[redacted]

Sponsor	VistaGen Therapeutics, Inc.
Protocol Title:	A US, Phase 3 Multicenter, Randomized, Double-blind, Placebo-controlled Clinical Trial of PH94B Nasal Spray for the Acute Treatment of Anxiety Induced by a Public Speaking Challenge in Adult Subjects with Social Anxiety Disorder (PaliSADe-2)
Protocol Number:	PH94B-CL032
NCT:	05011396
[redacted]	[redacted]
Document Version:	Final 1.0
Document Date:	28-Jun-2022

Approvals

Role	Signatures
Biostatistician	Print Name: [redacted]
	Sign Name: [redacted]
Peer Reviewer	Print Name: [redacted]
	Sign Name: [redacted]

[redacted]

Role	Signatures
VistaGen Therapeutics, Inc. Representative	Print Name: [redacted]
	Sign Name: [redacted]

Document History

Not applicable.

[redacted]

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Definition
AE	adverse event
ANCOVA	analysis of covariance
ASA	American Statistical Association
ATC	Anatomical Therapeutic Chemical
BMI	Body Mass Index
C-SSRS	Columbia-Suicide Severity Rating Scale
CGI-I	Clinical Global Impression – Improvement scale
CI	confidence interval
COVID-2019	coronavirus 2019
CSR	clinical study report
ECG	Electrocardiogram
eCRF	electronic case report form
EMA	European Medicines Agency
FCS	fully conditional specification
FDA	Food and Drug Administration
[redacted]	[redacted]
HAM-D	Hamilton Depression Scale
ICH	International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use

Abbreviation	Definition
IP	investigational product
ITT	Intent-to-Treat
IWRS	interactive web response system
LS	least squares
LSAS	Liebowitz Social Anxiety Scale
MAR	missing at random
MedDRA	Medical Dictionary for Regulatory Activities
MINI	Mini-International Neuropsychiatric Interview
PGI-C	Patient Global Impression of Change
Q1 / Q3	first quartile / third quartile
QTcF	QT interval corrected for heart rate using Fridericia's formula
RSS	Royal Statistical Society
SAD	social anxiety disorder
SAE	serious adverse event
SAP	statistical analysis plan
SD	standard deviation
SE	standard error
SOC	system organ class
SUDS	Subjective Units of Distress Scale

Abbreviation	Definition
TEAE	treatment emergent adverse event
WHO-DD	World Health Organization Drug Dictionary

1. Overview

This statistical analysis plan (SAP) describes the planned analysis and reporting for VistaGen Therapeutics, Inc. protocol number PH94B-CL032 (A US, Phase 3 Multicenter, Randomized, Double-blind, Placebo-controlled Clinical Trial of PH94B Nasal Spray for the Acute Treatment of Anxiety Induced by a Public Speaking Challenge in Adult Subjects with Social Anxiety Disorder [PALISADE-2]), dated 25-Jan-2022, Version 3.0. Reference materials for this statistical plan include the protocol and the accompanying sample data collection documents. Operational aspects related to collection and timing of planned clinical assessments are not repeated in this SAP unless relevant to the planned analysis.

The structure and content of this SAP provides sufficient detail to meet the requirements identified by the Food and Drug Administration (FDA), European Medicines Agency (EMA), and International Council for Harmonisation (ICH) of Technical Requirements for Pharmaceuticals for Human Use: Statistical Principles in Clinical Trials E9 (ICH, 1998). All work planned and reported for this SAP will follow internationally accepted guidelines, published by the American Statistical Association (ASA, 2018) and the Royal Statistical Society (RSS, 2014), for statistical practice.

The planned analyses identified in this SAP may be included in clinical study reports (CSRs), regulatory submissions, or future manuscripts. Also, post-hoc exploratory analyses not necessarily identified in this SAP may be performed to further examine study data. Any post-hoc or unplanned, exploratory analysis performed will be clearly identified as such in the final CSR.

The statistical plan described hereafter is an *a priori* plan. It will be approved before any unblinded inferential or descriptive analysis of data pertaining to VistaGen Therapeutics, Inc.'s study PH94B-CL032.

2. Study Objectives and Endpoints

2.1. Study Objectives

2.1.1. Primary Objective

The primary efficacy objective of the study is to evaluate whether the efficacy of PH94B to relieve acute anxiety induced during a public speaking challenge in adult subjects with social anxiety disorder (SAD) is greater than that of placebo as measured by the Subjective Units of Distress Scale (SUDS).

2.1.2. Secondary Objectives

The secondary efficacy objective of the study is to compare clinician-observed changes in subject response to an anxiety-provoking situation from Visit 2 (Baseline) to Visit 3 (Treatment) between [redacted] treated subjects and [redacted]-treated subjects, as measured by Clinical Global Impression - Improvement (CGI-I) scale.

The safety objective of the study is to determine safety and tolerability of PH94B compared to placebo in adult subjects with SAD from reported adverse events (AEs) and changes in vital signs, 12-lead electrocardiograms (ECGs), laboratory parameters, suicidality, level of depression, and physical examinations.

2.1.3. Exploratory Objectives

Exploratory efficacy objectives of the study are to compare PH94B-treated subjects with placebo-treated subjects with regard to the following:

- Subject self-evaluation of change in response to the anxiety provoking situation between Visit 2 (Baseline) and Visit 3 (Treatment) as measured by Patient Global Impression of Change (PGI-C)
- [redacted]
- To compare the proportion of subjects in each group with an improvement of 20 or more points from Visit 2 (Baseline) SUDS scores at Visit 3 (Treatment) ("responders").
- [redacted]

2.2. Study Endpoints

2.2.1. Efficacy Endpoints

2.2.1.1. Primary Efficacy Endpoint

The primary efficacy endpoint is mapped to the primary objective as follows:

Primary Objective	Primary Endpoint
The primary efficacy objective of the study is to evaluate whether the efficacy of PH94B to relieve acute anxiety induced during a public speaking challenge in adult subjects with SAD is greater than that of placebo as measured by the SUDS.	The primary endpoint for the study is the difference in average SUDS score during the [redacted] Visit 3 (Treatment) public speaking challenge versus the average SUDS scores during the [redacted] Visit 2 (Baseline) public speaking challenge for PH94B compared to placebo.

2.2.1.2. Secondary Efficacy Endpoint

The secondary efficacy endpoint is mapped to the secondary objective as follows:

Secondary Objective	Secondary Endpoint
The secondary efficacy objective of the study is to compare clinician-observed changes in subject response to an anxiety provoking situation from Visit 2 (Baseline) to Visit 3 (Treatment) between PH94B-treated subjects and placebo-treated subjects, as measured by CGI-I.	The secondary efficacy endpoint in the study is the difference between proportions of PH94B- and placebo-treated groups in CGI-I scores of 1 [redacted] or 2 [redacted] recorded at the end of Visit 3 (Treatment).

2.2.1.3. Exploratory Efficacy Endpoints

Exploratory endpoints are mapped to exploratory objectives as follows:

Exploratory Objectives	Exploratory Endpoints
Exploratory efficacy objectives of the study are to compare PH94B-treated subjects with placebo-treated subjects with regard to the following:	
• Subject self-evaluation of change in response to the anxiety provoking situation between Visit 2 (Baseline) and Visit 3 (Treatment) as measured by PGI-C	Difference between proportions of PH94B- and placebo-treated groups in PGI-C scores of 1 [redacted] or 2 [redacted] recorded at the end of Visit 3 (Treatment)
• [redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]

Exploratory Objectives	Exploratory Endpoints
• To compare the proportion of subjects in each group with an improvement of 20 or more points from Visit 2 (Baseline) SUDS scores at Visit 3 (Treatment) ("responders")	Proportion of subjects in each treatment group with 20-point improvement in SUDS scores from Visit 2 (Baseline) at Visit 3 (Treatment)
• [redacted]	[redacted]

2.2.2. Safety Endpoints

The safety endpoints are mapped to the safety objective as follows:

Safety Objective	Safety Endpoints
The safety objective of the study is to determine safety and tolerability of PH94B compared to placebo in adult subjects with SAD from reported AEs and changes in vital signs, 12-lead ECGs, laboratory parameters, suicidality, level of depression, and physical examinations.	Incidence and severity of AEs Changes in vital signs results Changes in clinical laboratory evaluation (hematology, chemistry, and urinalysis) results Changes in 12-lead ECG results Changes in physical examination findings Mean HAM-D scores Incidence of treatment-emergent suicidal ideation and behavior

3. Overall Study Design and Plan

3.1. Overall Design

The study is a Phase 3, multicenter, randomized, double-blind, placebo-controlled, parallel design, single-dose, clinical trial of the efficacy, safety, and tolerability of the acute administration of PH94B to relieve symptoms of acute anxiety in adult subjects with SAD.

Subject participation in the study will last a total of 3 to 7 weeks, depending on the duration of the screening period and intervals between visits. Subjects will be considered to have completed the study after they complete Visit 4 (Follow-Up).

Upon signing an informed consent form, all subjects will complete Visit 1 (Screening) and enter a screening period lasting between 3 and 35 days. If subjects meet all eligibility criteria at the end of the screening period, subjects will complete Visit 2 (Baseline), including participation in a 5-minute public speaking challenge [redacted] If the subject reports at least 1 SUDS score of 75 or greater during the Visit 2 (Baseline) public speaking challenge, then the subject will be instructed to come back for Visit 3 (Treatment) in 1 week (±2 days).

At Visit 3 (Treatment), the subject will be randomly allocated in a [redacted] manner to receive treatment with either PH94B or placebo. The subject will self-administer randomly allocated study treatment and will then undergo a 5-minute public speaking challenge, with [redacted] At the end of the Visit 3 (Treatment) public speaking challenge, the subject will complete the PGI-C questionnaire, [redacted] complete the CGI-I assessment.

One week (±2 days) after the completion of Visit 3 (Treatment) public speaking challenge, the subject will come back for Visit 4 (Follow-Up), which will involve a repeat of the safety and psychiatric assessments conducted at Screening, and [redacted]

Efficacy assessments for the primary endpoint will be completed at Visit 2 (Baseline) and Visit 3 (Treatment). Efficacy assessments for the secondary endpoint will be completed at Visit 3 (Treatment). Safety assessments are completed at Visit 1 (Screening) and Visit 4 (Follow-Up), with the exception of vital signs assessments, pregnancy tests, and Columbia-Suicide Severity Rating Scale (C-SSRS), which are completed at all visits, and AE recordings, which are completed starting at Visit 2 (Baseline). Refer to the Schedule of Events (Table 1) for more detail.

3.2. Sample Size and Power

The sample size calculation was based on the similarly designed Phase 2 randomized, double-blind, placebo-controlled clinical study of PH94B with the primary outcome variable of average subjective anxiety based on SUDS scores. [redacted]

[redacted]

Because this Phase 3 study will involve 15 to 18 sites [redacted] Based on these considerations, 208 subjects (104 in each arm) will be included in the study.

Figure 1 provides the power curve used to determine the sample size.

An estimated 400 subjects will be screened to randomize 208 subjects in a 1:1 manner to either PH94B or placebo.

[redacted]

3.3. Study Population

The study population consists of male or female subjects, 18 through 65 years of age, inclusive, with a current diagnosis of SAD and with no prior or current conditions that would increase risk of treatment or confound the results of the study.

3.4. Treatments Administered

Subjects will be randomized 1:1 to receive one of the treatments listed below at Visit 3 (Treatment).

[redacted]

3.5. Method of Assigning Subjects to Treatment Groups

Subjects who meet study entry criteria, and who complete the Visit 2 (Baseline) public speaking challenge with at least 1 SUDS score of 75 or higher will be randomly assigned in a 1:1 ratio to the PH94B or placebo treatment arms. The randomization schedule will be computer generated using a permuted block algorithm and will randomly allocate investigational product (IP) to randomization numbers. The randomization numbers will be assigned sequentially through a central interactive web response system (IWRS). Randomization will be stratified by study site; blocks will be allocated to sites by IWRS in real-time.

The randomization schedule will be prepared by [redacted] before the start of the study. No one involved in the study performance will have access to the randomization schedule before official unblinding of treatment assignment. No subject will be randomized into this study more than once.

Each study site will be provided with a supply of blinded IP, each vial individually numbered with a unique alpha-numeric identification code. For Visit 2 (Baseline) and Visit 3 (Treatment), the IWRS system will provide a unique vial identification code indicating the study drug vial to be dispensed to the subject. The vial identification codes will be recorded in the drug dispensing record and the electronic case report form (eCRF).

3.6. Blinding and Unblinding

[redacted] At Visit 3 (Treatment), treatment will be administered in a [redacted] manner: all subjects, investigators, and study personnel involved in the conduct of the study, including data management, will be blinded to treatment assignment with the exception of a specified unblinded statistician and a member of the Interactive Response Technology team from [redacted] who will have access to the randomization code. The unblinded study personnel will not participate in study procedures or data analysis prior to unblinding of the study data to all study-related personnel. If an interim analysis is to be conducted, then unblinded personnel who are not otherwise involved in the study will prepare the data for review.

Study personnel will make every effort to safeguard the integrity of the study blind to minimize bias in the conduct of the study. Treatment unblinding is discouraged if knowledge of the

treatment assignment will not materially change the planned management of a medical emergency. Unblinding will be permitted in a medical emergency that requires immediate knowledge of the subject's treatment assignment. The vial umber can be used to break the blind if necessary. Contact information will be provided to the study site.

Unblinding should be discussed in advance with the medical monitor, if possible. For emergency unblinding, study personnel will use the IWRS. If the investigator is not able to discuss treatment unblinding in advance, then he/she must notify the medical monitor as soon as possible about the unblinding incident without revealing the subject's treatment assignment.

The investigator or designee must record the date and reason for treatment unblinding on the appropriate eCRF for that subject. In all cases that are not emergencies, the investigator must discuss the event with the medical monitor prior to unblinding the subject's treatment assignment.

If treatment assignment is unblinded for an individual subject, study personnel will be notified of that subject's treatment assignment without unblinding the treatment assignments for the remaining subjects in the study. Thus, the overall study blind will not be compromised. If a subject's treatment assignment is unblinded, he/she may or may not be asked to withdraw from the study. The investigator will make this decision after consultation with the medical monitor.

Overall unblinding will take place at the end of the study only after database lock has been achieved.

3.7. Schedule of Events

The timing of all study assessments is outlined in the Schedule of Events (Table 1) below.

Table 1: Schedule of Events

Assessment	Screening *Day -35 to -3* Visit 1	Baseline *Day 0* Visit 2	Treatment *Day 7* Visit 3	Follow-Up *Day 14* Visit 4
Written Informed Consent	X			
Subject Demographics	X			
Urine Drug Screen (Instant Exam, on site)	X	X	X	X
Urine Pregnancy Test	X	X	X	X
MINI	X			
Medical and Psychiatric History	X			
Prior and Concomitant Medication Recording	X	X	X	X
[redacted]	[redacted]			
LSAS	X			X

Assessment	Screening *Day -35 to -3* Visit 1	Baseline *Day 0* Visit 2	Treatment *Day 7* Visit 3	Follow-Up *Day 14* Visit 4
Hamilton Depression Scale	X			X
[redacted]	[redacted]			
C-SSRS	X	X (end of visit)[c]	X (end of visit)[c]	X
Physical Examination and Examination of Nasal Passages	X			X
Vital Signs	X	X	X	X
Electrocardiogram	X			X
Blood Tests and Urinalysis	X			X
Inclusion/Exclusion Review	X	X		
Investigational Product Administration Training	X			
SUDS Introduction[a]		X		
Investigational Product Administration		X	X	
Public Speaking Challenge[b]		X	X	
Randomization			X	
PGI-C Scale	[redacted]			
CGI-I Scale			X	
Adverse Events Recording		X	X	X
[redacted]	[redacted]			

Abbreviations: C-SSRS = Columbia Suicide Severity Rating Scale; CGI-I = Clinical Global Impression of Improvement; LSAS = Liebowitz Social Anxiety Scale; MINI = Mini-International Neuropsychiatric Interview; PGI-C= Patient Global Impression of Change; SUDS = Subjective Units of Distress Scale

a. Introduce the SUDS to the subject and provide scores in relation to feelings of distress (see protocol).

b. Public Speaking Challenge involves the collection of SUDS scores from the subject and is outlined in the protocol.

c. The C-SSRS at Visits 2 and 3 should be administered at the end of the visit, i.e., after investigational product administration and public speaking challenge.

4. Statistical Analysis and Reporting

4.1. Introduction

Data processing, tabulation of descriptive statistics, calculation of inferential statistics, and graphical representations will primarily use SAS (release 9.4 or higher). If the use of other

software is warranted, the final statistical methodology report will detail what software was used for what purposes.

Unless otherwise stated, all continuous (quantitative) variable summaries will be presented by treatment group and include the number of subjects with non-missing values (n), mean, SD, minimum, first quartile (Q1), median, third quartile (Q3), and maximum.

Unless otherwise stated, all categorical (qualitative) variable summaries will be presented by treatment group and include the frequency and percentage of subjects who are in the particular category or each possible value. In general, the denominator for the percentage calculation will be based upon the total number of subjects in the relevant study population with non-missing values.

The minimum and maximum will be reported with the same degree of precision (i.e., the same number of decimal places) as the observed data. Measures of location (mean, median, Q1, and Q3) will be reported to 1 degree of precision more than the observed data and measures of spread (SD) will be reported to 2 degrees of precision more than the observed data.

Percentages <100% will be presented to 1 decimal place and percentages of 100% will be reported with no decimal place. Counts of zero will be presented without percentages.

Unless otherwise indicated, all statistical tests will be conducted at the 0.05 significance level using 2-tailed tests, and *P* values will be reported. Corresponding 95% confidence intervals (CIs) will be presented for statistical tests.

To control for the overall type I error rate, a fixed-sequence testing strategy will be used. If statistical significance is declared for the primary endpoint, formal hypothesis testing will be done for the secondary endpoint. If a non-significant *P* value is reached for the primary endpoint, *P* values for the secondary endpoint will be considered exploratory. The *P* values from the exploratory endpoints are not included in the fixed-sequence testing strategy and are included for informational purposes only. The planned sequence of hypothesis testing is detailed in Section 6.1.3.

4.2. Interim Analysis and Data Monitoring

No interim analysis is planned.

5. Analysis Populations

The following analysis populations are planned for this study:

- **All Subjects Population:** All subjects who sign an informed consent, participate in a Screening visit, and are entered into the eCRF will be included in the All Subjects population. The All Subjects population will be used for summaries of subject disposition and protocol deviations.

- **Safety Population:** All subjects who receive IP will be included in the Safety Population. For analysis of the Safety Population, each subject will be analyzed by the actual treatment received, even if it differs from the subject's randomized treatment assignment. Adverse events (AEs) will be classified by time of occurrence and Visit 3 (Treatment) group assignment to more clearly identify those arising after exposure to [redacted] and those arising after exposure to [redacted] The Safety Population will be used for safety analyses.
- **Intent-to-Treat (ITT) Population:** The ITT Population in the study includes all subjects who are randomized. For analyses of the ITT Population, each subject will be analyzed by their randomized treatment assignment, even if the subject did not receive the assigned treatment or did not follow the protocol until completion. The ITT Population will be used for efficacy analyses.
- **Per Protocol (PP) Population**: The Per Protocol (PP) Population includes all subjects who have non-missing average SUDS scores during the 5-minute public speaking challenges at Visit 2 (Baseline) and Visit 3 (Treatment) and no important protocol deviations that would have an impact on statistical analysis. For analyses of the PP Population, each subject will be analyzed by the actual treatment received, even if it differs from the subject's randomized treatment assignment. The PP Population will be used for a sensitivity analysis of the primary efficacy endpoint (Section 8.1.2).

Inclusion in the analysis populations will be determined prior to database lock. Assignment of subjects to populations will be confirmed at a blinded data review meeting to be held before the study database is locked.

6. General Issues for Statistical Analysis

6.1. Statistical Definitions and Algorithms

6.1.1. Baseline

Baseline for each study assessment is defined in Table 2.

Table 2: Baseline Definitions

Parameter	Baseline
Safety Variables (clinical laboratory assessments, vital signs, ECGs, physical examinations)	The last non-missing assessment before the first dose of study drug given at Visit 2 (Baseline).
C-SSRS Scores	Scores on the Baseline version of the C-SSRS given at Visit 1 (Screening).

Parameter	Baseline
Average SUDS Score	Sum of Visit 2 (Baseline) SUDS scores [redacted] divided by the number of SUDS scores recorded during the speech performance.
Average Anticipatory SUDS Score	Sum of Visit 2 (Baseline) SUDS scores [redacted] divided by the number of recorded [redacted] SUDS scores.
LSAS total score, HAM-D total score, [redacted]	Total score at Visit 1 (Screening).

Abbreviations: CGI-I = Clinical Global Impression of Improvement; ECG = electrocardiogram; C-SSRS = Columbia-Suicide Severity Rating Scale; [redacted]; HAM-D = Hamilton Depression Scale; LSAS = Liebowitz Social Anxiety Scale; PGI-C= Patient Global Impression of Change; SUDS = Subjective Units of Distress Scale

6.1.2. Adjustments for Covariates

Planned analysis of covariance (ANCOVA) analyses will adjust for analysis site (Section 6.1.6) and the baseline of the target endpoint by including them as terms in the model. The comparison of binomial proportions for the secondary endpoint will be presented overall and by strata.

6.1.3. Multiple Comparisons

The type I error rate of $\alpha = 0.05$ will be maintained for the study by using a fixed-sequence testing strategy for the primary and secondary endpoints. Testing of the secondary endpoint will proceed only if the primary endpoint achieves statistical significance. No multiple comparisons adjustments will be made for the exploratory endpoints as the analyses are included for descriptive purposes only. All *P* values for each comparison will be provided in the summary tables for informational purposes.

The testing will be done in the following order:

1. Primary efficacy endpoint: difference in average SUDS score during the [redacted] Visit 3 (Treatment) public speaking challenge versus the average SUDS score during the [redacted]

[redacted] Visit 2 (Baseline) public speaking challenge for PH94B compared to placebo

2. Secondary efficacy endpoint: difference between proportions of PH94B- and placebo-treated groups in CGI-I scores of 1 [redacted] or 2 [redacted] recorded at the end of Visit 3 (Treatment).

6.1.4. Handling of Dropouts or Missing Data

Data will be included in the analysis according to the analysis visit windows described in Section 6.1.5.

It is expected that all study subjects who are randomized at Visit 3 (Treatment) will complete the Visit 3 (Treatment) public speaking challenge immediately after randomization and be included in the primary efficacy analysis. If a subject has incomplete scores during the 5-minute interval, then their average SUDS score will be calculated using the non-missing scores. If all public speaking challenge Visit 3 (Treatment) SUDS scores are missing for unanticipated reasons (intercurrent events) that arise during Visit 3 (Treatment), the average SUDS score will be set to missing. If the fraction of missing change from baseline to Visit 3 (Treatment) average SUDS scores in the ITT population is greater than 0.05, multiple imputation of the change from baseline to Visit 3 (Treatment) average SUDS scores will be carried out for the primary efficacy analysis. The fraction of missing average change from baseline in average SUDS scores will be computed as described in Section 6.1.7.2. [redacted] only change from baseline SUDS scores will be imputed. Other efficacy and safety data will not be imputed and will be analyzed as observed.

The imputation approach has 2 broad components: i) the multiple imputation process for the placebo and PH94B data; and ii) the analysis method that will be used to draw inference regarding the primary causal estimand, along with the method for combining the results across the multiply-imputed datasets. The primary causal estimand is defined in Section 8.1.

The imputation approach for the primary efficacy endpoint will be carried out in SAS using the fully conditional specification (FCS) method with a regression model (Carpenter and Kenward, 2013). The method assumes data are missing at random (MAR). Thus, subjects with missing values will have imputed values similar to other subjects with comparable observed profiles in the regression variables. The imputation will be carried out separately within each treatment group. The regression model will include the following variables: baseline SUDS score, sex, age, baseline LSAS total score, and analysis site. For each imputation process, a total of 25 complete datasets will be created. The random seeds to be used in the imputation steps are 157340717 and 127581968 for the PH94B and placebo arms, respectively.

The analysis for the multiple imputation will match that described for the primary efficacy analysis (Section 8.1.1). Results from the multiply-imputed datasets will be combined using the standard Rubin's rules for multiple imputation (Little and Rubin, 1987) and will be done in SAS using PROC MIANALYZE.

6.1.5. Analysis Visit Windows

Analysis visits will be assigned based on the nominal visit recorded in the study database.

6.1.6. Pooling of Sites

Analysis sites (pooled sites) will be created if sites have fewer than 6 randomized subjects. Sites will be ordered by the number of randomized subjects. The smallest study site will be combined with the next smallest study site. These 2 sites may be combined with the next smallest study site if the combined number of subjects remains less than 6. This process will be continued until this first pooled site has 6 or more randomized subjects. Once pooling is completed for the first pooled site, the pooled site and all unpooled sites will be reordered by the number of randomized subjects. If the new smallest site has fewer than 6 randomized subjects, the pooling will proceed again as described above to create a second pooled site, followed by reordering of pooled and unpooled sites by the number of randomized subjects. The process will repeat until the smallest site has 6 or more randomized subjects. If, in combining the 2 smallest sites, the resulting new analysis site becomes the largest study site (i.e., more subjects than present in the initial largest site), the combining of sites will be halted and not performed. The final determination of pooled sites will be made prior to database lock.

6.1.7. Definitions and Derived Variables

6.1.7.1. General

- **Age** = Integer difference in years between the subject's date of informed consent and date of birth.
- **Change from Baseline to Visit X in Parameter Y** = Value at Visit X minus the value at Baseline (as defined in Table 2).
- **Treatment Duration** = Treatment end date minus treatment start date plus 1.
- **Height (cm) and Weight (kg)** = Height (cm) will be calculated as height (in) times 2.54. Weight (kg) will be calculated as weight (lb) times 0.4536.
- **Percent Change from Baseline to Visit X in Parameter Y** = Change from Baseline to Visit X in Parameter Y divided by the value at Baseline (as defined in Table 2) multiplied by 100.
- **Shift from Baseline to Visit X in Parameter Y Categories** = Cross-tabulation of below the lower limit of the normal range, within the limits of the normal range, and above the upper limit of the normal range at Baseline (as defined in Table 2) versus at Visit X.
- **Study Day** = Event date minus date of the first dose of study drug plus 1 for events occurring on or after the first dose. Event date minus the date of the first dose of study drug for events occurring before the first dose.

Temperature (C°) = Temperature (F°) times 5 divided by 9.

6.1.7.2. Subjective Units of Distress Scale (SUDS)

The SUDS, used at Visit 2 (Baseline) and Visit 3 (Treatment) as part of each public speaking challenge, is scored 0 to 100. It is a standard instrument for rating social and performance anxiety in patients with SAD during role-playing situations. Scores are operationalized for subjects in this study as 0=totally relaxed or no anxiety100=most distress or anxiety imaginable.

- **Average SUDS Score** = Sum of SUDS scores recorded at [redacted] divided by the number of SUDS scores recorded during the speech performance. For complete data, there should be [redacted] SUDS scores collected.
- [redacted]
- **Proportion SUDS Responders** = Proportion of subjects with an improvement of 20 or more points in average SUDS scores from Visit 2 (Baseline) to Visit 3 (Treatment).
- **Fraction of Missing Change from Baseline in Average SUDS Scores =** Number of missing values of change from baseline to Visit 3 (Treatment) average SUDS scores across all subjects in the ITT Population divided by the total number of expected change scores (1 per subject). The fraction of missing change from baseline in average SUDS scores is used to compute the threshold for determining whether to implement the multiple imputation procedure.

6.1.7.3. Clinical Global Impression Scale of Improvement (CGI-I)

The CGI-I is a clinician-rated instrument that measures the clinician's evaluation of change in subjects' overall improvement with treatment. Clinicians will respond to the following question:

"Please choose the response that best describes how anxious you felt the subject was during the second speech compared to how anxious they appeared in the first speech a week ago.

[redacted]

The derived variable from this assessment will be the following:

- **Proportion of CGI-I Responders** = Proportion of subjects with a score of 1 (Very much less anxious) or 2 (Much less anxious) on the CGI-I at Visit 3 (Treatment).

6.1.7.4. Patient Global Impression of Change (PGI-C)

The PGI-C is a self-administered instrument that measures change in subjects' overall improvement with treatment. Subjects will respond to the following question:

"Please choose the response that best describes how anxious you felt giving the second speech compared to how anxious you felt giving the first speech a week ago.

[redacted]

The derived variable from this assessment will be the following:

- **Proportion PGI-C Responders** = Proportion of subjects with a score of 1 [redacted] or 2 [redacted] on the PGI-C at Visit 3 (Treatment).

6.1.7.5. Liebowitz Social Anxiety Scale (LSAS)

The LSAS is a clinician-rated scale that has been shown to be sensitive to treatment-related change in social anxiety symptoms. The time frame for rating symptomatology is the past week. The scale consists of 24 items. Each item is given 2 ratings: fear or anxiety on scale of 0 to 3 and avoidance on a scale of 0 to 3, with a total maximum overall score of 144. The "Total Performance (P) Subscore – Fear or Anxiety" and the "Total Performance Subscore (P) – Avoidance" are the sums of the fear or anxiety or avoidance scores, respectively, on items 1-4, 6, 8, 9, 13, 14, 16, 17, 20, and 21. The "Total Social (S) Subscore – Fear or Anxiety" and "Total Social (S) Subscore – Avoidance" are the sums of the fear or anxiety or avoidance scores, respectively, on items 5, 7, 10-12, 15, 18, 19, and 22-24. The "Total Anxiety Subscore" and the "Total Avoidance Subscore" are the sums of performance and social subscores for fear or anxiety and avoidance scales, respectively. The LSAS total score for each subject is the sum of all ratings across both scales.

The derived variables for this assessment are the following:

- [redacted]

[redacted]

6.1.7.6. [redacted]

[redacted]

6.1.7.7. Hamilton Depression Scale (HAM-D)

At Visit 1 (Screening) and Visit 4 (Follow-Up), subjects will be asked to complete the clinician-rated HAM-D questionnaire to assess depression. The HAM-D questionnaire consists of 17 items, each scored on either a 3- or 5-point scale.

The Hamilton Depression Scale (HAM-D) total score is calculated as the sum of the 17 individual item scores and ranges from 0 to 52.

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

The C-SSRS is a semi-structured interview that captures the occurrence, severity, and frequency of suicide-related thoughts and behaviors during the assessment period. The interview includes definitions and suggested questions to solicit the type of information needed to determine if a suicide-related thought or behavior has occurred. The C-SSRS has a "baseline" version, which will be completed at Visit 1 (Screening), and a "since last visit" version that will be completed at all subsequent visits.

There are a maximum of 19 items to be completed: 7 that are required, 10 potential additional items if there is a positive response to a required item, and 2 items for suicide/suicide behavior present during the interview. The C-SSRS uses dichotomous scales (i.e., yes or no), Likert scales, and text or narrative are used to further describe the thoughts or behaviors.

Subjects are considered to have had suicidal ideation during a time period (lifetime, last 6 months, since last visit) if they respond "Yes" to any of the 5 suicidal ideation questions listed below for that time period:

- Wish to be dead (Category 1)
- Non-specific active suicidal thoughts (Category 2)
- Active suicidal ideation with any methods (Category 3)
- Active suicidal ideation with some intent to act, without specific plan (Category 4)
  Active suicidal ideation with specific plan and intent (Category 5)

Subjects are considered to have had suicidal behavior during a time period if they respond "Yes" to any of the 5 suicidal behavior questions listed below for that time period:

- Completed suicide
- Actual attempt
- Interrupted attempt
- Aborted attempt
- Preparatory acts or behavior

Subjects are considered to have suicidal ideation or behavior during a time period if they responded "Yes" to any of the suicidal ideation or behavior questions for that time period.

A subject's maximum suicidal ideation score is the maximum suicidal ideation category to which they responded "Yes". [redacted]

6.1.8. Data Adjustments/Handling/Conventions

All collected data will be presented in listings. Data not subject to analysis according to this plan will not appear in any tables or graphs but will be included only in the data listings.

All *P* values will be displayed in 4 decimals and rounded using standard scientific notation (e.g., 0.XXXX). If a *P* value less than 0.0001 occurs it will be shown in tables as < 0.0001; similarly, if a *P* value greater than 0.9999 occurs, it will be shown in tables as > 0.9999.

Medical history and AEs (all types) will be coded using the most recent version of the Medical Dictionary for Regulatory Activities (MedDRA), version 24.1. Prior and concomitant medications will be coded using the World Health Organization Drug Dictionary (WHO-DD) Enhanced Anatomical Therapeutic Chemical (ATC) classes and preferred terms, version WHODrug B3 Global Sep 2021.

For quantitative laboratory values reported as '<X' or '≤X', 1/2 times the lower limit of quantitation will be used for analysis. For quantitative laboratory values reported as '>X' or '≥X', 3/2 times the upper limit of quantitation will be used for analysis.

If partial dates occur for treatment emergent AEs, prior medications, or concomitant medications, the convention for replacing missing information for the purpose of statistical analysis is as follows.

Adverse Events

- If the event start date is completely missing, or if the subject was not treated, no imputation will be conducted.

- If the event start date is missing day and month, the treatment start date is not missing, and event start year is the same as the treatment start year, set the event start date to the treatment start date.

- If the event start date is missing day and month and the treatment start date is missing, the event start year is not the same as the treatment start year, or if there is information indicating that the event ended before the treatment start date (e.g., the event end date is before the treatment start date), then set the event start month and day to January 1st.

- If only the event start day is missing, the treatment start date is not missing, and the event start month and year are the same as the treatment start month and year, set the event start day to the treatment start day.

- If only the event start day is missing and the treatment start date is missing, or the event start month and year is not the same as the treatment start month and year, or if there is information to indicate that the event ended before the treatment start date, then set the event start day to the first day of the start month.

- End dates will not be imputed.

Prior and Concomitant Medications

Medication start and stop dates will not be imputed. Instead, available information will be used to assign the following categories to each medication: 1) prior versus concomitant versus both, 2) occurring before versus after Visit 3 (Treatment). If a missing date part leads to uncertainty in how to assign categories, the medication will be considered concomitant and to have been taken after Visit 3 (Treatment) dosing.

For example, if the medication start year or end year is after the start of treatment, the medication will be considered concomitant even if the start month and/or day are missing. Likewise, if the medication start year occurs in the same year as the treatment start year, and the medication start month or end month is after the treatment start month, then the medication will be considered concomitant even if the medication start or end day is missing.

7. Study Subjects and Demographics

7.1. Disposition of Subjects and Withdrawals

The numbers of subjects randomized, completing, withdrawing (including reasons), and assigned to each analysis population will be tabulated by treatment group and overall for the All Subjects population.

7.2. Protocol Violations and Deviations

Protocol deviations will be tracked, recorded, and reviewed prior to database lock, following the Protocol Deviation Guidance Plan for this study. The deviation list will be stored in the Remarque system and transferred to the biostatistics team for analysis.

Protocol deviations will be classified as "Important" or "Non-Important". An important deviation poses a possible safety issue to the participant, or it has a potential impact on the statistical analysis of the clinical data. A non-important deviation is identified as any protocol deviation that does not meet the criteria for an important deviation.

Important deviations include but are not limited to:

- Violation of inclusion/exclusion criteria
- Randomization or dosing error
- Use of prohibited medications
- Assessment conducted by unqualified personnel
- Missing baseline efficacy or safety assessments

Non-important deviations include but are not limited to:

- Lack of data collection due to subject non-compliance
- Subjects who fail to make scheduled visits
- Extension of screening period to clarify diagnoses or verify laboratory values
- Incorrect storage of study drug
- Failure to follow up on subjects who drop out

Subjects with at least 1 deviation classified as "Important" and suspected to have an impact on the statistical analysis of study endpoints will be excluded from the PP Population. The final decision regarding classification of deviations and inclusion and exclusion of participants from the analysis populations will be based on a final listing of protocol deviations. Subject inclusion/exclusion in analysis populations will be determined during a blinded review meeting before study data are unblinded. Clinical and Biostatistics team members will provide input to the decision and the final composition of the analysis populations will be approved by the Sponsor.

Protocol deviations will be summarized by treatment group, subjects not randomized and overall subjects for the All Subjects population. Data will also be presented in a by-subject listing.

Additionally, inclusion and exclusion criteria not met and reasons for screen failures will be listed.

7.3. Demographics and Other Baseline Characteristics

Demographic variables will include race, ethnicity, age (based on date of informed consent), age categories (18-35 years, 36-55 years, 56-65 years), age at onset (from SAD Diagnostic Checklist), sex, height, weight, and body mass index (BMI).. Baseline variables will include average SUDS score at Visit 2 (Baseline), pre-IP SUDS score at Visit 2 (Baseline), average anticipatory SUDS score at Visit 2 (Baseline), and LSAS total score at Visit 1 (Screening).

Demographic and baseline characteristic data will be summarized for the Safety, ITT and PP Populations by treatment group (including subjects not randomized for the Safety population only) and overall subjects. Data will also be presented in a by-subject listing.

Medical and psychiatric history will be summarized for the Safety Population by treatment group and overall subjects, classified by system organ class (SOC) and preferred term. Data will also be presented in a by-subject listing, which will include SOC, preferred term, and the verbatim term.

Prior medications will be summarized by treatment group, subjects not randomized and overall subjects, by the number and percentage of subjects taking each medication and classified by ATC class and preferred term. This analysis will be conducted for the Safety population. Data will also be presented in a by-subject listing, which will include ATC Level 2 classification, preferred name, and verbatim name.

Results from the Mini-International Neuropsychiatric Interview (MINI) and SAD Diagnostic Checklist will be tabulated by treatment group, subjects not randomized and overall subjects for the Safety population and presented in a by-subject listing.

7.4. Exposure and Compliance

Drug and placebo will be supplied in identical [redacted] Each spray delivers 1.6 µg per 100 µL spray. At Visit 2 (Baseline) and Visit 3 (Treatment), site personnel will instruct the subject to self-administer the IP by 1 spray into each nostril (right and left nasal passages), for 2 total sprays per dose; thus, 1 dose equals 2 sprays, or 3.2 µg. All vials dispensed at Visit 2 (Baseline) will contain placebo nasal spray solution.

Total number of doses of placebo and PH94B will be summarized by treatment group and subjects, by study visit for the Safety Population. Data will also be presented in a by-subject listing.

8. Efficacy Analysis

All efficacy analyses will be performed on the ITT Population (as defined in Section 5) unless otherwise noted. The order of testing for the primary and secondary endpoints is discussed in Section 6.1.3. Exploratory endpoints are not included in the testing strategy.

8.1. Primary Efficacy Analysis – Average Treatment SUDS Scores

The primary endpoint will be tested for the following hypothesis:

H_{10}: Mean change in average SUDS score from Visit 2 (Baseline) to Visit 3 (Treatment) does not differ between PH94B and placebo.

H_{11}: Mean change in average SUDS score from Visit 2 (Baseline) to Visit 3 (Treatment) does differ between PH94B and placebo.

The estimand for the primary efficacy endpoint has the following components:

Attribute	Definition
Treatment	The treatment of interest is PH94B, 3.2 µg administered as an intranasal solution (a 1.6 µg spray to each nostril per dose).
Population	The population of interest is adult subjects with SAD as defined by the protocol inclusion and exclusion criteria.
Variable of Interest	The variable of interest to be measured is the change from Visit 2 (Baseline) to Visit 3 (Treatment) in average SUDS score during a 5-minute public speaking challenge.
Intercurrent Events	Intercurrent events are expected to be rare. Any intercurrent event that does occur will be handled using the treatment policy strategy, i.e., included in the treatment regimen under evaluation.
Population Summary	The difference of means from Visit 2 (Baseline) to Visit 3 (Treatment) in SUDS will be estimated for each treatment group. PH94B will be compared to placebo using differences in LS group means.

8.1.1. ANCOVA Analysis of the Primary Endpoint

For each subject at each public speaking challenge, average SUDS scores and change from baseline (Visit 2 [Baseline] to Visit 3 [Treatment]) in average SUDS scores will be calculated as described in Section 6.1.7. Descriptive summaries of average SUDS scores and change from baseline in average SUDS scores will be provided for the primary endpoint based on the observed data.

An ANCOVA model will be used to test the null hypothesis that there is no difference in change from baseline average SUDS scores between PH94B- and placebo-treated subjects. Treatment group and analysis site will be included as factors and baseline average SUDS score as a covariate. The change from baseline LS means with standard error (SE), 95% CI for the LS means, *P* value for testing if the LS mean is 0, LS mean difference between treatment groups (PH94B minus placebo) with SE, 95% CI for the LS mean difference, and *P* value for testing if the treatment LS means are equal will be presented.

It is expected that all randomized study subjects will complete the Visit 3 (Treatment) assessments. If, for unanticipated reasons (intercurrent events), the fraction of subjects with missing change from baseline to Visit 3 (Treatment) average SUDS scores is greater than 0.05, change from baseline to Visit 3 (Treatment) average SUDS scores will be imputed using the fully conditional specification approach to multiple imputation as described in Section 6.1.4. If the fraction of missing change scores is ≤0.05, missing change from baseline to Visit 3 (Treatment) in average SUDS scores will be imputed with treatment group mean change scores.

8.1.2. Complete Case Sensitivity Analysis of the Primary Efficacy Endpoint

If imputation of missing data is required, a complete case analysis will be performed in the ITT population. The complete case analysis will be exactly parallel to the planned primary efficacy analysis, except it will exclude any participants with missing average SUDS scores at Visit 3 (Treatment), using only observed responses. Descriptive summaries will be provided for the primary endpoint based on the complete-case data.

8.1.3. Sensitivity Analysis of the Primary Endpoint in the Per Protocol Population

A sensitivity analysis of the primary endpoint will be performed in the PP Population. The analysis will be exactly parallel to that described in Section 8.1.1 except that imputation will not be necessary. All subjects in the PP Population will have non-missing average SUDS scores at Visit 2 (Baseline) and Visit 3 (Treatment).

8.2. Secondary Efficacy Analysis

The secondary endpoint will be tested for the following hypothesis:

H_{20}: The proportion of CGI-I responders at Visit 3 (Treatment) does not differ between PH94B and placebo.

H_{21}: The proportion of CGI-I responders at Visit 3 (Treatment) does differ between PH94B and placebo.

The estimand for the secondary efficacy endpoint has the following components:

Attribute	Definition
Treatment	The treatment of interest is PH94B, 3.2 µg administered as an intranasal solution (a 1.6 µg spray to each nostril per dose).
Population	The population of interest is adult subjects with SAD as defined by the protocol inclusion and exclusion criteria.
Variable of Interest	The variable of interest to be measured is the CGI-I score at Visit 3 (Treatment).
Intercurrent Events	Intercurrent events are expected to be rare. Any intercurrent event that does occur will be handled using the treatment policy strategy, i.e., included in the treatment regimen under evaluation.
Population Summary	The proportion of subjects with CGI-I scores of 1[redacted] or 2[redacted] for PH94B will be compared to placebo using a comparison of binomial proportions.

8.2.1. Comparison of Binomial Proportions Analysis of the Secondary Endpoint

The proportion of CGI-I responders in each treatment arm will be computed as described in Section 6.1.7.3.

The PH94B and placebo proportions of CGI-I responders will be compared using a normal approximation (Wald) test for the difference between 2 binomial proportions. The proportion of responders in each treatment group, difference in proportions of responders between treatment groups, 95% CIs for the difference in proportions, and *P* value for the Wald test for equality of proportions will be presented.

Additionally, CGI-I scores will be summarized using frequencies and percentages by treatment group.

8.3. Exploratory Efficacy Analysis

All exploratory endpoints will be summarized, and *P* values will be included for descriptive purposes only.

The following analyses will be performed on the observed data for the ITT Population. All derived variables will be computed as described in Section 6.1.7.

1. The PH94B and placebo proportions of PGI-C responders will be compared using a parallel analysis to that described for the CGI-I endpoint (Section 8.2.1).
2. Change in pre-IP, post-IP, and average anticipatory SUDS scores from Visit 2 (Baseline) [redacted]
3. [redacted]
4. [redacted]
5. Proportion of SUDS responders (subjects with a ≥20-point improvement in average SUDS score from Visit 2 [Baseline] to Visit 3 [Treatment]) will be compared between the PH94B and placebo arms using a normal approximation test for a difference in binomial using a parallel analysis to that described for the CGI-I endpoint (Section 8.2.1).
6. [redacted]

[redacted]

8.4. Subgroup Analyses of Efficacy Variables

The primary and secondary endpoints will be summarized by subgroups of interest [redacted]
Racial categories will be combined into larger categories should there be fewer than 5 subjects in a particular category within a treatment arm. Analyses will be performed on the observed data.

Subgroup Analyses of Primary Efficacy Endpoint

The primary efficacy endpoint of change in average SUDS scores from Visit 2 (Baseline) to Visit 3 (Treatment) will be compared between PH94B- and placebo-treated subjects [redacted]

Subgroup analyses of the primary efficacy endpoint will be analogous to the ANCOVA analysis outlined in Section 8.1.1 but imputation will not be performed. [redacted]

Subgroup Analyses of Secondary Efficacy Endpoint

The secondary efficacy endpoint of proportion of CGI-I responders will be compared between PH94B- and placebo-treated subjects within each of the subgroups listed above. Subgroup analyses will be compared using a normal approximation (Wald) test for the difference between 2 binomial proportions as described in Section 8.2.1.

9. Safety and Tolerability Analysis

Safety in this study will be evaluated based on data collected on AEs, clinical laboratory assessments, vital signs, 12-lead ECGs, suicidality, level of depression, and physical examinations including examination of the nasal passages.

All safety analysis reporting will be based on the Safety Population (as defined in Section 5). No formal statistical testing will be conducted for the safety analyses. Descriptive statistics will be used to evaluate safety data. Summaries will present data by treatment group and overall subjects. Listings will list safety data by treatment group and subject.

9.1. Adverse Events

Treatment emergent AEs (TEAEs) are defined as:

- AEs with onset at the time of or following the start of treatment with IP through Visit 4 (Follow-Up) or
- AEs starting prior to the start of treatment with IP but increasing in severity or relationship at the time of, or following, the start of treatment with IP through Visit 4 (Follow-Up).

Any medical condition that is present at the time that the subject is screened will be considered as baseline and not reported as an AE. However, if the study subject's condition deteriorates at any time during the study, it will be recorded as an AE.

The Investigator will record all reportable events with start dates occurring any time after informed consent is obtained until 7 days after the last day of study participation. Events will be followed for outcome information until resolution or stabilization. Missing and partially missing start dates will be imputed for the purpose of analysis as described in Section 6.1.8.

All AEs, TEAEs, and serious AEs (SAEs) will be coded using the MedDRA dictionary, as discussed in Section 6.1.8.

All AE data will be summarized for 2 periods: before Visit 3 (Treatment) dosing and during or after Visit 3 (Treatment) dosing. The reporting is done in this manner due to the difference in study drug exposure during these 2 periods. All subjects will receive placebo at Visit 2 (Baseline), and all subjects will receive their randomized treatment at Visit 3 (Treatment). Thus, between Visit 2 (Baseline) up to dosing at Visit 3 (Treatment), subjects will have only been exposed to placebo. From Visit 3 (Treatment) dosing and onward, subjects will have been exposed to either active drug (PH94B) or placebo.

All AE data will be categorized as Severe > Moderate > Mild. Relationship of all AEs to IP will be categorized as related or not related. In the summaries showing severity and relationship to study drug, the event with the maximum severity or strongest relationship to study drug will be reported. For the purposes of analysis, if a particular event is missing the severity, it will be categorized as severe, and it if it missing the relationship, it will be categorized as related.

An overall summary of TEAEs will be provided; this will present number and percent of subjects who reported at least 1 of the following: TEAE (including any TEAE, TEAEs by maximum severity, and related TEAEs), TEAE leading to discontinuation of the study drug, TEAE leading to discontinuation from the study, serious TEAEs, and SAEs leading to death.

In addition to the overall summary, summaries of the number and percentage of subjects with TEAEs will be displayed for each treatment group by SOC and preferred term. Summaries of TEAEs by severity (Severe > Moderate > Mild) and relationship to IP (Related > Not Related) will also be provided. Serious TEAEs and TEAEs resulting in discontinuation of IP will be summarized separately in a similar manner.

Subject listings of AEs, SAEs, and AEs causing discontinuation of IP will be produced. Adverse events that are treatment emergent will be flagged.

The AE summary tables will include counts of subjects. Therefore, if a subject experiences more

than 1 episode of a particular AE, the subject will be counted only once for that event. If a subject has more than 1 AE that is coded to the same preferred term, the subject will be counted only once for that preferred term. Similarly, if a subject has more than 1 AE within an SOC, the subject will be counted only once in that SOC. This will be handled separately for each time period (Before Visit 3 Dosing and During or After Visit 3 Dosing).

9.1.1. Adverse Events Leading to Withdrawal

A summary of incidence rates (frequencies and percentages) of TEAEs leading to discontinuation of study drug will be summarized by period (Before Visit 3 Dosing and During or After Visit 3 Dosing), treatment group and overall subjects, and by SOC and preferred term for the Safety Population. No inferential statistical tests will be performed.

A data listing of AEs leading to discontinuation of study drug will also be provided, displaying details of the event(s) captured on the eCRF.

9.1.2. Deaths and Serious Adverse Events

Any deaths that occur during the study will be listed. Serious AEs will be listed. Serious TEAEs and serious TEAEs related to IP will be tabulated by SOC and preferred term and presented by period (Before Visit 3 Dosing and During or After Visit 3 Dosing), treatment group, and overall subjects.

9.2. Clinical Laboratory Evaluations

Descriptive summaries of observed values and change from baseline values will be presented for continuous hematology, chemistry, and urinalysis results for each treatment group, subjects not randomized and overall subjects at each study visit. Categorical urinalysis results will be summarized using frequencies by treatment group, subjects not randomized and overall subjects and by study visit.

The number of subjects with clinical laboratory values categorized as below, within, or above normal ranges will be tabulated showing change from baseline (shift tables) for each clinical laboratory analyte by treatment group and overall subjects and by study visit.

Laboratory values that are outside the normal range will also be flagged in the data listings and presented with the corresponding normal ranges. Any out-of-range values that are identified by the investigator as being clinically significant will be listed as an adverse event.

For women only, pregnancy status will be determined by evaluation of urine pregnancy test. Data on pregnancy status will be listed by subjects. Subjects who are pregnant are excluded from the study. Subjects who become pregnant during the study will be discontinued.

9.3. Vital Signs

Descriptive summaries of observed values and changes from baseline will be calculated for systolic blood pressure (mmHg), diastolic blood pressure (mmHg), heart rate (beats/min), respiratory rate (breaths/min), and temperature (°C) by treatment group, subjects not randomized and overall subjects and by study visit.

The number of subjects with vital signs values categorized as below, within, or above normal ranges will be tabulated showing change from baseline (shift tables) for each parameter by visit by treatment group and overall.

9.4. 12-Lead Electrocardiograms

The number and percentage of subjects with normal and abnormal 12-lead ECG findings will be summarized for each treatment group, subjects not randomized and overall subjects at each study visit. Abnormal results will be grouped as clinically significant and not clinically significant.

Descriptive summaries for the observed results and change from baseline will be presented for 12-lead ECG measures of PR interval (msec), QRS interval (msec), and heart rate for each treatment group and overall subjects at each study visit. If triplicate reads are taken, the mean of the reads will serve as the observed value at the specific visit.

A comparison of QT and QTcF results will be presented. Summary statistics for baseline values at Visit 1 (Screening) and Visit 4 (Follow-Up) will be displayed by treatment group and overall subjects for QT interval (msec) and QT interval corrected for heart rate using Fridericia's formula (QTcF) (msec). In addition, the number and percent of subjects in each treatment group who experienced a change >30 msec or a change >60 msec will be presented.

9.5. Physical Examination

The number and percentage of subjects with normal and abnormal findings in the complete physical examination, including nasal passages examination, will be displayed for each treatment group, subjects not randomized and overall subjects by study visit for the Safety Population. Abnormal results will be grouped as clinically significant and not clinically significant.

9.6. Concomitant Medication

Prior and concomitant medications will be summarized descriptively by treatment using counts and percentages for the Safety Population. Medications will be summarized by treatment group, by the number and percentage of subjects taking each medication and classified by ATC class and preferred term. Data will also be presented in a by-subject listing, which will include ATC Level 2 classification, preferred term and verbatim name.

Prior medications will be presented separately from concomitant medications. All medications taken in the month prior to Visit 1 (Screening) or in the time interval between Visit 1 (Screening) and Visit 2 (Baseline) will be considered prior medications, whether or not they were stopped

before Visit 2 (Baseline). Any medications continuing or starting after Visit 2 (Baseline) will be considered concomitant. If a medication starts before Visit 2 (Baseline) and continues after Visit 2 (Baseline) it will be considered concomitant.

Medications will be coded using WHO-DD, as described in Section 6.1.8.

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

The number and percentage of subjects reporting "Yes" for any of the 5 suicidal ideation questions (Categories 1 to 5) or any of the 5 suicidal behavior questions (Categories 6 to 10) will be displayed for each treatment group and overall subjects, along with the number and percentage reporting suicidal ideation OR behavior (Categories 1 to 10). The number and percentage of subjects with treatment emergent suicidal ideation, treatment emergent suicidal behavior, and treatment emergent suicidal ideation or behavior as defined in Section 6.1.7.8 will also be presented.

9.8. Hamilton Depression Scale (HAM-D)

Total scores and change from baseline will be summarized by visit and by treatment group and overall subjects for the Safety Population.

10. Other Planned Analyses

[redacted]

10.2. Coronavirus 2019 (COVID-19) Impact

The SARS-CoV-2 virus and variants may still be a threat and safety precautions may still be in place during study enrollment. To assess the impact of the COVID-19 pandemic on study procedures, study staff will complete the COVID-19 Impact questionnaire. Should a study visit be affected by COVID-19, study staff will record how the visit was affected (missed, abbreviated, delayed, performed remotely, or otherwise deviating from the planned format in the protocol) and which procedures were impacted. All COVID-19 protocol deviations will be provided in a separate listing. This listing will include the subjects, affected visits, and affected assessments.

11. Changes from Planned Analysis

Study staff will complete the Coronavirus 2019 (COVID-19) Impact questionnaire to assess the impact of the COVID-19 pandemic on study procedures. The COVID-19 Impact questionnaire is described in Section 10.2. All COVID-19 protocol deviations will be provided in a separate listing.

The All Subjects Population was explicitly defined for programming purposes and will be used for summaries of subject disposition and protocol deviations.

The PP Population was defined and will be used for a sensitivity analysis of the primary endpoint.

Race was added as a subgroup of interest for the subgroup analyses.

12. References

ASA. (2018) Ethical Guidelines for Statistical Practice. Prepared by the Committee on Professional Ethics, April 2018. http://www.amstat.org/about/ethicalguidelines.cfm

Carpenter JR and Kenward MG. *Multiple Imputation and its Application (Statistics in Practice)*. Chichester, West Sussex, UK: John Wiley & Sons; 2013.

Graham JW, and Olchowski AE, Gilreath TD. (2007) "How Many Imputations are Really Needed? Some Practical Clarifications of Multiple Imputation Theory." *Prev Sci.* 2007;8(3): 206-213.

ICH (1998). ICH Harmonised Tripartite Guideline. Statistical Principles for Clinical Trials E9; 1998. https://database.ich.org/sites/default/files/E9_Guideline.pdf

Little R and Rubin D. (1987). Statistical Analysis with Missing Data, New York: Wiley.

RSS. (2014) The Royal Statistical Society: Code of Conduct, 2014. https://rss.org.uk/about/policy-and-guidelines/code-of-conduct/.

13. Tables, Listings, and Figures

All listings, tables, and graphs will have a header showing the sponsor company name and protocol and a footer showing the version of SAS, the file name and path, and the source of the data (e.g., listing number).

The following are planned summary tables for protocol number PH94B-CL032. The table numbers and page numbers are place holders only and will be determined when the tables are produced.

13.1. Demographic Data Summary Tables and Figures

Table 3: Demographic Data Summary Tables and Figures

Table Number	Population	Table Title/Summary
14.1 Demographic Data Summary Tables and Figures		
Table 14.1.1.1.1	All Subjects	Subject Enrollment and Disposition
Table 14.1.1.1.2	ITT	Subject Enrollment and Disposition
Table 14.1.1.2	All Subjects	Protocol Deviations
Table 14.1.2.1	Safety	Demographics and Baseline Characteristics
Table 14.1.2.2	ITT	Demographics and Baseline Characteristics
Table 14.1.2.3	PP	Demographics and Baseline Characteristics
Table 14.1.3	Safety	Medical and Psychiatric History by System Organ Class and Preferred Term
Table 14.1.4	Safety	Prior Medications by ATC Class and Preferred Term
Table 14.1.5	Safety	Baseline Summary of Mini International Neuropsychiatric Interview (MINI)
Table 14.1.6	Safety	Baseline Summary of Social Anxiety Disorder (SAD) Diagnostic Checklist
Table 14.1.7	Safety	Exposure to Investigational Product

13.2. Efficacy Data

Table 4: Efficacy Data

Table Number	Population	Table Title / Summary
Table 14.2. Efficacy Tables		
Table 14.2.1.1	ITT	ANCOVA Change from Baseline in Average SUDS Scores – Primary Analysis
Table 14.2.1.2	ITT	ANCOVA Change from Baseline in Average SUDS Scores – Sensitivity Analysis Using Complete Cases
Table 14.2.1.3	PP	ANCOVA Change from Baseline in Average SUDS Scores by Analysis Visit – Per Protocol Sensitivity Analysis
Table 14.2.2	ITT	Analysis of the Proportion CGI-I Responders – Secondary Analysis
Table 14.2.3.1	ITT	[redacted]
Table 14.2.3.2	ITT	[redacted]

Table Number	Population	Table Title / Summary
Table 14.2.3.3	ITT	[redacted]
Table 14.2.3.4	ITT	[redacted]
Table 14.2.4.1	ITT	[redacted]
Table 14.2.4.2	ITT	[redacted]
Table 14.2.4.3	ITT	[redacted]
Table 14.2.4.4	ITT	[redacted]
Table 14.2.5.1	ITT	Analysis of the Proportion PGI-C Responders – Exploratory Analysis
Table 14.2.5.2	ITT	[redacted]
Table 14.2.5.3	ITT	[redacted]
Table 14.2.5.4	ITT	[redacted]
Table 14.2.5.5	ITT	[redacted]
Table 14.2.5.6	ITT	[redacted]
Table 14.2.5.7	ITT	Analysis of the Proportion SUDS Responders – Exploratory Analysis

13.3. Safety Data

Table 5: Safety Data

Table Number	Population	Table Title / Summary
14.3.1 Displays of Adverse Events		
Table 14.3.1.1	Safety	Overall Summary of Treatment Emergent Adverse Events
Table 14.3.1.2.1	Safety	Treatment Emergent Adverse Events by System Organ Class and Preferred Term
Table 14.3.1.2.2	Safety	Treatment Emergent Adverse Events by Preferred Term
Table 14.3.1.3	Safety	Treatment Emergent Adverse Events by System Organ Class, Preferred Term, and Maximum Severity

Table Number	Population	Table Title / Summary
Table 14.3.1.4	Safety	Treatment Emergent Adverse Events by System Organ Class, Preferred Term, and Maximum Relationship to Investigational Product
14.3.2 Summary of Deaths, Other Serious and Significant Adverse Events		
Table 14.3.2.1	Safety	Treatment Emergent Adverse Events Leading to Study Discontinuation by System Organ Class and Preferred Term
Table 14.3.2.2	Safety	Serious Treatment Emergent Adverse Events by System Organ Class and Preferred Term
Table 14.3.2.3	Safety	Serious Treatment Emergent Adverse Events Related to Investigational Product by System Organ Class and Preferred Term
14.3.3 Narratives of Deaths, Other Serious and Certain Other Significant Adverse Events		
Table 14.3.3.1	Safety	Subject Listing of Serious Adverse Events
Table 14.3.3.2	Safety	Subject Listing of Adverse Events Leading to Discontinuation of Investigational Product
Table 14.3.3.3	Safety	Subject Listing of Adverse Events Leading to Death
14.3.4 Abnormal Laboratory Values		
Table 14.3.4.1	Safety	Subject Listing of Abnormal Laboratory Values: Clinical Chemistry and Thyroid
Table 14.3.4.2	Safety	Subject Listing of Abnormal Laboratory Values: Hematology
Table 14.3.4.3	Safety	Subject Listing of Abnormal Laboratory Values: Urinalysis
14.3.5 Laboratory Data Summary Tables		
Table 14.3.5.1.1	Safety	Clinical Chemistry and Thyroid: Observed Results and Change and Percent Change from Baseline by Study Visit
Table 14.3.5.1.2	Safety	Clinical Chemistry and Thyroid: Shifts from Baseline Relative to the Normal Range
Table 14.3.5.2.1	Safety	Hematology: Observed Results and Change and Percent Change from Baseline by Study Visit
Table 14.3.5.2.2	Safety	Hematology: Shifts from Baseline Relative to the Normal Range by Study Visit
Table 14.3.5.3.1	Safety	Quantitative Urinalysis: Observed Results and Change and Percent Change from Baseline by Study Visit
Table 14.3.5.3.2	Safety	Qualitative Urinalysis: Summary of Results by Study Visit
Table 14.3.5.3.3	Safety	Quantitative and Qualitative Urinalysis: Shifts from Baseline Relative to the Normal Range
14.3.6 Other Safety Data Summary Tables		
Table 14.3.6.1.1	Safety	Vital Signs: Observed Results and Change from Baseline by Study Visit

Table Number	Population	Table Title / Summary
Table 14.3.6.1.2	Safety	Vital Signs: Shifts from Baseline Relative to Normal Range by Study Visit
Table 14.3.6.2.1	Safety	12-Lead Electrocardiogram: Observed Results and Change from Baseline by Study Visit
Table 14.3.6.2.2	Safety	12-Lead Electrocardiogram: Summary of 12-Lead Electrocardiogram Interpretation by Study Visit
Table 14.3.6.2.3	Safety	12-Lead Electrocardiogram: Summary of QT and QTcF Results by Study Visit
Table 14.3.6.3	Safety	Summary of Physical Examination by Study Visit
Table 14.3.6.4	Safety	Summary of Concomitant Medications by ATC Class and Preferred Term
Table 14.3.6.5	Safety	Summary of Columbia Suicide Severity Rating Scale (C-SSRS) by Study Visit
Table 14.3.6.6	Safety	Hamilton Depression Scale (HAM-D): Observed Results and Change from Baseline by Study Visit

13.4. Other Data Summary Tables

Table 6: Other Data Summary Tables

Table Number	Population	Table Title / Summary
14.5 Other Data Summary Tables		
Table 14.5.1	Safety	Summary of Drug Prediction Questionnaire (DPQ)

13.5. Planned Listing Descriptions

The following are planned data and subject data listings for protocol number PH94B-CL032.

In general, 1 listing will be produced per eCRF domain. All listings will be sorted by treatment, site, and subject number. All calculated variables will be included in the listings.

In all listings a blank line will be placed between each subject. Within a data listing, if an item appears line after line (e.g., repetition of subject number), then only the first occurrence will be displayed. Screen failures will only be presented in Listings 16.2.1.1, 16.2.2.1, and 16.2.2.3; otherwise, only enrolled subjects will be listed.

In data listings, the information for 1 subject will be kept on 1 page, if at all possible, rather than splitting a subject's information across pages.

Table 7: Planned Listings

Listing Number	Population	Listing Title / Summary
16.2 Subject Data Listings		
16.2.1 Subject Discontinuations/Completions		
Listing 16.2.1.1	All Subjects	Subject Disposition
16.2.2 Protocol Deviations		
Listing 16.2.2.1	All Subjects	Inclusion and Exclusion Criteria
Listing 16.2.2.2	All Enrolled Subjects	Protocol Deviations
Listing 16.2.2.3	Screen Failures	Reasons for Screen Failures
16.2.3 Patients/Subjects Excluded from the Efficacy Analyses		
Table 16.2.3.1	All Subjects	Analysis Populations
16.2.4 Demographic Data and Other Baseline Characteristics		
Listing 16.2.4.1	All Subjects	Subject Consent and Demographics
Listing 16.2.4.2	All Subjects	Medical and Psychiatric History
Listing 16.2.4.3	All Subjects	Mini-International Neuropsychiatric Interview (MINI)
Listing 16.2.4.4	All Subjects	[redacted]

Listing Number	Population	Listing Title / Summary
16.2 Subject Data Listings		
16.2.5 Compliance and/or Drug Concentration Data		
Listing 16.2.5.1	All Subjects	Study Drug Dispensation
Listing 16.2.5.2	All Subjects	Study Drug Administration
16.2.6 Individual Efficacy Response Data		
Listing 16.2.6.1	All Subjects	Subjective Units of Distress Scale (SUDS)
Listing 16.2.6.2	All Subjects	Clinical Global Impression – Improvement scale (CGI-I)
Listing 16.2.6.3	All Subjects	Patient Global Impression of Change (PGI-C)
Listing 16.2.6.4	All Subjects	Liebowitz Social Anxiety Scale (LSAS)
Listing 16.2.6.5	All Subjects	[redacted]
16.2.7 Adverse Event Listings (by Subject)		
Listing 16.2.7.1	All Subjects	Adverse Events
16.2.8 Laboratory Values (by Subject)		
Listing 16.2.8.1	All Subjects	Hematology Laboratory Evaluations
Listing 16.2.8.2	All Subjects	Clinical Chemistry and Thyroid Laboratory Evaluations
Listing 16.2.8.3	All Subjects	Urinalysis Laboratory Evaluations
Listing 16.2.8.4	All Enrolled Female Subjects	Pregnancy Test Results
Listing 16.2.8.5	All Subjects	Laboratory Results for Drug Screening
16.2.9 Other Clinical Observations and Measurements (by Subject)		
Listing 16.2.9.1	All Subjects	Vital Signs
Listing 16.2.9.2	All Subjects	12-Lead Electrocardiogram Measurements
Listing 16.2.9.3	All Subjects	Physical Examination Measurements
Listing 16.2.9.4	All Subjects	Quick Olfactory Test (QOT)
Listing 16.2.9.5	All Subjects	Prior and Concomitant Medications
Listing 16.2.9.6	All Subjects	Columbia Suicide Severity Rating Scale (C-SSRS) Screening
Listing 16.2.9.7	All Subjects	Hamilton Depression Rating Scale (HAM-D)
Listing 16.2.9.8	All Subjects	[redacted]
16.2.10 Other Study Measurements or Assessments (by Subject)		
Listing 16.2.10.1	All Subjects	COVID-19 Impact Assessment