Statistical Analysis Plan: RIN-PRN-201

Study Title:	An Observational Study to Characterize Patient Global Impression Questions for Activity-induced Symptoms in Patients with Pulmonary Arterial Hypertension
Study Number:	RIN-PRN-201
Study Phase:	2
Product Name:	Treprostinil for PRN Inhalation Spray
Indication:	Treatment of Pulmonary Arterial Hypertension
IND Number:	138,142
Sponsor:	United Therapeutics Corp. 55 TW Alexander Drive Research Triangle Park, NC 27709 United States
Author:	[redacted]
Final Date:	30 October 2019

TABLE OF CONTENTS

MODIFICATION HISTORY

Unique Identifier for this Version	Date of the Document Version	Author(s)	Significant Changes from Previous Authorized Version
1.0	30 October 2019	[redacted]	Initial Version

LIST OF ABBREVIATIONS AND DEFINITION OF TERMS

AE	Adverse event
ANOVA	Analysis of variance
CSR	Clinical Study Report
DBP	Diastolic blood pressure
eCRF	electronic Case Report Form
EDC	Electronic data capture
ERA	Endothelin receptor antagonist
FAS	Full Analysis Set
FDA	Food and Drug Administration
HIV	Human immunodeficiency virus
HR	Heart rate
ICF	Informed Consent Form
ICH	International Council for Harmonisation
ISWT	Incremental Shuttle Walk Test
MedDRA	Medical Dictionary for Regulatory Activities
PAH	Pulmonary arterial hypertension
PAP	Psychometric Analysis Population
PDE5-I	Phosphodiesterase type 5 inhibitor
PGI-S	Patient Global Impression of Severity
PPROT	Per-Protocol
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SBP	Systolic blood pressure
SOC	System Organ Class
TEAE	Treatment-emergent adverse event
TFs	Tables and Figures
UPI	Unique Patient Identification
WHO	World Health Organization

1 INTRODUCTION

This Statistical Analysis Plan (SAP) describes the statistical methodology, as well as the rules and conventions, to be employed in the presentation and analyses of endpoint and safety data from Study RIN-PRN-201. The methods described herein conform to the recommendations outlined in the International Council for Harmonisation (ICH) E9 guideline entitled, "Guidance for Industry: Statistical Principles for Clinical Trials" (1998) and the ICH E3 guideline entitled, "Guidance for Industry: Structure and Content of Clinical Study Reports" (1996).

The various sections of this document describe the planned tabulated summaries, graphics, and other figures (termed Tables and Figures [TFs]) to be presented in the Clinical Study Report (CSR). A separate document will contain mock-ups of unique TFs.

This SAP is based on the United States RIN-PRN-201 Original Protocol (dated 31 January 2019). It supersedes the statistical considerations identified in the study protocol. Any major deviations in this SAP from the statistical considerations of the protocol will be described and justified within the CSR. Lung Biotechnology will follow this SAP in analyzing the study data. If during the course of the trial, the protocol or the set of electronic Case Report Forms (eCRFs) is modified in a way that affects the planned statistical analyses, a revised version of the SAP will be developed. Only minor changes in the analyses shall be applied following the final version of the SAP. The CSR shall document minor changes.

2 STUDY OBJECTIVES

2.1 PRIMARY OBJECTIVE

The primary objective of this study is to characterize the psychometric properties and performance of Patient Global Impression of Severity (PGI-S) questions, designed to measure the severity, character, and duration of patients' self-reported symptoms of pulmonary arterial hypertension (PAH) following effort.

2.2 SECONDARY OBJECTIVES

The secondary objectives of this study are:

- To characterize the relationship of PGI-S questions with:
  - Borg dyspnea score

- Oxygen saturation
- Heart rate (HR)
- HR recovery at 1 minute after the Incremental Shuttle Walk Test (ISWT)

- To characterize the recovery profile for activity-induced symptoms.

2.3 SAFETY OBJECTIVES

Safety will be evaluated based on adverse events (AEs) and vital sign measurements.

3 STUDY DESIGN

3.1 GENERAL DESCRIPTION

This is an observational, multicenter, single-day, Phase 2 study. This study will include a 14-day Screening Period and a Study Day 1 clinic visit.

Subjects will be required to perform an ISWT to induce symptoms of PAH. The total number of shuttles completed by a subject during the Screening ISWT will be the maximum targeted for that subject during the remaining ISWTs in the study. Subjects' severity of self-reported symptoms of PAH will be measured from pre-activity, immediately after the activity, and through the 30-minute recovery. Subjects will be asked about their PAH symptoms using 3 PGI-S questions that address their overall PAH symptoms, shortness of breath, and physical fatigue.

The subject population includes males and females 18 years and above, with a diagnosis of PAH that is either idiopathic or familial, collagen vascular disease associated PAH, PAH associated with human immunodeficiency virus (HIV) infection, PAH induced by anorexigens/toxins, or PAH associated with repaired congenital systemic-to-pulmonary shunts (repaired ≥1 year). Subjects must be World Health Organization (WHO) Functional Class I, II or III and have shortness of breath upon exertion (exhibits a ≥1-point change in Borg dyspnea score). Subjects are permitted to receive Food and Drug Administration (FDA)-approved PAH therapies if on a stable dose regimen for at least 60 days prior to the Screening Visit. Subjects must sign the Informed Consent Form (ICF) prior to participation in any study procedure.

After Screening, subjects will be assigned to 1 of 2 cohorts based on PAH medications as prescribed by their physician; Cohort A will include subjects who are currently prescribed and

using inhaled treprostinil for the treatment of PAH and Cohort B will include subjects who are taking other PAH medications (instead of inhaled treprostinil).

Subjects in Cohort A will be assigned to 1 of 2 sequences:

Sequence 1:

- Period 1: The ISWT will be initiated 3 to 4 hours after a subject's previous dose of inhaled treprostinil (expected trough level). Activity will be timed to allow all procedures, including recovery, to be completed prior to Period 2.
- Period 2: The ISWT will be initiated within approximately 30 minutes of the previous inhaled treprostinil dose (expected peak level). Following recovery and the Investigator's confirmation of return to baseline status, the subjects will be discharged from the study.

Sequence 2:

- Period 1: The ISWT will be initiated within approximately 30 minutes of the previous inhaled treprostinil dose (expected peak level). Activity will be timed to allow all procedures, including recovery, to be completed prior to Period 2.
- Period 2: The ISWT will be initiated 3 to 4 hours after a subject's previous dose of inhaled treprostinil (expected trough level). Following recovery and the Investigator's confirmation of return to baseline status, the subjects will be discharged from the study.

Subjects in Cohort B will also undergo 2 activity periods:

- Period 1: The ISWT will be initiated approximately 4 hours after the subject's morning dose of PAH medication.
- Period 2: The ISWT will be initiated at least 1 hour following the completion of the previous ISWT. Following recovery and the Investigator's confirmation of return to baseline status, the subjects will be discharged from the study.

The baseline PGI-S and Borg dyspnea score assessments will be conducted prior to (-15 and 0 minutes; pre-ISWT) an activity that is typically expected to induce symptoms of PAH. The subject will initiate activity (ISWT), and these same assessments will be performed immediately at the end of the ISWT (within 1 minute of completing the activity). The PGI-S assessments will also be performed throughout recovery following completion of the ISWT. Continuous pulse oximetry will be performed at each clinic visit starting at each Baseline and through recovery.

The study will consist of the following periods:

Screening: During the Screening Period, subjects who sign the ICF will first undergo all Screening assessments to evaluate eligibility as listed in the inclusion/exclusion criteria. A

window of up to 14 days is permitted to enroll a subject after Screening. When feasible, Screening assessments should be completed on the same day. Rescreening will be allowed with the Sponsor's Medical Monitor approval.

- The ISWT will be performed to determine eligibility and the ability to induce symptoms. Completion of a minimum of 3 shuttles (30 meters) of the test will be required for study eligibility. Pre-ISWT, ISWT conduct, and ISWT recovery will be performed and specific assessments will be conducted as outlined below.
- Pre-ISWT: Baseline PGI-S and Borg dyspnea score assessments completed at -15 and 0 minutes prior to activity inducement of PAH symptoms will be performed. Continuous pulse oximetry, for the assessment of oxygen saturation and HR, will be initiated at the beginning of pre-ISWT and performed until the end of recovery from the ISWT. Vital signs (systolic blood pressure [SBP], diastolic blood pressure [DBP]) will be assessed at the beginning of pre-ISWT.
- ISWT: PGI-S and Borg dyspnea scores will be assessed immediately upon completion of the ISWT. Continuous pulse oximetry, for the assessment of oxygen saturation and HR, will be assessed continuously during the activity.
- ISWT Recovery: After completion of the ISWT, PGI-S assessments will be performed every minute during the first 5 minutes followed by every 5 minutes for the remainder of the 30-minute recovery. Continuous pulse oximetry, for the assessment of oxygen saturation and HR, will be assessed during recovery.

Day 1: The same assessments mentioned above will be performed during 2 ISWTs. PGI-S assessments will be performed as outlined below. At least a 1-hour rest period between the conduct of the 2 ISWTs will be required.

- For subjects who are currently taking inhaled treprostinil (Cohort A), the ISWT will be initiated either 3 to 4 hours (expected trough level) or within 30 minutes (expected peak level) of a subject's previous dose in a cross-over manner. Baseline PGI-S and Borg dyspnea score assessments, completed at -15 and 0 minutes prior to activity inducement of PAH symptoms, will be performed. Continuous pulse oximetry will be performed from pre-ISWT until the end of recovery.
- For subjects who are taking other PAH medications (instead of inhaled treprostinil) (Cohort B), the ISWT will be initiated approximately 4 hours after the subject's morning dose of PAH medication.

Discharge: Following completion of all study procedures, recovery from study symptom-induced activities, and when the Investigator is satisfied that the subject has returned to the baseline status, per pulse oximetry readings, the subject will be discharged from the clinic.

Vital sign assessments (SBP and DBP) will be performed at the beginning and at the end of each clinic visit. All AEs will be documented from the time of informed consent until the time screen failure is documented, the subject discontinues, or the subject completes the study.

A subject may voluntarily withdraw from the study at any time for any reason. A withdrawal of consent would preclude data collection regarding that subject after the date of withdrawal.

3.2 TREATMENTS

This is an observational study; no treatment will be provided in the study.

3.2.1 Patient Identification

Each screened subject will receive a Unique Patient Identification (UPI) number (a unique 7-character alpha-numeric identifier) consisting of a 3-digit clinical site number and a 3-digit subject number, separated by a hyphen. The Sponsor will assign site numbers to each clinical site. The subject number will be assigned by the electronic data capture (EDC) system as a 3-digit number to each subject in sequential order by clinical site at the time of the Screening Visit. For example, the first subject who is screened at Site 301 is assigned the UPI of 301-001, the second, 301-002, etc. Subjects who discontinue from the trial will not be replaced.

3.2.2 Method of Assigning Subjects to Treatment Groups

Cohort A subjects are assigned 1:1 to either Sequence 1 or 2 via the EDC system based on responses during the Screening Visit (specifically, “Is the subject currently prescribed and using inhaled treprostinil for the treatment of PAH?” as “Yes”). A Sponsor biostatistician created and approved the allocation scheme.

3.2.3 Blinding

Not applicable.

4 DESIGN

The primary analysis endpoints are the 3 PGI-S items, measuring overall PAH symptoms, shortness of breath, and physical fatigue.

Approximately 40 subjects will be enrolled to ensure at least 36 evaluable subjects complete the study. A sample size of 36 is expected to provide sufficient power in the context of this study to evaluate change in the PGI-S items. Assuming a change from baseline endpoint with values that may range between -4 and 4, a common standard deviation as the 50th percentile of the range (~2 units) for each PGI-S item, a correlation between post-activity and baseline measurements of 0.75, and a standard deviation of the difference scores as approximately 1.3, a 2-sided 95% confidence interval for a paired mean difference, based on the large sample z statistic, will extend approximately 0.52 units. With a confidence interval length at the mid-point of PGI-S response ratings, reasonable confidence in the point estimates observed for the PGI-S items should be achieved.

5 ENDPOINTS

This section lists the study's observational endpoints and safety variables. Section 7 of the study protocol provides details regarding the description, timing, and collection of the variables. See the Overall Schedule of Times and Events in Section 3.2 of the study protocol for more details.

5.1 OBSERVATIONAL ENDPOINTS

5.1.1 Patient Global Impression of Severity

The primary objective is to characterize the psychometric properties and performance of the 3 PGI-S items, measuring overall PAH symptoms, shortness of breath, and physical fatigue during pre-ISWT, at the end of activity, and through recovery.

The focus of each item, separately, will be the subject's evaluation of:

1. Overall PAH Symptoms: [redacted]
2. Shortness of Breath: [redacted]
3. Physical Fatigue: [redacted]

Subjects will be asked to rate each item based on their level of symptom severity "at this moment" using a 5-point ordinal response scale (1=Not present, 2=Mild, 3=Moderate, 4=Severe, 5=Very severe). The timing of the assessments for Overall PAH Symptoms, Shortness of Breath, and Physical Fatigue scales will be at Screening/Baseline (-15 minutes and at 0 minutes

prior to activity-inducement of PAH symptoms), at completion of the ISWT, and every 1 minute for 5 minutes then every 5 minutes following completion of the ISWT during recovery. The average recovery time of symptoms is unknown but, based on clinician advice, it is expected that recovery should be within 15 to 20 minutes. Assessments for PGI-S will occur for 30 minutes following completion of the ISWT to ensure that the full recovery profile is captured across all subjects observed.

For each PGI-S item, a change from Baseline score will be calculated for the PGI-S items measured from the start of the activity-inducement of PAH symptoms.

5.1.2 Modified Borg Dyspnea Score

The Modified Borg Dyspnea score (see Section 9) will be completed by subjects as follows: Subjects will be asked to rate dyspnea "at this moment" on an 11-point numeric rating scale (0=Nothing at all, 10=Maximal) at Screening/Baseline (-15 minutes and at 0 minutes prior to activity-inducement of PAH symptoms) and at completion of the ISWT.

5.1.3 Oxygen Saturation and Heart Rate

Oxygen saturation and HR will be measured continuously throughout the observation period with measurements extracted and recorded into eCRFs at Screening/Baseline (-15 minutes and at 0 minutes prior to activity-inducement of PAH symptoms), each minute during the ISWT, at completion of the ISWT, and every 1 minute for 5 minutes then every 5 minutes following completion of ISWT during recovery. Key measurements used for analysis will be those that correspond to the measurement of the PGI-S and Borg Dyspnea scores.

5.2 SAFETY VARIABLES

The following safety variables will be assessed:

- AEs
- Vital sign measurements

6 STATISTICAL METHODS

6.1 DATASETS FOR ANALYSES

This study has the following analysis populations:

- All Subjects (All): Subjects who consented to the study protocol.
- Full Analysis Set (FAS) Population: Subjects who enrolled in the study protocol (initiated Day 1 procedures). Tabulations for endpoint analyses will be conducted on this population.
- Per-Protocol (PPROT) Population: Subjects who completed the study protocol, do not have significant protocol deviations in the study, and were appropriately included to investigate the research hypotheses. The Sponsor will assess inclusion of this population at the time of database lock. Endpoint analyses may be conducted from this population if the population is applicable after Sponsor review.
- Safety Population: Subjects who initiated the ISWT at Screening.
- Psychometric Analysis Population (PAP): Subjects who completed the PGI-S items or Modified Borg Scale at any visit or time point.

Summary tables will be presented with the number and percentages of subjects overall and by cohort.

6.2 GENERAL METHODOLOGY

Cohorts A and B will be labeled as “Treprostinil Users” and “Non-Treprostinil Users,” respectively.

Continuous variables will be summarized with standard descriptive statistics such as mean, standard deviation, median, 25th and 75th percentiles, mode, minimum, and maximum. For categorical data, descriptive analyses will be based on the number of subjects and related percentages.

Tabular presentations will be appropriate to the data and, in general, will display 3 columns of results - 1 “Overall” and 1 for each cohort, with the labels “Treprostinil Users” and “Non-Treprostinil Users.”

All tables will be presented in landscape format. For practical reasons, the point size may be reduced to (but not less than) 7 point for tables that contain too much information to fit into a single page. SAS version 9.3 or higher (2011) will be used. The Sponsor’s Biostatistics Department will analyze the endpoint and safety data.

6.3 DEFINITION OF ANALYSIS TIME POINTS

6.3.1 Baseline

For the purposes of computing change from baseline measurements in symptoms as a response to activity, the "baseline" is defined as the value(s) immediately prior to the intervention, eg, ISWT. During recovery, alternate baseline values may be as defined to compute change scores and time variables specific to recovery. For clarity, footnotes will specify the definition of the baseline value being used.

6.3.2 Change from Baseline

For parameters measured at Baseline, the variables of interest are the change from Baseline of the original measurements. Unless otherwise specified, all changes from baseline (for all variables where this is applicable) will be calculated as follows:

> Change from Baseline = Post-baseline – Baseline (average score from the PGI-S scores measured at -15 minutes and at 0 minutes prior to activity-inducement of PAH symptoms)

6.3.3 Study Day

Study day will be calculated relative to Day 1.

6.3.4 Visit Windows

Not applicable as this is a single-day study.

6.4 DATA HANDLING

6.4.1 Handling of Missing Data

Missing data will not be imputed. Ad hoc analyses may be conducted to assess the effects of missing data on the interpretation of the results.

6.5 DISPOSITION OF SUBJECTS

All subjects consented into the study will be tabulated for subject accountability and disposition. The number and percentage of subjects who completed and who discontinued (with reason for discontinuation) will be summarized overall and by cohort.

6.6 PROTOCOL DEVIATIONS

Protocol deviations occurring during the study, major or minor, will be presented for all subjects.

6.7 DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS

6.7.1 Demographic Characteristics

Descriptive statistics will be presented overall and by treatment group for the following demographic data:

- Age (in years)
- Sex (Male, Female)
- Race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino)

6.7.2 PAH Disease Characteristics

Descriptive statistics will be presented overall and by cohort for subjects' PAH history for the following:

- PAH etiology (idiopathic or familial, associated with collagen vascular disease, associated with HIV infection, induced by anorexigens/toxins, or associated with congenital systemic-to-pulmonary shunts [repaired ≥1 year])
- WHO Functional Class (I, II, or III)
- Background PAH therapies (endothelin receptor antagonist [ERA], phosphodiesterase type 5 inhibitor [PDE5-I inhibitor], inhaled treprostinil, riociguat)

6.7.3 Medical History

Significant past or present illnesses identified prior to informed consent will be summarized overall and by cohort. The number and percentage of subjects with any occurrence will be summarized by System Organ Class (SOC; alphabetically) and by preferred term in decreasing order of incidence.

For subjects experiencing the same medical condition multiple times, the summaries will count the condition only once for the corresponding preferred term. Similarly, the summaries will count a subject with multiple medical conditions within the same SOC only once for that SOC. SOCs will be tabulated alphabetically; within each SOC, preferred terms will be presented in decreasing order of incidence.

Medical conditions are coded according to the Medical Dictionary for Regulatory Activities (MedDRA) coding dictionary (the database will document the version current at the time of data lock).

6.7.4 Medications

Prescription or nonprescription medications (including vitamins and herbal products) will be coded using the WHO-Drug Global Dictionary (the database will document the version current at the time of data lock).

Medications will be summarized overall and by cohort in 2 separate analyses - 1 for prior medications and 1 for concomitant medications. The number and percentage of subjects with any medication will be summarized by decreasing order of incidence. The number and percentage of subjects who have taken a medication within a primary drug category will be summarized by cohort under that drug category by the most specific level available.

For these summaries, a subject taking the same medication multiple times will only be counted once for the corresponding most specific level available. Similarly, a subject who took multiple medications within the same drug category will be counted only once for that drug category. Prior and concomitant medications will be presented in alphabetical order of drug category; within each drug category, the most specific level available for drugs will be presented in decreasing order of incidence.

6.7.4.1 Prior Medications

Any medication or therapy reported before signing the ICF will be considered a prior medication.

6.7.4.2 Concomitant Medications

Any medication or therapy initiated or ongoing after signing the ICF will be considered a concomitant medication.

6.8 TREATMENT COMPLIANCE

Not applicable.

6.9 ANALYSIS OF ENDPOINTS

The primary objective is to assess the psychometric properties and performance of the PGI-S items.

The focus of the analyses will be to describe the responsiveness to change of the PGI-S items to the subjects' self-reported symptoms of PAH at the end of the activity and during recovery from the activity. The Sponsor will tabulate the relevant symptom scores and compute change from Baseline (pre-activity) scores for each item. Data may be displayed graphically and will be summarized overall and by cohort and sequence/period (as appropriate).

Additionally, for Cohort A, the sensitivity to a known inhaled treprostinil drug effect will be assessed. Tabulations will be made across sequences/periods for the PGI-S measurements with inhaled treprostinil to PGI-S measurements without inhaled treprostinil. Test-retest assessments for PGI-S measurements will be conducted for Cohort B.

Correlations between the 3 PGI-S questions and with Borg dyspnea scores, oxygen saturation, and HR will be assessed. Data for HR recovery at 1 minute after the ISWT will be examined in an exploratory fashion overall and as related to PGI-S measurements.

The onset and recovery profile for activity-induced symptoms will also be characterized.

6.10 HYPOTHESIS TESTING

Not applicable.

6.11 PRIMARY ANALYSIS FOR PGI-S

6.11.1 To Assess the PGI-S Items' Ability to Detect Change in Activity-induced Symptoms

Based on qualitative data from patient interviews, a 1-category change on a 5-point PGI-S scale was considered meaningful. The analysis of PGI-S data will describe the observed magnitude of the change from Baseline scores measured at completion of the ISWT for PGI-S Overall PAH Symptoms, PGI-S Shortness of Breath, and PGI-S Physical Fatigue.

Individual item-level data will be graphed to evaluate variability across subjects and summary statistics will be computed to characterize average change and variability overall and by cohort and sequence/period (as appropriate). The pre-ISWT (average of -15 and 0 minutes), post-ISWT at each time point, and pre-post difference for each of the PGI-S measurements (Overall PAH Symptoms, Shortness of Breath, Physical Fatigue) will be provided. Summary statistics will

include the mean, N, standard deviation, median, mode, minimum, maximum, the 25th and 75th percentile values, and 95% confidence intervals for the mean and difference values.

6.11.2 To Assess the PGI-S items' Ability to Detect Change in Activity-induced Symptoms Under Repeated Testing

For Cohort B, Period 1 and 2 will be compared for the difference in the change from Baseline scores measured at completion of the ISWT for PGI-S Overall PAH Symptoms, PGI-S Shortness of Breath, and PGI-S Physical Fatigue.

Summary statistics will be computed to characterize difference across periods in Cohort B, correlations between periods, variability measures, and period effects. Summary statistics will include the mean, N, standard deviation, median, mode, minimum, maximum, the 25th and 75^{th} percentile values, and 95% confidence intervals for the mean and difference values.

6.11.3 To Assess the PGI-S items' Ability to Detect Change in Activity-induced Symptoms Relative to a Known Drug Effect

For Cohort A, Periods 1 and 2 (at inferred peak and trough concentrations of inhaled treprostinil) will be compared for the inferred peak and trough concentrations of inhaled treprostinil difference in the change from Baseline scores measured at completion of the ISWT for PGI-S Overall PAH Symptoms, PGI-S Shortness of Breath, and PGI-S Physical Fatigue.

Summary statistics will be computed to characterize the difference between peak and trough conditions in Cohort A, correlations between conditions, variability measures, and sequence/period effects (as appropriate). Summary statistics will include the mean, N, standard deviation, median, mode, minimum, maximum, the 25th and 75th percentile values, and 95% confidence intervals for the mean and difference values.

6.11.4 To Assess the Ability of the PGI-S Response Scale to Detect Resolution in Activity-induced Symptoms

The analysis of PGI-S item-level data will assess the time to resolution of symptoms (return to baseline values) measured *after* completion of the ISWT (ie, change score ≤0) for PGI-S Overall PAH Symptoms, PGI-S Shortness of Breath, and PGI-S Physical Fatigue on Day 1. Summary statistics will include frequencies and percentages by protocol-specified evaluations, ie, 0 to 5 minutes, >5 to 10 minutes, etc, by cohort.

6.12 SUPPORTIVE ANALYSES

6.12.1 Borg Dyspnea Score

Summary statistics will be computed to characterize average change and variability overall and by cohort and sequence/period (as appropriate). The pre-ISWT (average of -15 and 0 minutes), post-ISWT at each time point, and pre-post difference for Borg Dyspnea score will be provided. Summary statistics will include the mean, N, standard deviation, median, mode, minimum, maximum, the 25th and 75th percentile values, and 95% confidence intervals for the mean and difference values.

6.12.2 Pulse Oximetry and Heart Rate

Summary statistics will be computed to characterize average change and variability overall and by cohort and sequence/period (as appropriate). The pre-ISWT (average of -15 and 0 minutes), post-ISWT at each time point, and pre-post difference for oxygenation and HR will be provided. Summary statistics will include the mean, N, standard deviation, median, mode, minimum, maximum, the 25th and 75th percentile values, and 95% confidence intervals for the mean and difference values.

6.12.3 ISWT

Tabulations will be provided by cohort and sequence/period (as appropriate) for the total number of shuttles completed, total distance walked (meters), use of oxygen during ISWT (Yes, No), reason for ISWT ending (Shortness of breath, Physical fatigue, Chest pain, Other), and time since most recent dose of PAH medication (Start time of ISWT - Time of most recent dose of PAH medication).

6.13 PSYCHOMETRIC ANALYSIS OF THE PGI-S ITEMS

The following analyses will be conducted on the PAP.

6.13.1 Validity of the PGI-S Items

A number of specific forms of validity exist, but one of the most important to the present study is construct validity, which describes the relationships among multiple indicators of a construct and the degree to which they follow predictable patterns. Correlations will be computed between PGI-S scores and in Borg Dyspnea scores, oxygen saturation, and HR. The goal is to

demonstrate stronger relationships among measures addressing similar constructs as compared with measures addressing more disparate constructs. Correlations will be computed using data collected at the same or nominally similar time points, specifically, at Baseline (pre-ISWT average of -15 minutes and 0 minutes) and post-ISWT (1, 2, 3, 4, and 5 minutes).

Inferences regarding validity are based on the patterns of correlations among the measures. Cohen (1992) provides the following guideline for interpreting correlation coefficients: absolute values of correlations of 0.50 or greater are considered strong, correlations that fall between 0.30 and 0.49 are moderate, and those between 0.10 and 0.29 are small.

Higher scores on the PGI-S items, Modified Borg scale, and HR indicate worse outcomes, while higher oxygen saturation is a better outcome. Therefore, positive correlations are expected between the PGI-S items and the Modified Borg scale and HR, and negative correlations are expected between the PGI-S items and oxygen saturation. Furthermore, the following are hypothesized:

- The PGI-S Overall item will be at least moderately correlated ($|r| > 0.30$) with oxygen saturation and with HR
- The PGI-S Shortness of Breath item will be at least moderately correlated with the Modified Borg score, oxygen saturation, and HR
- The PGI-S Physical Fatigue item will be at least moderately correlated with oxygen saturation and HR

To further support the construct validity of the PGI-S, correlations will also be computed between PGI-S change scores and change in Borg Dyspnea scores, oxygen saturation, and HR. These correlations will allow us to evaluate potential improvements (or declines) in the severity of symptoms commonly associated with PAH (as assessed by the PGI-S) and any corresponding improvements (or declines) in subjects' clinical measures. Correlations will be computed between changes from Baseline (pre-ISWT average of -15 minutes and 0 minutes) and post-ISWT (1, 2, 3, 4, and 5 minutes). Similar to the cross-sectional construct validity hypotheses, it is expected that:

- Change in the PGI-S Overall item will be at least moderately ($|r| > 0.30$) negatively correlated with change in oxygen saturation and at least moderately positively correlated with change in HR

- Change in the PGI-S Shortness of Breath item will be at least moderately positively correlated with change in the Modified Borg score and with change in HR, and at least moderately negatively correlated with change in oxygen saturation
- Change in the PGI-S Physical Fatigue item will be at least moderately negatively correlated with change in oxygen saturation and at least moderately positively correlated with change in HR

6.13.2 Response Frequency Distributions

To examine floor and ceiling effects for each PGI-S item, the percentage of subjects reporting the worst and best possible item scores will be examined in the context of the item-level response frequency distributions. A floor effect occurs if the worst response category (5=Very severe) is the mode of the empirical distribution and at least 2 times the expected percentage of subjects (given a uniform distribution) who select this response category. Similarly, a ceiling effect occurs if the best response category (Not present) is the mode of the empirical distribution and at least 2 times the expected percentage of subjects (given a uniform distribution) who select this response category. In the case of the PGI-S items, which have 5 response categories, the expected percentage of subjects given a uniform distribution is 20%; a floor or ceiling effect would require that more than 40% of the subjects select the worst or best response category, respectively.

PGI-S response frequency distributions will be tabulated and examined at Baseline (pre-ISWT average of -15 minutes and 0 minutes), when symptoms are expected at their minimal level, and post-ISWT (0 minutes), when symptoms are expected at their maximum level. Distributions will be provided for all subjects and visits/periods combined, as well as individually within Cohorts and visits/periods (Screening, Day 1 Period 1, and Day 1 Period 2).

6.13.3 Subgroup Analyses

Further analyses will assess outcomes presented above for subgroups based on demographic, cohort, and disease conditions.

Known-groups analyses comparing various subgroups of interest will provide support for the discriminating ability of the PGI-S items. Analysis of variances (ANOVAs) based on a priori hypotheses will examine mean differences in PGI-S item scores between subjects classified into subgroups, given sufficient subgroup sample sizes ($n \geq 5$):

- It is hypothesized that better PGI-S scores will be observed in WHO Functional Class I subjects compared with WHO Functional Class III subjects.
- It is hypothesized that better PGI-S scores will be observed in Cohort A subjects compared with Cohort B subjects.

6.13.4 Test-retest Reliability

Test-retest reliability will assess the stability of the PGI-S items using data from subjects collected at Baseline. During this 15-minute time interval, subjects are expected to remain relatively stable. The "test" data will be PGI-S item scores at -15 minutes pre-ISWT, and the "retest" data will be PGI-S item scores at 0 minutes pre-ISWT.

Because the PGI-S items are categorical, weighted kappa coefficients will be computed (Streiner and Norman, 1995). Kappa statistics can range from -1 to 1 and are interpreted such that ≤0 is poor, >0 to 0.2 indicates slight agreement, 0.21 to 0.4 indicates fair agreement, 0.41 to 0.6 indicates moderate agreement, 0.61 to 0.80 indicates substantial agreement, and 0.81 to 1.00 indicates almost perfect agreement (Landis and Koch, 1977).

6.13.5 Responsiveness to Change

It is important to provide evidence regarding the responsiveness of the PGI-S items—that is, their ability to detect change when change is expected. The responsiveness of the PGI-S items will be evaluated by computing effect-size estimates of change for each PGI-S item as the mean change from the first time point to the second time point divided by the standard deviation at the first time point. In addition, the observed score changes and associated t-tests (p-values) will be tabulated. Baseline (pre-ISWT -15 minutes and 0 minutes) to 0 minutes post-ISWT will be examined, as will 0 minutes post-ISWT to 30 minutes post-ISWT (end of recovery period).

Cohen (1992) provides a general rule of thumb for the interpretation of effect-size estimates: effect sizes of approximately 0.20 represent small effects, those of approximately 0.50 represent moderate effects, and those greater than approximately 0.80 represent large effects.

7 ANALYSIS OF SAFETY

The safety assessments will consist of AEs including serious adverse events (SAEs), and vital signs.

All safety analyses will be based on the Safety Population. No formal inferential testing will be reported for routine safety analyses.

7.1 EXTENT OF EXPOSURE

Not applicable.

7.2 ADVERSE EVENTS

AEs are recorded throughout the study and at early termination. They are coded using the MedDRA coding dictionary (the database will document the version current at the time of data lock).

7.2.1 Treatment-emergent Adverse Events

A treatment-emergent adverse event (TEAE) is defined as any event occurring after the first ISWT in the study or any event already present that worsens in either intensity or frequency.

Only TEAEs will be summarized. Any AE starting prior to the first ISWT in the study will be excluded from the summary analyses.

The following subsections note the planned presentations of TEAEs. Other presentations for TEAEs may be produced to support the Sponsor's review and analyses of events of special interest.

7.2.2 Date of Onset of Adverse Event

The onset date of each AE will be the date recorded on the eCRF.

7.2.3 Intensity Rating

The Investigator classifies AEs into 1 of the following intensity categories: "Mild," "Moderate," and "Severe."

7.2.4 Relationship to Study Drug

Not applicable.

7.2.5 Action Taken with Study Drug

Not applicable.

7.2.6 Outcome

For each AE, the investigator denotes the outcome of the AE under one of the following: "Fatal," "Not recovered/Not resolved," "Recovered/Resolved," "Recovered/Resolved with Sequelae," "Recovering/Resolving," or "Unknown."

7.2.7 Caused Study Discontinuation

The investigator denotes if the AE caused study discontinuation ("Yes," "No").

7.2.8 Summary of Treatment-emergent Adverse Events

TEAEs will be summarized overall and by cohort for the following:

- Total number of TEAEs, total number subjects with at least 1 TEAE
- Total number of serious TEAEs, total number subjects with at least 1 serious TEAE
- TEAEs by severity
- TEAEs leading to study discontinuation

TEAEs will be presented in alphabetical order of SOC and by decreasing order of incidence of the preferred terms.

For these summaries, a subject having the same TEAE multiple times will only be counted once for the corresponding preferred term. Similarly, a subject who experienced multiple TEAEs within the same SOC will be counted only once for that SOC.

A subject experiencing the same TEAE multiple times or at multiple severities will only be counted once under that preferred term at the worst severity rating. Similarly, a subject who experiences the same TEAE multiple times or at multiple severities across multiple preferred terms within the same SOC will only be counted once for the worst severity rating within the SOC.

7.2.9 Subgroup Analysis

TEAEs may be summarized within subgroups if suggested by the data.

7.3 VITAL SIGNS

Descriptive statistics for vital signs (SBP, DBP) will be presented by study visit, overall, and by cohort as raw data, and by change from Baseline within study visit.

7.4 INTERIM ANALYSES

This is an observational study and data may be evaluated at any time during the study.

8 PATIENT GLOBAL IMPRESSION OF SEVERITY QUESTIONS

[redacted]

9 MODIFIED BORG DYSPNEA SCALE

The person administering the test will obtain a rating of dyspnea using the Borg dyspnea scale. The person will use the following dialogue:

"I would like to use the following scale to indicate your shortness of breath at this moment (indicate the Borg dyspnea scale). If there is no shortness of breath at all you will point to 0; if the shortness of breath is not very great you will choose from 0.5 to 2; if you are somewhat more short of breath you will select 3; and if the breathing is getting very difficult, you will choose 4 to 9, depending on just how hard it is; 10 represents the greatest shortness of breath you have ever experienced in your life. If one of the numbers does not exactly represent how short of breath you are at this moment, then you can choose a fraction in between. For example, if you have shortness of breath somewhere between 4 and 5, you can choose 4.5."

Perceived Breathlessness (Borg Scale)

0 NOTHING AT ALL
0.5 VERY VERY SLIGHT (just noticeable)
1 VERY SLIGHT
2 SLIGHT
3 MODERATE
4 SOMEWHAT SEVERE
5 SEVERE
6
7 VERY SEVERE
8
9 VERY VERY SEVERE (almost maximum)
10 MAXIMUM

10 REFERENCES

2011. The SAS System, Version 9.3. *SAS Institute, Inc.* Cary, NC.

Cohen J. A power primer. Psychol Bull. 1992;112(1):155-159.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):159-174.

Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 2nd edition. New York: Oxford University Press; 1995.

11 LIST OF TABLES AND FIGURES

Table No.	Title	Population
14.1.1	Summary of Subject Accountability and Disposition	All
14.1.2	Summary of Analysis Populations	All
14.1.3	Summary of Demographic and Baseline Characteristics	All
14.1.4	Summary of Medical and Surgical History	All
14.1.5.1	Summary of Prior Medications	All
14.1.5.2	Summary of Concomitant Medications	All
14.1.6	Summary of Protocol Deviations	All
14.2.1.1	Summary of PGI-S Change from Baseline	All
14.2.1.2	Summary of PGI-S Change from Baseline	PPROT
14.2.1.3	Figures for PGI-S Change from Baseline	FAS
14.2.1.4	Figures for PGI-S Change from Baseline	PPROT
14.2.1.5	Summary of PGI-S Time to Recovery	FAS
14.2.1.6	Summary of PGI-S Time to Recovery	PPROT
14.2.2.1	Summary of Borg Change from Baseline	All
14.2.2.2	Summary of Borg Change from Baseline	PPROT
14.2.2.3	Summary of Pulse Oximetry and Heart Rate	All
14.2.2.4	Summary of Pulse Oximetry and Heart Rate	PPROT
14.2.2.5	Figures for Pulse Oximetry and Heart Rate	FAS
14.2.2.6	Summary of ISWT	All
14.2.3.1	PGI-S Response Frequency (%) Distributions	PAP
14.2.3.2	PGI-S Construct Validity Correlations	PAP
14.2.3.3	PGI-S Known-Groups ANOVAs	PAP
14.2.3.4	PGI-S Test-retest Reliability	PAP
14.2.3.5	PGI-S Responsiveness	PAP
14.3.1.1	Summary of Treatment-emergent Adverse Events by System Organ Class and Preferred Term in Decreasing Order	Safety
14.3.1.2	Summary of Treatment-emergent Adverse Events by Preferred Term in Decreasing Order	Safety
14.3.1.3	Summary of Serious Treatment-emergent Adverse Events by System Organ Class and Preferred Term in Decreasing Order	Safety
14.3.1.4	Summary of Treatment-emergent Adverse Events by Severity by System Organ Class and Preferred Term in Decreasing Order	Safety
14.3.1.5	Summary of Treatment-emergent Adverse Events Leading to Study Discontinuation by System Organ Class and Preferred Term in Decreasing Order	Safety

14.3.2.1	Listing of Deaths	Safety
14.3.2.2	Listing of Serious Adverse Events	Safety
14.3.3	Summary of Vital Signs	Safety