Janssen Research & Development *

Statistical Analysis Plan

A Phase 2a Randomized, Double-blind, Placebo-Controlled, Parallel-Group, Multi-center Study Investigating the Efficacy, Safety, and Tolerability of JNJ-42165279 in Subjects with Social Anxiety Disorder

Protocol 42165279SAX2001; Phase 2a

JNJ-42165279

Status:	Approved
Date:	17 September 2018
Prepared by:	Janssen Research & Development, LLC
Document No.:	EDMS -ERI-99935209

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

TABLE OF CONTENTS

ABBREVIATIONS

AE	adverse event
AEA	anandamide
ANCOVA	analysis of covariance
BMI	body mass index
CGI-I	Clinical Global Impression – Improvement
CRF	case report form
C-SSRS	Columbia Suicide Severity Rating scale
ECG	electrocardiogram
eCRF	electronic case report form
FAAH	fatty acid amide hydrolase
FAAs	fatty acid amides
GAD	generalized anxiety disorder
GAD-7	Generalized Anxiety Disorder 7
HAM-A_6	6-item subscale of the Hamilton Anxiety Rating scale
HAM-D_6	6-item subscale derived from the $HDRS_{17}$
$HDRS_{17}$	17-item Hamilton Depression Rating Scale
HR	heart rate
ICH	International Conference on Harmonization
ITT	Intent-to-Treat
IVRS	interactive voice response system
IWRS	interactive web response system
LSAS	Liebowitz Social Anxiety Scale
MDD	major depressive disorder
MedDRA	Medical Dictionary for Regulatory Activities
MOS Sleep-R	Medical Outcomes Study- 12-item Sleep Scale Acute - Revised
OEA	oleoylethanolamide
PD	pharmacodynamic
PEA	palmitoylethanolamide
PK	pharmacokinetic(s)
Q-LES-Q	Quality of Life, Enjoyment, and Satisfaction Questionnaire
SAD	social anxiety disorder
SAE	serious adverse event
SAP	statistical analysis plan
SD	standard deviation
SDS	Sheehan Disability Scale
SHAPS	Snaith-Hamilton Pleasure Scale
SI	standard international
SIGH-A	Structured Interview Guide for the Hamilton Anxiety Scale
TEAE	treatment-emergent adverse event
TEMA	treatment-emergent abnormality

1. INTRODUCTION

In January 2016, Janssen had decided to suspend the 42165279SAX2001 study (see the press release on 18 January 2016 on Janssen.com). The study was unblinded and the randomization codes were provided to the investigators so that the patients could be informed about their treatment. Safety information was analyzed, and a separate Statistical Analysis Plan (SAP) was written and approved for the interim safety analysis. No efficacy data was analyzed, and datasets containing efficacy information of the subjects who were unblinded at trial suspension will remain in a secure location at the data management department until final DB lock (see Partial Database Lock Notification form dated 24 March 2016 and Note to File dated 15 December 2016).

This SAP contains definitions of analysis sets, derived variables, and statistical methods for the final analysis of efficacy, biomarker and safety data. The final analysis will include subjects who were enrolled prior to study suspension and subjects who were enrolled after the study had resumed. See Section 2.5 for a definition of the analysis sets.

1.1. Trial Objectives

The primary objective of this study is to investigate the efficacy of JNJ-42165279 during 12 weeks of treatment in subjects with social anxiety disorder (SAD).

The secondary objectives of this study are:

- To assess the safety and tolerability of JNJ-42165279 in subjects with SAD.
- To assess the plasma pharmacokinetic (PK) profile of JNJ-42165279 administered as once daily (*qd)* in subjects with SAD using a population PK approach, and explore the relationship between exposure to JNJ-42165279 and efficacy and safety parameters.

The exploratory objectives are:

- To assess the efficacy of JNJ-42165279 on both anxiety and depression symptoms.
- To evaluate the impact of treatment with JNJ-42165279 compared to placebo on patient-reported assessments of symptoms of anxiety, depression, impairment in daily living and quality of life.
- To evaluate pharmacodynamic (PD) effects by the assessment of biomarkers of peripheral pharmacological activity after repeated doses of JNJ-42165279, including assessment of plasma concentrations of fatty acid amides (FAAs) (anandamide [AEA], palmitoylethanolamide [PEA], and oleoylethanolamide [OEA]) that are expected to rise as a consequence of the inhibition of their hydrolysis by fatty acid amide hydrolase (FAAH).
- To explore the relationship between plasma PK and plasma concentrations of FAAs (AEA, PEA and OEA) in subjects with SAD.
- To explore any potential differences between healthy subjects and subjects with SAD using a population PK approach.

1.2. Trial Design

This is a multi-center, double-blind, placebo-controlled, randomized, parallel-group study assessing the efficacy, safety, and tolerability of JNJ-42165279 during 12 weeks of treatment in subjects with SAD.

Approximately one-hundred twenty-two (122) subjects will be enrolled in this 12 week treatment study randomly assigned in a 1:1 ratio to either 25 mg of JNJ-42165279 or placebo (dosed once daily orally).

For all enrolled subjects, this study will consist of a 28-day eligibility screening period, a 12 week double-blind treatment period and a follow-up examination between 7 and 28 days after last dose. The study duration for each subject will be maximally 20 weeks.

The study will be an outpatient study.

1.3. Statistical Hypotheses for Trial Objectives

The primary efficacy endpoint for this study is the reduction of symptom burden associated with social anxiety disorder (SAD), as measured by the change from baseline to the 12-week endpoint in the Liebowitz Social Anxiety Scale (LSAS) total score. The null hypothesis to be tested to address the primary objective is that there is no difference between JNJ42165279 and placebo based on the primary efficacy endpoint.

1.4. Sample Size Justification

The sample size for the study is based on the assumption of a treatment difference of at least 10 points in the mean change from baseline to the endpoint in LSAS total score between JNJ-42165279 treatment group and placebo. A standard deviation of 24 in the change in LSAS total score from baseline is used based on published data.[7,10,11] To detect the treatment difference of 10 points (which is judged to be clinically relevant [7,10,11]) with a power of 90% at an overall 1-sided significance level of 0.20, 53 subjects in each group are required. When adjusted for a drop-out rate of approximately 15% of subjects, the required number of subjects is 61 per treatment group.

The impact of this sample size (N=122) on the JNJ– 42165279 effect size to be detected for the other continuous secondary and exploratory endpoints for different values of significance level and power, assuming a drop-out rate of 15%, are presented in Table 1.

Table 1: JNJ–42165279 effect size to be detected for different values of significance level and power, assuming a drop-out rate of 15%

1-sided significance level	Power	Effect size to be detected
0.10	80%	0.4
0.10	90%	0.5
0.20	80%	0.3
0.20	90%	0.4

1.5. Randomization and Blinding

Central randomization will be implemented in this study. At the start of the double-blind phase, subjects will be randomly assigned to one of two treatment groups based on the first of

two computer-generated randomization schedules prepared before the study by, or under the supervision of the sponsor. The randomization will be balanced by using randomly permuted blocks. Presence of comorbid major depressive disorder (MDD) and country will be used as stratification factors.

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 IVRS/IWRS, which has the functionality to allow the investigator to break the blind for an individual subject.

Data that may potentially unblind the treatment assignment (e.g., study medication 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.

Under normal circumstances, the blind should not be broken until all subjects have completed the study and the database is finalized. Otherwise, the blind should be broken only if specific emergency treatment/course of action would be dictated by knowing the treatment status of the subject. In such cases, the investigator may in an emergency determine the identity of the treatment by contacting the IVRS/IWRS. It is recommended that the investigator contact the sponsor or its designee if possible to discuss the particular situation, before breaking the blind. In the event the blind is broken, the sponsor must be informed as soon as possible. The date, time, and reason for the unblinding must be documented in the appropriate section of the case report form (CRF), and in the source document. The documentation received from the IVRS indicating the code break must be retained with the subject's source documents in a secure manner.

Subjects who have had their treatment assignment unblinded should continue to return for required follow up evaluations.

In general, randomization codes will be disclosed fully only if the study is completed and the clinical database is closed. However, if an interim analysis is specified, the randomization codes and, if required, the translation of randomization codes into treatment and control groups will be disclosed to those authorized and only for those subjects included in the interim analysis.

2. GENERAL ANALYSIS DEFINITIONS

2.1. Treatment Groups

Tables summarizing data in the double-blind treatment phase will have columns for 'Placebo' and 'JNJ-42165279 25 mg'.

Subjects who received an incorrect treatment will be analyzed under the treatment they were randomized to.

2.2. Pooling Algorithm for Analysis Centers

No pooling of centers will be carried out.

2.3. Other General Definitions

2.3.1. Study Day 1 or Day 1

Study Day 1 or Day 1 refers to the first day of double-blind study drug administration. All efficacy and safety assessments at all visits will be assigned a day relative to this date.

2.3.2. Study Day or Relative Day for a Visit

Study day or relative day for a visit is defined as:

- Visit date - Date of first dose of study medication (date of Day 1) +1, if visit date is ≥ date of Day 1
- Visit date - Date of Day 1, if visit date is < date of Day 1

By definition there is no 'Day 0'.

2.3.3. Baseline

Baseline is defined as the last non-missing observation prior to first dose of study drug.

2.3.4. End Point

End point is defined as the last non-missing post-baseline observation during the Double-Blind treatment phase (defined in Section 2.3.5).

2.3.5. Analysis Phases

All data collected from the Baseline visit to the End Point/Early Withdrawal visit (inclusive) will be assigned to the Double-Blind Treatment analysis phase. Data collected prior to the Baseline visit will be assigned to the Screening analysis phase, and data collected after the End Point/Early Withdrawal visit will be assigned to the Post Treatment analysis phase.

2.4. Visit Windows

As subjects do not always adhere to the protocol visit schedule, the following rules are applied to assign actual visits to protocol visits. Listed in Table 2 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 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. If a visit window has no scheduled visits but does have unscheduled visits, then the unscheduled visit closest to the scheduled visit will be used.

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

Table 2: Visit Windows

Parameter	Scheduled Visit Number	Time Interval (label on output)	Time Interval (Day)[1]	Target Time Point (Day)	Analysis Phase[5]
MINI Interview, Height, Serology	1	Screening	<1	-28 to -1	Screening
LSAS, $HDRS_{17}$	1	Screening	<1	-28 to -1	Screening
	2	Baseline	≤1	1	DB Treatment
	3	Week 1	2 to 10	7	DB Treatment
	4	Week 2	11 to 21	14	DB Treatment
	5	Week 4	22 to 42	28	DB Treatment
	7	Week 8	43 to 70	56	DB Treatment
	9	Week 12	71 to End of DB[4]	84	DB Treatment
		End Point[2]			DB Treatment
SIGH-A	2	Baseline	≤1	1	DB Treatment
	3	Week 1	2 to 10	7	DB Treatment
	4	Week 2	11 to 21	14	DB Treatment
	5	Week 4	22 to 42	28	DB Treatment
	7	Week 8	43 to 70	56	DB Treatment
	9	Week 12	71 to End of DB[4]	84	DB Treatment
		End Point[2]			DB Treatment
CGI - I	3	Week 1	2 to 10	7	DB Treatment
	4	Week 2	11 to 21	14	DB Treatment
	5	Week 4	22 to 42	28	DB Treatment
	7	Week 8	43 to 70	56	DB Treatment
	9	Week 12	71 to End of DB[4]	84	DB Treatment
		End Point[2]			DB Treatment
SDS, GAD-7, SHAPS, MOS Sleep – R, Q-LES-Q	2	Baseline	≤1	1	DB Treatment
	5	Week 4	2 to 42	28	DB Treatment
	7	Week 8	43 to 70	56	DB Treatment
	9	Week 12	71 to End of DB[4]	84	DB Treatment
		End Point[2]			DB Treatment
Self-assessment of reatment experience	9	Week 12	2 to End of DB[4]	84	DB Treatment
		End Point[2]			DB Treatment
12-Lead ECG, Body Weight	1	Screening	<1	-28 to -1	Screening
	2	Baseline	≤1	1	DB Treatment
	5	Week 4	2 to 56	28	DB Treatment
	9	Week 12	57 to End of DB[4]	84	DB Treatment
		End Point[2]			DB Treatment
	10	Follow Up[3]	> End of DB[4]	91 to 112	Post Treatment
Vital Signs, Body temperature	1	Screening	<1	-28 to -1	Screening
	2	Baseline	≤1	1	DB Treatment
	3	Week 1	2 to 10	7	DB Treatment
	4	Week 2	11 to 21	14	DB Treatment
	5	Week 4	22 to 42	28	DB Treatment
	7	Week 8	43 to 70	56	DB Treatment
	9	Week 12	71 to End of DB[4]	84	DB Treatment
		End Point[2]			DB Treatment
	10	Follow Up	> End of DB[4]	91 to 112	Post Treatment
CSSR-S	1	Screening	<1	-28 to -1	Screening
	2	Baseline	≤1	1	DB Treatment
	3	Week 1	2 to 10	7	DB Treatment
	4	Week 2	11 to 21	14	DB Treatment
	5	Week 4	22 to 35	28	DB Treatment
	6	Week 6	36 to 49	42	DB Treatment
	7	Week 8	50 to 63	56	DB Treatment
	8	Week 10	64 to 77	70	DB Treatment

Table 2: Visit Windows

Parameter	Scheduled Visit Number	Time Interval (label on output)	Time Interval (Day)[1]	Target Time Point (Day)	Analysis Phase[5]
	9	Week 12 End Point[2]	78 to End of DB[4]	84	DB Treatment DB Treatment
Clinical Laboratory Assessments	1	Screening	<1	-28 to -1	Screening
	2	Baseline	≤1	1	DB Treatment
	4	Week 2	2 to 21	14	DB Treatment
	5	Week 4	22 to 35	28	DB Treatment
	6	Week 6	36 to 49	42	DB Treatment
	7	Week 8	50 to 63	56	DB Treatment
	8	Week 10	64 to 77	70	DB Treatment
	9	Week 12 End Point[2]	78 to End of DB[4]	84	DB Treatment DB Treatment
	10	Follow Up[3]	> End of DB[4]	91 to 112	Post Treatment
Urine Drug Screen	1	Screening	<1	-28 to -1	Screening
	2	Baseline	≤1	1	DB Treatment
	5	Week 4	2 to 56	28	DB Treatment
	9	Week 12 End Point[2]	57 to End of DB[4]	84	DB Treatment DB Treatment
Alcohol Screen	1	Screening	<1	-28 to -1	Screening
	2	Baseline	≤1	1	DB Treatment
	4	Week 2	2 to 21	14	DB Treatment
	5	Week 4	22 to 42	28	DB Treatment
	7	Week 8	43 to 70	56	DB Treatment
	9	Week 12 End Point[2]	71 to End of DB[4]	84	DB Treatment DB Treatment
	10	Follow Up[3]	> End of DB[4]	91 to 112	Post Treatment
Pregnancy test	1	Screening	<1	-28 to -1	Screening
	2	Baseline	≤1	1	DB Treatment
	9	Week 12 End Point[2]	2 to End of DB[4]	84	DB Treatment DB Treatment
	10	Follow Up		91 to 112	Post Treatment
Physical Examination	1	Screening	<1	-28 to -1	Screening
	9	Week 12 End Point[2]	2 to End of DB[4]	84	DB Treatment DB Treatment
	10	Follow Up	> End of DB[4]	91 to 112	Post Treatment

[1] Relative to first day of study medication
[2] End Point is defined as the last non-missing post-baseline measurement during the double-blind treatment phase.
[3] Only in case of any clinical significant abnormality observed at Week 12 (except for weight)
[4] End of DB = the end of the double-blind treatment phase
[5] On output tables, DB Treatment will be presented as "Double-Blind Treatment".

2.5. Analysis Sets

2.5.1. All Subjects Analysis Set

The All Subjects analysis set includes all subjects with information entered into the clinical database. This set will be used for the listings unless specified otherwise.

2.5.2. All Randomized Analysis Set

The All Randomized analysis set includes all subjects who were randomized (i.e., subjects who reported a randomization date, or were assigned a randomization number) regardless of whether or not the treatment was received.

This analysis set will be used for summarizing the overall study completion/withdrawal information.

2.5.3. Intent-to-Treat (ITT) Analysis Set

The ITT analysis set is defined as all randomized subjects who received at least one dose of study medication (either placebo or JNJ-42165279), and have at least one assessment in the double-blind treatment phase on any of the efficacy parameters. Subjects who withdrew early from the study at the time of study suspension will not be included in the ITT analysis set (7 placebo subjects, 8 JNJ-42165279 subjects). Subjects who had already completed the study at the time it was suspended, however, will be included in the ITT set (9 placebo subjects, 9 JNJ-42165279 subjects).

The ITT analysis set is the analysis set for the efficacy analyses.

2.5.4. Safety Analysis Set

The Safety analysis set is defined as all randomized subjects who received at least one dose of study medication (either placebo or JNJ-42165279). Screen failures and randomized subjects who received no double-blind study medication will be excluded from the Safety analysis set.

The Safety analysis set is the analysis set for the safety analyses.

2.5.5. Per Protocol Analysis Set

The Per Protocol analysis set includes all ITT subjects who did not have a major protocol deviation during the study. In case there is a substantial amount of subjects with a major protocol deviation, a sensitivity analysis of the primary efficacy parameter will be performed for the per protocol analysis set.

2.6. Definition of Subgroups

Some of the secondary efficacy endpoints will be analyzed for the subgroup of subjects with comorbid MDD/GAD. See Section 5.4 for a definition of these exploratory endpoints.

Exploratory analyses of selected parameters will be performed for the subgroup of subjects who enrolled prior to study suspension versus those who enrolled after the study had resumed.

Exploratory sensitivity analyses of the primary efficacy parameter will be performed for the subgroup of subjects who did not meet criteria for sufficient compliance based on pharmacokinetic results. These criteria are defined by the clinical pharmacology department.

3. INTERIM ANALYSIS AND DATA MONITORING COMMITTEE REVIEW

A blinded interim analysis for the purpose of sample size re-estimation may be performed after 50% of the subjects are recruited if the observed standard deviation substantially deviates from the hypothesized standard deviation, or if the drop-out rate substantially deviates from the assumed drop-out rate.

No interim analysis for efficacy is planned.

4. SUBJECT INFORMATION

Subject information will be summarized for the Safety Analysis Set and listings will be created for the All Randomized analysis set.

4.1. Demographics and Baseline Characteristics

The following demographic and baseline characteristics will be listed and summarized by treatment group and overall:

> *Age (years), sex, race, ethnicity, weight (kg), height (cm), body mass index (BMI, calculated as weight (kg)/[height(m)]^2), country, presence of comorbid major depressive disorder (MDD), presence of comorbid generalized anxiety disorder (GAD).*

Continuous variables (age, weight, height and BMI) will be summarized using descriptive statistics (N, mean, standard deviation [SD], median and range [minimum and maximum]). Categorical variables (race, ethnicity, country, presence of comorbid MDD, presence of comorbid GAD) will be summarized using a frequency distribution with the number and percentage of subjects in each category.

Psychiatric history at baseline will be summarized by treatment group and overall using descriptive statistics for the following assessments: LSAS total score, SIGH-A total score, HAM-A_6 score, $HDRS_{17}$ total score, $HDRS_{17}$ anxiety/somatization factor score, HAM-D_6 score.

The summary of demographics and baseline characteristics and psychiatric history at baseline will be created for both the Safety Analysis Set and the ITT Analysis Set. The summary will also be presented by country and by enrollment prior to/after study suspension.

Listings will be created for the following assessments:

> *Randomization information, inclusion/exclusion criteria, childbearing potential, date of last menses, date of signature on informed consent, issue date of the protocol the subject consented to at study enrollment, preplanned surgery/procedures, psychiatric evaluation (M.I.N.I.), medical history, comments, study execution period.*

4.2. Disposition Information

The number and percentage of subjects in the All Randomized Subjects, ITT, Per Protocol and Safety analysis sets will be presented overall, and by country and center. In addition, summaries of the reasons for exclusion from each analysis set (i.e., Per Protocol set and ITT set) will be provided. The number of screen failures together with the withdrawal reason will be presented.

The distribution of the number of subjects who complete the entire course of treatment and who discontinue will be presented by treatment group and overall. The distribution of study

termination reasons will also be presented. In addition, the number and percentage of subjects who discontinue on or prior to each study visit will be summarized by treatment group and overall. The distribution of time to discontinuation will be presented graphically using Kaplan-Meier estimates. The time to treatment discontinuation will be calculated as the duration (in days) from Study Day 1 (defined in Section 2.3.1) to treatment discontinuation. Subjects who complete the double-blind treatment period will be censored at the end of treatment date.

Study completion and withdrawal information for each subject will also be presented in a listing.

4.3. Treatment Compliance

The number of tablets dispensed and returned at each visit will be listed.

4.4. Extent of Exposure

Study drug administration information will be presented for each subject in a data listing.

Treatment duration is defined as the number of days between the first and the last day of study drug in the double-blind treatment phase: last day of study drug administration – first day of study drug administration + 1.

Treatment duration will be listed and summarized by treatment group using descriptive statistics (N, mean, SD, median, and range). A frequency distribution will also be provided using the following categories: <1 week (<7 days), 1-<2 weeks (7-<14 days), 2-<4 weeks (14-<28 days), 4–<6 weeks (28-< 42 days), 6-<8 weeks (42-<56 days), 8-<10 weeks (56-< 70 days), 10-<12 weeks (70-<84 days) and ≥12 weeks (≥84 days).

4.5. Protocol Deviations

Deviations that occurred during the study will be listed by treatment group. Major deviations will be presented as grouped by the Data Management Group prior to the unblinding.

4.6. Prior and Concomitant Medications

The following groups of medication will be tabulated by treatment group, by medication class and generic term of the medication:

- Prior medications: medications with start date/time and end date/time prior to first dose intake
- Prior and ongoing medications: medications with start date/time prior to first dose intake and end date/time after first dose intake
- Concomitant medications: medications with start date/time after first dose intake

Data listings will be provided for all medications received from the signing of informed consent prior to study start and concomitantly during the study.

5. EFFICACY

All efficacy analyses will be based on the ITT analysis set. The efficacy variables in this study are listed in Table 3.

All efficacy data will be listed for the subjects in the ITT analysis set.

Table 3: Efficacy Variables

Endpoint	Scale*
Primary	LSAS total score
Secondary	LSAS fear/anxiety subscale
	LSAS avoidance subscale
	LSAS ≥50% improvement from baseline on total score
	LSAS ≥30% improvement from baseline on total score
	SIGH-A total score
	$HAM\text{-}A_6$ score
	$HDRS_{17}$ total score
	$HDRS_{17}$ anxiety/somatization factor score
	$HAM\text{-}D_6$ score
	CGI-I
	SIGH-A ≥50% improvement from baseline on total score
Exploratory	SDS
	GAD-7
	SHAPS
	MOS Sleep - R
	Q-LES-Q
	Self-Assessment of Treatment Experience
	$HDRS_{17}$ total score in subjects with comorbid MDD
	$HDRS_{17}$ anxiety/somatization factor score in subjects with comorbid MDD
	$HAM\text{-}D_6$ score in subjects with comorbid MDD
	SIGH-A total score in subjects with comorbid GAD
	$HAM\text{-}A_6$ score in subjects with comorbid GAD
	$HDRS_{17}$ ≥50% improvement from baseline on total score in subjects with comorbid MDD
	$HDRS_{17}$ ≥30% improvement from baseline on total score in subjects with comorbid MDD
	SIGH-A ≥50% improvement from baseline on total score in subjects with comorbid GAD

* CGI-I = Clinical Global Impression – Improvement; GAD-7 = Generalized Anxiety Disorders; $HAM\text{-}A_6$ = Hamilton Anxiety Rating Scale; $HAM\text{-}D_6$ = Hamilton Depression Rating Scale; $HDRS_{17}$ = Hamilton Depression Rating Scale; LSAS = Liebowitz Social Anxiety Scale; MOS Sleep – R: Medical Outcomes Study Sleep-Revised; SDS = Sheehan Disability Scale; Q-LES-Q = Quality of Life Enjoyment and Satisfaction Questionnaire; SHAPS = Snaith-Hamilton Pleasure Scale; SIGH-A = Structured Interview Guide for the Hamilton Anxiety Scale

5.1. Analysis Specifications

5.1.1. Level of Significance

The overall Type I error rate for testing the JNJ-42165279 group versus the placebo group for the primary efficacy analysis will be controlled at the 1-sided significance level of 0.20.

5.1.2. Imputation Methods for Missing Items

Imputation of missing individual items will apply only to the LSAS, as described in Section 5.2.1. For all other scales where multiple items are summed to create a total score, if any item of the scale is missing on a visit, the total score for that scale at that visit will be left blank.

5.2. Primary Efficacy Endpoint – LSAS Total Score

5.2.1. Definition

The primary efficacy endpoint will be the improvement in social anxiety symptoms, as measured by the change from baseline in the LSAS total score from baseline to the 12-week endpoint in the double-blind treatment phase.

The LSAS was designed to assess the range of social interaction and performance situations that individuals with social phobia may fear and/or avoid (Liebowitz, 1987)[6]. The 24 items in the scale are divided into two subscales that address social interactional (11 items) and performance (13 items) situations. The clinician asks the patient to rate fear and avoidance during the past week on 0–3 Likert-type scales; however, the clinician is given latitude to question the patient's responses and adjust the ratings accordingly. An example of the LSAS is provided in the Attachment.

An overall total score is calculated by summing the 24 fear and 24 avoidance scores. The LSAS total score has a maximum of 144, with higher scores indicating higher probability of social anxiety disorder.

If an LSAS item is missing, it will be imputed with the closest integer to the average of the remaining items within the LSAS subscale (fear/anxiety, avoidance) at that time point and then the total will be summed. If more than 1 item is missing, no imputation will be performed and the total LSAS score will be left missing. Imputation of item scores is performed prior to determining the endpoint for the LSAS.

5.2.2. Analysis Methods

5.2.2.1. Primary Analysis

Descriptive statistics of actual values and changes from baseline for the LSAS total score at each scheduled time point and endpoint will be provided by treatment group (mean, SD, median and range). A mean-SE plot over time will be presented for the LSAS total score observed values by treatment group. Listings will be created with LSAS item scores and the LSAS total score.

The JNJ-42165279 treatment group will be compared to placebo by means of a mixed-effects model using repeated measures (MMRM) with time, treatment (placebo, JNJ-42165279) and time-by-treatment interaction as factors, baseline LSAS total score and age as a continuous covariate and country and presence of comorbid MDD as categorical covariates. An unstructured variance-covariance matrix will be used. In case of convergence problems, simpler variance-covariance structures such as Toeplitz or AR(1) will be considered. The selection of any of these structures will be determined after exploration of the observed

correlation structure. The treatment-placebo differences will be obtained using the appropriate contrast in the MMRM model at the 12-week endpoint.

A one-sided p-value with level of significance of 20% will be derived from the SAS results based on the above mixed-effects model (two-sided p-values at 40% significance level). The reduction in LSAS total score is expected to be larger after dosing with JNJ, so the one-sided p-value will be calculated as described in Table 4.

Table 4: Calculation of one-sided p-value

Parameter	H_a	Estimated difference Mean $_{JNJ}$ – Mean $_{Pla}$	One-sided p-value
Change in LSAS total score at Week 12	Mean Chg $_{JNJ}$- Mean Chg $_{Pla}$< 0 (larger reduction expected in JNJ vs pla)	< 0	p-value $_{(2\text{-sided test})}$ / 2
		> 0	1 – (p-value $_{(2\text{-sided test})}$ / 2)

5.2.2.2. Additional Analysis

To assess the sensitivity of the results of the MMRM analysis of the primary endpoint, an analysis of covariance (ANCOVA) model for the change from baseline to the 12-week endpoint in LSAS total score will be carried out. The ANCOVA model will include factors for treatment, country and presence of comorbid MDD and baseline LSAS total score and age as a covariate. In addition, the same ANCOVA model will also be performed on observed case data at Week 12.

As a secondary analysis, the MMRM model described in Section 5.2.2.1 will be fitted with race, history of MDD and baseline SDS total score as additional covariates.

As a sensitivity analysis, the MMRM model described in Section 5.2.2.1 will be fitted on the set of ITT subjects excluding the subjects who had completed the study at the time it was suspended (9 subjects in each treatment group).

5.3. Secondary Efficacy Endpoints

5.3.1. Definitions

5.3.1.1. LSAS fear/anxiety subscale

The LSAS fear/anxiety subscale will be calculated by summing the 24 fear/anxiety item scores of the LSAS, and ranges from 0 to 72.

5.3.1.2. LSAS avoidance subscale

The LSAS avoidance subscale will be calculated by summing the 24 avoidance item scores of the LSAS, and ranges from 0 to 72.

5.3.1.3. LSAS ≥50% improvement from baseline on total score

The percentage change from baseline is calculated as

$$100 * \frac{(LSAS\ total\ score\ at\ any\ time\ point - LSAS\ total\ score\ at\ baseline)}{LSAS\ total\ score\ at\ baseline}$$

A ≥50% improvement on LSAS total score means that the percentage change from baseline is ≤ -50 in this calculation.

5.3.1.4. LSAS ≥30% improvement from baseline on total score

The percentage change from baseline is calculated as

$$100 * \frac{(LSAS\ total\ score\ at\ any\ time\ point - LSAS\ total\ score\ at\ baseline)}{LSAS\ total\ score\ at\ baseline}$$

A ≥30% improvement on LSAS total score means that the percentage change from baseline is ≤ -30 in this calculation.

5.3.1.5. Structured Interview Guide for the Hamilton Anxiety scale (SIGH-A) total score

The SIGH-A is included as a means to determine the frequency and severity of signs and symptoms of anxiety, including subjects with comorbid GAD and MDD, and determine both their influence on treatment and their responsiveness to treatment.

The original HAM-A scale assesses the severity of different anxiety-related symptoms (Hamilton 1959[4]; Hamilton 1969[3]). As the original HAM-A lacks instructions for administration and clear anchor points for the assignment of severity ratings, the structured interview guide version will be used in the current study (Shear 2001)[9]. The SIGH-A has been shown to have high inter-rater and test-retest reliability and produced similar but consistently higher (+ 4.2) scores compared to the original HAM-A. Correlation with a self-report measure of overall anxiety has also been shown to be high (Shear 2001)[9]. Subscales, such as the HAM-A_6 which focuses on psychic anxiety and may be more sensitive to certain treatments, can be derived from the SIGH-A.

The SIGH-A scale consists of 14 items with a score of 0 to 4. Higher scores indicate higher severity (0-absent, 1-mild, 2-moderate, 3-severe, 4-incapacitating). The SIGH-A total score will be calculated by summing the 14 item scores, and ranges from 0 to 56. Higher scores indicate worse results.

5.3.1.6. HAM-A_6 score

The HAM-A_6 is a 6-item subscale derived from the original Hamilton Anxiety scale (HAM-A). It comprises five psychic anxiety symptoms: anxious mood, psychic tension, fears, intellectual disturbances, and anxious behavior observed at the interview, as well as one somatic item, muscular tension. The HAM-A_6 score will be calculated by summing the 6 item scores, and ranges from 0 to 24. Higher scores indicate greater severity of symptoms.

5.3.1.7. $HDRS_{17}$ total score

The $HDRS_{17}$ is a clinician-administered rating scale designed to assess the severity of symptoms in subjects diagnosed with depression (Hamilton M 1960)[2]. It is the most widely

used symptom severity measure for depression. Each of the 17 items is rated by the clinician on either a 3- or a 5-point scale.

The $HDRS_{17}$ total score will be calculated by summing the 17 item scores, and ranges from 0 to 52. Higher scores indicate greater severity of depression.

The original $HDRS_{17}$ scale lacks instructions for administration and clear anchor points for the assignment of severity ratings. For this reason, the structured interview guide version of the $HDRS_{17}$ (the Structured Interview Guide for the Hamilton Depression Scale [SIGH-D]) will be used in the current study to facilitate and standardize gathering clinical information from the subject.

5.3.1.8. $HDRS_{17}$ anxiety/somatization factor score

The $HDRS_{17}$ anxiety/somatization factor derived from Cleary and Guy's factor analysis of the $HDRS_{17}$ scale, includes six items from the original 17-item version: the items for psychic anxiety, somatic anxiety, gastrointestinal somatic symptoms, general somatic symptoms, hypochondriasis, and insight. The $HDRS_{17}$ anxiety/somatization factor will be calculated as the sum of the 6 item scores, and ranges from 0 to 18, with higher scores indicating greater severity of symptoms.

5.3.1.9. HAM-D_6 score

A 6-item subscale from the $HDRS_{17}$ (HAM-D_6) will be analyzed as it has been shown to be a uni-dimensional scale that provides information to core depressive symptoms and is sensitive to treatment response (Bech 1975)[1]. The six items are: depressed mood, guilt feelings, work and interests, psychomotor retardation, psychic anxiety, and general somatics (tiredness and pains). The HAM-D_6 score will be calculated by summing the 6 items scores, and ranges from 0 to 22. Higher scores indicate greater severity of core symptoms.

5.3.1.10. CGI-I

The clinical global impression – improvement (CGI-I) is a 7-point scale that requires the clinician to assess how much the subject's illness has improved or worsened relative to a baseline state at the beginning of the intervention. The CGI-I is rated as: 1=very much improved; 2=much improved; 3=minimally improved; 4=no change; 5=minimally worse; 6=much worse; 7=very much worse.

Responders on the CGI-I are defined as subjects with a score of 1=very much improved or 2=much improved.

5.3.1.11. SIGH-A ≥50% improvement from baseline on total score

The percentage change from baseline is calculated as

$$100 * \frac{(SIGH\ A\ total\ score\ at\ any\ time\ point - SIGH\ A\ total\ score\ at\ baseline)}{SIGH\ A\ total\ score\ at\ baseline}$$

A ≥50% improvement on SIGH-A total score means that the percentage change from baseline is ≤ -50 in this calculation.

5.3.2. Analysis Methods

For all continuous secondary endpoints, descriptive statistics of actual values and changes from baseline at each scheduled time point and endpoint will be provided by treatment group (mean, SD, median and range). This includes LSAS fear/anxiety subscale, LSAS avoidance subscale, SIGH-A total score, HAM-A_6 score, $HDRS_{17}$ total score, $HDRS_{17}$ anxiety/somatization factor score and HAM-D_6 score. A mean-SE plot over time will be presented for these continuous endpoint observed values by treatment group.

For each of the continuous secondary endpoints, the JNJ-42165279 treatment group will be compared to placebo by means of a mixed-effects model using repeated measures (MMRM) with time, treatment (placebo, JNJ-42165279) and time-by-treatment interaction as factors, baseline score of the continuous endpoint and age as a continuous covariate and country and presence of comorbid MDD as categorical covariates. An unstructured variance-covariance matrix will be used. In case of convergence problems, simpler variance-covariance structures such as Toeplitz or AR(1) will be considered. The selection of any of these structures will be determined after exploration of the observed correlation structure. The treatment-placebo differences will be obtained using the appropriate contrast in the MMRM model at the 12-week endpoint. One-sided p-values with level of significance of 20% will derived as described in Section 5.2.2.1.

Frequency counts and percentages will be provided at each scheduled time point and endpoint by treatment group for the following variables: LSAS ≥50% improvement, LSAS ≥30% improvement, CGI-I, CGI-I response and SIGH-A ≥50% improvement.

5.4. Exploratory Efficacy Endpoints

5.4.1. Definition

5.4.1.1. SDS

The Sheehan Disability Scale (SDS) is a composite of three self-rated items designed to measure the extent to which three major sectors in the patient's life are impaired by panic, anxiety, phobic, or depressive symptoms. This anchored visual analog scale uses visual-spatial, numeric, and verbal descriptive anchors simultaneously to assess disability across three domains: work, social life, and family life. An SDS total score will be calculated as the sum of the three self-rates item scores, and ranges from 0 to 30. Higher scores indicates higher impairment of the patient's life.

5.4.1.2. GAD-7

Generalized Anxiety Disorder 7 (GAD-7) is a self-reported questionnaire for screening and severity measuring of generalized anxiety disorder (GAD) and is of particular interest because of the frequent comorbidity of GAD with SAD. GAD-7 has seven items, which measure severity of various signs of generalized anxiety disorder according to reported response categories of "not at all," "several days," "more than half the days," and "nearly every day." A GAD-7 total score is calculated as the sum of the 7 item scores, and ranges from 0 to 21, with higher scores indicating higher severity of symptoms.

5.4.1.3. SHAPS

The Snaith-Hamilton Pleasure Scale (SHAPS) is a 14-item, self report instrument developed for the assessment of hedonic capacity. The SHAPS was developed to minimize cultural, gender, and age biases in the evaluation of hedonic capacity. It not only measures hedonic tone, but also its absence, i.e. anhedonia. Anhedonia can be a core symptom of depression. Four major domains are covered in the scale, namely interest/pastimes, social interaction, sensory experience, and food/drink. Each of the SHAPS items has a set of four response categories: Strongly Disagree, Disagree, Agree and Strongly Agree, with either of the Agree responses receiving a score of 0 and either of the Disagree responses receiving a score of 1. A SHAPS total score will be calculated as the sum of the 14 item scores, and ranges from 0 to 14. A higher SHAPS total score indicates higher levels of present state of anhedonia.

5.4.1.4. MOS Sleep – R

Symptoms of poor sleep commonly occur in anxiety and mood disorders. The Medical Outcomes Study Sleep-Revised (MOS Sleep-R) is a subject-completed scale containing 12 items that addresses various dimensions of sleep. Ten items are answered on a 6-point scale, where 1="all of the time" and 6="none of the time". One item on sleep latency is answered on a 5 point Likert scale from 1="0-15 minutes" to 5="more than 60 minutes". One item on the duration of sleep allows the subject to write in the number of hours slept per night.

The instrument yields six subscales: sleep disturbance, snoring, shortness of breath or headache, sleep adequacy, sleep somnolence, and sleep quantity. The version to be used in this study has a recall period of the 4 past weeks. Quantity of sleep is scored as the average number of hours slept per night. Other subscales scores are converted to a T-score with a mean of 50, standard deviation (SD) of 10 and range of 0 to 100, where higher scores indicate fewer sleep-related problems.

5.4.1.5. Q-LES-Q

The Quality of Life, Enjoyment, and Satisfaction Questionnaire (Q-LES-Q) is a 16-item self-administered questionnaire designed to measure degree of enjoyment and satisfaction in general activities. The Q-LES-Q measures each of the items using a 5-point Likert-type scale as: 1=very poor; 2=poor; 3=fair; 4=good; 5=very good. The scoring of the Q-LES-Q involves summing only the first 14 items to yield a raw total score. The last 2 items are not included in the total score but are stand-alone items. The raw total score ranges from 14 to 70. The raw total score is transformed into a percentage maximum possible score using the following formula:

$$100 * \frac{(raw\ total\ score - minimum\ possible\ raw\ score)}{(maximum\ possible\ raw\ score - minimum\ possible\ raw\ score)}$$

The minimum possible raw score on the Q-LES-Q is 14, and the maximum possible raw score is 70. Higher Q-LES-Q scores reflect greater enjoyment and satisfaction with general activities.

5.4.1.6. Self-Assessment of Treatment Experience

The Self-Assessment of Treatment Experience questionnaire is a 4-item self-report scale designed to provide additional information regarding the subject's subjective experience while taking the treatment.

5.4.1.7. $HDRS_{17}$ total score in subjects with comorbid MDD

The $HDRS_{17}$ total score will be calculated as described in Section 5.3.1.7, for subjects with comorbid MDD.

5.4.1.8. $HDRS_{17}$ anxiety/somatization factor score in subjects with comorbid MDD

The $HDRS_{17}$ anxiety/somatization factor score will be calculated as described in Section 5.3.1.8, for subjects with comorbid MDD.

5.4.1.9. HAM-D_6 score in subjects with comorbid MDD

The HAM-D_6 score will be calculated as described in Section 5.3.1.9, for subjects with comorbid MDD.

5.4.1.10. SIGH-A total score in subjects with comorbid GAD

The SIGH-A total score will be calculated as described in Section 5.3.1.5, for subjects with comorbid GAD.

5.4.1.11. HAM-A_6 score in subjects with comorbid GAD

The HAM-A_6 score will be calculated as described in Section 5.3.1.6, for subjects with comorbid GAD.

5.4.1.12. $HDRS_{17}$ ≥50% improvement from baseline on total score in subjects with comorbid MDD

The percentage change from baseline is calculated as

$$100 * \frac{(HDRS17\ total\ score\ at\ any\ time\ point - HDRS17\ total\ score\ at\ baseline)}{HDRS17\ total\ score\ at\ baseline}$$

A ≥50% improvement on $HDRS_{17}$ total score means that the percentage change from baseline is ≤ -50 in this calculation. The $HDRS_{17}$ ≥50% improvement will be calculated for subjects with comorbid MDD.

5.4.1.13. $HDRS_{17}$ ≥30% improvement from baseline on total score in subjects with comorbid MDD

The $HDRS_{17}$ ≥30% improvement on total score will be calculated for subjects with comorbid MDD in a similar manner as described in Section 5.4.1.12

5.4.1.14. SIGH-A ≥50% improvement from baseline on total score in subjects with comorbid GAD

The SIGH-A ≥50% improvement will be calculated as described in Section 5.3.1.11, for subjects with comorbid GAD.

5.4.2. Analysis Methods

For all continuous exploratory endpoints, descriptive statistics of actual values and changes from baseline at each scheduled time point and endpoint will be provided by treatment group (mean, SD, median and range). A mean-SE plot over time will be presented for the observed values by treatment group. This includes SDS total score, GAD total score, SHAPS total score, MOS Sleep-R, Q-LES-Q percentage maximum possible score, $HDRS_{17}$ total score, $HDRS_{17}$ anxiety/somatization factor score, HAM-D_6 score, SIGH-A total score and HAM-A_6 score.

Frequency counts and percentages will be provided at each scheduled time point and endpoint by treatment group for the following variables: Q-LES-Q last 2 items, Self-Assessment of Treatment Experience scale, $HDRS_{17}$ $\geq$50% improvement, $HDRS_{17}$ $\geq$30% improvement and SIGH-A $\geq$50% improvement.

6. PHARMACOGENOMIC ANALYSES

It is recognized that genetic variation can be an important contributory factor to interindividual differences in drug distribution and response and can also serve as a marker for disease susceptibility and prognosis. Pharmacogenomic research may help to explain interindividual variability in clinical outcomes and may help to identify population subgroups that respond differently to a drug. The goal of the pharmacogenomic component is to collect DNA to allow the identification of genetic factors that may influence PK, PD, safety, and/or tolerability of JNJ-42165279 and to enable the development of safer, more effective and ultimately individualized therapies in the future. DNA samples will be analyzed for the FAAH gene. Additional analyses may be conducted if it is hypothesized that this may help resolve issues with the clinical data.

DNA samples will be used for research related to JNJ-42165279. They may also be used to develop tests/assays related to JNJ-42165279. Pharmacogenomic research may consist of the analysis of one or more candidate genes or of the analysis of genetic markers throughout the genome (as appropriate) in relation to JNJ-42165279 clinical endpoints. Analyses may be performed across multiple clinical studies.

The relationship between genetic subgroups for the FAAH gene (C/C or A/C) and JNJ-42165279 biomarker endpoints (AEA, PEAE and OEA) will be examined through descriptive statistics and graphical presentations. A similar analysis will be performed to explore the relationship between genetic subgroups and the LSAS total score.

7. BIOMARKERS

Biomarker analyses will be based on the safety analysis set, which will include all randomized subjects who receive at least one dose of study medication (either placebo or JNJ-42165279).

During the study, the following biomarker evaluations will be performed at Day 1, Day 28 and Day 84/ end of withdrawal visit: plasma concentrations of FAAs (AEA, PEA and OEA).

Descriptive statistics of actual values and changes and percent changes from baseline for concentrations of AEA, PEA and OEA at each scheduled time point and endpoint will be provided by treatment group (mean, SD, median and range). A mean-SE plot over time will be presented by treatment group, and results will be listed.

8. SAFETY

All safety analyses will be based on the safety analysis set, which will include all randomized subjects who receive at least one dose of study medication (either placebo or JNJ-42165279).

All safety data will be listed for the subjects in the safety analysis set.

8.1. Incomplete/Missing Dates for Adverse Events

A conservative approach will be used to handle the missing dates for adverse events. The rules for estimating incomplete AE onset dates will be as follows:

(1) The missing day of the month will be estimated as follows: If the month and year are known and double-blind study medication started during that month then the estimated date is the start date of double-blind study medication. If the month and year are known and double-blind study medication started prior to that month then the estimated date is the 1st day of the month. If the month and year are known and double-blind study medication started after the month, then no estimation will be done, and the AE will not be considered as treatment emergent for the double-blind phase.

(2) If both the day and the month are missing: No estimation will be performed. However, these AEs will be considered treatment emergent for the double-blind phase and will be included in the double-blind treatment summaries, except for the calculation of duration of the AE. Attempts will be made to get at least the month for the adverse events.

The rules for estimating incomplete AE onset times will be as follows:

The missing times will be estimated as follows: use time 00:00:00 of the given/imputed date for the start date/time of an AE and use 23:59:59 for the end date/time of an AE.

For incomplete AE resolution dates, the rules are:

(3) The missing day of the month will be estimated as follows: If the month and year are known and the study medication was stopped before, or during that month, the estimated date is the last day of the month or the end of the double-blind phase, whichever is earlier. If the study medication stopped after that month then the estimated date is the last day of the month.

If both the day and the month are missing: the estimated resolution date is the end of the double-blind phase.

8.2. Adverse Events

The verbatim terms used in the CRFs by the investigator to identify adverse events (AEs) will be coded using the current Medical Dictionary for Regulatory Activities maintained by the sponsor.

Therapeutic reach is the number of days after the last dose intake that a subject is still considered to be potentially affected by study drug. For JNJ42165279, therapeutic reach is defined as 4 days. AEs will be summarised using the therapeutic reach. Therapeutic reach will only be applied to the end of double-blind treatment period or to subjects who discontinue.

A treatment-emergent AE (TEAE) is an event that is new in onset or increased in severity following treatment initiation. Any event with a start date and time after the first dose intake will be considered treatment-emergent. An event that starts prior to and ends after the initiation of study medication will be considered treatment-emergent only if the severity increases after the start of medication. An event that starts in the follow up period but within the therapeutic reach of 4 days will be considered treatment-emergent. Adverse events with onset date during the follow-up phase outside the therapeutic reach of 4 days will be summarized separately.

The number (%) of subjects with TEAEs will be summarized by system organ class and preferred term for each treatment group. A summary will be created for the most common TEAEs (≥5% incidence in either treatment group). In addition, TEAEs will be summarized by severity and relationship to study drug for each treatment group by system organ class and preferred term. For the summaries of AEs by severity/relationship to study drug, the observation with the most severe occurrence/closest relationship to study drug will be chosen if there is more than one incident of an adverse event reported during the treatment phase by the subject.

TEAEs will also be summarized by sex, by country, and by enrollment prior to/after study suspension.

Serious AEs (SAEs) and TEAEs that lead to study discontinuation will be summarized separately by treatment group, system organ class and preferred term.

Treatment-emergent headaches, TEAEs related to liver function test abnormalities and TEAEs of special interest (diplopia, vision impairment, gait disturbance, severe headache) will be presented graphically over time for each individual subject.

Data listings will be generated for all AEs, deaths, other serious SAEs, severe AEs and discontinuations due to AEs. These listings will not be limited to TEAEs but will also include any adverse events with onset before the start of study treatment or after the end of study treatment.

8.3. Clinical Laboratory Tests

Descriptive statistics (N, mean, SD, median and range) for absolute values and changes from baseline will be provided by treatment group at each scheduled time point and endpoint for the clinical laboratory tests (hematology, chemistry and urinalysis). Laboratory summaries will be provided in Standard International (SI) units.

Cross-tabulations (with classes for below, within, and above the central laboratory range) showing the shift of laboratory values from baseline to each scheduled time point and endpoint will be presented by treatment group.

Clinical laboratory test values will be considered "treatment emergent markedly abnormal" (TEMA) using the criteria defined by the Sponsor (Johnson & Johnson Research & Development, LLC) listed in Table 5. The identification of TEMA laboratory values is based on the post-baseline (after the start of double-blind study medication) value being out of range while the baseline value is either missing or within the range given in Table 5. If post-baseline laboratory results are above the upper limit and the baseline value is below the lower limit, then the post-baseline abnormality will also be considered TEMA. The same applies to the post-baseline value being below the lower limit with the baseline value being above the upper limit.

The incidence of TEMA will be presented by treatment group at each scheduled time point, at endpoint, and at any time during the double-blind treatment phase.

Table 5: Laboratory Parameters: Criteria for Markedly Abnormal Values

Laboratory Parameter	Markedly Low	Markedly High
Albumin (g/L)	<20	> 2 x ULN
Alkaline phosphatase (U/L)	N/A	> 1.5 x ULN
Alanine aminotransferase (ALT) (U/L)	N/A	> 3 x ULN
Aspartate aminotransferase (AST) (U/L)	N/A	> 3 x ULN
Bicarbonate [mmol/L]	15.1	34.9
Bilirubin (umol/L)	N/A	> 1.5 x ULN
Blood Urea Nitrogen (mmol/L)	N/A	> 3 x ULN
Calcium (mmol/L)	<1.75	>3
Chloride (mmol/L)	<90	>120
Creatine kinase (U/L)	N/A	> 3 x ULN
Creatinine (umol/L)	N/A	> 1.5 x ULN
Gamma glutamyl transferase (U/L)	N/A	> 3 x ULN
Glucose (mmol/L)	<3.33	> 3 x ULN
Lactate dehydrogenase (U/L)	N/A	> 1.5 x ULN
Phosphate (mmol/L)	<0.484	> 1.5 x ULN
Potassium (mmol/L)	<2.8	>5.8
Protein (g/L)	N/A	> 2 x ULN
Prothrombin International Normalized Ratio	N/A	≥1.3
Sodium (mmol/L)	<125.0	>155.0
Hemoglobin (g/L)	<75	> 1.1 x ULN
Platelets (x10e9/L)	<100	> 3 x ULN
Erythrocytes (x10e12/L)	<3	> 2 x ULN
Leukocytes (x10e9/L)	<2.5	> 2.5 x ULN
Hematocrit [fraction]		
male	0.24	0.55
female	0.28	0.5

Graphical presentations will be created for the following liver function tests: alanine aminotransferase, aspartate aminotransferase, albumin, bilirubin, alkaline phosphatase, gamma glutamyl transferase. Liver function test results standardized to the upper limit of normal (based on central laboratory ranges) will be graphically presented over time for each subject. The incidence of liver function tests above the central laboratory normal limit will be presented by treatment group at each scheduled time point, at endpoint, and at any time during the double-blind treatment phase.

Graphical presentations will also be created for prothrombin international normalized ratio results.

Clinical laboratory tests that meet the criteria for abnormal and markedly abnormal will be listed by subject.

Results on urine drug screen, alcohol screen and pregnancy test will be presented in a listing.

8.4. Vital Signs and Physical/Neurological Examination Findings

Descriptive statistics (N, mean, SD, median and range) for absolute values and changes from baseline will be presented at each scheduled time point and endpoint by treatment group for height, body weight and BMI. Similar summary tables will be presented for systolic blood

pressure (mmHg), diastolic blood pressure (mmHg), pulse rate (beats per minute) and temperature (^{o}C).

Mean-SE plots over time will be presented for vital sign values by treatment group, and a subject data listing will be provided. In addition, scatter plots of post-baseline versus baseline values will be presented by treatment group.

Abnormal values are defined in Table 6, and will be summarized using frequency distribution tables over time by treatment group. A listing of subjects meeting any of the abnormality criteria will be provided.

Table 6: Vital Signs Abnormality Ranges (Supine and Standing)

	Low	Normal	High
SBP (mmJHg)	<90	90-140	>140
DBP (mmHg)	<50	50-90	>90
Pulse (bpm)	<45	45-90	>90
Temperature (C)	<35.5	35.5-37.5	>37.5

The incidence of subjects who have a treatment-emergent abnormality (as defined in Table 7 below) at each scheduled time point and endpoint and at any time during the double-blind treatment phase will be presented by treatment group. Shift from baseline tables at each scheduled time point and endpoint will be generated for the vital signs values using the classes (below, within, and above normal ranges) for each treatment group.

Table 7: Treatment-Emergent Abnormality Categories for Vital Signs

Vital Parameter	**Post-baseline value outside of normal limit if:**	
	Abnormally low	**Abnormally high**
Pulse (bpm)	A decrease from baseline of ≥ 15 to a value < 45	An increase from baseline of ≥ 15 to a value > 90
Systolic BP (mmHg)	A decrease from baseline of ≥ 20 to a value < 90	An increase from baseline of ≥ 20 to a value > 140
Diastolic BP (mmHg)	A decrease from baseline of ≥ 15 to a value < 50	An increase from baseline of ≥ 15 to a value > 90

BP = blood pressure

A listing of subjects meeting any of the abnormality criteria will also be provided.

The orthostatic hypotension measurements will be calculated as the standing measurement minus the supine measurement. Orthostatic hypotension is defined as a decrease in systolic (>20 mmHg), or diastolic (>10 mmHg) blood pressure after standing for at least 2 minutes with an increase in pulse rate of >15 beats per minute (Table 8).

Table 8: Abnormal limits for orthostatic parameters

Vital Signs Parameter	Outside of normal limit if difference (Standing minus supine)
(1) Pulse (bpm)	> 15 bpm
(2a) Systolic blood pressure (mmHg) (SBP)	< - 20 mmHg
(2b) Diastolic blood pressure (mmHg) (DBP)	< -10 mmHg

Orthostatic hypotension requires that conditions (1) and [(2a) or (2b)] are met.

For subjects who are unable to stand and have the vital signs measured in a sitting or supine position instead of the standing position, the difference between standing and supine value will remain missing.

Frequency table of the occurrence of orthostatic hypotension at each scheduled time point and endpoint and at any time during the double-blind treatment phase will be presented by treatment group. Listings will be provided for subjects who have orthostatic hypertension.

Physical examination abnormalities and neurological examination results will be listed.

8.5. Electrocardiogram

A triplicate 12-lead ECG will be obtained at Day 1, otherwise a single, 12-lead ECG will be obtained at each scheduled time point. The ECG variables that will be analyzed are heart rate (HR), ECG intervals (RR, PR, QRS, and QT) as well as corrected QT intervals according to Bazett's formula (QTcB) and Fridericia's formula (QTcF).

The corrected QTcF and QTcB intervals will be calculated according to the following formula:

- *Fridericia:* $QTcF\ (msec) = QT(msec) * (HR(bpm)/60)^{1/3}$
- *Bazett:* $QTcB\ (msec) = QT(msec) * (HR(bpm)/60)^{1/2}$

Summary tables for actual values and changes from baseline will be presented by treatment group at each scheduled time point and at endpoint. Mean of triplicate ECG values obtained on Day 1 (pre-dose) will be used as a baseline value. Mean-SE plots over time will be presented for ECG values by treatment group. Scatter plots of post-baseline versus baseline values will be presented by treatment group.

The frequency of subjects who are flagged with abnormal values will be summarized by treatment group. The values will be listed for all the subjects with abnormal flags if the ECG values are out of normal range. The following abnormality ranges will be used for ECG parameters (Table 9):

Table 9: ECG Abnormality Ranges

	Low	Normal	High
HR (bpm)	<45	45-90	> 90
PR Interval (msec)	<120	120-220	>220

QRS Interval (msec)	-	<120	≥120
QT Interval (msec)	-	<500	≥500

Criteria for abnormal corrected QTc interval values and changes from baseline are presented in Table 10. Abnormal values and changes will be summarized using frequency distribution tables over time by treatment group. A listing of subjects meeting any of the criteria will be provided.

Table 10: Criteria for Abnormal QTc Values and Changes From Baseline

Parameter	Classification	Criteria
QTc value	Normal	≤450
	> 450 – 480	>450 - ≤480
	> 480 – 500	>480 – ≤500
	> 500	> 500
QTc change from baseline	No concern	≤30
	Concern	>30 – 60
	Clear concern	> 60
These criteria are based on ICH E14 Guideline		

8.6. Other Safety Parameters

8.6.1. C-SSRS

The Columbia Suicide Severity Rating scale (C-SSRS) is a low-burden measure of the spectrum of suicidal ideation and behavior. It is a semi structured clinician-administered questionnaire designed to solicit the occurrence, severity, and frequency of suicide-related ideation and behaviors during the assessment period.

Using the C-SSRS, potentially suicide-related events will be categorized using the following scores:

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: Non-fatal suicide attempt

10: Completed suicide

A frequency distribution at each scheduled time point and endpoint will be provided by treatment group. A listing will be provided of C-SSRS items throughout the study for subjects with suicidal ideation or behavior at any time point.

REFERENCES

1. Bech P, Gram LF, Dein E, Jacobsen O, Vitger J, Bolwig TG. Quantitative rating of depressive states. Acta Psychiatr Scand 1975 Mar; 51(3): 161-170.
2. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960 Feb; 23: 56-62
3. Hamilton M. Standardised assessment and recording of depressive symptoms. Psychiatr Neurol Neurochir 1969 Mar-Apr; 72(2): 201-205.
4. Hamilton M. The assessment of anxiety states by rating. Br J Med Psychol 1959; 32(1): 50-55.
5. International Conference on Harmonization (ICH) Tripartite Guideline E14. The clinical evaluation of QT/QTc interval prolongation and proarrhythmic for non-antiarrhythmic drugs. London 2005.
6. Liebowitz MR. Social Phobia. Mod Probl Pharmacopsychiatry 1987;22:141-173
7. Pollack MH, Van Ameringen M, Simon NM, at al. A double-blind randomized controlled trial of augmentation and switch strategies for refractory social anxiety disorder. Am J Psychiatry. 2014;171(1):44-53.
8. R. G. Heimberg, K. J. Horner, H. R. Juster, S. A. Safren, E. J. Brown, F. R. Schneier, M. R. Liebowitz. Psychometric properties of the Liebowitz Social Anxiety Scale. Psychological Medicine, 1999, 29, 199–212
9. Shear MK, Vander Bilt J, Rucci P, et al. Reliability and validity of a structured interview guide for the Hamilton Axiety Rating Scale (SIGH-A). Depress Anxiety. 2001; 13(4): 166–178
10. Stein MB, Pollack MH, Bystritsky A, Kelsey JE, Mangano RM. Efficacy of low and higher dose extended-release venlafaxine in generalized social anxiety disorder: a 6-month randomized controlled trial. Psychopharmacology. 2005;177(3):280–288.
11. Stein DJ, Stein MB, Pitts CD, Kumar R, Hunter B. Predictors of response to pharmacotherapy in social anxiety disorder: An analysis of 3 placebo-controlled paroxetine trials. Journal of Clinical Psychiatry 2002:63(2):152–5

ATTACHMENT: LIEBOWITZ SOCIAL ANXIETY SCALE (LSAS)

Pt Name:		Pt ID #:
Date:	Clinic #:	Assessment point:

Fear or Anxiety:
0 = None
1 = Mild
2 = Moderate
3 = Severe

Avoidance:
0 = Never (0%)
1 = Occasionally (1—33%)
2 = Often (33—67%)
3 = Usually (67—100%)

	Fear or Anxiety	Avoidance	
1. Telephoning in public. (P)			1.
2. Participating in small groups. (P)			2.
3. Eating in public places. (P)			3.
4. Drinking with others in public places. (P)			4.
5. Talking to people in authority. (S)			5.
6. Acting, performing or giving a talk in front of an audience. (P)			6.
7. Going to a party. (S)			7.
8. Working while being observed. (P)			8.
9. Writing while being observed. (P)			9.
10. Calling someone you don't know very well. (S)			10.
11. Talking with people you don't know very well. (S)			11.
12. Meeting strangers. (S)			12.
13. Urinating in a public bathroom. (P)			13.
14. Entering a room when others are already seated. (P)			14.
15. Being the center of attention. (S)			15.
16. Speaking up at a meeting. (P)			16.
17. Taking a test. (P)			17.
18. Expressing a disagreement or disapproval to people you don't know very well. (S)			18.
19. Looking at people you don't know very well in the eyes. (S)			19.
20. Giving a report to a group. (P)			20.
21. Trying to pick up someone. (P)			21.
22. Returning goods to a store. (S)			22.
23. Giving a party. (S)			23.
24. Resisting a high pressure salesperson. (S)			24.