigned regimen.

Subjects will attend visits at Weeks 1, 2, 3, 4, 6, 8, 10, and 12/early termination (ET) during the randomization period.

3.1.4 Post-treatment Follow-up Period

For any subject who discontinues the trial early, the site should make every effort to complete the ET evaluations as soon as possible and prior to the subject starting any new medication or treatment. All subjects (completers and early withdrawals) will be monitored for safety events via telephone or clinic visit (at the investigator's discretion) 21 (+ 2) days after the last dose of IMP. For any subject who withdraws due to a serious adverse event (SAE), every effort should be made to ensure that the safety follow-up is performed face-to-face. Given that this may not be feasible in all circumstances, safety follow-up via phone is permitted in cases where face-to-face follow up cannot be performed.

3.2 Trial Treatments

During the placebo run-in period, all subjects will receive one brexpiprazole-matched placebo tablet and 2 sertraline-matched placebo capsules daily. During the randomization period, all subjects will receive one active or matched placebo brexpiprazole tablet and two active sertraline capsules.

3.2.1 Trial Treatment Administration

Treatment administration information is discussed in the full protocol.

3.2.2 Combination of Brexpiprazole plus Sertraline

All subjects will receive placebo from baseline to Week 1. At Week 1, subjects assigned to brexpiprazole + sertraline will begin dosing with active investigational medicinal product (IMP) and will have their dose increased up to the Week 3 visit in a fixed forced titration sequence. The Week 1 through Week 3 assigned doses will be as follows:

- Week 1: 0.5 mg/day brexpiprazole + 50 mg/day sertraline
- Week 2: 1 mg/day brexpiprazole + 100 mg/day sertraline
- Week 3: 2 mg/day brexpiprazole + 150 mg/day sertraline

At the Week 4 visit, the investigator will select whether the dose will be maintained at 2 mg/day brexpiprazole + 150 mg/day sertraline or increased to 3 mg/day brexpiprazole + 150 mg/day sertraline based on the subject's efficacy and tolerability. Subjects who maintain the 2 mg/day brexpiprazole + 150 mg/day sertraline dose will be assigned this dose at all subsequent trial visits (Weeks 4 through 12). Subjects who are assigned an increased dose of 3 mg/day brexpiprazole + 150 mg/day sertraline will maintain this dose at all subsequent trial visits (Weeks 4 through 12), but will be allowed a one-time dose

decrease at a scheduled or unscheduled visit for reasons of tolerability up to the Week 6 visit. After a dose reduction, subjects will maintain the decreased dose for the remainder of the trial. All other subjects unable to tolerate their assigned dose will be discontinued from the trial. The dosing schedule for subjects randomized to the brexpiprazole + sertraline arm is shown in Table 3.2.2-1.

Table 3.2.2-1 Dosing Schedule for Brexpiprazole + Sertraline

	Day 0	Week							
		1	2	3	4	6	8	10	12
Brexpiprazole (mg)	-	0.5	1	2	2 or 3	2 or 3	2 or 3	2 or 3	2 or 3
Brexpiprazole or matching placebo (number of tablets)	1	1	1	1	1	1	1	1	1
Sertraline (mg)	-	50	100	150	150	150	150	150	150
Sertraline or matching placebo (number of capsules)	2	2	2	2	2	2	2	2	2

Note: Subjects will receive 2 capsules + 1 tablet at all times.

3.2.3 Sertraline Administration

All subjects will receive placebo from baseline to Week 1. At Week 1, subjects assigned to sertraline will begin dosing with active IMP and will have their dose increased up to the Week 3 visit in a fixed forced titration sequence. The Week 1 through Week 3 assigned doses will be as follows:

- Week 1: 50 mg/day sertraline + placebo
- Week 2: 100 mg/day sertraline + placebo
- Week 3: 150 mg/day sertraline + placebo

At the Week 4 visit, the investigator can select whether the dose will be maintained or increased in a double-blind fashion, however no actual dose increase will occur as the dose is fixed at 150 mg/day sertraline. Subjects will be maintained at the 150 mg/day sertraline dose at all subsequent trial visits (Weeks 4 through 12). Subjects unable to tolerate their assigned dose will be discontinued from the trial. The dosing schedule for subjects randomized to the sertraline arm is shown in Table 3.2.3-1.

SAP 331-201-00071

Table 3.2.3-1 Dosing Schedule for Sertraline

	Day 0	Week							
		1	2	3	4	6	8	10	12
Sertraline (mg)	-	50	100	150	150	150	150	150	150
Sertraline or matching placebo (number of capsules)	2	2	2	2	2	2	2	2	2
Placebo for brexpiprazole (number of tablets)	1	1	1	1	1	1	1	1	1

Note: Subjects will receive 2 capsules + 1 tablet at all times.

4 Sample Size

A total of 284 randomized and evaluable subjects meeting the enrichment criteria (142 subjects each in the combination arm and the sertraline arm) will provide 85% power at a 2-sided alpha level of 0.05 to detect a treatment difference of −5.0 points in the primary efficacy endpoint with a standard deviation (SD) of 14. The difference of −5.0 points is based on phase 2 data and clinical experience with the population of PTSD patients. In order to have 284 subjects randomized in the Full Analysis Set (FAS) for Enriched Subjects Sample (defined in Section 5.1), and further adjusting for a 5% non-evaluability/dropout impact, a total of 422 subjects (211 subjects each in the combination arm and the sertraline arm) will be randomized.

It is expected that approximately 448 subjects (224 in each group) will be enrolled at Week 0 in order to randomize 422 subjects (211 in each group).

5 Data Sets for Analysis and Missing Data

5.1 Data Sets for Analysis

The following analysis samples are defined for this trial:

Enrolled Sample: comprises all subjects enrolled in placebo run-in period.

Randomized Sample: comprises all subjects randomized into this trial.

Enriched Randomized Sample: comprises all subjects who were randomized satisfying the Enriched Subjects Criteria, where the Enriched Subjects Criteria are defined as CAPS-5 total score is 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 is less

than 50% at end of the placebo run-in phase (from baseline (Day 0) to randomization visit (Week 1)).

Safety Sample: comprises all subjects in the randomized sample who were administered at least one dose of double-blind IMP

Full Analysis Set (FAS): comprises 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 post baseline evaluation for the Clinician-Administered PTSD Scale for DSM-5 (CAPS-5) total score.

FAS for Enriched Subjects - comprises those 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 Week 1 visit.

Randomization in this trial will be stratified by site, and whether or not they meet 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 sub-score 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 sub-scores 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.

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.

SAP 331-201-00071

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-IPF) and CCI 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] physical component summary score and mental component summary score and subscales of the CCI [redacted]

8 Efficacy Analysis

For analysis of Period B data, baseline is defined as the last available measurement prior to the first dose of double-blind IMP, scheduled at the Week 1 visit.

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 non-specific 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.

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

8.1.2 Primary Efficacy Analysis

The primary analysis will be performed on the FAS for Enriched Subjects. The primary efficacy analysis will be performed by fitting a mixed-effect model repeated measure (MMRM) analysis with an unstructured variance covariance structure in which the

change from the baseline in CAPS-5 total score during the double-blind treatment phase will be the dependent variable based on the observed cases (OC) data set. The OC data set will consist of actual observations recorded at each visit during the double-blind treatment period and no missing data will be imputed. The model will include fixed class effect terms for treatment, trial site, visit, previous pharmacological treatment intervention for PTSD (Yes/No), and an interaction term of treatment by visit, and include the interaction term of baseline values of CAPS-5 total score by visit as a covariate. All scheduled visits after baseline during Period B, including Week 12, will be included in the model but the primary comparison will be performed at the Week 10 visit. The primary comparison between the brexpiprazole + sertraline group and the sertraline group at the Week 10 visit during Period B will be estimated as the difference between Least Squares (LS) means utilizing the computing software SAS procedure PROC MIXED. The comparison between brexpiprazole + sertraline group and sertraline group will be tested at a significance level of 0.05 (2-sided).

In case there is a convergence problem with 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.

In the case of gross violations of the linear model assumptions, nonparametric van Elteren test stratified by center will be performed to compare treatment effect at Week 10 of the double-blind treatment period (Period B) on Multiple Imputation (MI) data.

8.1.3 Technical Computation Details for Primary Efficacy Analysis

The SAS code for the PROC MIXED procedure to carry out the above MMRM analysis with an unstructured variance covariance structure is illustrated as follows:

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

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

SAP 331-201-00071

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.

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

Most of MNAR methods[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 three 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

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 ANCOVA model includes treatment and study center as main effects, and baseline value as a covariate.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) 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.

8.1.5 Subgroup Analyses

Subgroup analyses of change from baseline (Week 1 visit) in CAPS-5 total score to every study week in Period B will be performed by the following factors:

- Sex (Based on the biological status at birth)
- Race (White and All Other Races)
- Age (Age<55, and Age ≥55)
- Previous pharmacological treatment intervention for PTSD (Yes/No)

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 Analysis

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

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 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.

To control the family-wise type I error when testing for both the primary efficacy endpoint and the key secondary efficacy endpoints, a stepwise hierarchical testing procedure is applied. The statistical testing will be performed in the following order. The statistical test between brexpiprazole and placebo group for an endpoint after the first one will be performed only when the nominal p-value reaches significance level at 0.05 (2-sided) for all the preceding endpoints:

1) Primary efficacy endpoint for the comparison of combination therapy versus sertraline based on the FAS for Enriched Subjects
2) The first key secondary endpoint of the change from baseline to Week 10 in the Clinical Global Impression - Severity (CGI-S) score for the comparison of combination therapy versus sertraline based on the FAS for Enriched Subjects
3) The second key secondary endpoint of the change from baseline to Week 12 in Brief Inventory of Psychosocial Function (B-IPF) score for the comparison of combination therapy versus sertraline based on the FAS for Enriched Subjects

The same sensitivity analyses as for the primary endpoint to evaluate the sensitivity of the results due to the impact of COVID-19 (see Section 8.1.4.3) will be performed for the key secondary endpoints.

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

CCI

[redacted]

8.7 Efficacy Outcome Measures in Period A

Summary statistics will be provided for efficacy outcome measures in Period A as applicable. No inferential statistical analysis will be performed.

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).

Analyses of Period B safety data will be performed on the Safety Sample unless indicated otherwise. 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. Prospectively defined criteria will be used to identify potentially clinically relevant abnormal values for clinical laboratory tests, vital signs, ECGs, and body weight.

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.

Treatment-emergent AEs (TEAEs) in Period B are defined as AEs with an onset date on or after the start of double-blind IMP (Week 1). In more detail, TEAEs are all adverse events which started after start of double blind IMP; or if the event was continuous from end of Period A and was worsening, serious, study drug related, or resulted in death, discontinuation, interruption or reduction of study therapy. Adverse events occurring up to 30 days after the last day of IMP will be included in the summary tables.

The incidence of the following events in Period B will be tabulated by treatment group and overall using the Safety Sample:

a) TEAEs

b) TEAEs by severity

c) TEAEs potentially causally related to the IMP as assessed by the investigator

d) TEAEs with an outcome of death

e) Serious TEAEs

f) TEAEs leading to discontinuations of the IMP

g) EPS-related TEAEs

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

In addition, incidence of TEAE during Period B of at least 5% in any treatment group other than placebo group, and also greater than placebo by SOC and PT will be provided.

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

SAP 331-201-00071

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.

Potentially clinically relevant laboratory measurement test results in Period B will be identified for the Safety Sample and will be summarized by treatment group and listed. Criteria for identifying laboratory values of potential clinical relevance are provided in Appendix 2.

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 ≥ three times the upper normal limits (ULN) or baseline.

Liver injury related laboratory test results will be summarized for subjects who met the following criteria during Period B.

- AST or ALT ≥ 3 x ULN or baseline and
- T_Bili ≥ 2 x ULN or baseline

The corresponding listing will be provided for the Safety Sample during Period B.

9.2.2 Metabolic Change

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

Table 9.2.2-1 Criteria for Treatment-Emergent Significant Change in Lipids and Glucose

LAB PARAMETER	BASELINE[a]	ANYTIME POST BASELINE
LDL Direct, Fasting (MG/DL)	Borderline 100-< 160 Normal/Borderline < 160 Normal < 100 Any Value	High ≥ 160 High ≥ 160 Borderline/High ≥ 100 Increased ≥ 30
HDL Cholesterol, Fasting (MG/DL)	Normal ≥ 40 Any Value	Low < 40 Decreased ≥ 20
Triglycerides, Fasting (MG/DL)	Normal < 150 Borderline 150-< 200	High 200-< 500 High 200-< 500

SAP 331-201-00071

Table 9.2.2-1 Criteria for Treatment-Emergent Significant Change in Lipids and Glucose		
LAB PARAMETER	**BASELINE[a]**	**ANYTIME POST BASELINE**
	Normal/Borderline < 200 Normal < 150 Any Value	High 200-< 500 Borderline/High/Very High ≥ 150 Increased ≥ 50
Glucose Fasting, Serum (MG/DL)	Normal < 100 Impaired 100-< 126 Normal/Impaired < 126 Any Value	High ≥ 126 High ≥ 126 High ≥ 126 Increased ≥ 10

[a] BASELINE IS CALCULATED FROM WEEK 1.

Table 9.2.2-2 Criteria for Treatment-Emergent Metabolic Syndrome	
DESCRIPTION	**ANYTIME POST BASELINE[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 ≥ 130mmHg and Diastolic ≥ 85mmHg
Glucose Fasting, Serum	≥ 100mg/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

Summary statistics for vital signs will be provided. For Period B vital signs, change from baseline (Week 1 visit) will be summarized for the Safety Sample by treatment group.

Potentially clinically relevant vital signs measurements identified in Period B for the Safety Sample will be summarized by treatment group. Criteria for identifying vital signs of potential clinical relevance are provided in Appendix 1. All potentially clinically relevant events or changes will be listed and included in summary tables.

Potentially clinically relevant vital signs measurements identified in Period A for the Safety Sample will also be summarized by treatment group and listed.

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

Potentially clinically relevant changes in the 12-lead ECG identified in Period B for the Safety Sample will be listed and summarized by treatment group. Criteria for identifying ECG measurements of potential clinical relevance are provided in Appendix 3.

Categorical changes in ECG parameters during Period B will be summarized based on the following criteria:

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

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

* QTc categorical change criteria apply to QTcB, QTcF and QTcN.

9.5 Physical Examinations

By-patient listings will be provided for physical examination.

9.5.1 Body Weight, Waist Circumference and BMI

Analyses of body weight, waist circumference and BMI will be performed for the Safety Sample. The mean change from end of Period A to Week 12 (OC) and last visit in Period B in body weight will be tabulated and analyzed using ANCOVA. The ANCOVA models for both the OC and last visit analyses will include the end of Period A body weight and the treatment group.

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

The mean change from baseline to every study week in Period B obtained from the SAS total score, AIMS total score (total of the first 7 item scores), and the BARS Global Clinical Results will be presented for every scheduled visit, including Week 12.

Assessment will be tabulated and analyzed using ANCOVA. Analyses will be performed on the OC data set. In addition, analyses will be performed using the maximum (i.e. the worst) value observed during Period B and the last visit data to determine the change from end of Period A score. The ANCOVA model for the OC data set will include the end of Period A measure and the treatment group. The ANCOVA model for change at the last visit and for change to the maximum value will include the end of Period A measure, study center, previous pharmacological treatment intervention for PTSD (Yes/No) and treatment group. The same analyses will be performed on the AIMS individual item scores 8, 9, and 10. In addition, incidence of BARS Global Clinical Assessment of Akathisia during Period B by severity category will be provided. Analyses of these EPS rating scales will be performed for the Safety Sample.

9.7 Suicidality Data

Suicidality will be monitored during the study using the C-SSRS and will be summarized as 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. Summary will be provided for Period B. Summary will also be provided for Period A.

Suicidality is defined as report of at least one occurrence of any type of suicidal ideation or at least one occurrence of any type of suicidal behavior during assessment period (count each person only once).

Treatment emergent suicidal behavior and ideation is summarized by four types: Emergence of suicidal ideation, Emergence of serious suicidal ideation, Worsening of suicidal ideation, Emergence of suicidal behavior.

Emergence of suicidal behavior/ideation is defined as report of any type of suicidal behavior/ideation during treatment when there was no baseline suicidal behavior/ideation.

Emergence of serious suicidal ideation is defined as observation of suicidal ideation severity rating of 4 or 5 during treatment when there was no baseline suicidal ideation.

Worsening of suicidal ideation is defined as a suicidal ideation severity rating that is more severe than it was at baseline.

For Period B analyses, End of Period A is being used as “Baseline”.

9.8 Concomitant Medications

Number and proportion of patients taking concomitant medications prior to Period B, during Period B, and after study therapy are tabulated by drug classification using the WHO drug dictionary. For Period B Randomized Sample, data will be presented by treatment group and overall.

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, etc. This summary will be performed on the Safety Sample.

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.

SAP 331-201-00071

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.

Table 10.1-1 shows classifications for study day intervals in Period B. The variable "target day" is defined using the number of days since the start of double-blind dosing in Period B. The first day of double-blind dosing is defined as "Day 1". An end of Period A observation is the last available observation in Period A on or before "Day 1".

If more than one 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 study visit windows and will be excluded from the analysis.

Table 10.1-1 Study Day and Visit Windows in Period B

Week	Target Day[a]	Study Day Interval[a]
2	7	2-10
3	14	11-17
4	21	18-27
6	35	28-41
8	49	42-55
10	63	56-69
12	77	70-84 [b]

[a]Relative to the first day of double-blind IMP in Period B.

[b]Evaluations occurring more than seven days after the last double-blind dosing date will be excluded from the Period B analyses.

10.2 Pooling of Small Centers

Primary efficacy analysis will be performed on the FAS for Enriched Subjects which comprises those subjects in the Randomized Sample who received at least 1 dose of double-blind IMP, have a baseline value and at least 1 valid post-randomization efficacy evaluation for CAPS-5 total score in Period B and meet the Enriched Subject Criteria. Small centers will be defined as centers that do not have at least one evaluable subject (evaluable with regard to the primary efficacy variable) in each treatment arm and each status of previous pharmacological treatment intervention for PTSD (Yes/No) in Period B. All small centers will be pooled to form "pseudo centers" for the purpose of analysis

according to the following algorithm. Small centers will be ordered from the largest to the smallest based on the number of evaluable subjects (i.e., subjects who have an End of Period A value and at least one post-randomization value for CAPS-5 total score in Period B). 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)

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.

SAP 331-201-00071

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.[11] 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 items. 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.

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

SAP 331-201-00071

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

[redacted]

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.

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

11 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

SAP 331-201-00071

Appendix 1 Criteria for Identifying Vital Signs of Potential Clinical Relevance

Variable	Criterion Value[a]	Change Relative to Baseline[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).

Appendix 2 Criteria for Identifying Laboratory Values of Potential Clinical Relevance

Laboratory Tests	Criteria
Chemistry	
Aspartate aminotransferase	≥ 3 × ULN
Alanine aminotransferase	≥ 3 × ULN
Alkaline phosphatase	≥ 3 × ULN
Lactate dehydrogenase	≥ 3 x ULN
Urea nitrogen	≥ 30 mg/dL
Creatinine	≥ 2.0 mg/dL
Urate	
Males	≥ 10.5 mg/dL
Females	≥ 8.5 mg/dL
Bilirubin	≥ 2.0 mg/dL
Creatine kinase	≥ 3 × ULN
Prolactin[a]	> ULN
Hematology	
Hematocrit	
Males	≤ 37% and decrease of ≥ 3 percentage points from baseline
Females	≤ 32% and decrease of ≥ 3 percentage points from baseline
Hemoglobin	
Males	≤ 11.5 g/dL
Females	≤ 9.5 g/dL
Leukocyte count	≤ 2,800 mm^3 or ≥ 16,000 mm^3
Eosinophils/Leukocyte	≥ 10%
Neutrophils/Leukocyte	≤ 15%
Neutrophil	≤ 1,500/mm^3
Platelet	≤ 75,000/mm^3 or ≥ 700,000/mm^3
Urinalysis	
Protein	Increase of ≥ 2 units
Glucose	Increase of ≥ 2 units
Casts	Increase of ≥ 2 units
Additional Criteria	
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
Nonfasting	≥ 200 mg/dL
Cholesterol (Full Analysis Set) ting	≥ 240 mg/dL
LDL cholesterol (Full Analysis Set) ting	≥ 160 mg/dL
HDL cholesterol (Full Analysis Set) ting	
Males	< 40 mg/dL
Females	< 50 mg/dL
Triglycerides (Full Analysis Set) ting	≥150 mg/dL

[a]Prolactin results will be blinded to the investigators and trial staff.

SAP 331-201-00071

Appendix 3 Criteria for Identifying ECG Measurements of Potential Clinical Relevance

Variable	Criterion Value[a]	Change Relative to Baseline[a]
Rate		
Tachycardia	≥ 120 bpm	increase of ≥ 15 bpm
Bradycardia	≤ 50 bpm	decrease of ≥ 15 bpm
Rhythm		
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
Conduction		
1 atrioventricular block	PR ≥ 200 msec	increase of ≥ 50 msec
2 atrioventricular block	all	not present → present
3 atrioventricular block	all	not present → present
Left bundle-branch block	all	not present → present
Right bundle-branch block	all	not present → present
Pre-excitation syndrome	all	not present → present
Other intraventricular conduction block[d]	QRS ≥ 120 msec	increase of ≥ 20 msec
Infarction		
Acute or subacute	all	not present → present
Old	all	not present → present ≥ 12 weeks post trial entry
ST/T Morphological		
Myocardial ischemia	all	not present → present
Symmetrical T-wave inversion	all	not present → present
Increase in QTc	QTcF > 450 msec (males) QTcF > 470 msec (females)	

[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

This page is a manifestation of an electronically captured signature

SIGNATURE PAGE

Document Name: 331-201-00071_Statistical_Analysis_Plan_V2

Document Number: CCI [redacted]

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy hh:min) - UTC timezone
PPD [redacted]	Clinical Approval	24-May-2023 22:23:41
	Biostatistics Approval	24-May-2023 22:30:19