Data Analysis Plan for Northwestern- Janssen R&D Collaboration Study

Protocol Title: Testing the ability of JNJ-18038683, a selective serotonin (5-HT)7 antagonist, to improve cognition and reduce residual depressive symptoms in stable bipolar patients

Protocol Identifier Number: 18038683BCD2001

Northwestern University IRB Number: STU00200919-MOD0042

Version: 1.13

Date: 08.04.2021

Goals:

The goals of this study are to evaluate the efficacy of JNJ-18038683 in an 8-week trial to ameliorate the cognitive deficit and reduce residual depressive symptoms in 60 stable bipolar outpatients receiving treatment for depression. JNJ-18038683 will be studied and compared with placebo as adjunctive treatment to standard pharmacologic treatment for bipolar disorder. Sixty patients will be enrolled and randomized with a 1:1 ratio to one dose of JNJ-18038683 or placebo providing 30 subjects for each treatment arm. The initial intention was to follow stratified randomization based on the presence/absence of atypical APDs with 5-HT7 antagonism so that there is a balance in the assignment of active treatment and placebo in individuals who are and are not, receiving atypical APDs with 5-HT7 antagonism. However, due to recruitment difficulties, the stratified randomization method was not able to be followed as anticipated and was amended.

Assessments and Events:

Demographics, medical history, physical examination, serum pregnancy test, urine drug test, SCID-I diagnosis, and screening neurocognitive assessments will be done only during the screening visit. Urine pregnancy testing is performed at the baseline visit.

Table 1: Assessments completed during the screening visit only

Event /Assessment	Screen
Medical History & Demographics	X
Physical Examination	X
Serum Pregnancy Test	X
Urine Drug Screen	X
Inclusion/Exclusion Criteria	X
SCID-I	X
Screening neurocognitive assessments	X

Neurocognitive battery will be the primary measure of interest and evaluated at baseline and 8 weeks. The clinical global severity scores (CGI-BP), Young Mania Rating (YMRS), Scale for Tardive Dyskinesia (AIMS), Montgomery Asberg Depression Rating Scale (MADRS), Electrocardiogram and Clinical Safety Laboratory Tests will be collected at screening or baseline, and at 4 and 8 weeks. A measure of functional ability (SLOF) also will be assessed at baseline and 8 weeks. The schedule of these assessments is listed in table 2.

Table 2: Assessments completed during the screen visits/baseline visits and at week 4 and/or week 8

Event/Assessment	Screen	Day 1*	Day 28	Day 56
Electrocardiogram	X		X	X
Clinical Safety Laboratory Tests	X		X	X
Montgomery Asberg Depression Rating Scale		X	X	X
SLOF		X		X
AIMS		X	X	X
Screening neurocognitive assessment	X			
Neurocognitive Assessments Cleveland Battery		X		X
CGI-BP		X	X	X
Young Mania Rating Scale		X	X	X

*Day 1 is equal to Baseline

Vital signs, C-SSRS (suicide tendency), concomitant medication, and adverse events experienced by patients will be collected during study visits at screening, baseline, and at weeks 1, 2,4, 6, and 8. In addition, data will be collected at weeks 3, 5, 7, and 12 via phone calls for C-SSRS (suicide tendency), concomitant medication, and adverse events.

Table 3: Assessments completed during the screen visits/baseline visits and at weeks 1,2,4,6 and 8

Event/Assessment	Screen	Day 1	Day 7	Day 14	Day 28	Day 42	Day 56
Vital Signs	X	X	X	X	X	X	X
C-SSRS*	X	X	X	X	X	X	X
Assessment of Concomitant Medication*	X	X	X	X	X	X	X
Assessment of Adverse Events*	X	X	X	X	X	X	X

* assessments will also be collected at 21, 35, 49 and 84 days via phone calls.

Statistical Analysis:

Screening data and drop out data:

Screen assessments listed in table 1 will be summarized for all the patients including reasons for screen failures. Concomitant medications were the reasons for most screen failures and will be documented. Included patients will be compared with excluded patients based on appropriate statistical methods (parametric or non-parametric) for the variables in table 1. During the study period, all dropouts and reasons for drop out will be tabulated.

In this blinded RCT, individuals with bipolar disorder received active drug (JNJ-18038683) or a placebo after screening and the completion of baseline assessments. Subjects in the two treatment arms will be compared on all the baseline efficacy measures, measures of psychopathology related to bipolar disorder, e.g., YMRS, MADRS, CGI-BP, SLOF measures, demographic variables, and suicide status.

Primary analysis for efficacy:

The primary outcome measure for this study is verbal fluency performance as assessed with the Rey Auditory Verbal Learning Test, Rey Verbal Learning Test – Recognition Form, Verbal Fluency Test "Letterforms", and Animal Naming Test. The 8-week evaluation of verbal fluency after randomization is the primary outcome measure. Since the main interest is to measure the treatment effect for verbal fluency improvement from baseline and there is only one follow-up time point for the primary outcome variable, an ANCOVA model will be best suited for this study to compare the treatment effect using baseline value as a covariate. Alternatively, an ANOVA model will be used to compare the percent improvement and changes in the verbal fluency scores across treatment groups. Also, we plan to compare the treatment effect on verbal fluency performance over time using a mixed model. The missing evaluations will be imputed with the last value carried forward method, while model fit statistics will be used to estimate the appropriate covariance structure for this model.

There is both preclinical and clinical evidence that age has an important influence on 5-HT7 receptor function, which is the target of JNJ-18038683. Capacity to respond to JNJ-18038683 may decline with age. Thus, we predict the age of the subject will be related to the extent of improvement in verbal fluency and other measurements in the cognition battery. The relationship of age and percent improvement and the direct change of verbal fluency from baseline to 8 weeks in each treatment arm will be assessed using correlation analysis. The age variable will be used as a continuous variable in the analysis. If the distribution of the age variable is skewed, then age groups will be formed appropriately. In the above described ANCOVA and repeated measures models, age will be use as a covariate to assess the treatment effect. Graphical methods will be used to check for any inflection points of improvement of verbal fluency score in the treatment group compared to placebo. In the treatment group, better improvement is expected in younger patients compared to older patients. This relationship is not expected in placebo-treated patients. Additionally, with the above-mentioned models, the effect of covariates, MADRS total score, YMRS and Clinical Global Severity score on cognition will be examined because baseline mania or depression severity may influence performance on cognitive testing.

Secondary analyses related to efficacy:

The secondary analyses regarding change in verbal fluency test scores will include identification of patients in each arm who improved in verbal fluency scores. The analysis plan includes exploring the effect of concomitant meds, gender or race that might contribute to treatment response. For this purpose, improvement in verbal fluency can be grouped as a dichotomous response and used as a dependent variable in a logistic regression model. The usual method of grouping will be improvement of ≥0.5 SD from the baseline scores. Using this model, the odds of improvement in verbal fluency can be compared in the active treatment group against the placebo group. In this model, the variables, age, gender, lab values and presence/absence of concomitant meds, baseline MADRS total score, YMRS and Clinical Global Severity score mania or depression severity may influence performance on cognitive testing and will be tested as potential covariates using one or few covariates at a time due to sample size restrictions.

Additional secondary analyses will study the effect of treatment group on other cognitive and behavioral measures including: 1) Digital Symbol Task, 2) Brown-Peterson Consonant Trigram Memory Task, 3) WISC-III Mazes; 4) Wisconsin Card Sorting Test; 5) MADRS total score,6) YMRS, and 7) Clinical Global Severity score using a mixed model repeated measure ANCOVA as described above. If indicated, variables such as age, gender, concomitant meds, will be examined, as will be done for the primary outcome measure.

Adverse Events:

A tabular summary of all anticipated and unanticipated serious or non-serious adverse events will be presented, including drop-outs due to any side effects or safety reasons. For each serious or other adverse events, the summary includes the adverse event term, affected organ system, number of participants at risk, and number of participants affected, by the study arm or comparison group. Using appropriate parametric or non-parametric methods, each type of adverse event will be compared between groups. All adverse events, e.g. development of suicidal ideation or behaviors or switch to mania/hypomania, will be examined thoroughly for impact on primary and secondary outcome measures. We will assess group differences, excluding subjects without these two side effects.