3111-302-001

A DOUBLE-BLIND, PLACEBO-CONTROLLED STUDY OF CARIPRAZINE AS AN ADJUNCT TO ANTIDEPRESSANTS IN THE TREATMENT OF PATIENTS WITH MAJOR DEPRESSIVE DISORDER WHO HAVE HAD AN INADEQUATE RESPONSE TO ANTIDEPRESSANTS ALONE

STATISTICAL ANALYSIS PLAN

Final: 16 JAN 2019
Amendment 1: 26 Mar 2021
Amendment 2: 28 Jul 2021

Sponsor:

Allergan Sales, LLC
5 Giralda Farms, Madison, NJ
07940

Allergan Ltd.
1st Floor, Marlow International,
The Parkway, Marlow
Buckinghamshire
SL7 1YL United Kingdom

1.0 TABLE OF CONTENTS

LIST OF TABLES

2.0 LIST OF ABBREVIATIONS

ADT	antidepressant therapy
AE	adverse event
AIMS	Abnormal Involuntary Movement Scale
ANCOVA	analysis of covariance
ATRQ	Antidepressant Treatment Response Questionnaire
CGI-I	Clinical Global Impressions–Improvement
CGI-S	Clinical Global Impressions–Severity
CRF	case report form
DB	Double-blind
ECG	electrocardiogram, electrocardiographic
HAM-A	Hamilton Rating Scale for Anxiety
HAMD-17	Hamilton Depression Rating Scale, 17 items
IP	Investigational Product
ITT	intent-to-treat
LOCF	last observation carried forward
MADRS	Montgomery-Åsberg Depression Rating Scale
MDD	major depressive disorder
MMRM	mixed-effects model for repeated measures
PCS	potentially clinically significant
PID	participant identification
QTc	QT interval corrected for heart rate
QTcB	QT interval corrected for heart rate using the Bazett formula ($QTcB = QT/(RR)^{1/2}$)
QTcF	QT interval corrected for heart rate using the Fridericia formula ($QTcF = QT/(RR)^{1/3}$)

SAE	serious adverse event
SAP	statistical analysis plan
SI	*Le Système International d'Unités* (International System of Units)
TEAE	treatment-emergent adverse event

3.0 INTRODUCTION

This statistical analysis plan (SAP) provides a more technical and detailed elaboration of the statistical analyses of the efficacy and safety data as outlined and/or specified in the final protocol of Study 3111-302-001 (version dated 23 Jul 2018) and the most recent amendment (amendment 3, dated 27 Jul, 2020). Specifications of tables, figures, and data listings are contained in a separate document. The SAP for pharmacokinetic/ pharmacodynamic will be prepared separately.

Study 3111-302-001 is a Phase 3, global, multicenter, randomized, double-blind, placebo-controlled, parallel-group fixed dose study in participants 18 to 65 years of age comparing cariprazine 1.5 mg/day and cariprazine 3 mg/day with placebo as an adjunctive treatment to ongoing ADT in outpatients with a diagnosis of MDD (via the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition [DSM-5]) who have an inadequate response, as measured by the modified Antidepressant Treatment Response Questionnaire (ATRQ), to 1 to 3 antidepressants administered during the current episode at an adequate dose (as defined per package insert) and duration of $\geq$ 6 weeks, with at least one dose escalation during the current depressive episode, and have a total score $\geq$ 22 on the Hamilton Depression Rating Scale-17 items (HAMD-17). Participants, who in the current episode, have had an inadequate response to an ADT taken above the minimum dose (per package insert), and for at least 6 weeks, can be enrolled. The study will consist of up to 14 days of screening and washout of prohibited medications followed by 6 weeks of DB treatment followed by a 4-week safety follow-up. Signed informed consent from the participant or the participant's legally authorized representative will be obtained before any study-related procedures are begun. Participants meeting the inclusion criteria will be randomized (1:1:1 ratio) to 1 of 3 treatment groups (cariprazine 1.5 mg/day +ADT or cariprazine 3 mg/day +ADT or placebo + ADT), stratified by ADT failure category (one ADT failure, more than one ADT failure) and the country.

During the DB treatment period (6 weeks), participants will take 1 capsule of IP, orally, per day in addition to their ongoing ADT as shown in Table 3-1.

The primary analysis will be performed after all participants have completed the 4-week safety follow-up or discontinued from the study and the database has been locked and unblinded.

Table 3-1 Treatment Regimen and Dosing

Drug/Dose	Investigational Product Frequency	Route of Administration
Double-Blind Treatment Period (Visit 2 through Visit 6)		
Placebo	Once daily	Oral (capsule)
Cariprazine 1.5 mg	Once daily	Oral (capsule)
Cariprazine 3 mg (1.5 mg/day for 2 weeks, starting at Visit 2 [Baseline]; 3 mg/day from Visit 4 [Week 2] through Visit 6 [Week 6])	Once daily	Oral (capsule)

All investigational products will be taken orally as a single daily dose at approximately the same time of day (morning or evening). The dosing time can be switched if there are tolerability problems. Any switch must allow at least 24 hours between 2 consecutive doses and must be documented in the eCRF.

COVID-19 pandemic emerged after March 11, 2020 when WHO declared COVID-19 a global pandemic. The Section 16.0 is added to specify analyses for evaluating the impact of COVID-19.

The COVID-19 pandemic has potential to impact the conduct of this study, due to possible effects on treatment duration and the collection, analysis and the interpretation of clinical trial data. Some protocol-specified visits in this study could possibly be impacted due to COVID-19 or logistical restrictions during the pandemic. For example, some scheduled visits may be missed due to self-quarantine or local government restrictions on travel; some visits may also be delayed or canceled due to healthcare resource constraints during the pandemic. Impacted visits due to COVID-19 will be recorded in the database. The probability of having missed visits and missing data due to COVID-19 infection or logistical restrictions related to the COVID-19 pandemic can be reasonably assumed to be unrelated to the unobserved values. Therefore, for the purpose of statistical analysis, it is reasonable to assume that these missing data are missing at random (MAR) and the statistical models that require MAR assumption are appropriate. The intent is to provide reliable estimates of the treatment effects targeted in the protocol under the scenario without the impact of COVID-19 pandemic. In addition to the planned sensitivity analyses for the primary and key secondary endpoint, one additional sensitivity analysis may be performed to assess the impact of missing data due to COVID-19 and the robustness of the conclusion. Details are provided in Section 16.1.

The schedule of evaluations for Study 3111-302-001 is presented in Table 3-2.

Table 3-2 Schedule of Visits and Procedures

Study Period	**Screening**	**Baseline**	**Double-blind Treatment**				**Safety Follow-up**
Visit	**1**	**2**	**3**	**4**	**5**	**6/ET**	**7**
Study Week	**Up to -2[k]**	**0**	**1**	**2**	**4**	**6[a,b]**	**10**
Study Day	**Up to -14[k]**	**1**	**8**	**15**	**29**	**43**	**71**
Visit Windows		**(within 14 days of start of screening procedures)[k]**	**± 3 Days**	**± 3 Days**	**± 3 Days**	**± 3 Days**	**± 3 Days**
Informed Consent	X						
Verified Clinical Trials Check[c]	X						
Medical and Psychiatric Histories	X						
Prior Medication History	X						
Inclusion/Exclusion Criteria	X	X					
Randomization		X					
Clinical Laboratory Tests[d,e]	X					X	
Serum Pregnancy Test[d]	X					X	X
Hepatitis Serology	X						
Hemoglobin A1c	X						
Urine Drug Screen[e]	X						
Blood Alcohol Concentration (by Breathalyzer)[e]	X						
Vital Signs[f]	X	X	X	X	X	X	X
Electrocardiogram	X					X	
Physical Examination	X					X	
SCID-5	X						
SAFER Remote Telephone Interview[g]	X						
Evaluate SAFER Score Qualification[g]		X					
Modified ATRQ	X						
HAMD-17	X	X				X	
YMRS	X	X				X	
MADRS	X	X	X	X	X	X	
CGI-S	X	X	X	X	X	X	

Study Period	Screening	Baseline	Double-blind Treatment				Safety Follow-up
Visit	**1**	**2**	**3**	**4**	**5**	**6/ET**	**7**
Study Week	**Up to -2[k]**	**0**	**1**	**2**	**4**	**6[a,b]**	**10**
Study Day	**Up to -14[k]**	**1**	**8**	**15**	**29**	**43**	**71**
Visit Windows		**(within 14 days of start of screening procedures)[k]**	**± 3 Days**	**± 3 Days**	**± 3 Days**	**± 3 Days**	**± 3 Days**
CGI-I			X	X	X	X	
HAM-A		X		X		X	
SF-12		X				X	
BARS/AIMS/SAS		X	X	X	X	X	
C-SSRS	X	X	X	X	X	X	X
Adverse Events	X	X	X	X	X	X	X
Concomitant Medications	X	X	X	X	X	X	X
Dispense IP		X	X	X	X		
Drug Return and Compliance			X	X	X	X	
ADT Compliance		X	X	X	X	X	
Pharmacogenetic Consent[h]				X			
Pharmacogenetic Sampling[i]				X			
Pharmacokinetic Sample[j]				X	X	X	

Note: After completion of Visit 6/ET, patients will be treated at the discretion of the investigator or designee.

ADT = antidepressant therapy; AIMS = Abnormal Involuntary Movement Scale; ATRQ = Antidepressant Treatment Response Questionnaire; BARS = Barnes Akathisia Rating Scale; CGI-I = Clinical Global Impressions–Improvement; CGI-S = Clinical Global Impressions–Severity; C-SSRS = Columbia–Suicide Severity Rating Scale; ET = early termination; HAM-A = Hamilton Anxiety Rating Scale; HAMD-17 = Hamilton Depression Rating Scale–17 items;
IP = investigational product; MADRS = Montgomery-Åsberg Depression Rating Scale; SAS = Simpson-Angus Scale; SF-12 = Short Form-12 v2 Health Survey; YMRS = Young Mania Rating Scale.

[a] Performed for all patients, including those prematurely discontinued after randomization (Visit 2).

[b] Clinical findings upon termination must be followed until the condition returns to prestudy status or can be explained as unrelated to investigational product. If necessary, an additional follow-up visit should be scheduled.

[c] Verified Clinical Trials database check to be performed, where applicable.

[d] Clinical laboratory tests include hematology, serum chemistry, urinalysis (including urine myoglobin in patients with creatine phosphokinase levels > 1000 U/L or as clinically indicated for any rise in creatine phosphokinase levels or as necessitated by symptoms), and blood alcohol, as well as serum pregnancy test in women of childbearing potential.

[e] Urine drug screen, blood alcohol concentration, and serum pregnancy test can be repeated at random at request of investigator.

[f] Height will be measured only at Visit 1 (Screening). Waist circumference will be measured at Visits 2 and 6/ET.

[g] SAFER/remote telephone interview will be scheduled and implemented before Visit 2 (Baseline) for patients who meet the screening criteria at Visit 1. The patient's SAFER evaluation score should be reviewed before Visit 2 to confirm eligibility.

[h] Pharmacogenetic consent may be obtained at any time between Visit 1 (Screening) and Visit 6 (Week 6).

[i] Pharmacogenetic sampling (1 per patient for the entire study) may be obtained from randomized patients at any time between Visit 2 (Baseline) and Visit 6 (Week 6).

[j] Can be taken at any time during visits 4, 5 and 6/ET.

[k] May be up to an additional 7 days (up to 21 days) if needed with Sponsor approval.

4.0 OBJECTIVES

The objective of this study is to evaluate the efficacy, safety, and tolerability of cariprazine 1.5 mg/day and 3 mg/day compared with placebo as an adjunctive treatment to ADT in patients with MDD who have had an inadequate response to antidepressants alone.

5.0 ANALYSIS POPULATIONS

Four populations will be considered in the statistical analysis of the study.

5.1 SCREENED POPULATION

The screened population will consist of all screened participants who sign informed consent.

5.2 RANDOMIZED POPULATION

The randomized population will consist of all participants in the screened population who were randomized to a treatment group.

5.3 SAFETY POPULATION

The safety population will consist of all participants in the randomized population who took at least 1 dose of double-blind investigational product.

Participants will be included in the treatment group corresponding to the double-blind investigational product they actually received in the safety population.

5.4 MODIFIED INTENT-TO-TREAT POPULATION

The modified intent-to-treat (mITT) population will consist of all randomized participants with ≥1 postbaseline assessment of the Montgomery-Åsberg Depression Rating Scale (MADRS) total score.

Participants will be included in the randomized treatment group for the mITT population.

5.5 STUDY-SPECIFIC DATA EXCLUDING ONE NON-COMPLIANT SITE

Due to multiple non-compliance issues at one study site (site #130), additional sensitivity analyses will be performed by excluding this site from the following:

- Demographic characteristics based on the Safety and mITT populations,
- Baseline characteristics based on the mITT population,
- Primary efficacy analysis (MMRM) based on the mITT population,
- Key secondary efficacy analysis (MMRM) based on the mITT population,
- Overall summary of safety (TEAEs, TESAEs, Death, TEAEs leading to discontinuation from study) based on the Safety population, and
- TEAEs by SOC and PT based on the Safety population.

6.0 PARTICIPANT DISPOSITION

The number and percentage of participants in the Randomized, Safety, and mITT Populations will be summarized by treatment group and study center; the number of participants screened will be summarized overall only by study center. The number and percentage of participants in the Randomized, Safety, and mITT Populations will be summarized by treatment group and background ADT as well.

Screen-failure participants (ie, participants screened but not randomized) and the associated reasons for failure to randomize will be tabulated overall for the screened participants. The number and percentage of participants who prematurely discontinued during the double-blind (DB) treatment period, and who entered the safety follow-up (SFU) period will be presented for each treatment group and pooled across treatment groups for the safety population. The reasons for premature discontinuation from the double-blind treatment period and the safety follow-up period as recorded on the disposition pages of the electronic case report forms (eCRFs) will be summarized (number and percentage) by treatment group and overall for the safety population. All participants who prematurely discontinue will be listed by discontinuation reason by period (the DB treatment period and the SFU period) for the safety population. The participants who were randomized but were not included in the safety population or the mITT population will be listed.

6.1 PROTOCOL DEVIATIONS

The number and percentage of participants with significant protocol deviations will be summarized by deviation category and treatment group for the Randomized Population. Supportive listings will also be provided.

These significant protocol deviations will be reviewed and documented before database lock and unblinding of treatment codes.

7.0 DEMOGRAPHICS AND OTHER BASELINE CHARACTERISTICS

Demographic parameters (eg, age, race, sex, ethnicity), baseline characteristics (weight, height, body mass index, ADT failure category, Country, ADT failure category by Country) will be summarized by treatment group for the safety and mITT populations, and baseline efficacy variables will be summarized by treatment group for the mITT population. Continuous variables will be summarized by number of participants and mean, SD, median, minimum, and maximum values. Categorical variables will be summarized by number and percentage of participants.

The ADT failure category based on eCRF will be considered primary data on ADT failure and will be used in all the statistical models rather than the one based on IWRS.

Abnormalities in participants' medical and surgical histories will be coded using the *Medical Dictionary for Regulatory Activities,* version 24.0. The number and percentage of participants with abnormalities in medical and surgical histories in each system organ class and preferred term will be summarized by treatment group for the Safety Population. Psychiatric history, non-drug psychiatric treatment, history of ocular events of special interest will be summarized by treatment group for the Safety Population.

Prior medication is defined as any medication taken before the date of the first dose of DB IP. *Concomitant medication* is defined as any medication taken on or after the date of the first dose of investigational product, and will be summarized for both double-blind treatment period and safety follow-up period.

The March 2017 version of the World Health Organization (WHO) Drug Dictionary Enhanced will be used to classify prior and concomitant medications by therapeutic class. The number and percentage of participants with prior and concomitant drug use will be summarized by treatment group and Anatomical Therapeutic Chemical 4 code for the safety population. Multiple drug use by a participant in the same category will be counted only once in the summary tables.

Rescue medications will be summarized by treatment group and generic terms for the following rescue medication categories in the safety population:

- Insomnia
- Extrapyramidal symptoms (EPS) or akathisia
- Agitation, restlessness, and hostility

In addition, the number of ADT failures for current episode of depression and the status of presently taking ADTs based on ATRQ eCRF will be summarized for the safety population by treatment group.

8.0 EXTENT OF EXPOSURE AND TREATMENT COMPLIANCE

8.1 EXTENT OF EXPOSURE

Exposure to the DB IP for the Safety Population during the double-blind treatment period will be summarized for treatment duration, calculated as the number of days from the date of the first dose of DB IP to the date of the last dose of DB IP, inclusive. Descriptive statistics (number of participants, mean, SD, median, minimum, and maximum) will be presented by treatment group.

Exposure to open-label ADT for the Safety Population during the double-blind treatment period will be summarized for treatment duration, calculated as the number of days from the date of the first dose taken to the date of the last dose taken, inclusive, during the DB treatment period.

Patient-years, defined as exposure to the DB IP in years, will be summarized by treatment for the safety population.

8.2 MEASUREMENT OF TREATMENT COMPLIANCE

Dosing compliance for the double-blind investigational product during a specified period is defined as the total number of capsules actually taken by a participant during that period divided by the number of capsules that were expected to be taken during the same period multiplied by 100. The total number of capsules actually taken during a specific period is calculated as the number of days in that period multiplied by the number of capsules taken each day during that period. The number of capsules expected to be taken for a specific treatment period will be calculated by multiplying the number of days in that period by the number of capsules to be taken per day. Descriptive statistics for investigational product compliance will be presented by treatment group for each period between 2 consecutive visits, as well as for the period from the first dose of DB IP actually taken to the last dose of DB IP actually taken for the Safety Population

Dosing compliance for the background ADT during a specified period is defined as the dose actually taken by a participant during that period divided by the dose expected to be taken during the same period multiplied by 100. Descriptive statistics for ADT compliance will be presented for each ADT and overall by the double-blind investigational product treatment group for the period from the first dose of DB IP actually taken to the last dose of DB IP actually taken for the Safety Population.

9.0 EFFICACY ANALYSES

All efficacy analyses will be based on the mITT population. Baseline for efficacy is defined as the last non-missing efficacy assessment before the date of the first dose of DB IP. All statistical tests will be conducted at the α=0.05 (2-sided) level for main effects. All confidence intervals (CIs) will be 2-sided 95% CIs, unless stated otherwise.

For efficacy analyses in which country is a factor, a *small country* will be defined as a country with fewer than 2 participants in at least one treatment group in the mITT Population. Small countries will be pooled to form pseudo-countries so that each treatment group includes at least 2 mITT participants within the country. Countries will be ranked on number of participants that are part of the mITT Population and lowest ranked countries will be pooled (sequentially, i.e., a new ranking is made after each step) until the smallest pooled country has reached at least 2 participants in each treatment group.

To compare with historic studies, some analyses based on the imputed data using the last-observation-carried-forward (LOCF) approach, will be presented for all efficacy parameters. Only the postbaseline total score of a parameter will be imputed; individual item scores will not be carried forward. The baseline value will be carried forward only for the intermittent missing values immediately after baseline. If all the postbaseline values are missing, the baseline value will not be carried forward.

9.1 PRIMARY EFFICACY PARAMETER

The primary estimand is provided below through specification of the population, the variable, the handling of intercurrent events, and the population-level summary.

Population

The target population are participants with MDD who have had an inadequate response to antidepressant therapy in the current episode and who satisfy the inclusion and exclusion criteria as specified in the protocol.

The analysis population is defined to be the mITT population consisting of all randomized participants with $\geq$ 1 postbaseline assessment of the MADRS total score.

Variable

The variable is the primary efficacy endpoint, change from baseline to Week 6 in the MADRS total score.

Accounting of Intercurrent Events

Intercurrent events and their handling rules are described as follows:

- Regardless of whether allowed rescue medications are taken or not, data are included in the analysis. Participants are still taking assigned drugs as specified in the protocol while taking allowed rescue medications. The clinical objective is to assess the efficacy of the treatment regardless of allowed rescue medication use.
- To evaluate the efficacy at Week 6 in the mITT population, participants are assumed to adhere to the assigned treatment for the duration of the study. As a result, data after the discontinuation from the study treatment due to all reasons will not be included in the primary analysis and they will be assumed as missing at random.

Population-level Summary

The population-level summary for the primary endpoint is the difference in primary variable means between each cariprazine group and placebo.

To address the primary hypotheses that each dose of cariprazine is superior to placebo in the reduction of MADRS total score after 6 weeks of treatment, the change from baseline in MADRS total score will be analyzed using a mixed-effects model for repeated measures (MMRM) with treatment group, pooled country, ADT failure category (per CRF), visit, and treatment group-by-visit interaction as fixed effects, and the baseline value and baseline-by-visit interaction as the covariates. An unstructured covariance matrix will be used to model the covariance of within-participant scores. The Kenward-Roger approximation (Kenward and Roger, 1997) will be used to estimate denominator degrees of freedom. In the event of non-convergence of the model using unstructured covariance matrix, a structured covariance matrix will be used in combination with empirical variance estimate (i.e., sandwich estimator) to address the potential mis-specified situation. The following sequence of alternative covariance structures (First-order ante-dependence [ANTE (1)], Toeplitz [TOEP], First-order autoregressive [AR (1)] and compound symmetry[CS]) will be considered in the MMRM until convergence.

The analysis will be performed based on all postbaseline scores using only the observed cases without imputation of missing values. The treatment difference in the primary endpoint at Week 6 for each active treatment group versus placebo in the mITT population will be estimated and reported along with the corresponding 95% CI and the p-value.

The two-stage mixture parallel gatekeeping procedure (Dmitrienko 2011) (truncated Hochberg with truncation parameter of 0.9 for the primary endpoint and regular Hochberg for the secondary endpoint) will be used to control the overall type I error rate at a 0.025 level (1-sided) for multiple comparisons of 2 active doses with placebo for the primary endpoint and key secondary endpoint. Details are provided in Section 9.2.

The study will be considered positive if at least 1 dose arm of cariprazine is statistically superior to placebo for change from baseline in MADRS total score at Week 6 after multiplicity adjustment.

A sensitivity analysis using a pattern-mixture model based on non–future-dependent missing value restrictions (Kenward et al, 2003) will be performed to assess the robustness of the primary MMRM results to the possible violation of the missing-at-random assumption. The details of this sensitivity analyses are as follows:

The pattern for the pattern-mixture model will be defined by the participant's last visit with an observed value. The observed MADRS total score at a visit is assumed to have a linear relationship with the participant's prior measurements. The missing values will be imputed under the assumption that the distribution of a missing observation differs from the observed only by a shift parameter value Δ. The dataset with missing values imputed will be analyzed using an analysis-of-covariance (ANCOVA) model with treatment group, pooled country, and ADT failure category (per CRF), as factors and baseline MADRS total score as a covariate for between–treatment-group comparison at Week 6. The imputation of missing values and the analysis will be performed multiple times, and the inference of this sensitivity analysis will be based on the combined estimates using the standard multiple imputation technique. The values for Δ will be selected as 0 to 6, inclusively. This range is selected because the mean reduction of MADRS score from baseline within a treatment group at Week 6 is likely to be within 15 points (Durgam et al, 2016; El-Khalili et al, 2010), and a Δ value of 6 accounts for 40% of treatment efficacy.

The second sensitivity analysis will consider dropout reasons while imputing missing values after the discontinuation. Participants who discontinued due to lack of efficacy in the cariprazine arms are assumed to have no treatment effect after the discontinuation. These participants are assumed to copy the profile of placebo arm and missing values are imputed based on the distribution estimated from the placebo group using copy-reference approach (Carpenter et al, 2013). The rest of missing values in the placebo arm and cariprazine arms will be imputed using the observed data in their respective group under the MAR assumption.

Least squares means and their standard errors for change in MADRS total score based on the MMRM analysis will be plotted by treatment group and visit. Mean treatment difference (raw mean difference ± standard error) in change in primary endpoint between each cariprazine dose and placebo with missing values imputed using the LOCF approach will be plotted against country. In addition, summary statistics will be provided by center to examine the consistency of treatment effect across the study centers. The impact of dropouts on the efficacy outcomes will be explored graphically by plotting the time courses of mean changes by dropout reason.

To compare with historic studies, the LOCF approach will be used with an ANCOVA model that has treatment group, pooled country and ADT failure category (per CRF) as factors and the baseline value as the covariate. Results of analyses from MMRM and from ANCOVA models will also be reported for each visit between baseline and Week 6.

To determine whether the treatment effect is consistent across various subgroups, the estimate of the between-group treatment effect (with a 95% CI) for each cariprazine dose versus placebo for the primary efficacy variable will be estimated within each category of the following classification variables:

- ADT failure category (one ADT failure, more than one ADT failure)
- Region (US, Non-US)

The treatment effect within each category of a subgroup factor will be assessed using MMRM with terms including treatment group, visit, and treatment group-by-visit interaction, the baseline value and baseline-by-visit interaction, in addition to one term defined as following:

- Add term ADT failure category while conducting region subgroup analyses
- Add term Country while conducting ADT failure category subgroup analyses.

Treatment effects and nominal 95% CI by category for the classification variables listed above will be reported. Because the sample size for a specific subgroup category might be small, the MMRM might not converge, in which case only summary statistics using the LOCF data at Week 6 will be reported.

9.2 KEY SECONDARY EFFICACY PARAMETER

The key secondary endpoint is the change from baseline to Week 6 in the CGI-S score.

To test the secondary hypotheses that each dose of cariprazine is superior to placebo in the reduction of CGI-S score after 6 weeks of treatment, the change from baseline in CGI-S score will be analyzed using the MMRM approach similar to the one used for the primary analysis. A sensitivity analysis will also be performed using ANCOVA (with LOCF imputation) as used for the analysis of primary endpoint.

To control the overall familywise type I error rate at the 0.025 level (one-sided), a truncated Hochberg will be used for the primary endpoint, and the regular Hochberg procedure for the key secondary endpoint. The truncation parameter will be set to $\gamma = 0.9$ (Dmitrienko 2011).

The cariprazine dose groups will be labeled as Dose 1 (1.5 mg) and Dose 2 (3.0 mg).

The null hypotheses can be grouped into two families:

(i) Family 1 includes the null hypotheses $H_1 and\ H_2$ for the primary endpoint:

H_1 denotes the null hypothesis of the difference in mean change from baseline in MADRS at Week 6 between cariprazine Dose 1 and placebo is ≥ 0.

H_2 denotes the null hypothesis of the difference in mean change from baseline in MADRS at Week 6 between cariprazine Dose 2 and placebo is ≥ 0.

(ii) Family 2 includes the null hypotheses $H_3 and\ H_4$ for the key secondary endpoint:

H_3 denotes the null hypotheses of the difference in mean change from baseline in CGI-S at Week 6 between cariprazine Dose 1 and placebo is ≥ 0.

H_4 denotes the null hypotheses of the difference in mean change from baseline in CGI-S at Week 6 between cariprazine Dose 2 and placebo is ≥ 0.

The gatekeeping procedure utilizes the following decision rules:

(i) Step 1. The null hypotheses in Family 1 are tested using the truncated Hochberg procedure at the full α level. Let $p_{(1)} < p_{(2)}$ denote the ordered p-values in Family 1 and let $H_{(1)}$ and $H_{(2)}$ denote the associated null hypotheses. The truncated Hochberg procedure rejects both $H_{(1)}$ and $H_{(2)}$ if $p_{(2)} \leq \alpha(1+\gamma)/2$ or it rejects only $H_{(1)}$ if $p_{(1)} \leq \alpha/2$ and $p_{(2)} > \alpha(1+\gamma)/2$.

(ii) Step 2. The null hypotheses in Family 2 are tested as follows:

  (i) The null hypotheses are tested using the regular Hochberg procedure at the full α level if both $H_{(1)}$ and $H_{(2)}$ are rejected. Let $p_{(3)} < p_{(4)}$ denote the ordered p-values in Family 2 and let $H_{(3)}$ and $H_{(4)}$ denote the associated null hypotheses. The Hochberg procedure rejects both $H_{(3)}$ and $H_{(4)}$ if $p_{(4)} \leq \alpha$ or it rejects only $H_{(3)}$ if $p_{(3)} \leq \alpha/2$ and $p_{(4)} > \alpha$.

  (ii) The null hypothesis logically related to $H_{(1)}$ is tested at $\alpha(1-\gamma)/2$ if $H_{(1)}$ is rejected and $H_{(2)}$ is not rejected.

(iii) Step 3. If only one null hypothesis is rejected in Family 1 and only one null hypothesis is rejected in Family 2, the remaining null hypothesis in Family 1 is re-tested at the full α level. If this null hypothesis is rejected, the associated null hypothesis in Family 2 is tested at the full α level.

9.3 ADDITIONAL EFFICACY PARAMETER(S)

Additional efficacy parameters will include the following at postbaseline visits described in Table 3-2:

- Change from baseline in MADRS total score at Weeks 1, 2, 4
- Change from baseline in the CGI-S score at Weeks 1, 2, 4
- CGI-I score at Weeks 1, 2, 4, 6

- CGI-I response (CGI-I score ≤ 2) at Weeks 1, 2, 4, 6
- MADRS response (≥ 50% reduction from baseline in MADRS total score) at Weeks 1, 2, 4, 6
- MADRS remission (MADRS total score ≤ 10) at Weeks 1, 2, 4, 6
- Change from baseline in the HAMD-17 total score at Week 6
- Change from baseline in the HAM-A total score at Week 2, 6
- HAM-A response (≥ 50% reduction from baseline in HAM-A total score) at Weeks 2, 6
- HAM-A remission (HAM-A total score ≤ 7) at Weeks 2, 6
- Change from baseline in MADRS individual item scores at Weeks 1, 2, 4, 6
- Change from baseline in HAM-A individual item scores at Weeks 2, 6

Additional quantitative efficacy parameters will be analyzed in the following way:

Analysis of CGI-I score, change from baseline in HAM-A total score, change from baseline in MADRS and HAM-A individual item scores, and change from baseline in CGI-S will be performed using a similar MMRM to that used for the primary analysis. Baseline CGI-S score will be used as a covariate for the analysis of CGI-I score. In addition, these parameters will be analyzed using ANCOVA (with LOCF imputation) as used for the analysis of primary efficacy parameters.

Change from baseline in HAMD-17 total score will be analyzed based on ANCOVA model using observed cases, with treatment group, pooled country, and ADT failure category (per CRF) as factors and baseline HAMD-17 total score as a covariate for between–treatment-group comparisons at Week 6.

Additional categorical efficacy parameters will be analyzed in the following way:

Rates for categorical parameters (response and remission) will be reported by treatment group and by visit; a logistic regression model (with LOCF imputation) will be used to model the probability of a response or the probability of a remission as a function of a treatment group, pooled country, ADT failure category (per CRF) and the corresponding baseline score (CGI-S for the analysis of the CGI-I response) as explanatory variables.

10.0 SAFETY ANALYSES

The safety analysis will be performed using the safety population for the double-blind treatment period and safety follow-up period. The double-blind treatment period starts with the first dose of double-blind investigational product and ends with the last scheduled assessment at Visit 6 (Week 6) or early termination (before Week 6). The safety follow-up period starts the day after the last scheduled assessment at Visit 6 (or early termination) and ends with the last safety follow-up visit. The safety parameters will include adverse events (AEs), clinical laboratory parameters, vital signs, electrocardiogram (ECG) parameters, EPS scales, the C-SSRS and the YMRS. For each safety parameter, the last assessment made before the first dose of double-blind investigational product will be used as the baseline for all analyses of that safety parameter. The summarization will be by treatment group for both double-blind and safety follow-up periods. Continuous variables will be summarized by number of participants and mean, SD, median, minimum, and maximum values. Categorical variables will be summarized by number and percentage of participants.

10.1 ADVERSE EVENTS

Adverse events will be coded by system organ class and preferred term using the *Medical Dictionary for Regulatory Activities,* version 24.0.

An adverse event (AE) will be considered a treatment-emergent adverse event (TEAE) if the AE began or worsened (increased in severity or became serious) on or after the first dose of DB investigational product. Per case report form instructions, a new AE record will be created with a new AE onset date for any AE that worsens. Therefore, TEAEs can simply be identified as those AEs with recorded onset date (and time, if known) on or after the date of the first dose of study intervention.

An AE will be considered a TESAE if it is a TEAE that additionally meets any SAE criterion.

For the DB treatment period, the number and percentage of participants reporting TEAEs in each treatment group will be tabulated by system organ class and preferred term; by system organ class, preferred term, and severity; and by system organ class, preferred term, and relationship to the IP. If more than 1 AE is coded to the same preferred term for the same participant, the participant will be counted only once for that preferred term using the most severe and most related occurrence for the summarization by severity and by relationship to the IP.

The distribution of TEAEs by severity and relationship to the IP will be summarized by treatment group.

The TEAEs during the safety follow-up period will be summarized by body system, preferred term, and treatment group for the safety population.

An AE that occurs more than 30 days after the date of the last dose of DB IP or occurs after the last safety follow-up visit (whichever comes last) will not be summarized except for AEs related to reported pregnancies. AEs associated with pregnancies reported in the protocol-required timeframe up to approximately 12 weeks following the last dose of DB IP will be summarized.

The number and percentage of participants with common (≥ 2% of participants in any treatment group) TEAEs, will be summarized by preferred term and treatment group for the double-blind period.

Summary tables will be provided for participants with TESAEs and participants with TEAEs leading to discontinuation if these occurred in 5 or more participants. The number and percentage of participants with TESAEs will be summarized for both the double-blind and safety follow-up periods.

Listings will be presented for participants with SAEs, participants with AEs leading to discontinuation, and participants who die (if any). All participants with SAEs, including SAEs reported during the screening period and the safety follow-up period, and participants discontinuing because of AEs occurring before the start of DB IP will be included in these listings.

The number and percentage of participants reporting TEAEs of ocular events of special interest during the double-blind treatment period will be summarized. A listing of all reported ocular events of special interest will be provided.

The number and percentage of participants with EPS TEAEs, EPS treatment-emergent SAEs, and the number and percentage of participants with EPS TEAEs leading to premature discontinuation during the double-blind treatment period will be summarized for the double-blind treatment period.

For EPS AEs, the preferred terms include:

Akathisia, Restlessness, Dystonia cluster (Dystonia, Myoclonus, Oculogyric crisis, Oromandibular dystonia, Tongue spasm, Trismus, Torticollis), Parkinsonism cluster (Akinesia, Bradykinesia, Cogwheel rigidity, Drooling, Dyskinesia, Extrapyramidal disorder, Hypokinesia, Reduced facial expression, Muscle rigidity, Muscle tightness, Parkinsonism, Salivary hypersecretion, Tremor), Musculoskeletal stiffness, and Tardive dyskinesia.

For EPS AEs excluding Akathisia/Restlessness, the preferred terms include all the terms listed above except Akathisia and Restlessness.

Any EPS TEAEs including Akathisia/Restlessness, any EPS TEAEs excluding Akathisia/Restlessness, and any TEAEs of Akathisia/Restlessness will also be summarized.

10.2 CLINICAL LABORATORY PARAMETERS

Descriptive statistics for clinical laboratory values (in SI units) and changes from the baseline values at each assessment time point will be presented by treatment group for the following laboratory parameters:

Hematology: Absolute and differential white blood cell count, erythrocyte count, hemoglobin, hematocrit, platelet count, and red blood cell indices (mean corpuscular volume, mean corpuscular hemoglobin, and mean corpuscular hemoglobin concentration)

Chemistry: Sodium, potassium, calcium, chloride, bicarbonate, glucose, blood urea nitrogen, creatinine, total protein, alkaline phosphatase, albumin, total bilirubin, direct bilirubin, indirect bilirubin, AST, ALT, free T3, free T4, TSH, lactate dehydrogenase, creatine phosphokinase, γ-glutamyl transpeptidase, uric acid, phosphate, lipid panel (total cholesterol, triglycerides, low-density lipoproteins, high-density lipoproteins), prolactin, insulin, and magnesium

Urinalysis: Specific gravity, pH

In addition, descriptive statistics for values and changes from the baseline values in conventional units at each assessment time point will be presented for selected clinical laboratory parameters listed in Appendix II:. A description of reporting the lab values in conventional units in participant narratives (along with the standard reporting in SI units) is presented at the end of Appendix II:. Only participants with clinical laboratory data at baseline and at least one postbaseline visit during the double-blind period will be included in the summary.

Laboratory tests values are considered to be potentially clinically significant (PCS) if they meet either the low or high PCS criteria listed in Table 10-1. The number and percentage of participants with PCS postbaseline clinical laboratory values will be tabulated by treatment group. The percentages will be calculated relative to the number of participants with available non-PCS baseline values and at least 1 postbaseline assessment. The numerator will be the total number of participants with available non-PCS baseline values and at least 1 PCS postbaseline value. A supportive listing of participants with PCS postbaseline values will be provided, including the participant identification number, study center number, and baseline and postbaseline values. A listing of all AEs occurring in participants who have PCS laboratory values will also be provided.

The number and percentage of participants with treatment-emergent significant changes in lipid parameters (total cholesterol, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, and triglycerides) and glucose will be tabulated by treatment group for the DB treatment period. The criteria for treatment-emergent significant changes in lipids and glucose is provided in Table 10-2. Percentages will be calculated relative to the number of participants with baseline values meeting the specified baseline criteria and with at least 1 postbaseline assessment. The change in lipids and glucose from baseline to the highest (lowest for high-density lipoprotein cholesterol) postbaseline measurement will be summarized. Supportive listings of participants with treatment-emergent changes in lipids and glucose values will be provided.

The number and percentage of participants meeting potential Hy's Law criteria (elevation of alanine aminotransferase [ALT] or aspartate aminotransferase [AST] ≥ 3x ULN with total bilirubin [TBL] ≥ 2x ULN and alkaline phosphatase [ALP] < 2x ULN in a 24-hour period) and Concurrent Elevations criteria in a 24-hour period will be tabulated by DB treatment group for the safety population. A supportive listing will be provided.

Concurrent Elevations in a 24-hour period are defined as:

1. ALT or AST ≥ 3x ULN and TBL≥ 1.5 x ULN
2. ALT or AST ≥ 3x ULN and TBL≥ 2 x ULN

Table 10-1 Clinical Laboratory PCS Criteria

Laboratory Parameter	***SI Unit***	***PCS Criteria Low Values***	***PCS Criteria High Values***
Hematology			
Hemoglobin	g/L	< 0.9 × LLN	—
Hematocrit	%	< 0.9 × LLN	—
Eosinophils	%	—	> 10
Neutrophils	%	< 30	> 90
Basophils	%	—	> 6
Monocytes	%	—	> 20
Lymphocytes	%	< 10	> 60
Absolute neutrophils	10^9/L	< 1.0	—
Platelet count	10^9/L	≤ 75	≥ 700
White cell count	10^9/L	≤ 2.5	≥ 15
Chemistry			
Albumin	g/L	< 0.9 × LLN	> 1.1 × ULN
Alkaline phosphatase	U/L	—	≥ 3 × ULN
ALT	U/L	—	≥ 3 × ULN
AST	U/L	—	≥ 3 × ULN
GGT	U/L	—	≥ 3 × ULN
LDH	U/L	—	≥ 3 × ULN
Blood urea nitrogen	mmol/L	—	> 1.2 × ULN
Calcium	mmol/L	< 0.9 × LLN	> 1.1 × ULN
Chloride	mmol/L	< 0.9 × LLN	> 1.1 × ULN
Cholesterol	mmol/L	—	> 1.3 × ULN
HDL	mmol/L	< 0.8 × LLN	—
LDL	mmol/L	—	> 1.2 × ULN
CPK	U/L	—	> 1.5 × ULN
Creatinine	µmol/L	—	> 1.3 × ULN
Glucose, fasting	mmol/L	< 0.8 × LLN	> 1.2 × ULN
Magnesium	mmol/L	< 0.9 × LLN	> 1.1 × ULN
Potassium	mmol/L	< 0.9 × LLN	> 1.1 × ULN
Sodium	mmol/L	< 0.9 × LLN	> 1.1 × ULN
Total bilirubin	µmol/L	—	> 1.5 × ULN
Total protein	g/L	< 0.9 × LLN	> 1.1 × ULN
Triglycerides, fasting	mmol/L	—	> 1.2 × ULN
Uric acid	µmol /L	—	> 1.1 × ULN
Urinalysis			
Protein	—	—	At least 2 +
Glucose	—	—	At least 2 +
Blood	—	—	At least 2 +

ALT = alanine aminotransferase; AST = aspartate aminotransferase; CPK = creatine phosphokinase; GGT = gamma glutamyl transferase; HDL = high-density lipoprotein; LDH = lactate dehydrogenase; LDL = low-density lipoprotein; LLN = lower limit of normal laboratory reference range; PCS = potentially clinically significant; SI = *Le Système International d'Unités* (International System of Units); ULN = upper limit of normal laboratory reference range.

Laboratory Parameter	***SI Unit***	***PCS Criteria Low Values***	***PCS Criteria High Values***

Fasting: Includes both, fasting and unspecified fasting status.

Table 10-2 The Criteria for Treatment-emergent Significant Changes in Lipids and Glucose

Laboratory Parameter	**Measurement**	**Category**	**Criteria**
Total cholesterol (Fasting or Non-fasting)	Actual	Normal	< 200 mg/dL
		Borderline	≥ 200 and < 240 mg/dL
		Normal/Borderline	< 240 mg/dL
		High	≥ 240 mg/dL
		Borderline/High	≥ 200 mg/dL
	Change from baseline	Increase ≥ 40 mg/dL	Increase ≥ 40 mg/dL
LDL Cholesterol* (Fasting)	Actual	Normal	< 100 mg/dL
		Borderline	≥ 100 and < 160 mg/dL
		Normal/Borderline	< 160 mg/dL
		High	≥ 160 mg/dL
		Borderline/High	≥ 100 mg/dL
	Change from baseline	Increase ≥ 30 mg/dL	Increase ≥ 30 mg/dL
HDL Cholesterol (Fasting and Non-Fasting)	Actual	Normal	≥ 40 mg/dL
		Low	< 40 mg/dL
	Change from baseline	***Decrease*** ≥ 20 mg/dL	***Decrease*** ≥ 20 mg/dL
Triglycerides (Fasting)	Actual	Normal	< 150 mg/dL
		Borderline	≥ 150 and < 200 mg/dL
		Normal/ Borderline	< 200 mg/dL
		High	≥ 200 mg/dL
		Very High	≥ 500 mg/dL
		Borderline/High/Very High	≥ 150 mg/dL
		≥ 1000 mg/dL	≥ 1000 mg/dL
	Change from baseline	Increase ≥ 50 mg/dL	Increase ≥ 50 mg/dL
Triglycerides (Non-Fasting)	Actual	Very High	≥ 500 mg/dL
		≥ 1000 mg/dL	≥ 1000 mg/dL
Serum Glucose (Fasting)	Actual	Normal	< 100 mg/dL
		Impaired	≥ 100 and < 126 mg/dL
		Normal/ Impaired	< 126 mg/dL
		High	≥ 126 mg/dL
	Change from baseline	Increase ≥ 10 mg/dL	Increase ≥ 10 mg/dL
Serum Glucose (Non-Fasting)	Actual	Normal	< 140 mg/dL
		Borderline	≥ 140 and < 200 mg/dL
		Normal/Borderline	< 200 mg/dL
		High	≥ 200 mg/dL
		Borderline/High	≥ 140 mg/dL
	Change from baseline	Increase ≥ 20 mg/dL	Increase ≥ 20 mg/dL

* LDL direct and LDL calculated are combined.
Fasting: Includes both, fasting and unspecified fasting status.

10.3 VITAL SIGNS

Descriptive statistics for vital signs (supine radial pulse rate, supine systolic and diastolic BP, body weight, waist circumference, BMI and temperature) and changes from baseline values at each visit and at the end of the double-blind treatment period will be presented by treatment group.

Vital sign values will be PCS if they meet both the observed-value criteria and the change from baseline value criteria. The criteria for PCS vital sign values are provided in Table 10-3. The percentages will be calculated relative to the number of participants with baseline values and at least 1 postbaseline assessment. The numerator will be the total number of participants with available baseline values and at least 1 PCS postbaseline value. A supportive listing of participants with postbaseline PCS values will be provided, including the participant identification number, study center number, and baseline and postbaseline values. A listing of all AEs occurring in participants who have PCS vital sign values will also be provided.

Table 10-3 Criteria for Potentially Clinically Significant Vital Signs

Parameter	***Flag***	***Criteria***	
		Observed Value	*Change From Baseline*
Supine systolic blood pressure, mm Hg[a]	High	≥ 180	Increase of ≥ 20
	Low	≤ 90	Decrease of ≥ 20
Supine diastolic blood pressure, mm Hg[a]	High	≥ 105	Increase of ≥ 15
	Low	≤ 50	Decrease of ≥ 15
Supine pulse rate, bpm[a]	High	≥ 120	Increase of ≥ 15
	Low	≤ 50	Decrease of ≥ 15
Weight, kg	High	—	Increase of ≥ 7%
	Low	—	Decrease of ≥ 7%

a A postbaseline value is considered potentially clinically significant if it meets both the observed-value and the change-from-baseline criteria.

bpm = beats per minute.

The number and percentage of participants with PCS weight increase or decrease will also be provided by baseline BMI categories (underweight [< 18.5 kg/m^2], normal [≥ 18.5 and < 25 kg/m^2], overweight [≥ 25 and < 30 kg/m^2], and obese [≥ 30 kg/m^2]) by treatment group for the safety population.

The number and percentage of participants with orthostatic hypotension will be provided by treatment group. Orthostatic hypotension is defined as a reduction of ≥20 mm Hg in systolic BP or a reduction of ≥ 10 mm Hg in diastolic BP measured after the participant stands up after resting in the supine position. Standing measurements should be taken after a sufficient amount of time has passed to allow the BP to equilibrate in the standing state. A supportive listing will be provided including the participant identification number, study center number, and baseline and postbaseline systolic and diastolic BP values (supine and standing). A listing of all AEs occurring in participants who have orthostatic hypertension will also be provided.

As specified in the Seventh Report of the Joint National Committee on prevention, detection, evaluation and treatment of high blood pressure (NIH Publication No. 04-5230, 2004), hypertension status is defined as:

- Normotensive: supine SBP < 120 mm Hg and DBP < 80 mm Hg.
- Prehypertension: supine SBP 120 to 139 mm Hg or DBP 80 to 89 mm Hg.
- Stage I Hypertension: supine SBP 140 to 159 mm Hg or DBP 90 to 99 mm Hg.
- Stage II Hypertension: supine SBP ≥ 160 mm Hg or DBP ≥ 100 mm Hg.

Tabulations showing the number and percentage of participants with hypertension status changes from baseline will be provided for:

- Shift of hypertension status from baseline to end of the DB treatment period
- Shift of hypertension status from baseline to highest category during the DB treatment period

Supportive listings of participants who have a shift in hypertension status from normotensive/prehypertension at Baseline to stage I/stage II hypertension will be provided.

10.4 ELECTROCARDIOGRAM

Descriptive statistics for ECG parameters (heart rate, RR interval, PR interval, QRS interval, QT interval, QTcB interval, and QTcF interval) and changes from baseline values at each assessment time point to the end of the DB treatment period will be presented by treatment group. The QTc will be calculated using both the Bazett and Fridericia corrections (if the vendor does not provide).

Electrocardiographic parameter values are considered PCS if they meet or exceed the higher-limit PCS criteria listed in Table 10-4. The number and percentage of participants with PCS postbaseline ECG values will be tabulated by treatment group for the double-blind treatment period. The percentages will be calculated relative to the number of participants with available non-PCS baseline values and at least 1 postbaseline assessment for the double-blind treatment period. The numerator is the total number of participants with available non-PCS baseline values and at least 1 PCS postbaseline value for the double-blind treatment period. A supportive tabular display of participants with PCS postbaseline values will be provided, including the participant identification (PID) number, baseline, all postbaseline (including non-PCS) values, and change from baseline.

In addition, a tabular display showing all AEs that occurred in participants who had postbaseline PCS ECG values will be provided.

A shift table from baseline to the end of study in the Investigator's overall interpretation of the ECG will be presented by treatment group for the following categories: normal; abnormal, not clinically significant; abnormal, clinically significant. A tabular display showing participants with postbaseline clinically significant ECG abnormalities according to the Investigator's overall interpretation will be provided.

Table 10-4 Criteria for Potentially Clinically Significant Electrocardiograms

Parameter	***Unit***	***Higher Limit***
QRS interval	msec	≥ 150
PR interval	msec	≥ 250
QTc (QTcB or QTcF) interval	msec	>500
QTc (QTcB or QTcF) interval	msec	Increase of > 60

QTc = QT interval corrected for heart rate.

10.5 OTHER SAFETY PARAMETERS

Other safety parameters include YMRS, EPS scales (AIMS, BARS, and SAS scores) and C-SSRS.

10.5.1 YMRS Scale

The number and percentage of participants with treatment-emergent mania will be presented by treatment group. Treatment-emergent mania will be defined as an YMRS total score of 16 or greater at any postbaseline visit. Descriptive statistics for YMRS total score and change from baseline values at each assessment time point will be presented by treatment group.

10.5.2 EPS Scales

A participant will be considered to have treatment-emergent parkinsonism if the participant's SAS score was ≤ 3 at baseline and > 3 at any DB assessment. A participant will be considered to have treatment-emergent akathisia if the participant's BARS score was ≤ 2 at baseline and > 2 at any DB assessment. The number and percentage of participants reporting treatment-emergent parkinsonism or treatment-emergent akathisia will be tabulated by treatment group. Listings of participants with treatment-emergent parkinsonism and participants with treatment-emergent akathisia will be provided and will include the participant identification number, study center number, and baseline and postbaseline values. Listings of all AEs occurring in participants who have treatment-emergent parkinsonism or treatment-emergent akathisia will also be provided.

Descriptive statistics for EPS scale parameters (AIMS, BARS, and SAS) and changes from baseline values at each assessment timepoint in this study will be presented.

10.5.3 Suicidality Assessment

The number and percentage of patients with suicidal ideation or suicidal behavior as recorded on the C-SSRS will be summarized by treatment group. The distribution of responses for most severe suicidal ideation and most severe suicidal behavior during the patient's lifetime, during the DB treatment period, and during the safety follow-up period will also be presented by treatment group for the safety population. Supportive listings will be provided and will include the patient identification number, study center number, lifetime history, and postbaseline values. Intensity of suicidal ideation, suicidal behavior type, and lethality of suicidal behavior will also be included in these listings. A listing of all AEs occurring in patients who have suicidal ideation or suicidal behavior will also be provided.

11.0 HEALTH OUTCOMES ANALYSES

The SF-12 will be administered at Visit 2 and Visit 6/ET.

Analysis of change from baseline in the PCS-12 and MCS-12 score at Week 6 will be analyzed using the ANCOVA (based on observed cases) that has treatment group, pooled country, ADT failure category (per CRF) as factors and the baseline value as the covariate.
The detailed algorithms for derivation of the PCS-12 and MCS-12 score are presented in Section 15.2.1.

12.0 INTERIM ANALYSIS

No interim analysis is planned for this study.

13.0 DETERMINATION OF SAMPLE SIZE

The study will randomize approximately 750 participants in a 1:1:1 ratio to cariprazine 1.5mg/day, cariprazine 3mg/day, and placebo groups. A sample size of 250 participants per arm will provide approximately 83% statistical power to show statistically significantly higher effect in each dose of cariprazine versus placebo based on the mITT analysis set. The study has approximately 90% statistical power to show that at least 1 of the 2 cariprazine doses is statistically significantly more efficacious than placebo in the primary endpoint. These calculations assumed an effect size of 0.286. All statistical powers presented in this section were calculated adjusting for multiple comparisons across the two cariprazine doses and the primary and secondary endpoints using the Hochberg-based gatekeeping procedure with re-testing with the family-wise type I error rate being controlled at a 0.05 level (2-sided). The dropout rate is assumed to be 15% at Week 6. Within-person correlation for the primary endpoint is assumed to be 0.58. This value is used in the sample size calculation to calculate an inflation factor that accounts for information loss due to the missing data at Week 6 for longitudinal data collection.

Assumptions of effect size, intracorrelation and dropout rate are based on cariprazine Study RGH-MD-75.

14.0 STATISTICAL SOFTWARE

Statistical analyses will be performed using version 9.4 (or newer) of SAS on a Linux operating system.

15.0 DATA HANDLING CONVENTIONS

15.1 VISIT TIME WINDOWS

Table 15-1 presents the visits assigned for efficacy and safety analyses and the corresponding range of treatment days (window) during which an actual visit may occur.

Table 15-1 Efficacy and Safety Analysis Visit Definitions

Analysis Phase	Derived Visit	Scheduled Visit Day [a]	Window
Pretreatment	Baseline	Day 1	Days ≤ 1
Double-blind Treatment	Week 1	Day 8	Days [2, 11]
	Week 2	Day 15	Days [12, 21]
	Week 4	Day 29	Days [22, 35]
	Week 6	Day 43	Days ≥ 36 and within Double-blind Treatment Period
	End of Double-Blind Treatment Period [b]	Final double-blind treatment visit (expected Day 43) or ET visit during the double-blind treatment period	Final non-missing assessment during the double-blind treatment Period
Safety Follow-up	Week 10	Day 71	day of final double-blind or ET visit during the double-blind treatment period +1 ≤ Days ≤ last visit

a Relative to the date of the first dose of double-blind investigational product. Day 1 = the date of the first dose of double-blind investigational product.

b Presented in analysis tables for safety parameters, including but not limited to electrocardiograms, clinical laboratory values, and vital signs.

ET = early termination.

If the visit date is on or after the date of the first dose of double-blind investigational product, the study day is calculated by visit date – date of the first dose of double-blind investigational product + 1. If the visit date is before the date of the first dose of double-blind investigational product, the study day is calculated by visit date – date of the first dose of double-blind investigational product. Therefore, a negative day indicates a day before the start of the double-blind investigational product.

If a participant has 2 or more visits within the same window, the last visit with a non-missing value will be used for analysis.

15.2 DERIVED VARIABLES

Some efficacy, safety variables and SF-12 health outcome measure are derived as follows:

- MADRS total score is the sum of the 10 items from the MADRS. If more than 2 items of MADRS are missing, then the total score will be set to missing
- HAMD-17 total score is the sum of the first 17 items from the 24-item Hamilton Depression Rating (HAMD) scale. If more than 2 items of the first 17 items of HAMD scale are missing, then the score will be set to missing
- HAM-A total score is the sum of all the 14 items from the HAM-A. If more than 2 items of the HAM-A are missing, then the total score will be set to missing
- YMRS total score is the sum of the 11 items of the YMRS. If more than 1 item of the YMRS are missing, then the total score will be set to missing.
- AIMS total score is the sum of the first 7 items of the AIMS. If more than 1 item for the AIMS (items 1-7) is missing, then the total score will be set to missing. Items 8 through 12 will be summarized separately
- SAS total score is the sum of the 10 items of the SAS. If more than 1 item for the SAS is missing, then the total score will be set to missing
- BARS total score is the sum of the first 3 items of the BARS. If any item for the BARS (items 1-3) is missing, then the total score will be set to missing. Item 4 will be summarized separately
- MADRS responders are defined as participants with a $\geq$ 50% reduction from baseline in MADRS total score. The value is 1 for MADRS responders and 0 otherwise
- MADRS remission is defined as a MADRS total score of $\leq$ 10. The value is 1 for MADRS remission and 0 otherwise
- CGI-I responders are defined as participants with a CGI-I score $\leq$ 2. The value is 1 for CGI-I responders and 0 otherwise
- PCS-12 and MCS-12 derivation are described in the Section 15.2.1.

The total score for a specific variable at a particular visit will be calculated using (sum of non-missing items) × (total number of items) / (number of non-missing items) only if the number of missing items is less than the specified number for this safety variable. Otherwise, the total score will be set to missing. For the LOCF analysis, if a participant misses a postbaseline visit or if the postbaseline visit occurs outside the visit time window, a record for the scheduled visit will be imputed using the last non-missing value immediately before the missing value. If the missing value occurs at Week 1, the baseline value will be carried forward for Week 1, provided that at least 1 subsequent postbaseline assessment is available. For a composite scale such as MADRS total score, individual items of a rating scale will not be carried forward. Only total scores will be carried forward using the last-observation-carried-forward approach.

15.2.1 Scoring Algorithm of SF-12V2

SF-12v2 (4-week recall) consists of 12 items. Two composite summary scores, including Physical Component Summary (PCS) and Mental Component Summary (MCS), and 8 scale scores, including Physical Functioning (PF), Role Physical (RP), Bodily Pain (BP), General Health (GH), Vitality (VT), Social Functioning (SF), Role Emotional (RE) and Mental Health (MH), will be derived using the following scoring algorithm. Within each scale, if there are less than or equal to 50% of the item scores missing, the missing item score(s) will be imputed using the mean score of the non-missing items and used to calculate the scale score. If more than 50% of the item scores are missing, no imputation will be performed and the scale score will be set to missing.

Step 1: Item coding

Code the response to each item as follows such that higher score indicates better health state.

Response to Item 1

Response Choices	Precoded Item Value	Final Item Value
Excellent	1	5.0
Very Good	2	4.4
Good	3	3.4
Fair	4	2.0
Poor	5	1.0

Response to Items 2a and 2b

Respond Choices	Precoded Item Value	Final Item Value
Yes, limited a lot	1	1
Yes, limited a little	2	2
No, not limited at all	3	3

Response to Items 3a, 3b, 4a, 4b and 7

Response Choices	Precoded Item Value	Final Item Value
All of the time	1	1
Most of the time	2	2
Some of the time	3	3
A little of the time	4	4
None of the time	5	5

Response to Item 5

Response Choices	Precoded Item Value	Final Item Value
Not at all	1	5
A little bit	2	4
Moderately	3	3
Quite a bit	4	2
Extremely	5	1

Response to Items 6a, 6b and 6c

Response Choices	Precoded Item Value	Final Item Value
All of the time	1	5
Most of the time	2	4
Some of the time	3	3
A little of the time	4	2
None of the time	5	1

Step 2: Computing raw scale scores

After item recoding, a raw score is computed for each scale which is the simple algebraic sum of the final item values for all items in that scale as shown in the below table. For example, the raw score for PF scale is the sum of the final item values of items 2a and 2b.

Scale	Sum Final Item Values	Lowest and Highest Possible Raw Scores	Possible Raw Score Range
PF	Items 2a + 2b	2, 6	4
RP	Items 3a + 3b	2, 10	8
BP	Item 5	1, 5	4
GH	Item 1	1, 5	4
VT	Item 6b	1, 5	4
SF	Item 7	1, 5	4
RE	Items 4a + 4b	2, 10	8
MH	Items 6a + 6c	2,10	8

Step 3: Transforming raw scale scores

Transforming the raw scale score to a 0-100 scale using the formula below. The table above include the lowest possible and possible range of raw score for each scale.

$$Transformed\ scale = \frac{Actual\ raw\ score - Lowest\ possible\ raw\ score}{Possible\ raw\ score\ range} \times 100$$

For example, the transformed score for PF scale is (Raw PF score – 2) / 4.

Step 4: Standardizing transformed scale scores

Each transformed scale is standardized to a z-score using scale means and standard deviations from the 1998 general United States population as shown in the below table.

Scale	Mean	Standard Deviation
PF	81.18122	29.10558
RP	80.52856	27.13526
BP	81.74015	24.53029
GH	72.19795	23.19041
VT	55.59090	24.84380
SF	83.73973	24.75775
RE	86.41051	22.35543
MH	70.18217	20.50597

The formula for z-score transformation is:

Z-score standardized scale = (Transformed scale score – Mean) / Standard Deviation

For example, the z-score of PF scale = (Transform PF score – 81.18122) / 29.10558.

Step 5: Aggregating scale scores to compute raw component summary scores

The standardized scales are aggregated using weights (factor score coefficients) from the 1998 general United States population to derive raw component summary scores (PCS and MCS). Formulae for aggregating z-score standardized scales are:

$$Raw\ PCS = PF \times .42402 + RP \times .35119 + BP \times .31754 + GH \times .24954 +$$

$$VT \times .02877 + SF \times -.00753 + RE \times -.19206 + MH \times .22069$$

$$Raw\ MCS = PF \times -.22999 + RP \times -.12329 + BP \times -.09731 + GH \times -.01571 +$$

$$VT \times .23534 + SF \times .26876 + RE \times .43407 + MH \times .48581$$

Step 6: Standardizing composite summary scores

The raw PCS and MCS scores are standardized as follows such that the final PCS and MCS scores have values with mean 50 and standard deviation 10.

$$PCS = 50 + Raw\ PCS \times 10$$

$$MCS = 50 + Raw\ MCS \times 10$$

15.3 REPEATED OR UNSCHEDULED ASSESSMENTS OF SAFETY PARAMETERS

If a participant has repeated assessments before the start of the first treatment, the results from the final non-missing assessment made prior to the start of the DB investigational product will be used as baseline. If end-of-study assessments are repeated or if unscheduled visits occur, the last non-missing postbaseline assessment will be used as the end-of-study assessment for generating summary statistics. However, all postbaseline assessments will be used for PCS value determinations, and all assessments will be presented in the data listings.

15.4 MISSING DATE OF THE LAST DOSE OF STUDY TREATMENT

When the date of the last dose of study treatment is missing for a participant in the Safety Population, all efforts should be made to obtain the date from the Investigator. If after all efforts are made it is still missing, the last available dosing record date will be used as the last dose date.

15.5 MISSING SEVERITY ASSESSMENT FOR ADVERSE EVENTS

If severity is missing for an AE that started before the date of the first dose of study treatment, an intensity of mild will be assigned. If severity is missing for an AE that started on or after the date of the first dose of study treatment, an intensity of severe will be assigned. The imputed values for severity assessment will be used for the incidence summary; the values will be shown as missing in the data listings.

15.6 MISSING CAUSAL RELATIONSHIP TO STUDY DRUG FOR ADVERSE EVENTS

If the causal relationship to the study treatment is missing for an AE that started on or after the date of the first dose of study treatment, a causality of yes will be assigned. The imputed values for causal relationship to study treatment will be used for the incidence summary; the values will be shown as missing in the data listings.

15.7 MISSING DATE INFORMATION FOR ADVERSE EVENTS

The following imputation rules only apply to cases in which the start date for AEs is incomplete (ie, partly missing).

Missing month and day

- If the year of the incomplete start date is the same as the year of the first dose of study treatment, the month and day of the first dose of study treatment will be assigned to the missing fields

- If the year of the incomplete start date is before the year of the first dose of study treatment, *December 31* will be assigned to the missing fields

- If the year of the incomplete start date is after the year of the first dose of study treatment, *January 1* will be assigned to the missing fields

Missing month only

- If only the month is missing, the day will be treated as missing and both the month and the day will be replaced according to the above procedure

Missing day only

- If the month and year of the incomplete start date are the same as the month and year of the first dose of study treatment, the day of the first dose of study treatment will be assigned to the missing day

- If either the year of the incomplete start date is before the year of the date of the first dose of study treatment or if both years are the same but the month of the incomplete start date is before the month of the date of the first dose of study treatment, the last day of the month will be assigned to the missing day

- If either the year of the incomplete start date is after the year of the date of the first dose of study treatment or if both years are the same but the month of the incomplete start date is after the month of the date of the first dose of study treatment, the first day of the month will be assigned to the missing day

If the stop date is complete and the imputed start date as above is after the stop date, the start date will be imputed by the stop date.

If the start date is completely missing and the stop date is complete, the following algorithm will be used to impute the start date:

- If the stop date is after the date of the first dose of study treatment, the date of the first dose of study treatment will be assigned to the missing start date

- If the stop date is before the date of the first dose of study treatment, the stop date will be assigned to the missing start date

15.8 MISSING DATE INFORMATION FOR PRIOR OR CONCOMITANT MEDICATIONS

For prior or concomitant medications, including rescue medications, incomplete (ie, partly missing) start dates and/or stop dates will be imputed. When the start date and the stop date are both incomplete for a participant, the start date will be imputed first.

15.8.1 Incomplete Start Date

The following rules will be applied to impute the missing numeric fields for an incomplete prior or concomitant medication start date. If the stop date is complete (or imputed) and the imputed start date is after the stop date, the start date will be imputed using the stop date.

Missing month and day

- If the year of the incomplete start date is the same as the year of the first dose of study treatment, the month and day of the first dose of study treatment will be assigned to the missing fields

- If the year of the incomplete start date is before the year of the first dose of study treatment, *December 31* will be assigned to the missing fields

- If the year of the incomplete start date is after the year of the first dose of study treatment, *January 1* will be assigned to the missing fields

Missing month only

- If only the month is missing, the day will be treated as missing and both the month and the day will be replaced according to the above procedure

Missing day only

- If the month and year of the incomplete start date are the same as the month and year of the first dose of study treatment, the day of the first dose of study treatment will be assigned to the missing day

- If either the year of the incomplete start date is before the year of the date of the first dose of study treatment or if both years are the same but the month of the incomplete start date is before the month of the date of the first dose of study treatment, the last day of the month will be assigned to the missing day.

- If either the year of the incomplete start date is after the year of the date of the first dose of study treatment or if both years are the same but the month of the incomplete start date is after the month of the date of the first dose of study treatment, the first day of the month will be assigned to the missing day

15.8.2 Incomplete Stop Date

The following rules will be applied to impute the missing numeric fields for an incomplete prior or concomitant medication stop date. If the date of the last dose of study treatment is missing, impute it as descripted in Section 15.4. If the imputed stop date is before the start date (imputed or nonimputed start date), the imputed stop date will be equal to the start date.

Missing month and day

- If the year of the incomplete stop date is the same as the year of the last dose of study treatment, the month and day of the last dose of study treatment will be assigned to the missing fields

- If the year of the incomplete stop date is before the year of the last dose of study treatment, *December 31* will be assigned to the missing fields

- If the year of the incomplete stop date is after the year of the last dose of study treatment, *January 1* will be assigned to the missing fields

Missing month only

- If only the month is missing, the day will be treated as missing and both the month and the day will be replaced according to the above procedure

Missing day only

- If the month and year of the incomplete stop date are the same as the month and year of the last dose of study treatment, the day of the last dose of study treatment will be assigned to the missing day

- If either the year of the incomplete stop date is before the year of the date of the last dose of study treatment or if both years are the same but the month of the incomplete stop date is before the month of the date of the last dose of study treatment, the last day of the month will be assigned to the missing day

- If either the year of the incomplete stop date is after the year of the date of the last dose of study treatment or if both years are the same but the month of the incomplete stop date is after the month of the date of the last dose of study treatment, the first day of the month will be assigned to the missing day

15.9 CHARACTER VALUES OF CLINICAL LABORATORY PARAMETERS

If the reported value of a clinical laboratory parameter cannot be used in a statistical summary table because, for example, a character string is reported for a parameter of the numeric type, a coded value must be appropriately determined for use in the statistical analyses. The actual values, however, as reported in the database will be presented in the data listings.

16.0 COVID-19 RELATED ANALYSES

To eliminate immediate potential hazards to study participants and site staff due to the COVID-19 pandemic while ensuring participant safety and maintaining data integrity, the Change in Research Memo, followed by COVID-19 global protocol addendum, and protocol amendment #3 were sent to sites during the pandemic.

The following protocol modifications (see protocol amendment #3, Section 12.2) can be followed for active patients who have completed all assessments through Visit 2 (inclusive) per Table 3-2, Schedule of Visits and Procedures. The following procedures apply in cases where patients are either unable or unwilling to attend study visits as a result of the pandemic.

- Allowing wider visit windows when necessary
- Replacing protocol mandated in-person study visits with one or more of the following:
  - home visits
  - telemedicine virtual visits (preferred method if/when home visits are not feasible)
  - telephone/video calls (no recording will be performed)
- Allowing blood draws at alternative or commercial laboratories (where available)
- Study sites shipping investigational products (IP) to research patients (where permissible by local/statutory or country law) and where approved by Sponsor
- Extending the window for consecutive missed doses of IP specifically resulting from COVID-19-related circumstances from 4 or more days to more than 7 days, prior to requiring withdrawal of the patient from the study (see protocol amendment #3, Section 12.2.6.)

This section specifies analyses for evaluating the impact of COVID-19.

16.1 EFFICACY EVALUATION

Efficacy Endpoints

For efficacy assessments that are conducted remotely (eg, by telephone, telemedicine virtual visit, video call) or in-home by qualified site staff, the following assessments are required to be completed in the course of the remote visit. Note: The source documents and eCRF should clearly denote which assessments have been completed remotely.

- Montgomery-Åsberg Depression Rating Scale (MADRS)
- Hamilton Anxiety Rating Scale (HAM-A)
- Clinical Global Impression-Severity scale (CGI-S)
- Clinical Global Impression-Improvement scale (CGI-I)
- Hamilton Depression Rating Scale–17 Items (HAMD-17)

To evaluate the impact of COVID-19, the rate of missing due to COVID-19 and assessment done by alternative modalities will be summarized by study visit for each assessment.

For the primary endpoint and key secondary endpoint, one additional sensitivity analysis may be conducted to investigate the impact of COVID-19 dropout cases on the estimation results by treating the dropout cases due to COVID-19 as MAR.

- Participants who discontinued due to lack of efficacy (**but not due to COVID-19**) in the cariprazine arms are assumed to have no treatment effect after the discontinuation. These participants are assumed to copy the profile of placebo arm and missing values are imputed based on the distribution estimated from the placebo group under the MAR using copy-reference approach (Carpenter et al, 2013).
- The rest of missing values in the placebo arm and cariprazine arms (**including dropouts due to COVID-19**) will be imputed using the observed data in their respective group under the MAR assumption.

In addition, sensitivity analyses for the primary and key secondary endpoints by pre-COVID and post-COVID randomized participants will be performed:

- pre-COVID randomized participants: date of randomization ≤ March 11, 2020
- post-COVID randomized participants: date of randomization > March 11, 2020

Lastly, to explore the COVID impact further, the primary analysis for primary and secondary endpoints will be repeated on the subset of participants who completed or discontinued from the study prior to the COVID cutoff date of March 11, 2020.

16.2 SAFETY AND OTHER EVALUATIONS

This section specifies analyses related to COVID-19 pandemic from the following aspects:

- COVID-19 Status
- Disposition
- Demographics and Other Baseline Characteristics
- Study visit (missing entire visit due to COVID-19 or missing assessments due to COVID-19)
- Protocol deviation
- Study drug interruption(s) due to COVID-19
- TEAEs related to COVID-19 infection

Safety Population will be used for the planned analyses in this section described above, unless stated otherwise.

The number of participants impacted by COVID-19 during the study and COVID-19 status will be summarized by treatment group. In addition, the number of participants impacted by COVID-19 and their corresponding disposition status in the double-blind treatment period and the follow-up period will be summarized respectively.

Baseline characteristics (including demographic, baseline disease characteristics, and relevant medical history) will be summarized relative to the onset of pandemic (date of randomization <= Mar 11, 2020 vs. date of randomization > Mar 11, 2020).

The number of participants who missed at least one entire visit due to COVID-19 will be summarized by treatment group. Furthermore, the number of participants who missed at least one assessment due to COVID-19 will be summarized by assessment category (laboratory, vital signs, ECG, C-SSRS, YMRS, EPS scales) and by visit.

The number of participants with significant protocol deviation due to COVID-19 will be provided. The number of participants with any COVID-19 related treatment interruption(s) will be provided as well.

The number of participants with TEAEs related to COVID-19 infection will be provided.

Supporting listings for the above described analyses will be provided. Furthermore, a listing of TEAEs leading to study discontinuation related to COVID-19 will be provided.

17.0 CHANGES TO ANALYSES SPECIFIED IN PROTOCOL

In Section 11.0 (Health Outcome Analyses), the term "Study center" in the ANCOVA model for analyzing PCS-12 and MCS-12 will be replaced with terms country and ADT failure category to align with the analyses for other efficacy endpoints.

In Section 10.1 (Adverse Events), AE summary timeframe was modified to align with AE reporting period specified in the protocol amendment 1. The new time frame is defined below:

> An AE that occurs more than 30 days after the date of the last dose of DB IP or occurs after the last safety follow-up visit (whichever comes last) will not be summarized except for AEs related to reported pregnancies. AEs associated with pregnancies reported in the protocol-required timeframe up to approximately 12 weeks following the last dose of DB IP, will be summarized.

The summary of SAEs was modified in Section 10.1. The definition of on-therapy SAE was updated to TESAE, and summary timeframe will align with general AE summary timeframe as specified above.

18.0 APPENDICES

Appendix I Pattern-Mixture Model Details

For repeated measures with the monotone missing mechanism, the pattern-mixture model with non-future dependent missing assumption, proposed by Kenward et al. (2003), provides a feasible solution to accommodate certain missing not at random (MNAR) mechanism. The methodology relies on constructing unidentifiable conditional densities using identifiable densities and borrows techniques from standard multiple imputation.

1. Non-Future Dependent Missing Assumption

Assume there are T designed visits in a longitudinal study and let $y_i (i = 1,2,...,T)$ represent participant's measurement at Visit i. When the missing mechanism is monotone, the pattern of missing data can be defined by the number of measurements (L) actually observed from the participant. Let $f(y_i,...,y_j \mid L = t)$ denote the conditional density of
y_i ,... y_j, given that the last observed measurement is at Visit t. Then the overall density function for Pattern t can be written as

$$f(y_1,..., y_T \mid L = t) = f(y_1,..., y_t \mid L = t) f(y_{t+1} \mid y_1,..., y_t, L = t) \times \prod_{s=t+2}^{T} f(y_s \mid y_1,..., y_{s-1}, L = t) \quad (1)$$

Note on the right hand side of (1) the first factor is clearly identifiable from the observed data, while the second and the beyond are not, due to lack of available data. The second factor $f(y_{t+1} \mid y_1,...,y_t, L = t)$ could be identifiable based on an assumed relationship between $f(y_{t+1} \mid y_1,...,y_t, L = t)$ and $f(y_{t+1} \mid y_1,...y_t, L \geq t + 1)$. The third and beyond factors $f(y_s \mid y_1,...,y_{s-1}, L = t)$ (with all $s \geq t + 2$) could be identifiable with the help of non-future dependent missing assumption.

For longitudinal data with dropouts, non-future dependent missing (NFD) mechanism (Kenward et al., 2003) assumes that the unidentifiable conditional distributions of y_s ($s \geq t + 2$), given earlier measurements, in Pattern t, is equal to the corresponding distribution in patterns $L \geq s - 1$:

$$f(y_s \mid y_1,..., y_{s-1}, L = t) = f(y_s \mid y_1,..., y_{s-1}, L \geq s - 1) \quad (2)$$

The right hand side of (2) can further be partitioned into

$$f(y_s \mid y_1,..., y_{s-1}, L \geq s-1) = \sum_{j=s-1}^{T} \omega_{s-1,j} \cdot f(y_s \mid y_1,..., y_{s-1}, L = j) \qquad (3)$$

Where mixture probabilities $\omega_{s-1,j}$ are:

$$\omega_{s-1,j} = \frac{\alpha_j f(y_1,...,y_{s-1} \mid L = j)}{\sum_{t=s-1}^{T} \alpha_t f(y_1,...,y_{s-1} \mid L = t)}, \text{ and } \alpha_j \text{ represents the fraction of} \qquad (4)$$

participants from Pattern j.

Each factor of the unidentifiable conditional distribution of y_s ($s \geq t+2$) on the right side of (1) can be expressed using the following:

- $f(y_s \mid y_1,..., y_{s-1}, L = s-1)$, the unidentifiable conditional distribution of the first missing in pattern $s-1$,

- $f(y_s \mid y_1,..., y_{s-1}, L = j)$, the identifiable conditional distributions of y_s given $y_1,..., y_{s-1}$ of pattern j ($j \geq s$), and

- α_j, the fraction of participants from pattern j ($j \geq s - 1$).

So under NFD, all the unidentifiable conditional distribution on the right side of (1) can be estimated and missing value could be therefore imputed based on the assumption for unidentifiable conditional distribution of the first missing.

We re-formulate the partition in (3), for $s \geq t+2$, as the following:

$$\begin{aligned} f(y_s \mid y_1,..., y_{s-1}, L = t) = {} & \delta_{s-1} f(y_s \mid y_1,..., y_{s-1}, L = s-1) \\ & + (1-\delta_{s-1}) f(y_s \mid y_1,..., y_{s-1}, L \geq s) \end{aligned} \qquad (5)$$

for $s \geq t+2$ with $\delta_{s-1} = \omega_{s-1,s-1}$.

Therefore, under monotone missing and NFD assumption, the unidentifiable conditional densities for Visit s in Pattern t ($s \geq t+2$) can be expressed as a mixture distribution of $f(y_s \mid y_1,...,y_{s-1}, L = s-1)$ - the unidentifiable conditional distribution of the first missing measurement y_s in Pattern $s-1$, and $f(y_s \mid y_1,...,y_{s-1}, L \geq s)$ - the identifiable conditional distribution of y_s from all the patterns with observed data at Visit s or beyond:

$$f(y_s \mid y_1,..., y_{s-1}, L \geq s) = \sum_{j=s}^{T} \lambda_{s-1,j} f(y_s \mid y_1,..., y_{s-1}, L = j) \qquad (6)$$

where the mixture probability $\lambda_{s-1} = \omega_{s-1,j} / (1 - \omega_{s-1,\, s-1}) =$

$$\frac{\alpha_j f(y_1,...,y_{s-1} \mid L = j)}{\sum_{t=s}^{T} \alpha_t f(y_1,...,y_{s-1} \mid L = t)} \text{ for } j \geq s, \text{ where } \alpha_j \text{ is the fraction of} \qquad (7)$$

participants from Pattern j.

The conditional densities for the first missing are selected as:

$$f(y_s \mid y_1,..., y_{s-1}, L = s-1) = f(y_s - \Delta \mid y_1,..., y_{s-1}, L \geq s) \text{ for } s = 2, ..., T \qquad (8)$$

Note that the two distributions only differ by a shift (Δ) parameter. When $\Delta = 0$, the missing value y_s in Pattern s-1 is imputed based on the distribution of all observed data up to Visit s, as a result, leading to missing at random (MAR) missingness. When $\Delta \neq 0$, (8) will introduce a scenario of MNAR. The similar idea was also presented in the recent publication "The Prevention and Treatment of Missing Data in Clinical Trials" by the National Academies Press. The selection of the plausible values for the shift parameter (Δ) is discussed in Section 3.0.

Note that per recommendation in Wang and Daniels (2011), only the observed data within pattern is assumed to be multivariate normal. The observed data distribution can be expressed in terms of the marginal distribution of baseline measurement and the conditional distributions of postbaseline measurements given earlier measurements. Assuming that these distributions are normal, the linear regression of each observation on prior observations will yield least-squares estimates of model parameters that can be utilized for independent posterior draws of model parameters for observed data. Multiple imputation approach will be used to estimate the overall mean at the final time point.

2. Imputation Procedure

All the missing data will be imputed to create complete datasets, then statistical analysis can be performed using appropriate techniques such as ANCOVA. The imputation can accommodate MNAR missing data mechanisms, based on the theory discussed in the previous section.

The model parameters for each dropout pattern, ie, the mean, variance and proportions of observations in each pattern, are drawn from their posterior distributions prior to the imputation of missing data for a single imputation.

The details of imputation within a pattern, say Pattern t, are as the following:

Step 1. Impute the first missing value y_{t+1} for each participant in Pattern t (t = 1,..., T – 1):

a. Compute estimates of mixture probabilities $\lambda_{s-1,j}$ in (7) with s = t+1 given the posterior draw of proportions of observations in each pattern and the posterior draw of regression parameters for the observed data.

b. Draw a random integer from {s,..., T} to index a component distribution on the right hand side of (6), using mixture probabilities obtained in a). Draw y^*_{t+1} from the identified component normal distribution. Impute the missing y_{t+1} as $\tilde{y}_{t+1} = y^*_{t+1} + \Delta$.

Step 2. Impute the rest of the missing values of $y_{t+2}, y_{t+3},..., y_T$ for participants in Pattern t:

Starting with imputation for y_{t+2}, first, similar to Step 1, draw y^*_{t+2} from the normal mixture (6) based on the observed $y_1,...,y_t$ and the already imputed $\tilde{y}_{t+1}$ for the participant. Then the missing y_{t+2} is imputed as $\tilde{y}_{t+2} = y^*_{t+2} + \Delta$ with probability δ_{t+1} and as $\tilde{y}_{t+2} = y^*_{t+2}$ with probability $1 – \delta_{t+1}$, where the mixture probability $\delta_{t+1} = \omega_{t+1,t+1}$ is obtained from (4) given the posterior draw of proportions of observations in each pattern and the posterior draw of regression parameters for the observed data.

Missing values of y_{t+3} through y_T can be imputed similarly as y_{t+2}.

To summarize, the imputations of y_{t+1} through y_T is done recursively within each Pattern t (for all t = 1,...,T – 1) to create a complete dataset after imputation is done for all patterns with missing values.

The above imputation procedure is applied to all participants in each missing data pattern to create a single imputed data set. Repeating the process of drawing parameters from the posterior distribution and imputing missing data given the posterior draw m times will yield m imputed data sets. The observed or imputed values at the final data point are averaged to obtain the overall mean estimate for each imputed data set, and the multiple imputation estimate is obtained by averaging the estimates across m imputations.

In this sensitivity analysis, m is set to equal to 200.

3. Determination of the Shift Parameter Values

The common shift parameter Δ is the difference between the mean of y_{t+1} among those who drop out at Visit t and those who remain beyond Visit t ($1 \leq t \leq T - 1$). The exact value of Δ is unknown and can't be estimated from data because of missingness. The magnitude of Δ depends on the medical aspects of the trial. Using relevant historical data, one may select Δ as a proportion of the sample standard deviation or a proportion of observed treatment efficacy.

Appendix II: Reporting Selected Laboratory Parameters in Conventional Units

All clinical laboratory parameters will be reported in the International System (SI) units as standard practice. In addition, descriptive statistics for values and changes from baseline in conventional units at all assessed visits will be reported for selected laboratory parameters as listed in Table 18-1 below.

Table 18-1 List of Selected Parameters to be Reported in Conventional Units

Number	Laboratory Parameter	Conventional Unit	Decimal Places
1	Alanine Aminotransferase (SGPT)	U/L	0
2	Albumin	G/dL	1
3	Alkaline Phosphatase	U/L	0
4	Aspartate Aminotransferase (SGOT)	U/L	0
5	Bilirubin, Direct (Conjugated)	mg/dL	1
6	Bilirubin, Indirect (Unconjugated)	mg/dL	1
7	Bilirubin, Total	mg/dL	1
8	Blood Urea Nitrogen	mg/dL	0
9	Calcium	mg/dL	1
10	Cholesterol, HDL	mg/dL	0
11	Cholesterol, LDL	mg/dL	0
12	Cholesterol, LDL direct and calculated (combined) (This lab parameter could be the same as #11)	mg/dL	0
13	Cholesterol, Total	mg/dL	0
14	Creatine Kinase	U/L	0
15	Creatinine	mg/dL	1
16	Glucose	mg/dL	0
17	Insulin	uIU/mL	1
18	Triglycerides	mg/dL	0
19	Uric Acid	mg/dL	1
20	Hemoglobin	G/dL	1

Participant narratives will also include the values in conventional units for the selected lab parameters. That might be accomplished by presenting the values in conventional units within the parentheses next to the values in SI units.

Appendix III: SUMMARY OF MAJOR CHANGES FOR SAP AMENDMENT 2

Amendment #2 specifies the following changes to the SAP Amendment #1 for Study 3111-302-001 dated 26 Mar 2021. The major changes are summarized below:

1. Sections 9.2 and 9.3: updated the key secondary efficacy endpoint (from HAM-A to CGI-S at Week 6).
2. Section 16.1: added additional analysis related to COVID-19 impact on primary efficacy endpoint for completed or discontinued patients prior to the COVID-19 pandemic onset, per the FDA feedback.
3. Section 11.0: replaced MMRM by ANCOVA, for consistency with the protocol.
4. Section 7.0: added a clarification that ADT failure based on CRF is primary and will be used for all statistical models, rather than per IWRS.
5. Section 9.3: replaced MMRM by ANCOVA for change from baseline in HAMD-17 total score, per the HAMD-17 data collection at baseline and Week 6.
6. Section 13.0: updated the power that was re-calculated using the mixture parallel gatekeeping procedure.
7. Section 10.2: added a footnote to Tables 10-1 and 10-2 for the parameters with fasting status.
8. Section 10.1: added a paragraph for EPS TEAEs excluding akathisia/restlessness, with any Akathisia/Restlessness, including a list of preferred terms.
9. Sections 7.0 and 10.1: updated the Medical Dictionary for Regulatory Activities version to 24.0.
10. Section 9.1: updated subgroup analyses model, to align with ISE.
11. Section 5.5: added additional selected analyses excluding one non-compliant site.

Amendment #1 specifies the following changes to the final SAP for Study 3111-302-001 dated 16 Jan 2019. The major changes are summarized below:

1. Added “The primary analysis will be performed after all participants have completed the 4-week safety follow-up or discontinued from the study and the database has been locked and unblinded.” in Section 3.0.
2. Added the key secondary efficacy endpoint in Section 9.2 (HAM-A at Week 6).
3. Added additional efficacy endpoints in Section 9.3: MARDS individual items and HAM-A individual items, HAM-A response and HAM-A remission.
4. Added SF-12 analysis in Section 11.0.
5. Added Section 16.0 to specify analyses related to COVID-19 to follow protocol amendment #3.

19.0 REFERENCES

Carpenter JR, Roger JH, Kenward MG. Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. J Biopharm Stat. 2013;23:1352-1371.

Dmitrienko A, Millen B, Brechenmacher T, Paux G. Development of gatekeeping strategies in confirmatory clinical trials. Biometrical Journal 2011; 53:875-893.

Durgam S, Earley W, Guo H, Li D, Németh G, Laszlovszky I, Fava M, Montgomery SA. Efficacy and safety of adjunctive cariprazine in inadequate responders to antidepressants: A randomized, double-blind, placebo-controlled study in adult MDD patients. J Clin Psychiatry, 2016; 77: 371-378.

El-Khalili N, Joyce M et al. Extended-release quetiapine fumarate (quetiapine XR) as adjunctive therapy in major depressive disorder (MDD) in patients with an inadequate response to ongoing antidepressant treatment: a multicentre, randomized, double-blind, placebo-controlled study. International Journal of Neuropsychopharmacology, 2010; 13: 917-932.

Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. Biometrika. 1988;75:800-802.

Kenward MG and Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics.1997;53:983–997. Kenward MG, Molenberghs G, Thijs H. Pattern-mixture models with proper time dependence. Biometrika.2003;90(1):53-71.

Kenward MG, Molenberghs G, Thijs H. Pattern-mixture models with proper time dependence. Biometrika. 2003;90(1):53-71.

National Institutes of Health, National Heart, Lung, and Blood Institute, National High Blood Pressure Education Program. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. NIH Publication No. 04-5230, 2014

Wang C, Daniels MJ. A note on MAR, identifying restrictions, model comparison, and sensitivity analysis in pattern mixture models with and without covariates for incomplete data. *Biometrics,* 2011; 67(3):810-818.