Janssen Research & Development

Statistical Analysis Plan

A Randomized, Stratified, Double-blind, Placebo-Controlled Study to Investigate the Efficacy, Safety and Tolerability of JNJ-55308942 in Bipolar Depression

Protocol 55308942BIP2001; Phase 2a

JNJ-55308942

Status:	Approved
Date:	07 June 2024
Prepared by:	Janssen Research & Development, a division of Janssen Pharmaceutica NV
Document No.:	EDMS-RIM-736880, 2.0

Compliance: The study described in this report was performed according to the principles of Good Clinical Practice (GCP).

TABLE OF CONTENTS

VERSION HISTORY

Table 1: SAP Version History Summary

SAP Version	Approval Date	Change	Rationale
1.0	21 November 2023	Not Applicable	Initial release
2.0	07 June 2024	1. Section 5.1.2: The end of Week 6 visit window was updated to the day before the first Follow-up/Early Withdrawal visit for safety assessments. 2. Section 5.4.2: Summary statistics and frequency tables were added for individual items of patient-reported outcomes, including PROMIS- Ability to Participate in Social Roles and Activities, PHQ-9, and GAD-7. 3. Section 5.4.2.7: MADRS subgroup analysis by country was added. 4. Section 6.2: The method for handling the assessments that were not performed predose on Treatment Day 1 as per protocol in the analyses was added as a change to protocol planned analyses. 5. Section 6.3: Specifications on the analysis sets to be used for summarizing demographics and baseline characteristics were added. 6. Throughout the SAP: Minor grammatical, formatting, or spelling changes were made.	1. To clarify the visit window start and end dates. 2. To evaluate the efficacy of JNJ-55308942 compared to placebo on the individual items of the patient-reported outcomes. 3. To evaluate the efficacy of JNJ-55308942 compared to placebo on symptoms of depression in different countries. 4. To utilize all collected data in analyses. A significant change in these assessments is not expected on the same day. 5. To clarify the analysis sets to be used for summarizing demographics and baseline characteristics. 6. Minor errors were noted.

1. INTRODUCTION

This statistical analysis plan (SAP) contains definitions, data handling conventions, algorithms, and statistical methods for the planned analyses for the clinical study report (CSR) of Study 55308942BIP2001 (Protocol Amendment 5, 18 August 2023).

Population pharmacokinetics (PK) and pharmacokinetic/pharmacodynamic (PK/PD) analyses, if conducted, will be specified in separate documents as appropriate.

With the exception of the biomarkers used to define subgroups in the analysis of the secondary endpoints of special interest, analyses of other biomarkers and pharmacogenomics data, if conducted, will be specified in separate documents as appropriate.

1.1. Objectives and Endpoints

Objectives	Endpoints
Primary	
To evaluate the efficacy of JNJ-55308942 compared to placebo on symptoms of depression in participants with bipolar disorders (BD) in a major depressive episode (MDE) at Week 6.	Change in Montgomery-Åsberg Depression Rating Scale (MADRS) total score from baseline at Week 6.
Secondary Objectives of Special Interest	
To evaluate the efficacy of JNJ-55308942 compared to placebo on symptoms of anhedonia.	Change in Snaith-Hamilton Pleasure Scale (SHAPS) total score from baseline at Week 6.
To evaluate the efficacy of JNJ-55308942 compared to placebo on symptoms of depression in participants with BD who are heterozygous or homozygous for the rs[CCI] P2RX7 Gain of Function (GoF) single-nucleotide polymorphism (SNP) (genetic subgroup analysis)	Change in MADRS total score from baseline at Week 6.
To evaluate the efficacy of JNJ-55308942 compared to placebo on symptoms of depression in participants with BD type I or BD type II (diagnosis subgroup analysis)	Change in MADRS total score from baseline at Week 6.
To evaluate the efficacy of JNJ-55308942 compared to placebo on symptoms of depression in subgroups of patients with specific biomarker profiles (biomarker subgroup analysis).	Change in MADRS total score from baseline at Week 6.

Objectives	Endpoints
Secondary	
To evaluate the overall safety and tolerability of treatment with JNJ-55308942 as compared to placebo in participants with symptomatic BD over a treatment period of 6 weeks.	Vital signs (pulse/heart rate [HR], systolic blood pressure [SBP], diastolic blood pressure [DBP], respiratory rate), clinical labs (chemistry, hematology, urinalysis), Adverse Events, ECG, Young Mania Rating Scale (YMRS) score, Columbia Suicide Severity Rating Scale (C-SSRS) score.
To evaluate the effect of JNJ-55308942 compared to placebo on disease severity and improvement on the Clinical Global Impression-Severity scale (CGI-S)	CGI-S
To evaluate the pharmacokinetics (PK) of JNJ-55308942 in participants with BD.	Plasma concentrations of JNJ-55308942.
To assess the effect of JNJ-55308942 on reduction of symptoms associated with depression compared to placebo as measured by self-rated outcomes.	PROMIS – Ability to Participate in Social Roles and Activities, Patient Health Questionnaire-9 (PHQ-9), Generalized Anxiety Disorder 7 (GAD-7) score.
To evaluate the impact of treatment with JNJ-55308942 compared with placebo on response (≥ 50% improvement in MADRS total score from baseline) and remission (MADRS total score ≤12) rates at Week 6.	Response (≥ 50% improvement in MADRS total score from baseline) and remission (MADRS total score ≤12) rates at Week 6.
To evaluate the efficacy of JNJ-55308942 compared to placebo on symptoms of depression in a subgroup of patients with messenger ribonucleic acid (mRNA) transcript levels at baseline that exceed the median level for both P2RX7 and IL-1β.	Change in MADRS total score from baseline at Week 6.
To evaluate the efficacy of JNJ-55308942 compared to placebo on symptoms of depression in subgroups of participants with BD not taking any mood stabilizer or antipsychotic, taking a mood stabilizer alone, taking an antipsychotic alone, and taking a combination of a mood stabilizer and an antipsychotic (concomitant medication subgroup analysis).	Change in MADRS total score from baseline at Week 6.

Objectives	Endpoints
Exploratory	
To explore the pharmacokinetic/pharmacodynamic (PK/PD) relationship between plasma exposure of JNJ-55308942, and efficacy and safety parameters in patients with BD.	Exposure-response between exposure and MADRS and selected safety events, as appropriate.
To explore the effect of JNJ-55308942 on PD-related biomarkers.	Including, but not limited to, blood biomarkers related to immune, inflammatory, hormonal, metabolic, hypothalamic-pituitary-adrenal (HPA) axis, P2X7 receptor and monocyte activity.
To explore the effect of pharmacogenomic variations on the PK, efficacy, and safety of JNJ-55308942.	Influence of genetic polymorphisms (e.g., CCI -1) on plasma exposure, as well as on efficacy and safety measures.
To explore the effect of JNJ-55308942 compared to placebo on the well-being of participants.	Self-Assessment of Well-Being (SAWB) outcome.

1.2. Study Design

This is a double-blind, randomized, stratified, placebo-controlled, parallel-group, multicenter study to investigate the efficacy, safety, and tolerability of JNJ-55308942 in bipolar depression. Approximately 164 participants will be enrolled in this study to obtain approximately 132 completers.

The target study population includes participants of any gender, between 18 and 64 years of age inclusive, with a Diagnostic and Statistical Manual of Mental Disorders - 5th edition (DSM-5) diagnosis of BD (types I or II) without current psychotic features as confirmed by the Mini International Neuropsychiatric Interview (MINI) with Borderline Personality Disorder module. The participants should be experiencing an MDE that does not require either hospitalization or antidepressant medication. Each potential participant must satisfy one of the three criteria listed below to meet eligibility for the study population.

- Participants who are currently unmedicated and have not received any mood stabilizers or antipsychotics for four weeks before screening.
- Participants who are on a stable monotherapy regimen of one of the following mood stabilizers (lithium, valproate forms [e.g., divalproex sodium], or lamotrigine) or one of the following antipsychotics (lurasidone, cariprazine, quetiapine, lumateperone, or olanzapine) for four weeks before screening.

- Participants who are on a stable adjunctive regimen of one of the following antipsychotics (lurasidone, quetiapine, lumateperone, or olanzapine) and one of the following mood stabilizers (lithium or valproate forms [e.g., divalproex sodium]) may be considered for eligibility based on the approved therapies in each market with an eligibility assessment conducted between the investigator and sponsor. Assessment of the stable adjunctive regimen of an antipsychotic and a mood stabilizer will be conducted by the investigator and sponsor's clinical judgement and documented for each patient. This regimen must be in place 4 weeks before screening.

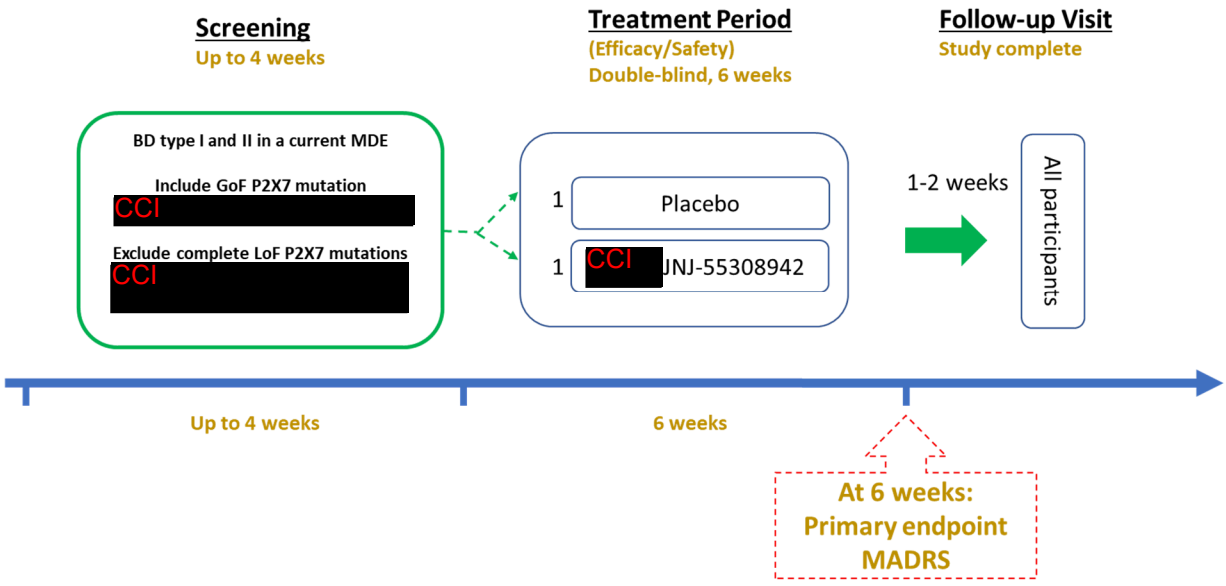
The patient must be symptomatic as assessed using the 17-item Hamilton Depression Rating Scale (HDRS17). Symptom intensity over the past week will be evaluated.

In addition, the participants must carry the GoF P2RX7 mutation (for inclusion, must be CCI or CCI at rsCCI 9/nucleotide CCI; CCI) and must not have the Loss of Function (LoF) P2RX7 mutations (must be CCI or CC at rsCCI 3/nucleotide CCI; CCI and must be CCI at rsCCI /nucleotide CCI CCI).

Participants who are enrolled in the study will continue on their current medication regimen, if applicable, throughout the study. No dose changes are permitted from screening through the end of the study (until after the follow-up visit). No concurrent use of antidepressants will be allowed until after the follow-up visit.

For each participant, the study will consist of 3 phases: a screening phase of up to 4 weeks, a double-blind treatment phase of 6 weeks, and a post-treatment follow-up phase of up to 2 weeks (Figure 1). After giving written informed consent for screening, participants will be screened between 28 days and 1 day before the double-blind treatment phase to ascertain their eligibility for the study according to the inclusion and exclusion criteria. Eligible participants will enter the double-blind treatment phase where they will be randomly assigned with a 1:1 ratio to once daily administration with C CI mg JNJ-55308942 or placebo for 6 weeks, Day 1 up to and including the last treatment visit, Day 43 +/- 1 day. Between approximately 1 to 2 weeks after administration of the last dose, participants will return to the study site for a follow-up visit. The total study duration for each participant will be up to 12 weeks. There will be 7 scheduled visits to the study site during the study, including screening and follow-up visit.

Figure 1: Schematic Overview of the Study



1.2.1. Randomization and Blinding

Procedures for Randomization

Central randomization will be implemented in this study. Participants will be randomly assigned to 1 of 2 groups (JNJ-55308942 or placebo) based on a computer-generated randomization schedule prepared before the study by or under the supervision of the sponsor. The randomization will be balanced by using randomly permuted blocks and will be stratified by 4 factors: BD type (I or II), country, P2RX7 GoF SNP mutation genotype (homozygous or heterozygous), as well as concomitant medication status (no mood stabilizer or antipsychotic, a mood stabilizer alone, an antipsychotic alone, or a combination of a mood stabilizer and an antipsychotic). The interactive web response system (IWRS) will assign a unique intervention code, which will dictate the intervention assignment and matching study intervention kit for the participant.

Blinding

Blinded treatment will be used to reduce potential bias during data collection and evaluation of clinical endpoints.

The investigator will not be provided with randomization codes. The codes will be maintained within the IWRS, which has the functionality to allow the investigator to break the blind for an individual participant.

Data that may potentially unblind the intervention assignment (e.g., study intervention plasma concentrations, plasma biomarkers) will be handled with special care to ensure that the integrity of the blind is maintained and the potential for bias is minimized. This can include making special provisions, such as segregating the data in question from view by the investigators, clinical team,

or others as appropriate until the time of database lock and unblinding. Also, the site staff should not discuss insights on unblinding of randomization with the participants.

Under normal circumstances, the blind should not be broken until all participants have completed the study and the database is finalized/locked. The investigator may, in an emergency, determine the identity of the intervention by contacting the IWRS. While the responsibility to break the intervention code in emergency situations resides solely with the investigator, it is recommended that the investigator contact the sponsor or its designee if possible, to discuss the particular situation, before breaking the blind.

Participants who have had their study intervention assignment unblinded should continue to return for scheduled evaluations.

In general, randomization codes will be disclosed fully only if the study is completed and the clinical database is closed/locked.

2. STATISTICAL HYPOTHESES

The primary objective of this study is to evaluate the efficacy of JNJ-55308942 compared to placebo on symptoms of depression in participants with BD in an MDE. The null hypothesis that will be tested to address the primary objective of the study is that there is no difference between JNJ-55308942 ([redacted] mg once daily) and placebo in reducing the symptoms of depression in participants with BD in an MDE as assessed by change in MADRS total score from baseline to Week 6.

3. SAMPLE SIZE DETERMINATION

The estimated sample size of 164 participants (82 participants per group) was determined based on the assumption of an effect size of at least 0.45 for the MADRS total score (difference in mean change from baseline at Week 6 between the JNJ-55308942 and placebo groups of 3.9 units with a standard deviation of 8.6). This is considered to be a clinically relevant difference in a population with BD types I and II. The standard deviation of 8.6 in the change in MADRS total score from baseline is a reasonable assumption based on data in a published phase 3 study of quetiapine in BD type I and type II depression (Calabrese 2005). Power is set at 90%, with a 1-sided alpha of 0.10 and a 6-week drop-out rate of 20%. The estimated number of completers to be included in the primary analysis is 132.

4. POPULATIONS (ANALYSIS SETS) FOR ANALYSIS

Population	Description
Enrolled	All participants who signed the ICF at screening
Randomized	All participants who were randomized in the study
Full Analysis Set (FAS)	All randomized participants who received at least 1 dose of study intervention and have both the baseline and at least 1 post-baseline MADRS measurement
Safety	All randomized participants who take at least 1 dose of study intervention
Pharmacokinetics (PK) Analysis Set	All participants who received at least 1 dose of JNJ-55308942 and have at least 1 valid blood sample drawn for PK analysis

The randomized analysis set will be used for summarizing the overall study completion/withdrawal information. Efficacy analyses will be based on the full analysis set unless otherwise specified. Safety analyses will be based on the safety analysis set and PK analyses will be based on the PK analysis set.

5. STATISTICAL ANALYSES

5.1. General Considerations

Continuous variables will be summarized by descriptive statistics including the number of participants (N), mean, standard deviation (SD), median, and range (minimum; maximum) unless otherwise specified. Categorical variables will be summarized by frequency counts and percentages.

5.1.1. Analysis Phases

Screening

The screening phase begins on the date the first informed consent is obtained and ends 1 day before the first dose of study intervention in the double-blind treatment phase. The screening phase end date is left missing for those participants who did not receive study intervention.

Double-Blind Analysis Phase

The analysis reference start date of the double-blind analysis phase is the date of the first dose of the double-blind study intervention. The analysis reference end date of the double-blind analysis phase is the date of study disposition (last contact).

5.1.2. Visit Windows

As participants do not always adhere to the protocol visit schedule, the following rules are applied to assign actual visits to analysis visits. Listed below are the visit windows and the target days for each visit. The reference day is Study Day 1 (which is the first day the study drug was taken in the double-blind treatment phase). If a subject has 2 or more actual visits in one visit window, the visit closest to the target day will be used as the protocol visit for that visit window. The other additional

visit(s) will not be used in the summaries or analyses. If 2 actual visits are equidistant from the target day within a visit window, the later visit is used.

All assignments will be made in chronological order. Once a visit date is assigned to a visit window, it will no longer be used for a later time point except for the endpoint. Listed below (Table 2) are the analysis visit windows and the target days for each visit defined in the protocol.

The baseline is defined as the last observation before the start of the first study intervention administration in the double-blind treatment phase. Assessments prior to Day 1 of the double-blind treatment phase that are not considered baseline assessments will be labeled as 'Screening'.

Table 2: Visit Windows

Parameter	Analysis Phase	Scheduled Visit Number	Time Interval (Label on Output)	Time Interval (Day)*	Target Time Point (Day)*
MADRS, CGI-S	Double-Blind	2	Baseline	≤1	1
		3	Week 1	[2,11]	8
		4	Week 2	[12,21]	15
		5	Week 4	[22,35]	29
		6	Week 6	≥36	43
SHAPS	Double-Blind	2	Baseline	≤1	1
		6	Week 6	≥2	43
PROMIS, PHQ-9	Double-Blind	2	Baseline	≤1	1
		3	Week 1	[2,11]	8
		4	Week 2	[12,18]	15
			Week 3	[19,25]	22
		5	Week 4	[26,32]	29
			Week 5	[33,39]	36
		6	Week 6	≥40	43
GAD-7	Double-Blind	2	Baseline	≤1	1
		4	Week 2	[2,21]	15
		5	Week 4	[22,35]	29
		6	Week 6	≥36	43
Clinical Laboratory tests (Hematology, Serum, Urinalysis)	Screening	1	Screening	<1	<1
	Double-Blind	2	Baseline	≤1	1
		4	Week 2	[2,21]	15
		5	Week 4	[22,35]	29
		6	Week 6	[36, Day of FU/EW)	43
		7	FU/EW		49 to 56
Vital signs, 12-lead ECG, C-SSRS, Urine Drug Screen	Screening	1	Screening	<1	<1
	Double-Blind	2	Baseline	≤1	1
		3	Week 1	[2,11]	8
		4	Week 2	[12,21]	15
		5	Week 4	[22,35]	29
		6	Week 6	[36, Day of FU/EW)	43
		7	FU/EW		49 to 56
Physical examination, Body weight	Screening	1	Screening	<1	<1
	Double-Blind	2	Baseline	≤1	1
		6	Week 6	[2, Day of FU/EW)	43
		7	FU/EW		49 to 56
YMRS	Screening	1	Screening	<1	<1
	Double-Blind	2	Baseline	≤1	1
		3	Week 1	[2,11]	8

Table 2: Visit Windows

Parameter	Analysis Phase	Scheduled Visit Number	Time Interval (Label on Output)	Time Interval (Day)*	Target Time Point (Day)*
		4	Week 2	[12,21]	15
		5	Week 4	[22,35]	29
		6	Week 6	≥36	43

*Relative to Study Day 1; FU/EW = Follow-up/Early Withdrawal

5.1.3. Imputation of Efficacy

Imputation for missing data will include the following methods (Table 3).

Table 3: Imputation of Missing Efficacy Data

Imputation	Text
Multiple Imputation (MI) method	1) Copy Reference 2) Delta Adjustment.
Non-Responder	Participants with missing values will be imputed as non-responders.
Non-Remitter	Participants with missing values will be imputed as non-remitters.

Imputation of total scores will apply only to the MADRS. Imputation of the MADRS total score will be performed only when 1 item score is missing. If 2 or more items are missing, the total score will be left missing. The total score will be imputed by calculating the sum of the scores of the non-missing items and multiplying it by the ratio of the maximum possible number of items (i.e., 10) to the number of non-missing items (i.e., 9).

For all other efficacy scales where multiple items are summed to create a total score, if any item of the scale is missing on one visit, the total score for that scale at that visit will be left missing.

5.2. Participant Dispositions

The number of participants in the following disposition categories will be summarized throughout the study by intervention group and overall:

- Participants randomized
- Participants who received study intervention
- Participants who completed, discontinued study intervention and reasons for discontinuation
- Participants who completed, terminated study prematurely and reasons for termination

Listings of participants will be provided for the following categories:

- Participants who discontinued study intervention
- Participants who terminated study prematurely (trial disposition page)
- Participants who were unblinded during the study period
- Participants who were randomized yet did not receive study intervention.

These summaries and listings will be provided for the randomized analysis set.

5.3. Primary Endpoint(s) Analysis

The primary endpoint analysis will be based on the full analysis set using the MADRS total scores collected during the double-blind treatment phase.

5.3.1. Definition

The MADRS is a clinician-administered scale designed to measure depression severity and detects changes due to antidepressant intervention (Montgomery 1979). The scale consists of 10 items, each of which is scored from 0 (item not present or normal) to 6 (severe or continuous presence of the symptoms). Higher scores represent a more severe condition. The MADRS evaluates reported sadness, apparent sadness, inner tension, reduced sleep, reduced appetite, concentration difficulties, lassitude, inability to feel, pessimistic thoughts, and suicidal thoughts. The test exhibits high inter-rater reliability. The recall period for the MADRS is 7 days.

The MADRS total score is the sum of scores from individual question items at a given time point and ranges from 0 to 60. Higher scores represent a more severe condition. Imputation of the total score is presented in section 5.1.3.

Negative changes in the MADRS total score indicate improvement.

5.3.2. Estimand

Primary Trial Objective: To evaluate the efficacy of JNJ-55308942 compared to placebo on symptoms of depression in participants with BD in an MDE

Estimand Scientific Question of Interest: What is the effect of assigning BD patients in an MDE to JNJ-55308942 vs. placebo in reducing the symptoms of depression after 6 weeks of treatment if no participant discontinued the treatment?

The estimand for the primary efficacy analysis is defined by the following 5 components:

Study Intervention: JNJ-55308942 [redacted] mg once daily versus placebo

Population: participants with BD types I and II and in an MDE

Variable: change from baseline in the MADRS total score to Week 6

Population-level Summary: the difference in mean change from baseline in MADRS total score to Week 6 between JNJ-55308942 [redacted] mg once daily and placebo

Intercurrent events and their corresponding strategies:

Intercurrent Events	Strategy for Addressing Intercurrent Events and Its Description
Discontinuation of study intervention	**Hypothetical strategy:** as if no participant discontinued the study intervention. After study intervention discontinuation, similar efficacy is assumed for participants who discontinued as those

Intercurrent Events	Strategy for Addressing Intercurrent Events and Its Description
	participants from the same intervention group who did not discontinue the study intervention.
Change or discontinuation of background therapy (mood stabilizer and/or antipsychotic)	**Hypothetical strategy:** same as above. Change or discontinuation of background therapy will result in discontinuation of study intervention.

5.3.2.1. Analysis Methods

5.3.2.1.1. Primary Analysis

The primary efficacy analysis will be performed utilizing the hypothetical estimand defined above. The change from baseline in MADRS total score will be compared between the JNJ-55308942 C
Ci mg once daily group and the placebo group at Week 6 in a mixed-effects model for repeated measures (MMRM), with time, intervention (placebo or JNJ-55308942), BD type (I or II), country, P2RX7 GoF SNP mutation genotype (homozygous or heterozygous), concomitant medication status (no mood stabilizer or antipsychotic, a mood stabilizer alone, an antipsychotic alone, or a combination of a mood stabilizer and an antipsychotic), and time-by-intervention interaction as factors, with baseline MADRS total score as a continuous covariate. Data after treatment discontinuation will be considered missing. The within-subject covariance between visits will be estimated via an unstructured variance-covariance matrix. In case of convergence problems, alternative variance-covariance structures will be attempted in the following order, with the first structure that converges being used in the analysis: heterogeneous Toeplitz, standard Toeplitz, and first order autoregressive. Kenward-Roger approximation will be used to estimate the denominator degrees of freedom. Sandwich estimator will be used to address the potential misspecification of the covariance matrix when a structured covariance matrix is used due to the convergence problem with the unstructured covariance matrix.

The comparison of JNJ-55308942 versus placebo will be performed using the appropriate contrasts directly from the MMRM analysis. A 2-sided 80% confidence interval (CI) for the difference in least squares means will be presented over time. A 1-sided p-value will be calculated at Week 6. Least square means of change from baseline (+/- SE) will be presented graphically over time.

Descriptive statistics of the actual values and the change from baseline to each postbaseline time point will be presented for MADRS total score by intervention group.

5.3.2.1.2. Sensitivity Analysis

The primary analysis based on MMRM is valid and produces unbiased estimates of the treatment effect if missing data are missing at random (MAR), meaning that the probability of a value being missing, conditional on the observed data in the statistical model, is random and not dependent on the unknown value of missing data point. The MAR assumption cannot be verified vs. missing not at random (MNAR) using observed data. Therefore, a sensitivity analysis using copy-reference multiple imputation will be conducted under the MNAR assumption to evaluate the robustness of efficacy results and the effect of the missing data. In addition, a tipping-point analysis will be conducted if the primary analysis results show a significantly greater improvement in the MADRS

total score at Week 6 in JNJ-55308942 compared to placebo (i.e., 1-sided p-value ≤ 0.1 in favor of JNJ-55308942).

The copy reference multiple imputation approach does not assume a sustained benefit of experimental intervention for the efficacy data that is either not used or missing after the intercurrent event and uses an imputation method that is based on the control group distribution and the estimated correlations between time points in the control group. MAR is assumed for intermediate missing data (i.e., missing data between non-missing observations). MNAR is assumed for monotone missing (i.e., missing data after the participant experienced an intercurrent event) in participants in the experimental intervention group, where efficacy scores are assumed as if the participant had always been in the control group. MAR is assumed for missing data in the control group.

This method will employ the following 3 steps:

Step 1 – Multiple imputation

If there are participants with a non-monotone missing data pattern, datasets with only monotone missing data patterns will be created first by imputing the intermediate missing values using the Markov Chain Monte Carlo (MCMC) method, which assumes a multivariate normal distribution over all variables included in the imputation model. This imputation will be done by intervention group. The variables to be used in the imputation are treatment, corresponding baseline values, and values observed at all double-blind visits (Weeks 1, 2, 4, 6). 50 imputations will be performed to create 50 datasets that now have monotone missing (i.e., missing data after the participant experienced an intercurrent event) data pattern.

The efficacy data that is either not used or missing after the intercurrent event at a given time point will be imputed using the imputation model of the control group, i.e., conditional on the data observed or imputed at previous time points relative to the mean of the model for the control group.

Step 2 – Analysis

The same MMRM analysis as described for the primary efficacy analysis will be performed for each of the adjusted fully imputed datasets.

Step 3 – Pooling

Rubin's methodology (Rubin 1987) will be applied to the MMRM results from the 50 imputed datasets to produce final inferences.

An additional sensitivity analysis using the delta adjustment multiple imputation (tipping-point) approach will be conducted if the results from the primary analysis show a significantly greater improvement in the MADRS total score at Week 6 in JNJ-55308942 compared to placebo (i.e., 1-sided p-value ≤ 0.1 in favor of JNJ-55308942). This approach imputes missing data over a range of possible scenarios for the treatment effect and finds a "tipping point" where the treatment effect in participants with missing data overturns the significant treatment effect. This method will

employ the same 3 steps as in the copy-reference multiple imputation sensitivity analysis except during the first step monotone missing data will first be imputed by MAR-based MI, and the imputed values in the experimental intervention group will be adjusted using a range of delta values to generate adjusted fully imputed datasets. The adjusted fully imputed datasets will be analyzed and combined using the same methods as Steps 2 and 3 of the copy-reference multiple imputation approach. The delta will be set in a range from 0 to Δ^* in increments of 1 as adding positive values results in higher (worse) scores. Δ^* represents the adjustments leading to the "tipping point". It is the smallest delta adjustment value at which conclusions change from favorable (i.e., statistically significant, 1-sided p-value $\leq$ 0.1 in favor of JNJ-55308942) to unfavorable. Between-group comparisons to placebo at Week 6 (e.g., 1-sided p-values, point estimates for intervention difference) will be displayed graphically for each considered delta, up to the "tipping point" adjustment. Clinical judgement will be applied to evaluate the plausibility of the assumptions underlying this "tipping point".

5.4. Secondary Endpoint(s) Analysis

All secondary endpoints analysis will be based on the full analysis set.

5.4.1. Secondary Endpoints of Special Interest

5.4.1.1. Snaith-Hamilton Pleasure Scale (SHAPS)

5.4.1.1.1. Definition

The SHAPS is a reliable, valid, and unidimensional instrument to assess hedonic capacity in adults with Major Depressive Disorder in the last few days. It is a 14-item, self-report tool with a completion time below 5 minutes. Each of the items has a set of 4 response categories: 1=Definitely Agree/Strongly Agree, 2=Agree, 3=Disagree, and 4=Strongly Disagree. The SHAPS total score is the sum of the 14 item scores, ranging from 14 to 56. A higher SHAPS total score indicates higher levels of current anhedonia (Nakonezny 2010).

Negative changes in the SHAPS total score indicate improvement.

5.4.1.1.2. Analysis Methods

Descriptive statistics of the actual values at Baseline, Week 6, and the change from baseline to Week 6 will be presented for SHAPS total score by intervention group.

An analysis of covariance (ANCOVA) model will be used to test the difference of change from baseline at Week 6 in SHAPS total score between JNJ-55308942 [redacted] mg once daily and placebo. The model will include intervention (placebo or JNJ-55308942), BD type (I or II), country, P2RX7 GoF SNP mutation genotype (homozygous or heterozygous), and concomitant medication status (no mood stabilizer or antipsychotic, a mood stabilizer alone, an antipsychotic alone, or a combination of a mood stabilizer and an antipsychotic) as factors, and baseline SHAPS total score as a covariate. Comparison between JNJ-55308942[redacted] mg once daily and placebo will be performed using appropriate contrast. Difference of least square means and 2-sided 80% CI will

be presented. Least square means of change from baseline (+/- SE) to Week 6 will be presented graphically.

Frequency distributions of the SHAPS individual items will be provided at each assessment time point.

5.4.1.2. MADRS Subgroup Analysis

5.4.1.2.1. Definition of Subgroups of Special Interest

MADRS subgroup analysis will be conducted according to the definition below.

Subgroup	Variant	Definition
P2RX7 GoF SNP mutation genotype	1	Heterozygous / Homozygous
BD type	1	I / II
Biomarker profile	1	Yes / No

5.4.1.2.2. Analysis Methods

Changes in MADRS total score at Week 6 from baseline, will be assessed in subgroups for BD type (I or II), P2RX7 GoF SNP mutation genotype (heterozygous vs. homozygous), and patients with specific biomarker profiles (yes or no). Participants will be divided into subgroups based on the definition in Section 5.4.1.2.1 to evaluate efficacy within these subgroups (i.e., evaluating change from baseline to the end of Week 6 in the MADRS total score in participants within each subgroup).

Descriptive statistics of the actual values and change from baseline at each postbaseline time point in the MADRS total score will be presented by intervention group for these subgroups. The change from baseline in the MADRS total score will be analyzed using the same MMRM described in Section 5.3.2.1.1 for each subgroup. Comparison between JNJ-55308942 [redacted]mg once daily and placebo will be performed using the appropriate contrasts. Difference of least square means and 2-sided 80% CI will be presented over time. Least square means of change from baseline (+/- SE) will be presented graphically over time.

A forest plot summarizing the differences of least square means and 2-sided 80% CI between JNJ-55308942 [redacted] mg once daily and placebo at Week 6 in the subgroups will be presented.

5.4.2. Other Secondary Endpoints

5.4.2.1. Clinical Global Impression – Severity (CGI-S)

5.4.2.1.1. Definition

The CGI-S provides an overall clinician-determined summary measure of the severity of the participant's illness that takes into account all available information, including knowledge of the participant's history, psychosocial circumstances, symptoms, behavior, and the impact of the symptoms on the participant's ability to function (Guy 1991). The CGI-S evaluates the severity of

psychopathology on a scale of 1 to 7. Considering total clinical experience with the depression population, a participant is assessed on severity of illness at the time of rating according to 0=not assessed; 1=normal (not at all ill); 2=borderline ill; 3=mildly ill; 4=moderately ill; 5=markedly ill; 6=severely ill; 7=among the most extremely ill patients.

A score of 0 indicates that the participant was not assessed and will be treated as missing. The score from 1 to 7 will be summarized as recorded.

Negative changes in CGI-S score indicate improvement.

5.4.2.1.2. Analysis Methods

Descriptive statistics of the actual values and the change from baseline will be presented by intervention group for observed case data.

Frequency distribution of the CGI-S scores at each scheduled time point will be provided over time by intervention group. The frequencies of CGI-S scores will be plotted over time by intervention group. In addition, frequency of shifting scale at Week 6 from baseline will be provided.

5.4.2.2. PROMIS – Ability to Participate in Social Roles and Activities

5.4.2.2.1. Definition

The PROMIS - Ability to Participate in Social Roles and Activities item bank assesses the perceived ability to perform one's usual social roles and activities. The item bank does not use a time frame (e.g., over the past seven days) when assessing the ability to participate in social roles and activities. The Short Form 4a includes 4 items that represent this concept. Each question has 5 response options ranging in value from 1 to 5 with higher scores indicating better social function. The total raw score for the short form is calculated by summing the values of the response to each question, so for the 4-item form, the lowest possible raw score is 4; the highest possible raw score is 20. The total raw score can also be converted to a T score with a mean of 50 and a standard deviation of 10 (Appendix 6.9).

5.4.2.2.2. Analysis Methods

Descriptive statistics of the actual values and the change from baseline at each post-baseline time point will be presented for PROMIS - Ability to Participate in Social Roles and Activities total T-score, total raw score, and individual items by intervention group and visit.

The change from baseline in the PROMIS - Ability to Participate in Social Roles and Activities total T-score will be analyzed using a similar MMRM as described in Section 5.3.2.1.1, with the covariate "baseline MADRS total score" changed to "baseline PROMIS - Ability to Participate in Social Roles and Activities total T-score". Comparison between JNJ-55308942 [redacted] mg once daily and placebo will be performed using the appropriate contrasts. Difference in least square means and 2-sided 80% CI will be presented over time. Least square means of change from baseline (+/- SE) for PROMIS T-score will be presented graphically over time.

Frequency distributions of the individual items of the PROMIS - Ability to Participate in Social Roles and Activities will be provided at each assessment time point.

5.4.2.3. Patient Health Questionnaire – 9 (PHQ-9)

5.4.2.3.1. Definition

The PHQ-9 scale scores each of the 9 symptom domains of the DSM-5 major depressive disorder criteria and it has been used both as a screening tool and a measure of response to intervention for depression. Each item is rated on a 4-point scale (0=Not at all, 1=Several Days, 2=More than half the days, and 3=Nearly every day). The participant's item responses are summed to provide a total score (range of 0 to 27), with higher scores indicating greater severity of depressive symptoms. The recall period is 2 weeks.

Negative changes in the PHQ-9 total score indicate improvement.

5.4.2.3.2. Analysis Methods

Descriptive statistics of the actual values and the change from baseline to each post-baseline time point will be presented for the PHQ-9 total score and individual items by intervention group.

The analysis for changes from baseline in PHQ-9 total score at each week will be carried out using a similar MMRM as described in Section 5.3.2.1.1 for the primary efficacy endpoint, with the covariate "baseline total MADRS score" changed to "baseline PHQ-9 total score". Comparison between JNJ-55308942 [redacted] mg once daily and placebo will be performed using the appropriate contrasts. Least square means of change from baseline (+/- SE) for the PHQ-9 total score will be presented graphically over time.

Frequency distributions of the PHQ-9 individual items will be provided at each assessment time point.

5.4.2.4. Generalized Anxiety Disorder – 7 (GAD-7)

5.4.2.4.1. Definition

The GAD-7 is a brief and validated 7-item self-report assessment of overall anxiety. Participants respond to each item using a 4-point scale with response categories of 0=not at all, 1=several days, 2=over half the days, and 3=nearly every day. The total score is the sum of item responses, ranging from 0 to 21. Higher scores indicate more anxiety.

Negative changes in the GAD-7 total score indicate improvement.

5.4.2.4.2. Analysis Methods

Descriptive statistics of the actual values and the change from baseline at weeks 2, 4, and 6 will be presented for the GAD-7 total score and individual items by intervention group for observed case data.

The analyses for changes from baseline in GAD-7 total score at weeks 2, 4, and 6 will be carried out using a similar MMRM as described in Section 5.3.2.1.1 for the primary efficacy endpoint, with the covariate "baseline total MADRS score" changed to "baseline GAD-7 total score". Comparison between JNJ-55308942 [redacted]mg once daily and placebo will be performed using the appropriate contrasts. Least square means of change from baseline (+/- SE) for the GAD-7 total score will be presented graphically over time.

Frequency distributions of the GAD-7 individual items will be provided at each assessment time point.

5.4.2.5. Response Based on MADRS Total Score

5.4.2.5.1. Definition

A participant is defined as a responder at a given time point if the percent improvement from baseline in MADRS is ≥50% at that time point (i.e., percent change ≤-50%). Participants who do not meet such a criterion will be considered as non-responders. Imputation of missing response status is presented in Section 5.1.3.

5.4.2.5.2. Analysis Methods

The number and percentage of participants who achieve a response will be summarized at each time point during the double-blind treatment phase by intervention group. The analysis will be performed on observed as well as imputed data (participants with missing values will be imputed as non-responders).

A Chi-square test will be used to test the difference between JNJ-55308942[redacted] mg once daily and placebo for response rate at each time point. A 2-sided 80% CI for the intervention difference will be calculated based on the Wald statistic. The point estimate and 2-sided 80% CI will be provided for the relative risk of response.

Response rate will be plotted over time by intervention group.

The cumulative response rate, defined as the percentage of participants experiencing at least a given value of percent reduction from baseline to Week 6 in the MADRS total score, will be presented graphically for observed data.

5.4.2.6. Remission Based on MADRS Total Score

5.4.2.6.1. Definition

A participant is defined as a remitter at a given time point if the MADRS total score is ≤12 at that time point. Participants who do not meet such a criterion will be considered as non-remitters. Imputation of missing remission status is presented in Section 5.1.3.

5.4.2.6.2. Analysis Methods

The number and percentage of participants who achieve remission will be summarized at each time point during the double-blind treatment phase by intervention group. The analysis will be

performed on observed as well as imputed data (participants with missing values will be imputed as non-remitters).

A Chi-square test will be used to test the difference between JNJ-55308942 CC mg once daily and placebo for remission rate at each time point. A 2-sided 80% CI for the intervention difference will be calculated based on the Wald statistic. The point estimate and 2-sided 80% CI will be provided for the relative risk of remission.

The remission rate will be plotted over time by intervention group.

5.4.2.7. Additional MADRS Subgroup Analysis

Changes in MADRS total score to Week 6 from baseline, will be assessed in subgroups for concomitant medication status (no mood stabilizer or antipsychotic, a mood stabilizer alone, an antipsychotic alone, or a combination of a mood stabilizer and an antipsychotic) and country. Descriptive statistics of the actual values and change from baseline at each postbaseline time point in the MADRS total score will be presented by intervention group for the subgroups. Mean changes from baseline (+/- SE) will be presented graphically over time. A forest plot summarizing the differences of mean changes and 2-sided 80% CI between JNJ-55308942 CC mg once daily and placebo at Week 6 in the subgroups will be presented. The change from baseline in the MADRS total score will be analyzed using the same MMRM described in Section 5.3.2.1.1 for each subgroup when feasible.

5.5. Exploratory Endpoint(s) Analysis

5.5.1. Self-Assessment of Well-Being (SAWB)

5.5.1.1. Definition

The SAWB questionnaire is a multiple-item self-report questionnaire designed to provide additional information regarding the participant's subjective experience while taking study intervention. This is an internal Janssen questionnaire, and the questions will be asked to the participant daily by the Q1.6 app during the double-blind treatment phase. The SAWB questionnaire has a total of 11 questions. The first 8 questions monitor the participant's manic symptoms, pattern of sleep, energy level, social activities, and concentration level. Each question is answered with a visual analog scale (VAS) between 0 and 100. Question 9 asks the participant's subjective experience of overall depression level since study start with 3 possible answers: a) improved, b) not changed, or c) worsened. If the participant answers "improved", Question 10 will be asked to assess how much depression has improved (slightly improved, much improved, or very much improved). If the participant answers "worsened" to Question 9, Question 11 will be asked to assess how much depression has worsened (slightly worse, much worse, or very much worse). The 8 VAS questions will be rotated with 2 questions being asked on a single day. The remaining questions will be asked weekly. The Q1.6 app will send an alert to the investigator when the participant has indicated that her/his depression is "much worse" or "very much worse".

5.5.1.2. Analysis Methods

Analysis of the SAWB data will be based on the safety analysis set.

Actual values will be averaged by study week for each of the 8 VAS items. Descriptive statistics of the weekly average of each of the 8 VAS items will be presented by intervention group over time. Mean (+/- SE) of the weekly average of each of the 8 VAS items will be presented graphically over time.

A frequency distribution at each time point by intervention group will be provided for Questions 9 to 11. The number and percentage of participants with a reply that her/his depression is 'much worse' or 'very much worse' will be summarized by intervention group. A listing of participants with a reply that her/his depression is 'much worse' or 'very much worse' will be provided.

5.6. Safety Analyses

All safety analyses will be based on the safety analysis set based on actual intervention received, unless otherwise specified.

5.6.1. Extent of Exposure

Duration of study intervention and total number of capsules taken will be summarized by intervention group using descriptive statistics.

Study intervention duration is defined as (date of last dose of study intervention – date of first dose of study intervention) +1.

Total number of capsules taken will be calculated based on drug accountability data.

In case a medication packet is not returned by a participant or the drug accountability data is missing, the actual number of capsules taken may be inferred based on Digital Healthy Assessment app data if available.

5.6.1.1. Intervention Compliance

Compliance will be summarized descriptively for each intervention.

The percent compliance will be categorized and the number and percentage of participants in each category will be summarized by intervention group.

Compliance will be calculated as follows:

Compliance (%) = (actual number of capsules taken/total number of capsules supposed to be taken) x100.

5.6.2. Adverse Events

The verbatim terms used in the CRF by investigators to identify adverse events (AEs) will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). Any AE occurring at or after the initial administration of study intervention through the day of the last dose plus 30 days

is considered to be treatment-emergent. If the event occurs on the day of the initial administration of study intervention, and either event time or time of administration is missing, then the event will be assumed to be treatment-emergent. If the event date is recorded as partial or completely missing, then the event will be considered to be treatment-emergent unless it is known to be prior to the first administration of study intervention based on the partial onset date or resolution date. All reported treatment-emergent adverse events (TEAEs) will be included in the analysis. For each AE, the number and percentage of participants who experience at least 1 occurrence of the given event will be summarized by intervention group.

Summary tables will be provided for:

- Overall summary of TEAEs, including any TEAEs, serious TEAEs, TEAEs leading to discontinuation of study intervention, related TEAEs, and TEAEs by worst severity
- TEAEs by system organ class and preferred term
- Treatment-emergent serious adverse events (SAEs) by system organ class and preferred term
- TEAEs leading to discontinuation of study intervention by system organ class and preferred term
- TEAEs by system organ class, preferred term, and worst severity
- TEAEs by system organ class, preferred term, and relationship to study intervention.

In addition to the summary tables, listings will be provided for participants who had:

- AEs
- SAEs
- AEs leading to discontinuation of study intervention

A listing of participants who died will be provided.

5.6.3. Additional Safety Assessments

5.6.3.1. Clinical Laboratory Tests

Descriptive statistics of chemistry, hematology, and urinalysis (pH and specific gravity) laboratory tests and changes from baseline will be presented at each scheduled time point by intervention group.

Abnormality criteria based on normal ranges will be applied to baseline and postbaseline values. Shift tables will be provided summarizing the shift in laboratory values from baseline to scheduled time points with respect to abnormality criteria (low, normal, high). A listing of participants with any laboratory results outside of the reference ranges will be provided.

A listing of participants meeting any of the following hepatic toxicity stopping criteria (confirmed by repeat testing) will be provided:

- Aspartate aminotransferase (AST) exceeds 5x the upper limit of normal (ULN)

- Alanine aminotransferase (ALT) exceeds 5x the ULN
- AST exceeds 3x the ULN and total bilirubin exceeds 1.5x ULN
- ALT exceeds 3x the ULN and total bilirubin exceeds 1.5x ULN

The number and percentage of participants with positive urine drug screen results will be presented over time by intervention group.

5.6.3.2. Vital Signs and Physical Examination Findings

Vital sign parameters that will be analyzed are temperature (oral, tympanic, or temporal), respiration rate, pulse, blood pressure (systolic and diastolic), weight, and Body Mass Index (BMI). BMI will be calculated as weight (kg)/(height (m))^2. The height measurement collected at screening will be used in the calculation. Change from baseline will be summarized at each assessment time point by intervention group.

Vital sign parameters and changes from baseline will be summarized using descriptive statistics at each scheduled time point by intervention group.

Abnormality criteria (as defined in Table 4) will be applied to baseline and postbaseline values. For baseline values, increase or decrease criteria are not applied.

Postbaseline values will be considered treatment-emergent if they meet both value and change criteria in Table 4.

For criteria that do not include an increase or decrease from baseline:

- Treatment-emergent will be concluded if the postbaseline value is above the upper limit and the baseline value is below the upper limit (e.g., Normal or Low). The same applies to the postbaseline value being below the lower limit with the baseline value being above the lower limit (e.g., Normal or High).
- If the baseline value is missing, a postbaseline abnormality will always be considered as treatment-emergent.

Incidence of treatment-emergent clinically important abnormalities in vital signs during intervention will be summarized at each scheduled time point by intervention group. A listing of participants with clinically important abnormalities in vital signs will be presented.

Table 4: Clinically Important Abnormalities in Vital Signs

Vital Sign	Abnormal Category	Criteria
Pulse	Abnormally high	>100 bpm and with >15 bpm increase from baseline
	Abnormally low	<50 bpm and with >15 bpm decrease from baseline
Systolic blood pressure	Abnormally high	>180 mm Hg and with >20 mm Hg increase from baseline
	Abnormally low	<90 mm Hg and with >20 mm Hg decrease from baseline

Vital Sign	Abnormal Category	Criteria
Diastolic blood pressure	Abnormally high	>105 mm Hg and with >15 mm Hg increase from baseline
	Abnormally low	<50 mm Hg and with >15 mm Hg decrease from baseline
Temperature	Abnormally high	>37.5°C
	Abnormally low	<35.5°C
Weight	Abnormally high	increase >7% from baseline
	Abnormally low	decrease >7% from baseline
Respiratory rate	Abnormally high	>20 breaths per minute

In addition, a by-subject listing of the abnormal physical examination data will be presented.

5.6.3.3. Electrocardiogram

The ECG parameters that will be analyzed are heart rate (HR), PR interval, QRS interval, QT interval, and corrected QT(QTc) interval using the following correction methods: Bazett's formula (QTcB), and Fridericia's formula (QTcF).

Bazett's formula: $QTcB\ (msec) = QT\ (msec) * (HR(bpm)/60)^{0.5}$;

Fridericia's formula: $QTcF\ (msec) = QT\ (msec) * (HR(bpm)/60)^{0.33}$

If ECG measurements are repeated at a visit, they will be averaged. The averaged value will considered the 'Visit' ECG result.

Descriptive statistics of ECG parameters and change from baseline will be summarized at each scheduled time point by intervention group.

The number and percentage of participants with categorized QTc interval values will be summarized at each scheduled time point. The number and percentage of participants with categorized QTc interval increases from baseline to the maximum postbaseline value will be summarized. A shift table will be provided summarizing the shift from baseline to maximum QTc interval classification. Refer to Table 5 for summary categories.

Table 5: Criteria for Abnormal QTc Values and Changes from Baseline

QTc value (msec)	Normal QTc	≤450 for male, ≤470 for female
		>450 to ≤480 for male, >470 to ≤480 for female
		>480 to ≤500
		>500
QTc change from baseline (msec)	No concern	≤30
	Concern	>30 to ≤60
	Clear concern	>60

Abnormality criteria (Table 6) will be applied to baseline and postbaseline values. Postbaseline abnormalities will be compared with their corresponding baseline result:

- Treatment-emergent will be concluded if the postbaseline value is above the upper limit and the baseline value is below the upper limit (e.g., Normal or Low). The same applies to the postbaseline value being below the lower limit with the baseline value being above the lower limit (e.g., Normal or High).
- If the baseline value is missing, a postbaseline abnormality will always be considered as treatment-emergent.

The number and percentage of participants with treatment-emergent abnormal postbaseline values (relative to baseline) will be presented by intervention group over time:

Table 6: Abnormal Limits for ECG Parameters

ECG Parameter	Abnormally low	Abnormally high
Heart Rate (bpm)	<50 bpm	>100 bpm
PR interval (msec)	<120 msec	>200 msec
QRS interval (msec)	<60 msec	>120 msec
QT interval (msec)	<200 msec	>500 msec

The interpretation of the ECGs as determined by a qualified physician (investigator or qualified designee) will be displayed by the number and percentage of participants meeting the normality criteria. The interpretation will be summarized over time.

A listing of clinically relevant ECG abnormalities will also be provided.

5.6.3.4. Other Safety Parameters

5.6.3.4.1. Young Mania Rating Scale (YMRS)

The YMRS is one of the most frequently utilized rating scales to assess manic symptoms (Young 1978). The scale has 11 items and is based on the patient's subjective report of his or her clinical condition over the previous 48 hours. Additional information is based on clinical observations made during the clinical interview. The items are selected based on published descriptions of the core symptoms of mania. There are 4 items (irritability, speech, thought content, and disruptive/aggressive behavior) that are graded on a 0 to 8 scale, while the remaining 7 items are graded on a 0 to 4 scale. These 4 items are given twice the weight of the others to compensate for poor cooperation from severely ill patients. There are well-described anchor points for each grade of severity. The YMRS total score, with a range of 0 to 60, is the sum of each of the 11 individual scores, with higher total YMRS scores reflecting greater symptom severity.

Descriptive statistics of the actual values and the change from baseline to each postbaseline time point will be presented for YMRS total score and individual items by intervention group. A listing of YMRS items throughout the study for individual participants at any time point will be provided.

5.6.3.4.2. Columbia Suicide Severity Rating Scale (C-SSRS)

The C-SSRS is a low-burden measure of the spectrum of suicidal ideation and behavior that was developed to assess severity and track suicidal events through any intervention (Posner 2007). It is a clinical interview providing a summary of both suicidal ideation and behavior that can be

administered during any evaluation or risk assessment to identify the level and type of suicidality present. The C-SSRS has been used frequently in clinical studies and is a validated, standard measure for suicidal ideation assessment. Using the C-SSRS, the outcomes will be categorized using the scoring for the 11 categories:

Suicidal Ideation (1-5)	
1	Wish to be dead
2	Non-specific active suicidal thoughts
3	Active suicidal ideation with any methods (not plan) without intent to act
4	Active suicidal ideation with some intent to act, without specific plan
5	Active suicidal ideation with specific plan and intent
Suicidal Behavior (6-10)	
6	Preparatory acts or behavior
7	Aborted attempt
8	Interrupted attempt
9	Actual attempt
10	Suicide
Non-suicidal self-injurious behavior (11)	
11	Non-suicidal self-injurious behavior

If no events qualify for a score of 1 to 10, a score of 0 will be assigned (0 = "no suicidal ideation or behavior that can be assessed on the basis of C-SSRS"). A participant with a score of 11 will be considered as not having suicidal ideation or behavior.

Shifts from baseline to the maximum postbaseline score pertaining to suicidal ideation or suicidal behavior (i.e., scores 1 to 10) will be summarized by intervention group.

The maximum score (of scores 1 to 10) assigned to each participant will be grouped into one of three broad categories: No suicidal ideation or behavior (0), Suicidal ideation (1-5), and Suicidal behavior (6-10). Shifts from baseline to the maximum postbaseline category will be summarized by intervention group.

Frequency distribution of the scores for the 11 categories will be provided at each time point by intervention group.

A listing of C-SSRS items throughout the study for participants with Suicidal Ideation or Behavior at any time point will be provided.

5.7. Other Analyses

5.7.1. Pharmacokinetics

PK analyses will be performed on the PK analysis set, defined as participants who have received at least 1 dose of JNJ-55308942 and have at least 1 valid blood sample drawn for PK analysis.

Based on the individual plasma concentration-time data, using the binned nominal sampling times at Visits 2, 3, 4, 5, and 6, descriptive statistics (N, arithmetic mean, SD, coefficient of variation, geometric mean, median, range) will be used to summarize JNJ-553098942 plasma concentrations. All concentrations below the lowest quantifiable (BLQ) concentration will be treated as zero in the summary statistics. Mean and/or median (+/- SD) plasma concentration-time profiles of JNJ-55308942 will be plotted.

5.8. Interim Analyses

No interim analysis is planned.

6. SUPPORTING DOCUMENTATION

6.1. Appendix 1 List of Abbreviations

AE	adverse event
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
AST	aspartate aminotransferase
ATC	anatomic and therapeutic class
BMI	body mass index
BD	bipolar disorders
CGI-S	Clinical Global Impression – Severity
CI	confidence interval
CRF	case report form
CSR	Clinical Study Report
C-SSRS	Columbia Suicide Severity Rating Scale
DSM-5	Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition
ECG	electrocardiogram
GAD-7	Generalized Anxiety Disorder – 7 Items
GoF	Gain of Function
HDRS-17	Hamilton Depression Rating Scale – 17 Items
HR	heart rate
iARBM	Integrated Analytical Risk-Based Monitoring
IWRS	interactive web response system
LoF	Loss of Function
MADRS	Montgomery-Asberg Depression Rating Scale
MAR	missing at random
MCMC	Markov Chain Monte Carlo
MDE	major depressive episode
MedDRA	Medical Dictionary for Regulatory Activities
MINI	MINI International Neuropsychiatric Interview
MMRM	mixed-effects model for repeated measures
MNAR	missing not at random
mRNA	messenger ribonucleic acid
PD	pharmacodynamic(s)
PHQ-9	Patient Health Questionnaire – 9 Item
PK	pharmacokinetic(s)
PROMIS	Patient-Reported Outcomes Measurement Information System
QTcB	QT interval corrected using Bazett's formula
QTcF	QT interval corrected using Fridericia's formula
QTL	Quality Tolerance Limit
SAE	serious adverse event
SAP	Statistical Analysis Plan
SAWB	Self-Assessment of Well-Being
SD	standard deviation
SHAPS	Snaith-Hamilton Pleasure Scale
SNP	single nucleotide polymorphism
TEAE	treatment-emergent adverse event
ULN	upper limit of normal
VAS	Visual Analog Scale
WHO-DD	World Health Organization Drug Dictionary
YMRS	Young Mania Rating Scale

6.2. Appendix 2 Changes to Protocol-Planned Analyses

Protocol Planned Analysis	**Changes and Rationale**
Analysis for the following secondary objective: To evaluate the efficacy of JNJ-55308942 compared to placebo on symptoms of depression in a subgroup of patients with messenger ribonucleic acid (mRNA) transcript levels at baseline that exceed the median level for both P2RX7 and IL-1β.	This analysis may be conducted in the future when validated assays for mRNA transcript levels of P2RX7 and IL-1β become available. If conducted, the analysis may be reported separately.
Analyses of the patient-reported outcomes including SHAPS, PROMIS – Ability to Participate in Social Roles and Activities, PHQ-9, and GAD-7; analysis of the safety scale YMRS.	Several participants had their Treatment Day 1 assessment of these scales performed postdose instead of predose as per protocol. The assessments performed postdose on Treatment Day 1 will be included in the analyses as the baseline for these participants.

6.3. Appendix 3 Demographics and Baseline Characteristics

The number of participants in each analysis set will be summarized and listed by intervention group, and overall. In addition, the distribution of participants by country and site ID will be presented unless otherwise noted.

Table 7 presents a list of the demographic variables that will be summarized by intervention group and overall for the safety and the full analysis sets.

Table 7: Demographic Variables

Continuous Variables	**Summary Type**
Age (years)	Descriptive statistics (N, mean, standard deviation [SD], median, and range [minimum and maximum]).
Weight (kg)	
Height (cm)	
Body Mass Index (BMI) (kg/m^2)	
Categorical Variables	Frequency distribution with the number and percentage of participants in each category.
Age (18-25 years, 26-50 years, 51-64 years)	
Sex (male, female, undifferentiated)	
Race[a] (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or other Pacific Islander, White, Multiple)	
Ethnicity (Hispanic or Latino, not Hispanic or Latino)	
Country	

[a]If multiple race categories are indicated, the Race is recorded as 'Multiple'

Additional summaries will be presented by the randomization stratification factors (BD type [I or II], country, P2RX7 GoF SNP mutation genotype [homozygous or heterozygous], as well as concomitant medication status [no mood stabilizer or antipsychotic, a mood stabilizer alone, an antipsychotic alone, or a combination of a mood stabilizer and an antipsychotic]) for the full analysis set.

Table 8 presents a list of the baseline disease characteristics variables that will be summarized by intervention group and overall for the safety and the full analysis set.

Table 8: Baseline Disease Characteristics

Continuous Variables	**Summary Type**
Baseline HDRS-17 total score	Descriptive statistics (N, mean, standard deviation [SD], median, and range [minimum and maximum]).
Baseline MADRS total score	
Baseline CGI-S score	
Categorical Variables	Frequency distribution with the number and percentage of participants in each category.
Number of prior nonresponse in the current episode (0, 1, 2, 3, 4)	
BD Type (I or II)	
Baseline CGI-S score (0, 1, 2, 3, 4, 5, 6, 7)	
P2RX7 GoF SNP mutation genotype (homozygous or heterozygous)	
Specific biomarker profile (Yes, No)	
Concomitant medication status (no mood stabilizer or antipsychotic, a mood stabilizer alone, an antipsychotic alone, a combination of a mood stabilizer and an antipsychotic)	

6.4. Appendix 4 Protocol Deviations and Quality Tolerance Limits (QTL)

In general, the following list of major protocol deviations may have the potential to impact participants' rights, safety or well-being, or the integrity and/or result of the clinical study. Participants with major protocol deviations will be identified prior to database lock and the participants with major protocol deviations will be summarized by the following categories.

- Developed withdrawal criteria but not withdrawn
- Entered but did not satisfy criteria
- Received a disallowed concomitant treatment
- Received wrong treatment or incorrect dose
- Other medically important protocol deviations identified during medical and statistical review

Quality Tolerance Limit parameters and thresholds are defined and will be monitored in this study. QTL parameters will be summarized. More details are described in the Integrated Analytical Risk-Based Monitoring (iARBM) Plan (VTMF-17567624, 03 August 2022).

6.5. Appendix 5 Prior and Concomitant Medications

Prior and concomitant medications will be coded using the World Health Organization Drug Dictionary (WHO-DD). Prior medications are defined as any therapy used before the day of the first dose (partial or complete) of study intervention. Concomitant medications are defined as any therapy used on or after the same day as the first dose of study intervention, including those that started before and continued on after the first dose of study intervention.

Summaries of concomitant medications will be presented by ATC term and intervention group. The proportion of participants who receive each concomitant medication will be summarized as well as the proportion of participants who receive at least 1 concomitant medication.

Prior medications will be summarized by ATC term and intervention group.

6.6. Appendix 6 Medical History

A listing of medical history abnormalities will be provided.

6.7. Appendix 7 Adverse Events of Special Interest

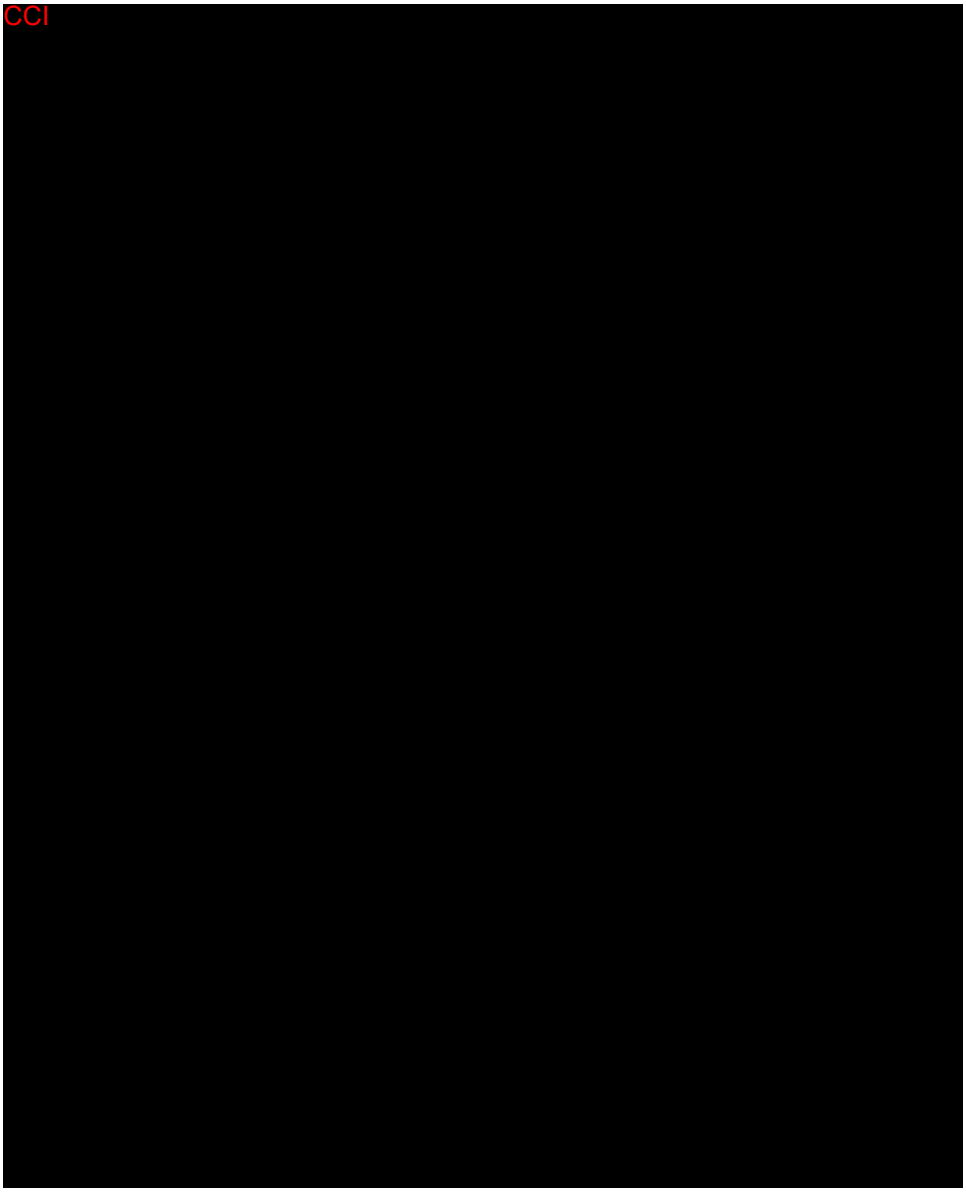
Not applicable

6.8. Appendix 8 Medications of Special Interest

Not applicable.

6.9. Appendix 9 Conversion of Raw Score to T-Score for PROMIS - Ability to Participate in Social Roles and Activities

CCI



7. REFERENCES

Calabrese JR, Keck PE Jr, Macfadden W, Minkwitz M, Ketter TA, Weisler RH, Cutler AJ, McCoy R, Wilson E, Mullen J. A Randomized, Double-Blind, Placebo-Controlled Trial of Quetiapine in the Treatment of Bipolar I or II Depression. Am J Psychiatry. 2005 Jul;162(7):1351-60.

Guy W. CGI: Clinical Global Impressions. In: ECDEU Assessment Manual for Psychopharmacology. Rockville, MD: US Department of Health and Human Services; printed 1976 (reprinted 1991):217-222.

Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry. 1979;134:382-389.

Nakonezny PA, Carmody TJ, Morris DW, Kurian BT, Trivedi MH. Psychometric evaluation of the Snaith-Hamilton pleasure scale in adult outpatients with major depressive disorder. Int Clin Psychopharmacol. 2010;25(6):328-333.

Posner K, Oquendo MA, Gould M, Stanley B, Davies M. Columbia Classification Algorithm of Suicide Assessment (C-CASA): classification of suicidal events in the FDA's pediatric suicidal risk analysis of antidepressants. Am J Psychiatry. 2007;164:1035-1043.

Rubin, D. (1987), Multiple Imputation for Nonresponse in Surveys, Wiley, New York, XXIX+258 pp.

Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. Br J Psychiatry. 1978;133:429-435.