Otsuka Pharmaceutical Development & Commercialization, Inc.

Investigational New Drug

Brexpiprazole (OPC-34712)

A Multicenter, Randomized, Flexible-dose, Double-blind Trial of Brexpiprazole versus Placebo for the Treatment of Adults With Borderline Personality Disorder

Protocol No. 331-201-00242
IND No. 141091
EudraCT No. 2019-002813-20

Statistical Analysis Plan

Version: 3.0

Date: June 21, 2021

CONFIDENTIAL – PROPRIETARY INFORMATION

Table of Contents

List of In-text Tables

List of In-text Figures

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 331-201-00242. All amendments and addendums (including the COVID-19 specific protocol addendum) to the protocol 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 compare the efficacy of brexpiprazole versus placebo for the treatment of subjects with a diagnosis of borderline personality disorder (BPD).

Secondary: To evaluate the safety and tolerability of brexpiprazole for the treatment of subjects with a diagnosis of BPD.

3 Trial Details

3.1 Study Design

This will be a 12-week, multicenter, randomized, double-blind, placebo-controlled trial of brexpiprazole in subjects diagnosed with BPD according to the Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM-5) and confirmed by a valid diagnostic instrument (Structured Clinical Interview for DSM-5 Personality Disorders [SCID-5-PD]). See Figure 3.1-1 for a schematic of the trial design.

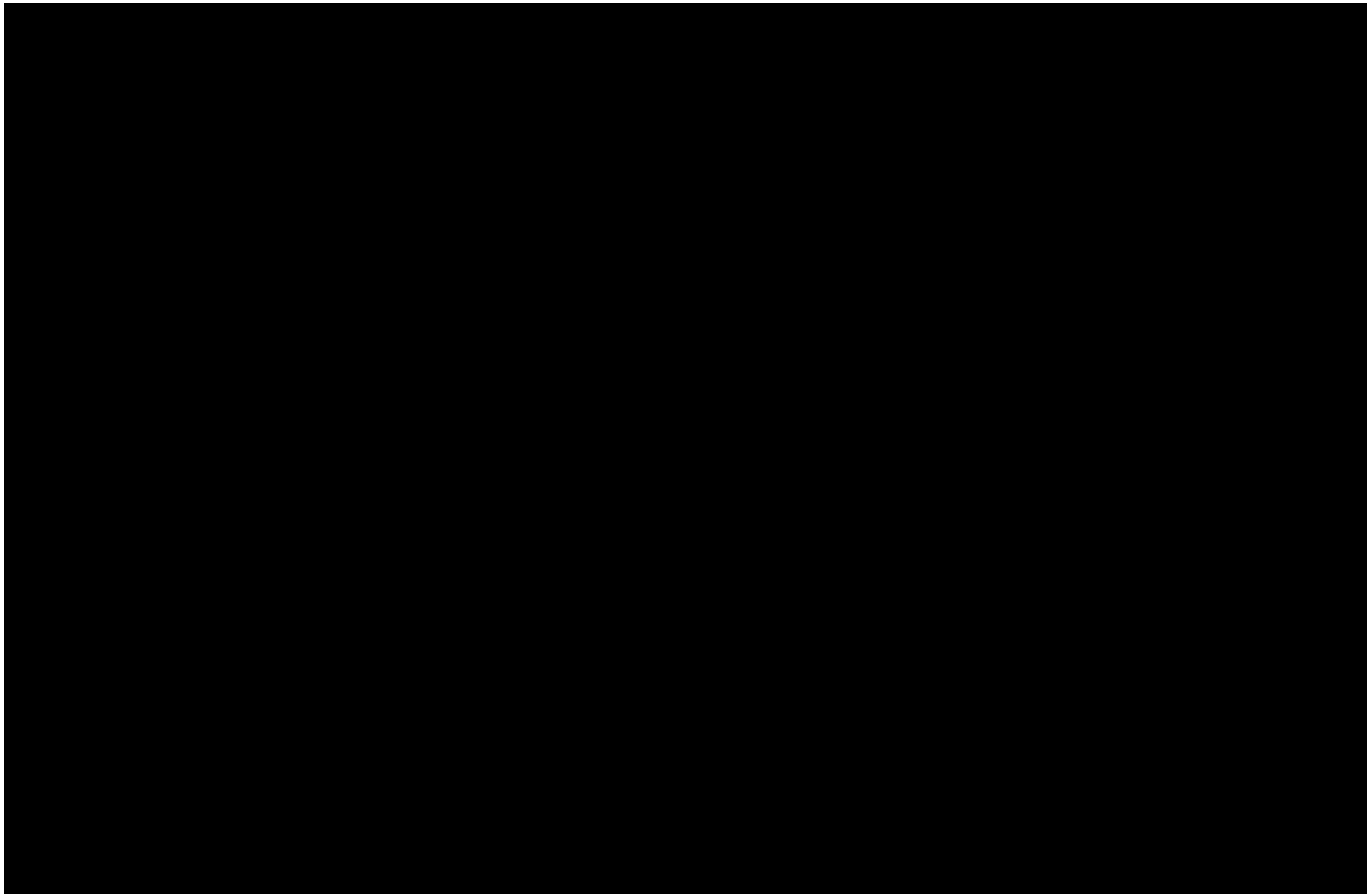
The trial will be organized as follows:

Screening Period: The screening period will begin after written informed consent has been obtained and will take place between Day −21 and Day −1 prior to enrollment. Eligible subjects are required to meet all inclusion criteria at both screening and the start [redacted] including a total score of ≥ 12 on the Zanarini Rating Scale for Borderline Personality Disorder (ZAN-BPD). In addition, a score of ≥ 2 in at least 2 of the following 4 ZAN-BPD scale items will be required at screening and Day 0: 1) “inappropriate, intense anger or difficulty controlling anger” (hereafter referred to as inappropriate anger), 2) “transient stress-related paranoid ideation or severe dissociative symptoms” (paranoid ideation), 3) “affective instability due to a marked reactivity of mood” (affective instability), and 4) “impulsivity in at least other two areas that are potentially self-damaging” (impulsivity).

[redacted]

Double-blind Treatment Period [redacted] subjects will be randomly assigned to treatment with brexpiprazole 2 to 3 mg/day or placebo in a 1:1 fashion. Subjects will receive their assigned treatment during an 11-week, double-blind treatment period [redacted] The timing of the primary and key secondary endpoints will be blinded to the investigator; these endpoints will be assessed after 9 weeks of assigned treatment (Week 10). Subjects will continue treatment through Week 12 of the trial. The primary endpoint and the key secondary endpoint will be the change from baseline to Week 10 in the clinician administered ZAN-BPD total score and Clinical Global Impression Scale - Severity of Illness (CGI-S) score, respectively. Visits will occur at the end of Weeks 1, 2, 4, 6, 8, 10, and 12/early termination (ET).

Post-treatment (Safety) Follow-up Period: Subjects who do not continue treatment in the open-label extension trial will be followed up for safety reasons via telephone contact or in clinic visit 21 (± 2) days after the last dose of investigational medicinal product (IMP). This contact also applies to subjects who are withdrawn prematurely from the trial.



3.2 Trial Treatments

Subjects will have their dose increased, according to the blinded titration schedule described below, based on the treatment group to which they are randomized:

- Brexpiprazole 2 to 3 mg/day
- Placebo

[redacted] the daily dosing schedule will be followed as shown below. Subjects assigned to brexpiprazole will start with 1 mg/day [redacted] followed by an increase to 2 mg/day [redacted] and another increase to 3 mg/day [redacted]

The dose of IMP can be adjusted to optimize efficacy and safety/tolerability as shown in Table 3.2-1, according to the following rules:

- Subjects who had their dose increased to 3 mg/day can have this reduced back to 2 mg/day due to tolerability issues at any time during the trial after [redacted] If a subject requires a dose reduction for tolerability issues outside of a regularly scheduled visit, they need to complete an unscheduled visit.
- Subjects unable to tolerate 2 mg/day will be discontinued from the trial.
- [redacted]
- Subjects who had their dose decreased to 2 mg/day can have this increased back to 3 mg/day until [redacted] Dose increases are only permitted at scheduled visits. [redacted]

[redacted]

4 Sample Size and Power Justification

Sample size is calculated based on an expected between-group difference of −2.6 (standard deviation = 6.5) in the mean change from baseline to Week 10 in ZAN-BPD total score. A

total of 200 subjects (100 subjects each in the brexpiprazole treatment arm and the placebo treatment arm) will yield 80% power to detect the treatment effects at a 2-tailed significance level of 0.05. In order to have 200 subjects randomized and evaluable (assuming a 5% non-evaluable/dropout rate) in the Full Analysis Set for Enriched Subjects (FAS for Enriched Subjects) sample (defined in Section 5.1), approximate 240 subjects (120 subjects each in the brexpiprazole treatment arm and the placebo treatment arm) will be randomized in this trial.

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 who signed an informed consent form (ICF) for the trial and enrolled [redacted]

Enriched Randomized Sample: comprises all subjects who were randomized satisfying the ITT for Enriched Subjects Criteria, where the ITT for Enriched Subjects Criteria are defined as the ZAN-BPD total score ≥ 10 at baseline, and a score ≥ 2 in at least 2 of the following 4 ZAN-BPD subscale items at baseline: inappropriate anger, paranoid ideation, affective instability, and impulsivity [redacted]

Randomized Sample: comprises all subjects who were randomized [redacted] Subjects are considered randomized when they are assigned a treatment number by the interactive response technology (IRT) [redacted] A subject receiving investigational medicinal product (IMP) outside of the IRT will not be considered randomized, but safety will be reported.

Safety Sample: comprises those randomized subjects [redacted] Randomized Sample who received at least 1 dose of double-blind IMP.

FAS for Enriched Subjects: comprises those subjects in the Enriched Randomized Sample who received at least 1 dose of double-blind IMP, and have a baseline value and at least 1 valid post-randomization efficacy evaluation for ZAN-BPD total score [redacted]

Full analysis Set (FAS): comprises all subjects in the Safety Sample who have a baseline value and at least 1 valid post-randomization efficacy evaluation for ZAN-BPD total score [redacted]

In general, baseline 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, status of background ADT therapy (with or without background ADT), and whether or not they meet the ITT for Enriched Subjects Criteria.

5.2 Handling of Missing Data

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 (MAR). The OC dataset consists of actual observations recorded at each visit [redacted] and no missing data will be imputed.

The last observation carried forward (LOCF) analysis will include data recorded at a scheduled double-blind treatment phase visit or, if no observation is recorded at that visit, data carried forward from the previous scheduled double-blind treatment phase visit. Baseline data will not be carried forward to impute missing values for the LOCF analysis.

6 Study Conduct

6.1 Subject Disposition, Completion Rate and Reasons for Discontinuation

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

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

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. 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. In addition, protocol deviations affected by the COVID-19 will be summarized. Listing of subjects with protocol deviations affected by the COVID-19 will also be provided.

7 Baseline Characteristics

7.1 Baseline Definition

For analyses of the double-blind treatment period [redacted] 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.

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

A summary of medical, psychiatric, and borderline personality disorder history will 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 Structured Clinical Interview for DSM-5 Personality Disorders (SCID-5-PD) will be presented for the Randomized Sample (by treatment group and overall). Summarized will be the number and percentage of patients who meet criteria of each personality disorder.

7.5 Baseline Psychiatric Evaluation

For the Randomized Sample, baseline psychiatric 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: ZAN-BPD Total Score, Clinical Global Impression – Severity

of Illness Scale (CGI-S), PGI-S, [redacted], ZAN-BPD sector scores, [redacted].

8 Efficacy Analysis

For analysis [redacted], 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 levels.

8.1 Primary Efficacy Endpoint

The primary efficacy endpoint is the change from baseline to Week 10 visit in the ZAN-BPD total score.

8.1.1 Primary Estimand

The primary estimand defining the treatment effect of interest in the trial uses the hypothetical strategy specified in the International Conference for Harmonisation (ICH) E9 (R1) Addendum. The objective of the primary analysis is to compare the efficacy of brexpiprazole versus placebo for the treatment of subjects with a diagnosis of BPD. 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 BPD or other prohibited medications. In the case of terminal or lost to follow-up events no ET evaluations would be expected, and only scheduled assessments 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 ZAN-BPD total score
- Intercurrent Events: Premature treatment discontinuation
- Measure of Intervention Effect: Difference in endpoint means between Brexpiprazole and Placebo arm.

The National Emergency Announcement concerning the coronavirus disease 2019 (COVID-19) pandemic was declared on March 13, 2020. The pandemic had a significant impact on many aspects of clinical trials. There are occasionally virtual visits (i.e., virtual assessments) and possibly early discontinuation of treatment directly or indirectly related to the pandemic. However, subjects were required to attend the screening, Day 0, Week 1, Week 10, and the end of trial visits in person (i.e., face-to-face). The details of trial conduct during the COVID-19 pandemic are described in the COVID-19 specific protocol addendum. Note that virtual visits in the pandemic environment will not be treated as an intercurrent event for the primary analysis. Subjects who drop out with a reason relating to the COVID-19 pandemic will be handled as they would have dropped out for another reason if the pandemic had not happened.

The hypothetical strategy of handling intercurrent events will be used to clarify the efficacy of the brexipiprazole had there be no occurrence of intercurrent events, regardless of being COVID-19 related or not. In other words, the estimand as described above will use the hypothetical strategy to address the treatment effect of interest that would be envisioned under the hypothetical setting of no occurrence of intercurrent events in the planned 12-week treatment period.

The estimator will be the Mixed Model Repeated Measurements (MMRM) estimate for treatment difference at Week 10, based on all observed case (OC) data until discontinuation from the trial. This reflects the chosen strategies for the identified intercurrent events. Details of the model are provided in the next section.

In this hypothetical strategy, the event of withdrawing IMP is considered 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 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 (MI) 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 ZAN-BPD 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 phase and no missing data will be imputed. The model will include fixed class effect terms for treatment, trial site, visit, ADT Status (with/without background ADT), and an interaction terms of treatment by visit, and gender by visit. The model will also include the interaction term of baseline values of ZAN-BPD Total score by visit and age by visit as covariates. All scheduled visits after baseline [redacted] 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 group and the placebo group at the Week 10 visit [redacted] will be estimated as the difference between Least Squares (LS) means utilizing the computing software SAS procedure PROC MIXED. The comparison between brexpiprazole group and placebo 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 [redacted] 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 ADT_Status sex subjid;
   model change=treatment center visit ADT_Status treatment*visit baseline*visit  age*visit
sex*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 ZAN-BPD Total Score at baseline (Week 1 visit of the double-blind treatment period).

8.1.4 Sensitivity Analyses

8.1.4.1 Sensitivity Analyses for Missing at Random (MAR) Assumption

The mixed-model repeated measures (MMRM) assume data are missing at random (MAR), which is a reasonable assumption in longitudinal clinical trials[1]. However, the possibility of "missing not at random" (MNAR) data can never be ruled out. As sensitivity analyses, pattern-mixture model[2,3,4,5] and shared parameter model[6] 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 (Little, 1995): (1) Missing Completely at Random (MCAR), in which the probability of dropout doesn't depend on the observed data and the missing data; (2) Missing at Random (MAR), in which the probability of dropout depends on the observed data, and (3) Missing Not at Random (MNAR), where the probability of dropout depends on the missing data and possibly the observed data.

Most of MNAR methods (Diggle P, Kenward MG, 1994) have treated all observations with dropout as if they fall within the same dropout type. In practice, we would find that different dropout reasons may be related to the outcomes in different ways, for example, detailed dropout reasons for this study are: adverse event (AE), death, lack of efficacy (LOE), lost to follow-up, non-compliance with IMP, pregnancy, protocol deviation, withdrawal by subject, trial terminated by sponsor, site terminated by sponsor, physician decision, and other. Dropout due to an AE, death 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. Missing data due to COVID-19 will also be assumed as MAR.

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 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%, .., of the expected treatment difference of 2.6 points and/or the observed treatment difference between Brex and Placebo from the primary analysis of MMRM model until conclusion of the primary analysis is overturned. That is, until the primary efficacy results will tip over with p-value ≥ 0.05. 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 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.

8.1.4.2 Supplemental 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[7] (van Elteren, 1960) stratified by center 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.

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, and 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 and the FAS, respectively, 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 and the FAS, respectively.
2. An MMRM analysis using the non-COVID data set based on the FAS for Enriched Subjects and the FAS, respectively. 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 (or FAS) who did not have any virtual assessments nor COVID-19 related protocol deviations.
4. To explore the impact of assessments before versus after the COVID breakout, a subgroup analysis for subjects who completed or discontinued from the study before March 13, 2020 and subjects who completed or discontinued from the study on or after March 13, 2020 will be performed. An MMRM analysis based on the FAS for Enriched Subjects and the FAS will be performed to the extent where data allow or the summary statistics will be provided.

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 in ZAN-BPD Total Score to every study week in the double-blind treatment period will be performed by the following factors:

- Sex (Based on the biological status)
- Race (White and All Other Races)
- Age group (Age<55 and Age≥55)
- Region (North America and Europe)
- Status of Background ADT (With ADT and Without ADT)

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 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. The same MMRM model will be used as for the primary efficacy analysis with the addition of terms for subgroup-by-week and treatment-by-subgroup-by-week. These treatment-by-subgroup interaction analyses will be presented in statistical documentation.

8.2 Key Secondary Endpoint Analysis

The key secondary efficacy endpoint is the change from baseline to Week 10 in the double-blind treatment period [redacted] in CGI-S score. This endpoint will be analyzed by fitting the similar MMRM model described in the primary analysis.

This endpoint will be analyzed by similar analysis described in Section 8.1.4.3 to evaluate the impact of COVID.

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 endpoint, 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 based on the FAS for Enriched Subjects
2) Key secondary endpoint based on the FAS for Enriched Subjects
3) Primary efficacy endpoint based on the FAS
4) Key secondary endpoint based on the FAS

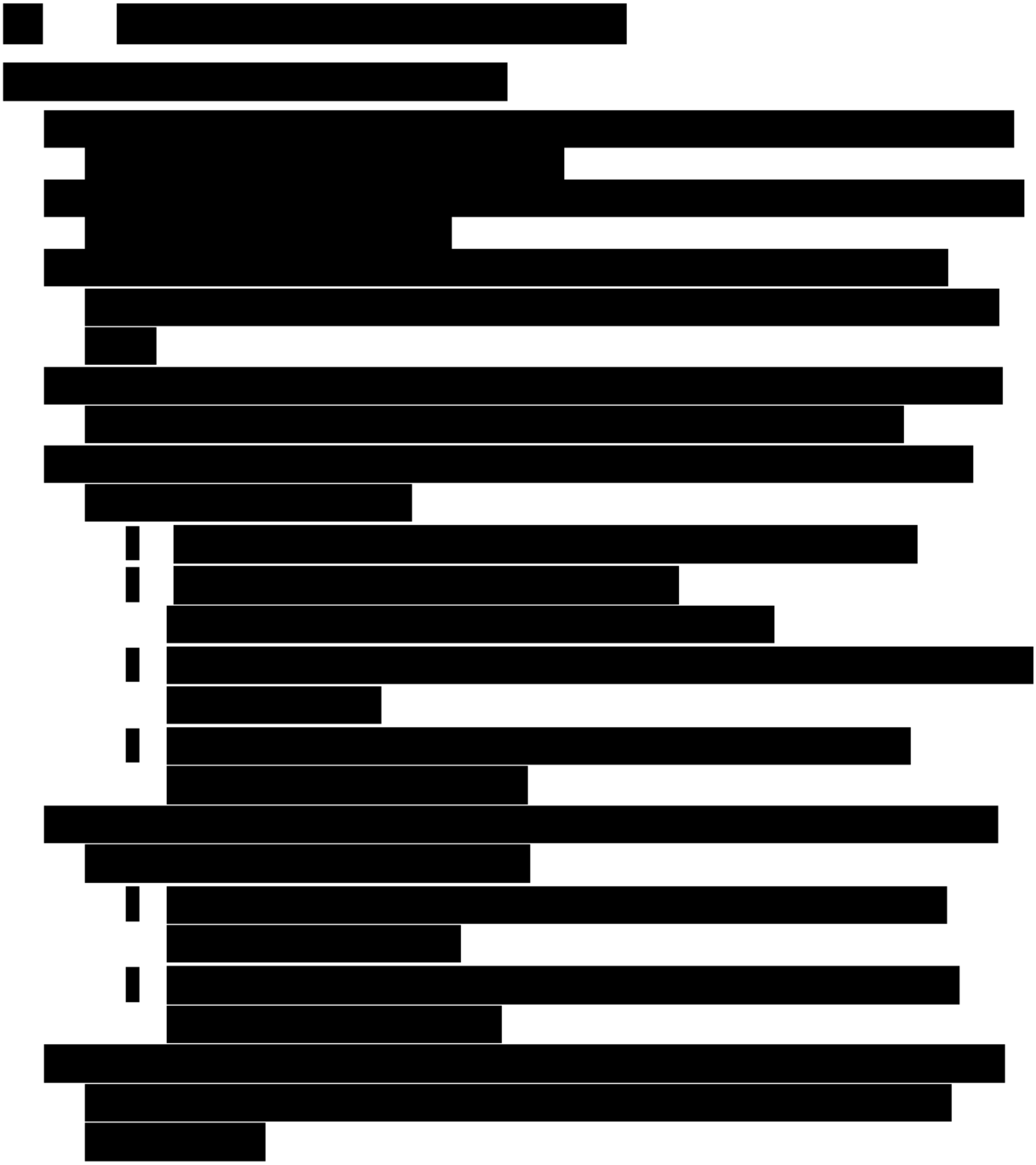
8.4 Secondary Efficacy Endpoint Analysis

Secondary efficacy endpoints are as follows:

1) Change from baseline in the Patient's Global Impression of Severity (PGI-S) to each trial visit during the double-blind treatment period
2) Patient's Global Impression of Change (PGI-C) score at each trial visit during the double-blind treatment period

3) Clinical Global Impression-Improvement of Illness (CGI-I) score at each trial visit during the double-blind treatment period

Variable (1) will be evaluated using the same MMRM model described in the primary analysis. Variables (2) and (3) will be evaluated by the Cochran-Mantel-Haenszel (CMH) Row Mean Score Differ Test controlling, in last-observation-carried-forward (LOCF) analysis, for trial center.



[redacted]

9 Safety Analysis

Standard safety variables to be analyzed include AEs, clinical laboratory tests, vital signs, ECGs, and physical examinations. In addition, data from the following safety scales will be evaluated: assessments of suicidality (C-SSRS) and EPS (eg, the SAS, AIMS, and BARS). Safety analysis will be conducted based on the Safety Sample defined in Section 5.1. 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) are defined as AEs with an onset date on or after the start of double-blind treatment period. In more detail, TEAEs are all adverse events which started after start of double blind IMP; or if the event was continuous from baseline and was

worsening, serious, study drug related, or resulted in death, discontinuation, interruption or reduction of study therapy. Adverse events occurring up to 30 days after the last day of IMP will be included in the summary tables.

The incidence of the following events in the double-blind treatment period 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

d) TEAEs with an outcome of death

e) Serious TEAEs

f) TEAEs leading to discontinuations of the IMP

g) AESI

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 the double-blind treatment period 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, age and region subgroups.

Extrapyramidal symptoms (EPS)-related AEs will be grouped into five categories.

1) Dystonic Events, which include cervical spasm, dystonia, emprosthotonos, muscle contractions involuntary, muscle rigidity, muscle spasms, muscle spasticity, myotonia, nuchal rigidity, oculogyration, opisthotonos, pleurothotonus, risus sardonicus, torticollis, and trismus;

2) Parkinsonian Events, which include akinesia, asterixis, athetosis, bradykinesia, cogwheel rigidity, essential tremor, extrapyramidal disorder, freezing phenomenon, gait festinating, hypertonia, hypokinesia, hypokinesia neonatal, intention tremor, masked facies, parkinson's disease, parkinsonian crisis, parkinsonian gait, parkinsonian rest tremor, parkinsonism, tremor, and tremor neonatal;

3) Akathisia Events, which include akathisia, hyperkinesia, and psychomotor hyperactivity;

4) Dyskinetic Events, which include ballismus, buccoglossal syndrome, choreoathetosis, clumsiness, dyskinesia, dyskinesia neonatal, dyskinesia oesophageal, fumbling, nodding of head, on and off phenomenon, and tardive dyskinesia;

5) Residual Events, which include chorea, Huntington's chorea, muscle twitching, and myoclonus.

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

Adverse Events of Special Interest

The new onset or exacerbation of "Pathological Gambling and Other Compulsive Behaviors" will be analyzed as an AESI.

9.2 Clinical Laboratory Tests

Summary statistics for mean and mean change from baseline in the routine clinical laboratory measurements, prolactin concentrations, coagulation parameters (PT, aPTT, and INR), HbA1c, and TSH will be provided by treatment and by visit.

Potentially clinically relevant laboratory measurement test results in the double-blind treatment period [redacted] 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 Drug Induced Liver Injury (DILI)

Total bilirubin level should be checked for any subject with increased ALT or AST levels ≥ three times the upper normal limits (ULN) or baseline.

- Reporting all DILI as SAE to the FDA based on Hy's Law:
  - AST or ALT ≥ 3 x ULN or baseline and
  - T_Bili ≥ 2 x ULN or baseline

A separate incidence table will be provided for DILI cases, and the corresponding listing will be provided for Safety Sample during the double-blind treatment period.

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.

Criteria for Treatment-Emergent Significant Change in Lipids and Glucose		
LAB PARAMETER	**BASELINE[1]**	**ANYTIME POST BASELINE**
LDL Direct (FAS)ting (MG/DL)	Borderline 100-<160	High >=160
	Normal/Borderline <160	High >=160
	Normal <100	Borderline/High >=100
	Any Value	Increased >=30
HDL Cholesterol (FAS)ting (MG/DL)	Normal >=40	Low <40
	Any Value	Decreased >=20
Triglycerides (FAS)ting (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

Criteria for Treatment-Emergent Metabolic Syndrome	
DESCRIPTION	**ANYTIME POST BASELINE[1]**
Central Obesity	Waist Circumference >=102cm(MALE), >=88cm (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

9.3 Vital Signs

Summary statistics for vital signs will be provided. For the double-blind treatment period [redacted] vital signs, change from baseline will be summarized for the Safety Sample by treatment group.

Potentially clinically relevant vital signs measurements identified in the double-blind treatment period [redacted] 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.

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 the double-blind treatment period 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 the double-blind treatment period will be summarized based on the following criteria:

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

Categorical Change Criteria in QT/QTc Parameters		
Classification	**Category**	**Criteria**
	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 baseline-to Week 10 (OC) and last visit in the double-blind treatment period in body weight will be tabulated and analyzed using ANCOVA. The ANCOVA models for both the OC and last visit analyses will include the baseline as a covariate and the treatment group as fixed effect. The mean change from screening to Week 12 in waist circumference will be analyzed using ANCOVA.

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 baseline to Week 10 (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 trial visit in the double-blind treatment period obtained from the SAS total score, AIMS total score (total of the first 7 item scores), and the BARS Global Clinical 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 the double-blind treatment period and the last visit data to determine the change from baseline score. The ANCOVA model for the OC data set will include the baseline measure and the treatment group. The ANCOVA model for change at the last visit and for change to the maximum value will include the baseline measure, study center 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 the double-blind treatment period 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 also be monitored during the study using the C-SSRS [redacted] 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 the double-blind treatment period.

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.

9.8 Concomitant Medications

Number and proportion of patients taking concomitant medications prior to study therapy, during the double-blind treatment period [redacted] and after study therapy are tabulated by drug classification using the WHO drug dictionary.

9.9 Extent of Exposure

The start date of double-blind treatment period study therapy after randomization - brexpiprazole or placebo - will be the first day of double-blind treatment period dosing after randomization. The number and percentage of patients who receive double-blind treatment period study medication, 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 study medication`, and the mean and range of the mean daily dose for each week.

10 Conventions

10.1 Study Visit Windows

Study visit windows will be used to map visits using study day intervals. Observations at each scheduled visit and Early Termination will be assigned to Week 1, Week 2, Week 4, Week 6, Week 8 , Week 10 and Week 12 visits based on their visit windows as shown in Table 10.1A. This visit window convention applies to tables and listings for all efficacy and safety scales (ZAN-BPD, CGI-S, CGI-I, [redacted] PGI-S, PGI-C, [redacted], SAS, AIMS and BARS). This derived study window variable will be named as DAY and will be footnoted. In listings, it will be listed along with the CRF study visit.

Table 10.1A shows classifications for study day intervals in the double-blind period. The variable “target day” is defined using the number of days since the start of double-blind dosing. The first day of double-blind dosing is defined as “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 10.1A: Study Day and Visit Windows[redacted]

Week	Target Day [a]	Study Day Interval [a]
2	7	2-13
4	21	14-27
6	35	28-41
8	49	42-55
10	63	56-69

12	77	70-84 [b]

[a] Relative to the first day of IMP in the double-blind treatment period.

[b] Evaluations occurring more than seven days after the last dosing date of IMP in the double-blind treatment period will be excluded from the efficacy 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 have a baseline value and at least one post-baseline value for ZAN-BPD Total Score in the double-blind treatment period. 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 ADT Status (with/without background ADT) in FAS for Enriched Subjects in the double-blind treatment period. 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 a Baseline value and at least one post-randomization value for ZAN-BPD Total Score in the double-blind treatment period). The process will start by pooling the largest of the small centers with the smallest of the small centers until a non-small center is formed. This process will be repeated using the centers left out of the previous pass. In case of ties in center size, the center with the smallest center code will be selected. If any centers are left out at the end of this process, they will be pooled with the smallest pseudo centers, or if no pseudo 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 ZAN-BPD

The ZAN-BPD is utilized as the primary efficacy assessment of a subject's severity of disease symptoms in patients with BPD. The questions for the ZAN-BPD reflect a 1-week time frame. Each of the 9 criteria for BPD is rated on a 5-point anchored rating scale of 0 to 4. These scores are clustered into 4 sector scores (akin to domains) and a total score. The 3 affective symptoms (anger, moodiness, and emptiness scores) have a sector score range from 0 to 12. The 2 cognitive symptoms (identity disturbance and distrust/suspiciousness/dissociation scores) have a sector score range from 0 to 8. The 2 impulsive symptoms (self-mutilation/suicidality and other forms of impulsivity scores) have a sector range from 0 to 8. The 2 interpersonal symptoms (efforts to avoid abandonment and unstable relationships scores) have a sector score range of 0 to 8. These 4 sector scores add up

to provide the overall total score for the ZAN-BPD, which ranges from 0 to 36. The total score is missing if any element comprising toward the total score is missing.

10.3.2 CGI-S

The severity of illness for each subject will be rated using the CGI-S. To perform this assessment, the investigator (or designee) will answer the following question: "Considering your total clinical experience with borderline personality disorder, how ill is the subject at this time?" Response choices are 0 = not assessed; 1 = normal, not at all ill; 2 = borderline ill; 3 = mildly ill; 4 = moderately ill; 5 = markedly ill; 6 = severely ill; and 7 = among the most extremely ill subjects

10.3.3 CGI-I

Change from baseline in the subject's condition will be assessed using the CGI-I scale. The investigator (or designee) will answer the following question: "Rate total improvement whether or not in your judgment it is due entirely to drug treatment. Compared to his/her condition at Day 0, how much has the patient changed?". All responses will be compared with the subject's condition at baseline (Day 0). Response choices are 0 = not assessed, 1 = very much improved, 2 = much improved, 3 = minimally improved, 4 = no change, 5 = minimally worse, 6 = much worse, and 7 = very much worse.

[redacted]

10.3.5 PGI-S

The PGI-S is a 7-point single-item self-report scale for the patient to rate the severity of symptoms of BPD. Subjects answer the following question: "Taking into account all of your symptoms, how severe is your Borderline Personality Disorder at this time?" Scores range from 1 "no symptoms" to 7 "very severe".

10.3.6 PGI-C

The Patient's Global Impression of Change (PGI-C) is a 7-point single-item self-report scale depicting a patient's rating of overall change in their condition since starting trial medication. Subjects answer the following question: "Since starting study medication, how much have your symptoms of Borderline Personality Disorder changed?" Scores range from 1 "very much improved" to 7 "very much worse".

[redacted]

10.3.9 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.

10.3.10 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.

10.3.11 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.

10.3.12 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. The “Since Last Visit” C-SSRS form will also be completed at all visits after screening.

11 References

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

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

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

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

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

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

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

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

9 Cortina, J. M., & Nouri, H. (2000). *Effect size for ANOVA designs*. Thousand Oaks, Calif.: Sage Publications.

[redacted]

12 Potential Clinical Relevance Criteria from Protocol

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	
AST (SGOT)	≥ 3 x upper limit of normal (ULN)
ALT (SGPT)	≥ 3 x ULN
Alkaline phosphatase	≥ 3 x ULN
Lactate dehydrogenase (LDH)	≥ 3 x ULN
Blood urea nitrogen (BUN)	≥ 30 mg/dL
Creatinine	≥ 2.0 mg/dL
Uric Acid	
Men	≥ 10.5 mg/dL
Women	≥ 8.5 mg/dL
Bilirubin (total)	≥ 2.0 mg/dL
Creatine phosphokinase (CPK)	≥ 3 x ULN
Prolactin	> ULN
Hematology	
Hematocrit	
Men	≤ 37 % and decrease of ≥ 3 percentage points from Baseline
Women	≤ 32 % and decrease of ≥ 3 percentage points from Baseline
Hemoglobin	
Men	≤ 11.5 g/dL
Women	≤ 9.5 g/dL
White blood count	≤ 2,800/ mm^3 or ≥ 16,000/ mm^3
Eosinophils	≥ 10%
Neutrophils	≤ 15%
Absolute neutrophil count	≤ 1,500/ mm^3
Platelet count	≤ 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
Non-Fasting	≥ 200 mg/dL
Total Cholesterol (FAS)ting	≥ 240 mg/dL
LDL Cholesterol (FAS)ting	≥ 160 mg/dL
HDL Cholesterol (FAS)ting	
Men	< 40 mg/dL
Women	< 50 mg/dL
Triglycerides (FAS)ting	≥ 150 mg/dL

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 study entry
ST/T Morphological		
Myocardial Ischemia	all	not present → present
Symmetrical T-wave inversion	all	not present → present
Increase in QTc	QTcF ≥ 450 msec (men) QTcF ≥ 470 msec (women)	

[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.

Appendix 4 Details and SAS codes for MI Procedure in Delta Adjustment Imputation Methods and Placebo Based Imputation Methods

Delta Adjustment Imputation Methods

- Step 1: Use MCMC to obtain monotone missing data pattern. Dataset *mi_indata* contains the original OC data.

```
proc mi data=mi_indata out=mi_mono nimpute=100 seed=12345;
  var y0 y2 y4 y6 y8 y10 ;
  by treatment;
  mcmc chain=multiple impute=monotone;
run;
```

- Step 2: Use a standard MAR-based regression method to impute monotone missingness data:

```
proc mi data=mi_mono out=mi_reg seed=54321 nimpute=1;
  by _Imputation_;
  var treatment y0 y2 y4 y6 y8 y10 ;
  class treatment;
  monotone regression;
run;
```

Placebo Based Imputation Methods (Use MI for Copy Placebo MNAR)

- Method 1: Use MCMC methodology to impute Placebo arm as MAR

```
proc mi data=mi_indata (where=(treatment in ("pbo"))) out=mi_mcmc_pbo nimpute=100
    seed=12345;
  var y0 y2 y4 y6 y8 y10;
  mcmc chain=multiple outest=mcmc_pbo;
run;

proc mi data=mi_indata (where=(treatment in ("drug"))) out=mi_mcmc_drug;
  var y0 y2 y4 y6 y8 y10;
 mcmc inest=mcmc_pbo;
run;
data mi_mcmc ;
   set mi_mcmc_drug mi_mcmc_pbo;
run;
```

- Method 2: Use MI for Copy Placebo MNAR

Step 1: Perform the same steps 1 and 2 in the delta adjusted imputation method as described above.

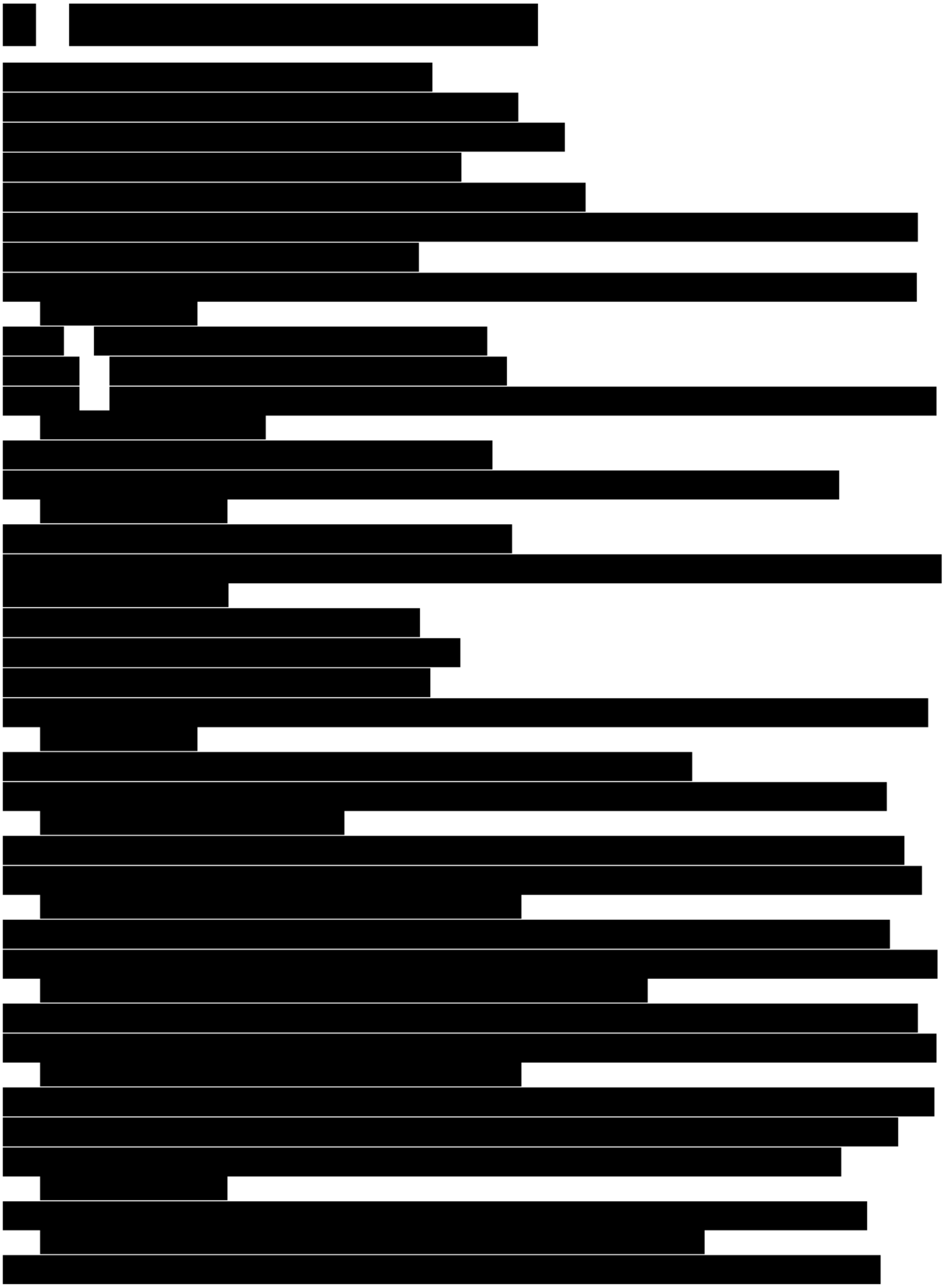
```
proc mi data=mi_indata out=mi_mono nimpute=100 seed=12345;
  var y0 y2 y4 y6 y8 y10;
  by treatment;
  mcmc chain=multiple impute=monotone ;
run;
```

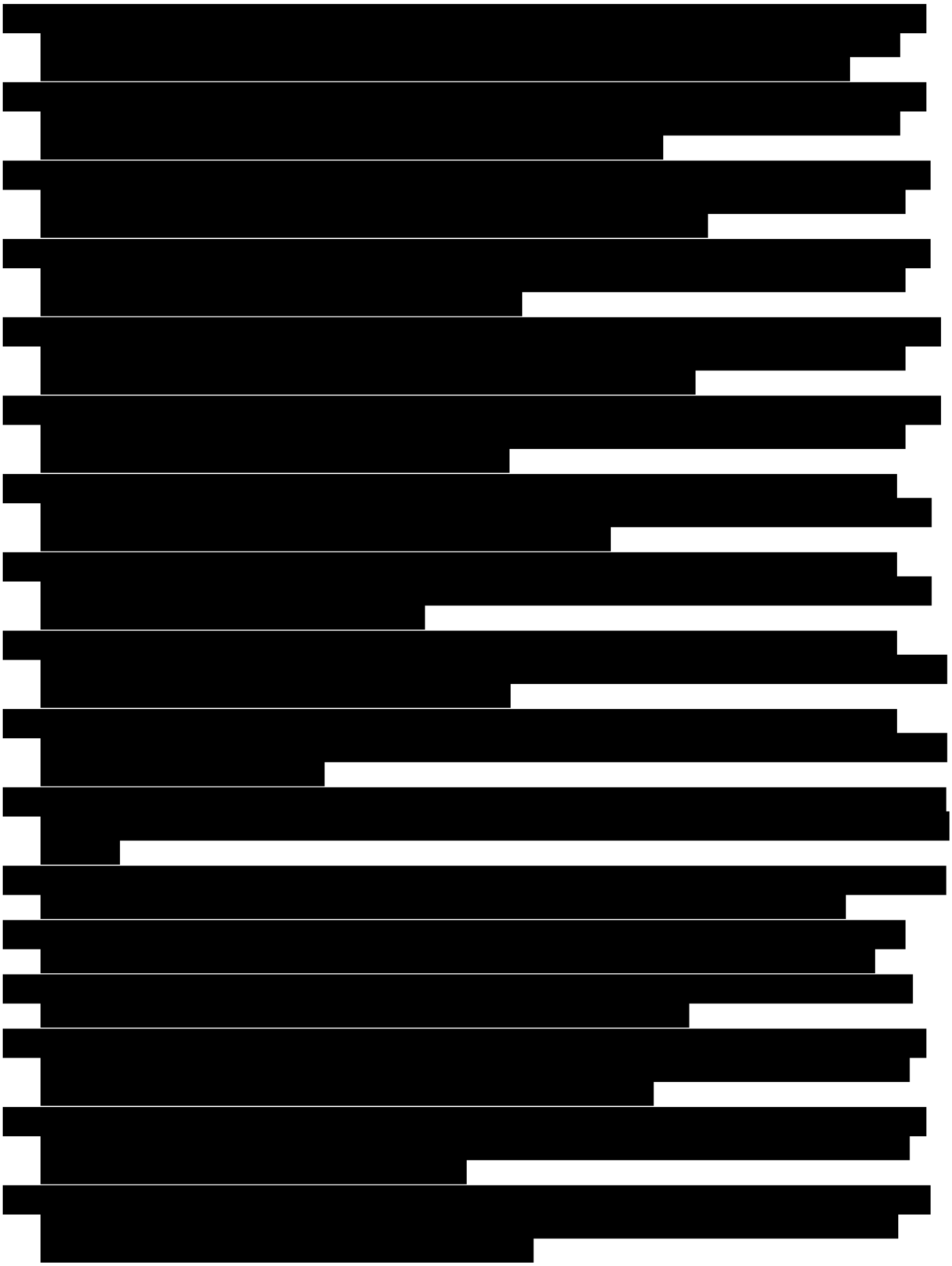
Step 2: Use MNAR Copy Placebo Imputation

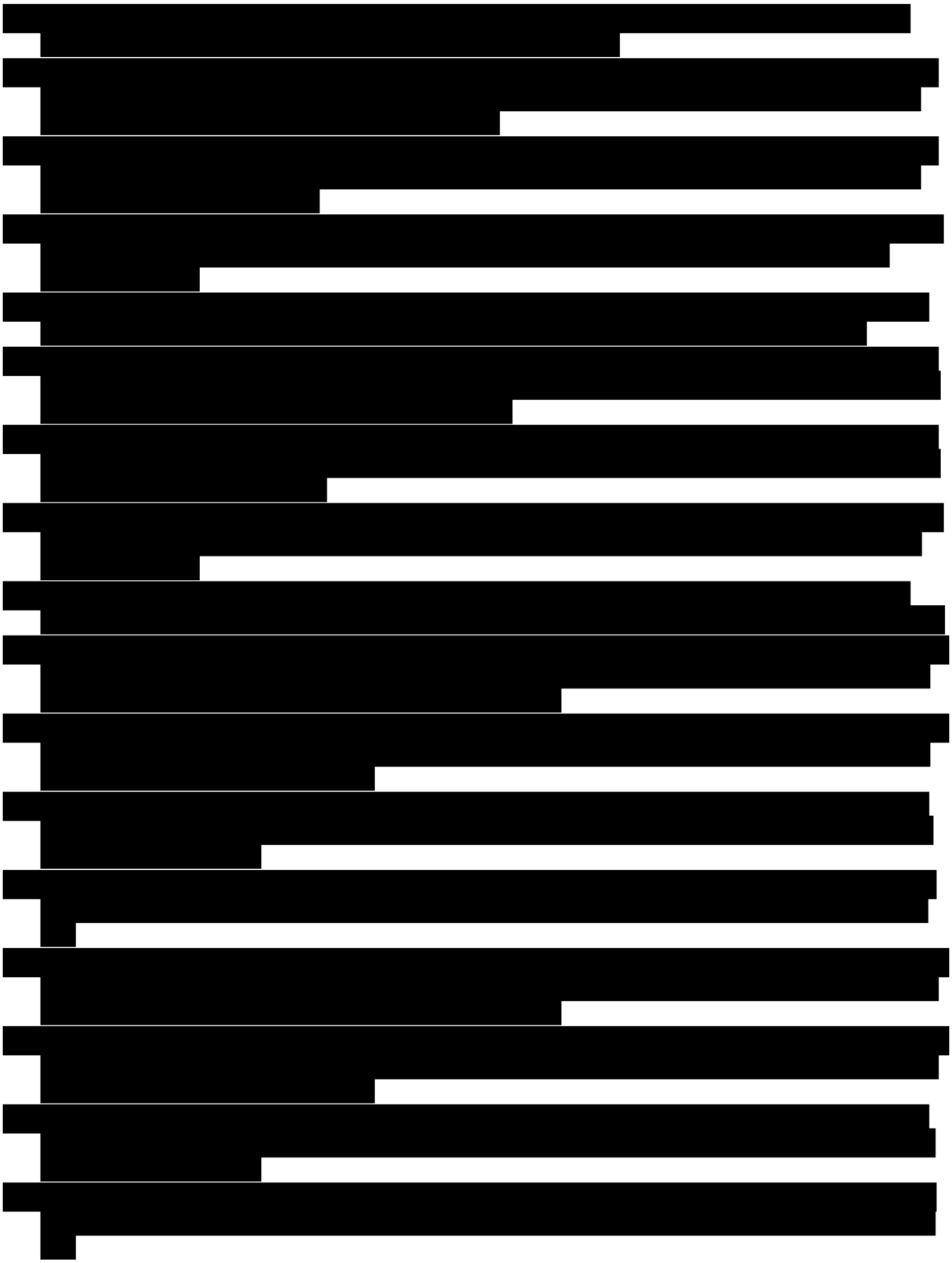
```
data mi_imp0;
  set mi_mono;
  by _imputation_;
run;
```

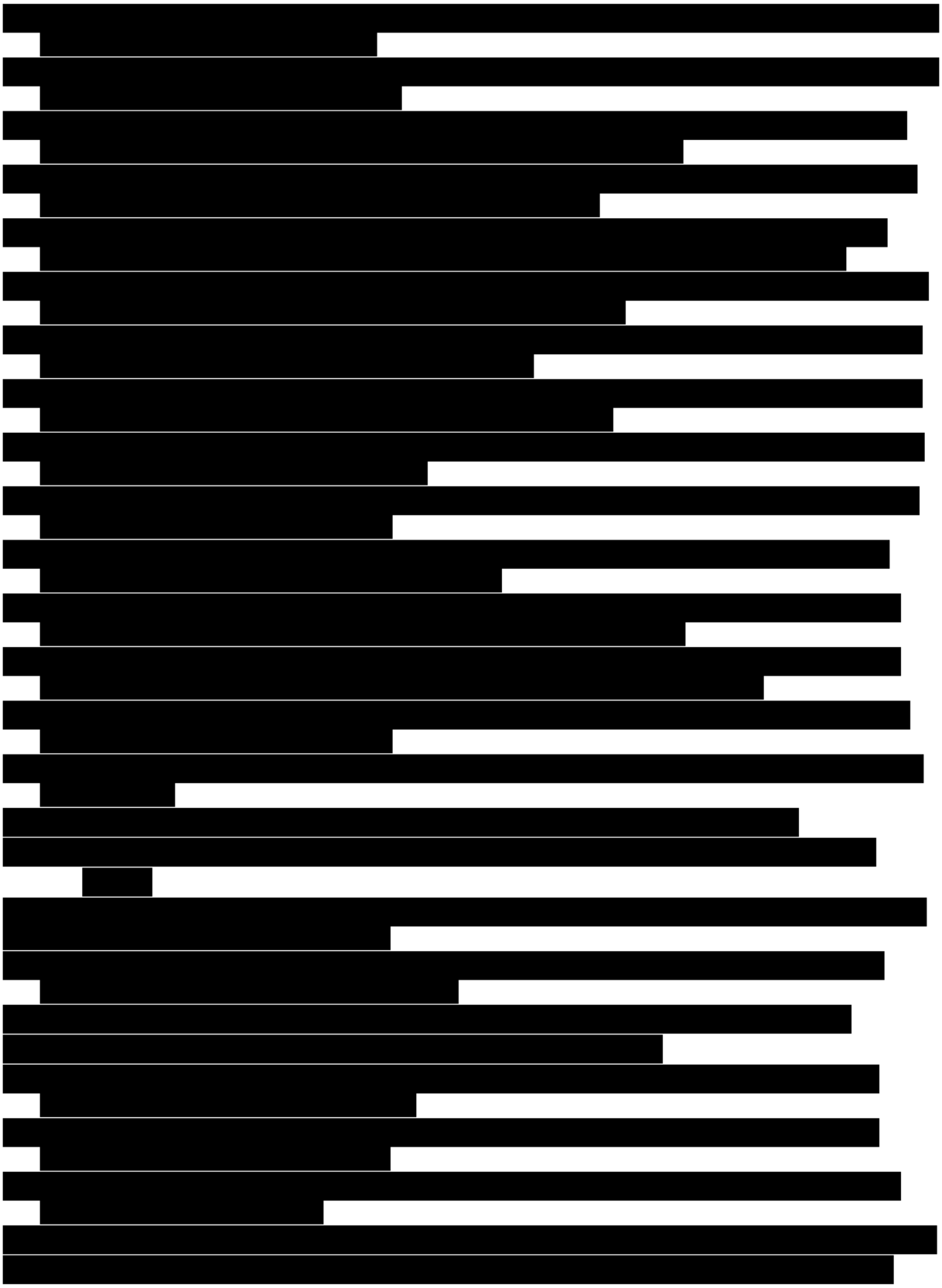
The following are repeated for *i*=1 to 4 (corresponding to Weeks 4,6,8,10, denoted by *k*=4, 6, 8, 10 and *j*=*i*-1).

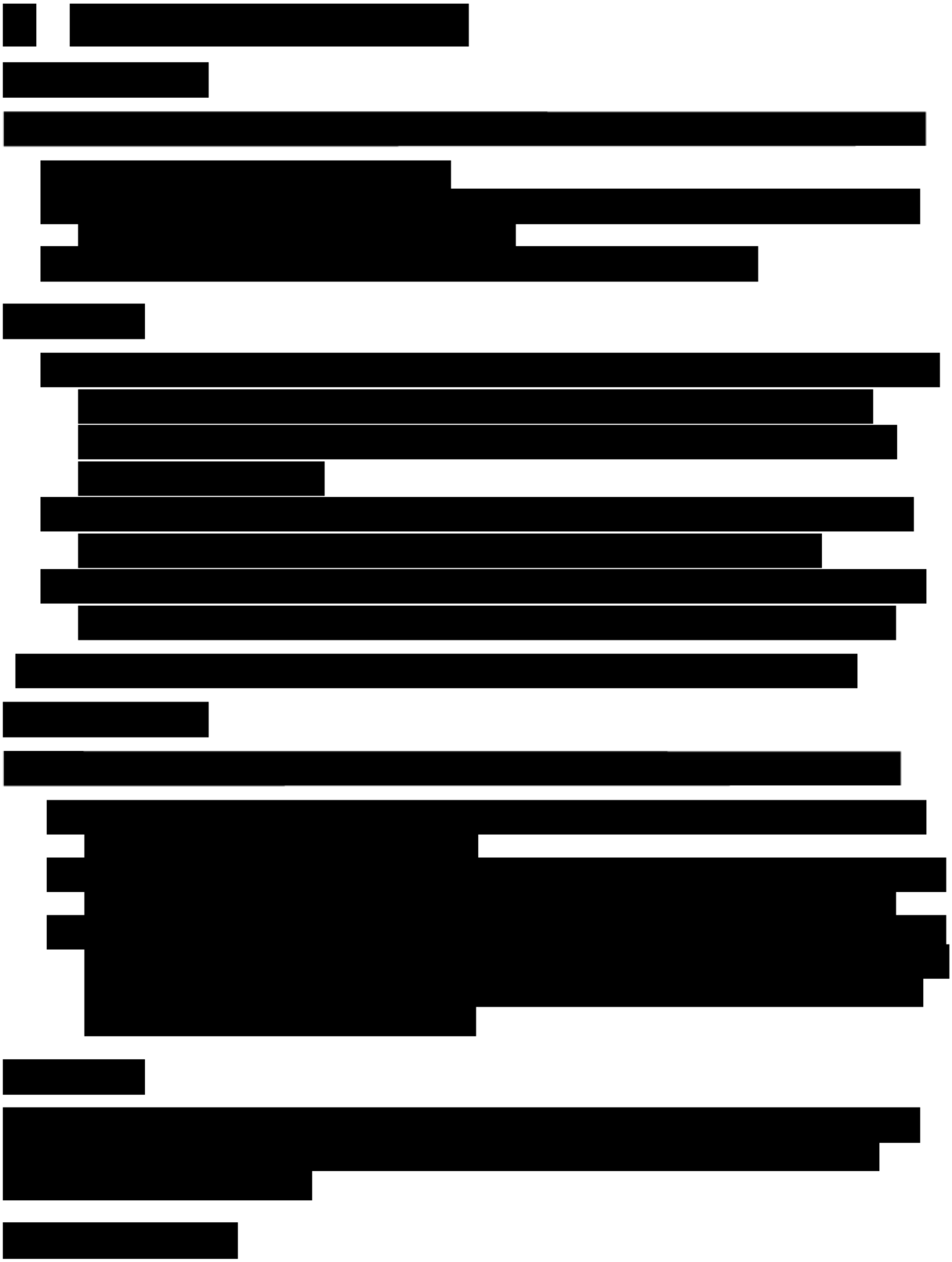
```
data mono_imp&i mono_rest&i;
  set mi_imp&j;
  if treatment in (“drug”) and lastvisit >=&k then output mono_rest&i;
  else output mono_imp&i;
run;

proc mi data=mono_imp&i out=reg_imp&i nimpute=1 seed=xxxxx;
   by _Imputation_;
   %if &k =4 %then %do;
    var Y0 Y2 Y4;
   %end;
   %if &k =6 %then %do;
    var Y0 Y2 Y4 Y6;
   %end;
   %if &k =8 %then %do;
    var Y0 Y2 Y4 Y6 Y8;
   %end;
    %if &k =10 %then %do;
    var Y0 Y2 Y4 Y6 Y8 Y10;
   %end;
   monotone reg(Y&k);
run;
data mi_imp&i;
  set mono_rest&i reg_imp&i;
run;
```

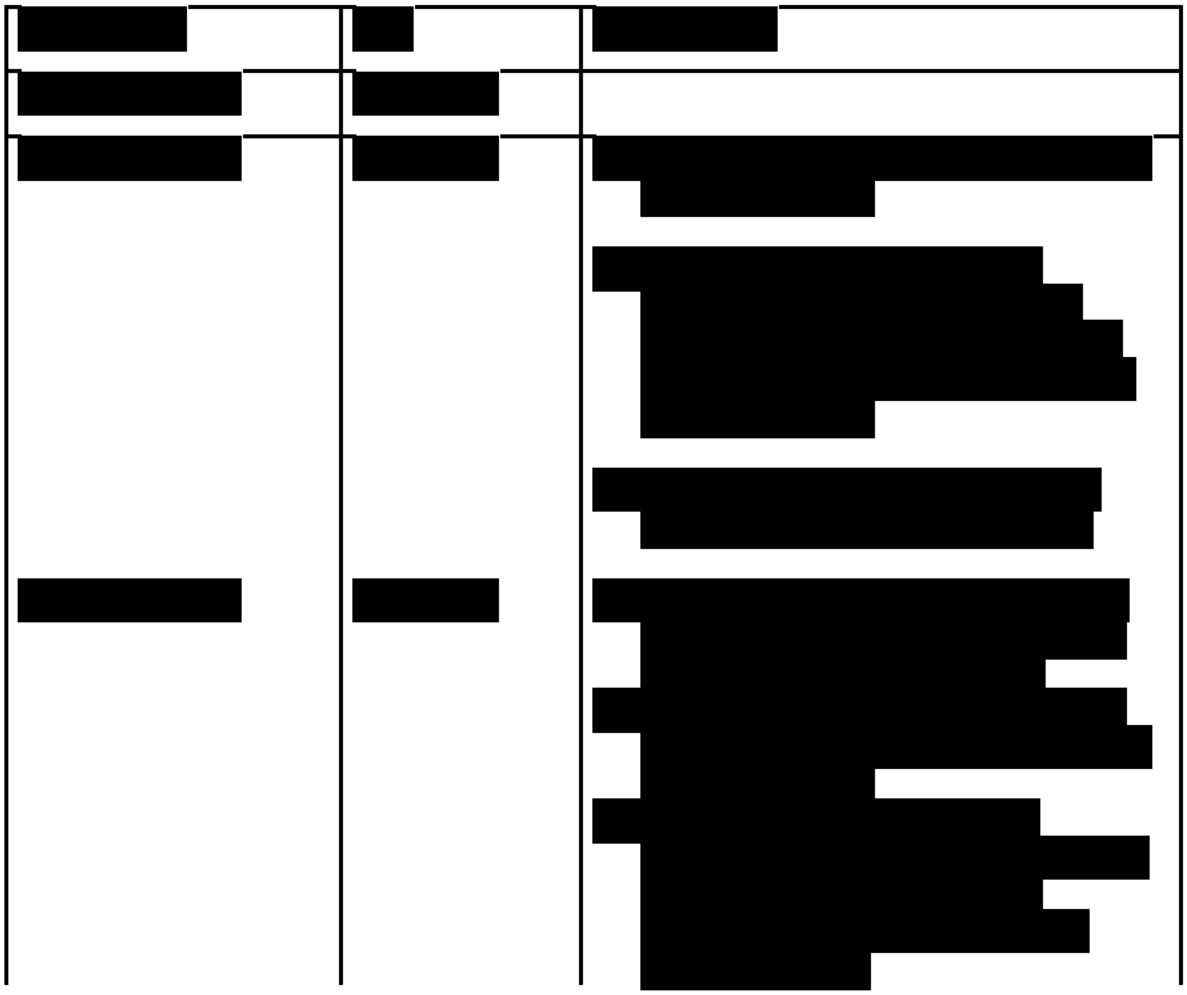












This page is a manifestation of an electronically captured signature

SIGNATURE PAGE

Document Name: P331-201-00242_Statistical_Analysis_Plan_Amendment 2

Document Number: 1000144391

Document Version: 2.0

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy hh:min) - UTC timezone
[redacted]	Clinical Approval	21-Jun-2021 17:08:22
[redacted]	Biostatistics Approval	21-Jun-2021 16:57:17