DAIT/Rho STATISTICAL ANALYSIS PLAN

08 March 2019 FINAL

Protocol # ASC01

A Randomized, Double-blind, Placebo-Controlled, Phase II Multicenter Trial of a Monoclonal Antibody to CD20 (Rituximab) for the Treatment of Systemic Sclerosis-Associated Pulmonary Arterial Hypertension (SSc-PAH)

Short Title: Rituximab for Treatment of SSc-PAH

VERSION: 1.0

DATE: 08 March 2019

SPONSOR:
Division of Allergy, Immunology, and Transplantation
National Institute of Allergy and Infectious Diseases – NIH
5601 Fishers Lane
Rockville, MD 20852
Telephone: (301) 496-5717
Fax: (301) 402-3573

PREPARED BY:
Rho
Rho Building
6330 Quadrangle Drive
Chapel Hill, NC 27517
Telephone: (919) 408-8000
Fax: (919) 408-0999

DAIT/Rho STATISTICAL ANALYSIS PLAN ACKNOWLEDGMENT AND SIGNATURE SHEET

A Randomized, Double-blind, Placebo-Controlled, Phase II Multicenter Trial of a Monoclonal Antibody to CD20 (Rituximab) for the Treatment of Systemic Sclerosis-Associated Pulmonary Arterial Hypertension (SSc-PAH)

Approved: _____ Date: _____
Mark Nicolls, MD
Protocol Chair

Approved: _____ Date: _____
David Badesch, MD
Protocol Chair

Approved: _____ Date: _____
Ellen Goldmuntz, MD, PhD
NIAID Medical Monitor

Approved: _____ Date: _____
Lynette Keyes-Elstein, DrPH
Rho Scientist

Approved: _____ Date: _____
Ashley Pinckney, MS
Rho Statistician

Document History

Version	Date	Change(s)	Author
0.1	16Jan2019	Initial Draft	Ashley Pinckney
0.2	01Mar2019	Second Draft	Ashley Pinckney
1.0	08Mar2019	Final Draft	Ashley Pinckney

Table of Contents

1. PROTOCOL SYNOPSIS

Title of the Protocol: A Randomized, Double-blind, Placebo-Controlled, Phase II Multicenter Trial of a Monoclonal Antibody to CD20 (Rituximab) for the Treatment of Systemic Sclerosis-Associated Pulmonary Arterial Hypertension (SSc-PAH)

ACE Protocol Number: ASC01

Protocol Chair(s): Mark Nicolls, MD; David Badesch, MD

IND Holder: DAIT/NIAID/NIH

Primary Objective: The primary objective of this study is to compare patients with SSc-PAH treated with rituximab to those on placebo for change in exercise capacity, as determined by six minute walk distance (6MWD).

Secondary Efficacy Objectives:

- To compare patients treated with rituximab to those on placebo for change in exercise capacity, as determined by 6MWD, at time points other than 24 weeks.
- To compare patients treated with rituximab to those on placebo for change in PVR as assessed by right heart catheterization.
- To compare patients treated with rituximab to those on placebo for time to clinical worsening, defined as first occurrence of:
    a. death,
    b. hospitalization for SSc-PAH,
    c. lung transplantation,
    d. atrial septostomy,
    e. addition of other PAH therapeutic agents, or
    f. worsening of 6MWD by > 20% and a decrease in functional class.
- To compare patients treated with rituximab to those on placebo who changed or added other PAH therapeutic agents.
- To compare patients treated with rituximab to those on placebo for change in quality of life as determined by the short form 36 (SF-36) and the disability index of the Scleroderma Health Assessment Questionnaire (SHAQ).
- To compare patients treated with rituximab to those on placebo for change in number of new digital ulcers.
- To compare patients treated with rituximab to those on placebo for change in severity of Raynaud phenomenon as measured by the VAS scale of the SHAQ.
- To compare patients treated with rituximab to those on placebo for change in DLCO and room air oxygen saturation at rest.

A secondary objective of this study is to determine whether the effects on clinical disease progression are paralleled by changes in selected biomarkers.

Secondary Safety Objectives:

- To assess the safety and tolerability of rituximab for the treatment of SSc-PAH in patients receiving stable background medical treatment with prostanoid, endothelin receptor antagonist, PDE-5 inhibitor therapy, and/or guanylate cyclase stimulators.

Title of the Protocol: A Randomized, Double-blind, Placebo-Controlled, Phase II Multicenter Trial of a Monoclonal Antibody to CD20 (Rituximab) for the Treatment of Systemic Sclerosis-Associated Pulmonary Arterial Hypertension (SSc-PAH)

Mechanistic Study Objectives:

- CD19 Studies
  - To determine whether the extent and/or duration of CD19+ B cell depletion correlates with treatment response as measured by exercise capacity (6MWD), PVR (right heart catheterization), and quality of life (SF-36 and SHAQ).
  - To monitor and correlate the reconstitution patterns of B cell subpopulations with treatment response
- To determine if the biomarkers anti-U1 RNP, anti-U3 RNP, anti-B23, anti-cardiolipin, anti-Th/To, anti-CENP-B, RF, and other autoantibodies, and quantitative immunoglobulin levels, including IgG subclasses, correlate with treatment response as measured by PVR (right heart catheterization), exercise capacity (6MWD), and quality of life (SF-36 and SHAQ).
- To determine if the biomarkers BNP/NT-proBNP, serum Na^+, IL-1, IL-6, and other cytokines correlate with treatment response as measured by PVR (right heart catheterization), exercise capacity (6MWD), and quality of life (SF-36 and SHAQ).
- To determine if B cell clones are present in SSc-PAH and what impact rituximab may have on these clones.
- To determine if B cell clones present in SSc-PAH are correlated with clinical outcomes.

Exploratory Objectives:

- To evaluate agreement between PVR calculated by site-based standards and a standardized Fick-based calculation of PVR.
- To assess sensitivity of the secondary efficacy analysis to standardization of the PVR calculation.
- To assess the utility of a standardized derived pulmonary artery capacitance as a primary efficacy endpoint.
- To evaluate the relationship between 6MWD, PVR, and other clinical measures.

Study Arms:

Subjects will be randomized in a 1:1 ratio to receive either:

- 2 infusions of rituximab, 1000 mg each, given 2 weeks apart, or
- 2 infusions of placebo given 2 weeks apart

Study Design: This is a prospective, double-blind, placebo-controlled, multi-center, randomized trial to evaluate the effect of rituximab on disease progression in subjects with SSc-PAH receiving concurrent stable-dose standard medical therapy with a prostanoid, endothelin receptor antagonist, phosphodiesterase 5 (PDE-5) inhibitor, and/or guanylate cyclase stimulators. The study will focus on assessment of clinical response and safety measures longitudinally. In addition, the effects of treatment with rituximab on the underlying immune mechanisms associated with B-cell dysregulation and pathogenic autoantibody response in this disease will be investigated.

Subjects with SSc-PAH with elevated mean pulmonary artery pressure as measured by right heart catheterization will be enrolled. The diagnosis of SSc-PAH should be confirmed

Title of the Protocol: A Randomized, Double-blind, Placebo-Controlled, Phase II Multicenter Trial of a Monoclonal Antibody to CD20 (Rituximab) for the Treatment of Systemic Sclerosis-Associated Pulmonary Arterial Hypertension (SSc-PAH)
by a rheumatologist experienced in the diagnosis and treatment of systemic sclerosis in conjunction with a pulmonologist or cardiologist specializing in management of PAH. Both specialists will be part of the study team at each site. Rituximab/placebo will be administered as two IV infusions of 1000 mg each, given two weeks apart at Day 0 and Week 2. All subjects will receive 40mg of prednisone orally the night before and morning of each infusion, as well as 100 mg methylprednisolone or equivalent corticosteroid intravenously thirty minutes prior to each infusion, with diphenhydramine and acetaminophen orally thirty to sixty minutes prior to each infusion of rituximab or placebo. Subjects will remain on their baseline standard medical regimen. Sixty eligible subjects will be accrued. Each potential study subject will provide written informed consent prior to screening procedures. All inclusion and exclusion criteria (Section 4, *Selection of Subjects*) must be met at time of randomization prior to receipt of the first dose of rituximab/placebo (Day 0, Treatment Initiation). Clinical assessments and sample collection will occur regularly through Week 48 with telephone follow-up conducted intermittently. If a subject has not recovered B cells by Week 48, B cell studies will be conducted quarterly until recovery is documented or for 2 years after initial treatment (whichever occurs first). Recovery is defined as the time point at which B cells recover to within 10% of baseline levels; if a baseline value is not available then recovery is defined as a B cell level of at least 90% of the lower limit of normal. During this quarterly monitoring period of B cell recovery, AEs and SAEs will be assessed, providing the subject has not withdrawn consent, to capture any infectious event ≥ Grade 3 using the National Cancer Institute—Common Terminology Criteria for Adverse Events (NCI-CTCAE). No additional study-related data will be collected. The primary efficacy endpoint is the change in 6MWD from baseline to 24 weeks after treatment initiation. Hemodynamic measures will be assessed at baseline and Week 24, contributing to the understanding of the relationship between PVR and clinical endpoints. Time to clinical worsening will be assessed as a secondary outcome measure through Week 48, contributing to the understanding of the relationship between PVR and clinical endpoints. Initiation of new therapy due to disease worsening prior to Week 24 will prompt an endpoint visit and right heart catheterization prior to initiation of the new therapy. B cell recovery for all will be monitored by serial assessments of B cell counts as determined by flow cytometry. The total duration of the study is anticipated to be approximately 8 years.
Study Population: Any individual 18 to 75 years of age, who has been diagnosed with SSc-PAH within the past 5 years (as defined by a clinical diagnosis of SSc with a mPAP ≥ 25 mmHg), a NYHA Functional Class of II, III, or IV, a mean PVR of > 3 Wood units, and who meets all entry criteria is eligible for randomization.
Primary Efficacy Endpoint: Change in exercise capacity, assessed by 6MWD, from baseline to 24 weeks after treatment initiation.

Title of the Protocol: A Randomized, Double-blind, Placebo-Controlled, Phase II Multicenter Trial of a Monoclonal Antibody to CD20 (Rituximab) for the Treatment of Systemic Sclerosis-Associated Pulmonary Arterial Hypertension (SSc-PAH)

Secondary Efficacy Endpoints:

Secondary efficacy endpoints include the following clinical endpoints:

1. Change in PVR measured by right heart catheterization, from baseline to 24 weeks after treatment initiation.
2. Exercise capacity, as determined by 6MWD, assessed longitudinally over the 48 weeks following treatment initiation.
3. Assessment of time to clinical worsening, censored at 48 weeks, defined as first occurrence of:
    a. death,
    b. hospitalization for SSc-PAH,
    c. lung transplantation,
    d. atrial septostomy,
    e. addition of other PAH therapeutic agents, or
    f. worsening of 6MWD by > 20% and a decrease in functional class.
4. Percentage of subjects who changed or added PAH therapeutic agents at 24 and 48 weeks.
5. Change in quality of life as determined by the SF-36 and disability index of the SHAQ from baseline to 24 and 48 weeks after treatment initiation.
6. Number of new digital ulcers assessed longitudinally over both the 24 and 48 weeks following treatment initiation.
7. Severity of Raynaud phenomenon, as measured by the VAS scale of the SHAQ, assessed longitudinally over both the 24 and 48 weeks following treatment initiation.
8. DLCO and oxygen saturation at rest on room air assessed longitudinally over both the 24 and 48 weeks following treatment initiation.
9. Change in PVR from baseline to 24 weeks after treatment initiation as a percentage of the baseline value.

Secondary efficacy endpoints also include the following mechanistic endpoints:

1. CD19 Studies
    a. CD19+ B cells (absolute number and percentage of gated lymphocytes) at time of B cell depletion and longitudinally over the 48 weeks following treatment initiation, where B cell depletion is defined as <5 CD19+ cells/μL.
    b. Time to repopulation of CD19+ B cells, defined as the time point at which B cells recover to within 10% of baseline levels; if a baseline value is not available, then recovery is defined as a B cell level of at least 90% of the lower limit of normal.
    c. Reconstitution pattern of B cell subpopulations assessed longitudinally over the 48 weeks following treatment initiation.
2. Level of anti-U1 RNP, anti-B23, anti-cardiolipin, anti-CENP-B, RF autoantibodies, and immunoglobulins, including IgG subclasses, measured by ELISA and assessed longitudinally over the 48 weeks following treatment initiation.
3. Presence or absence of anti-U3 RNP, anti-Th/To, and other autoantibodies assessed longitudinally over the 48 weeks following treatment initiation.

Title of the Protocol: A Randomized, Double-blind, Placebo-Controlled, Phase II Multicenter Trial of a Monoclonal Antibody to CD20 (Rituximab) for the Treatment of Systemic Sclerosis-Associated Pulmonary Arterial Hypertension (SSc-PAH)

4. Level of BNP/NT-proBNP, serum Na^+, IL-1, IL-6, and other cytokines assessed longitudinally over the 48 weeks following treatment initiation.
5. Presence or absence of B cell clones assessed at baseline and longitudinally over the 48 weeks following treatment initiation.

Secondary Safety Endpoints:

1. All NCI-CTCAE Grade 3-5 AEs, which are defined as possibly, probably, or definitely related to rituximab infusion.
2. All NCI-CTCAE Grade 3-5 adverse events.
3. All infection related events.
4. Treatment-related mortality, defined as death occurring at any time after randomization and possibly, probably, or definitely resulting from treatment given in the study. Study investigators will make an initial determination of the cause of death of any subject and if uncertainty remains, a final decision will be made by an independent blinded panel after reviewing all clinical and, if available, autopsy data.
5. All-cause mortality, defined as any death occurring at any time after randomization.
6. CD19+ B cell levels post treatment (note that analyses will be performed regularly per Table 6.1 through 48 weeks and quarterly thereafter as documented in Section 3.1, *Description of Study Design*).
7. Peripheral blood cell counts, liver function, and creatinine levels after treatment.

Exploratory Endpoints:

Exploratory analysis of standardization of the PVR calculation will be based on PVR calculated from a standardized Fick calculation of cardiac output, which is a function of hemoglobin, arterial and mixed-venous saturations, and oxygen consumption according to the La-Farge and Miettinen formula. The cardiac output calculation is a function of assumed oxygen consumption according to the La-Farge calculation of values normalized for age, gender, heart rate. Calculation of a standardized Fick cardiac output will also allow for determination of the pulmonary artery capacitance (PACf). See Appendix E, *Exploratory Analysis: Standardized Fick-based Pulmonary Vascular Resistance and Pulmonary Artery Capacitance* for details.

Additionally, an exploratory analysis of the relationship between change in PVR and change in 6MWD will be addressed. Relationships with other clinical endpoints, like change in NTproBNP, will also be explored.

Sample Size: The study will include 30 subjects treated with rituximab and 30 subjects treated with placebo.

Data Analyses: The primary efficacy analysis of change in 6MWD from baseline to 24 weeks will be performed on the intention-to-treat (ITT) population. and will be tested using a repeated measures random effects model to model change from baseline in 6MWD as the primary outcome and treatment as the primary predictor, controlling for visit day, limitations by musculoskeletal disease, baseline levels of 6MWD, and a treatment by visit day interaction. A random slope and intercept will be fit for each subject. An unstructured covariance matrix will be assumed for the random effects. Additional covariates will be considered for this model as part of secondary analysis.

Title of the Protocol: A Randomized, Double-blind, Placebo-Controlled, Phase II Multicenter Trial of a Monoclonal Antibody to CD20 (Rituximab) for the Treatment of Systemic Sclerosis-Associated Pulmonary Arterial Hypertension (SSc-PAH)

Secondary analyses will include repeating the primary analysis in the per protocol population, subgroup analyses of the primary outcome, safety analyses, exploratory analyses of the secondary endpoints, and exploratory analyses of the mechanistic endpoints.

Safety Stopping Rules: In addition to the pre-scheduled data reviews and planned safety monitoring, the DSMB may be called upon for ad hoc reviews or emergency meetings. The DSMB will review any event that potentially impacts safety at the request of the protocol chair or DAIT/NIAID. In addition, the following events will trigger both a comprehensive DSMB Emergency Safety Review and a temporary halt in enrollment:

a. Any immediately life threatening event or death that occurs in the study which is unexpected and possibly, probably, or definitely related to study intervention.
b. Any event, including those listed in Section 3.3.2, *Discontinuation of Protocol-Specified Treatment Requirements*, that results in permanent discontinuation of infusion of study intervention occurring in 2 of the first 5 subjects, 4 of the first 15 subjects, 6 of the first 30 subjects, or a total of 8 events after 30 subjects have been enrolled in the trial.
c. Any occurrence of PML.
d. Clinical Grade 3 or higher events or SAEs which are possibly, probably, or definitely related to study treatment but do not result in discontinuation of study treatment will be reviewed whenever the ratio of events to enrolled subjects is 20% (e.g. 9 events for 47 subjects, 10 events for 48-52 subjects, 11 events for 53-57 subjects, etc.).
    - Note: clinical events include clinical signs/symptoms and diagnoses and laboratory abnormalities with clinical consequence (defined as the requirement for intervention, correction, increased monitoring, or further evaluation).

The DSMB will have discretion to recommend actions regarding study conduct and continuation as a consequence of any planned or unplanned monitoring activity.

A temporary halt in enrollment will occur for the events listed above with the exception of event d. For events falling under d, a halt in enrollment will occur if the DSMB Emergency Safety Review is not completed within 2 weeks. In the event that the study temporarily halts enrollment, no new subjects will be consented or start on therapy with rituximab or placebo; and subjects already on rituximab or placebo will continue on therapy unless they are the focus of the DSMB review. Subjects in the screening phase of the study may continue to undergo minimal risk procedures (e.g. blood tests), but more than minimal risk procedures should be deferred. Randomization will not occur until the DSMB review is complete. The FDA will be notified of any halt in enrollment. After careful review of the data, the DSMB will make recommendations regarding study conduct and/or continuation.

2. INTRODUCTION

This statistical analysis plan includes pre-planned analyses related to the study objectives outlined in the protocol.

3. GENERAL ANALYSIS AND REPORTING CONVENTIONS

The following analyses and reporting conventions will be used:

- Categorical variables will be summarized using counts (n) and percentages (%) and will be presented in the form "n (%)." Percentages will be rounded to one decimal place.

- Numeric variables will be summarized using n, mean, standard deviation (SD), median, minimum (min), maximum (max). The min/max will be reported at the same level of significance as original data. The mean and median will be reported at one more significant digit than the precision of the data, and SD will be reported at two more significant digits than the precision of the data.

- The median will be reported as the average of the two middle numbers if the dataset contains an even number of observations.

- Test statistics including *t* and *z* test statistics will be reported to two decimal places.

- *P*-values will be reported to three decimal places if greater than or equal to 0.001. If less than 0.001, the value will be reported as "<0.001." A *p*-value can be reported as "1.000" only if it is exactly 1.000 without rounding. A *p*-value can be reported as "0.000" only if it is exactly 0.000 without rounding.

If departures from these general conventions are present in the specific evaluations section of this SAP, then those conventions will take precedence over these general conventions.

4. ANALYSIS SAMPLES

4.1. Safety Population

The Safety Population (SP), which will be used for all safety analyses, will include all subjects for whom study treatment is initiated.

4.2. Modified Intent-to-Treat Population

The Modified Intent-to-Treat (ITT) population will include all randomized subjects who initiate rituximab therapy and meet entry criteria. Randomized subjects who withdraw from the trial prior to starting rituximab infusion will be excluded from the ITT analysis set and replaced. The primary efficacy analyses will be based on the ITT population, as will key secondary efficacy and mechanistic analyses. Subjects who, for whatever reason, do not complete their assigned therapy will be included in the ITT population in the groups to which they were randomized.

4.3. Per Protocol Populations

The Per Protocol (PP) population will be defined as those subjects who comply with the assigned treatment protocol by completing two IV infusions of 1000mg of Rituximab/Placebo at Visit 3 and Visit 4 as well as completing primary endpoint evaluation (6MWD) at Visit 10 with no major protocol deviations that would affect their efficacy outcome. Subjects who are withdrawn from follow-up prior to Visit 10 will not be included in the PP population. A masked data review panel will evaluate deviations from the protocol including, for example, violations of entry criteria, departures from assigned treatment regimen, or administration of study procedures outside the specified visit windows. The panel may exclude subjects with serious protocol deviations from the PP population. Both the primary analyses and all secondary analyses will be replicated on the PP population.

A second Per Protocol (PP2) population will be also be defined. This population will be defined as eligible subjects who comply with the assigned treatment protocol by completing two IV infusions of 1000mg of Rituximab/Placebo at Visit 3 and Visit 4.

5. STUDY SUBJECTS

5.1. Disposition of Subjects

The disposition of all enrolled subjects will be summarized in tables and listed.

The numbers and percentages of subjects randomized, in each analysis sample, and completing selected study weeks (Day 0; Weeks 2, 24 and 48), as well as reasons for early termination from the study will be presented. For subjects discontinuing study treatment early, the reasons for discontinuing study treatment early will also be presented.

Subject status will also be shown graphically by subject. Each subject will be represented by a line with events such as treatment, early study withdrawal, first addition/change in PAH medications, death, etc. shown over time in the study.

5.2. Demographic and Other Baseline Characteristics

Summary descriptive statistics for baseline and demographic characteristics will be reported for all analysis samples. Characteristics to be summarized include: age, race, ethnicity, sex, body weight and height at screening, duration of SSc-PAH disease, PVR (by type), right atrial mean pressure, mean pulmonary artery pressure, cardiac output, pulmonary capillary wedge mean pressure, 6MWD, NYHA Functional Class, digital ulcer count, DLCO, FVC, TLC, HAQ-DI score, oxygen saturation, autoantibody status, number of PAH medications, oxygen use, limited or diffuse SSc, duration of SSc, smoking history, and BNP/pro-NTBNP.

Demographic and baseline characteristic data will also be presented in data listings by treatment group and subject.

6. STUDY OPERATIONS

6.1. Protocol Deviations

Major protocol deviations will be listed by site with information such as type of deviation, whether notification to the IRB was required, date of occurrence, and the reason for the deviation. Protocol deviations will be summarized in tabular format by type of deviation.

6.2. Treatment Adherence

Subjects will be randomized into either the rituximab or placebo arm. Rituximab or placebo will be administered as two IV infusions of 1000 mg each, given two weeks apart at Day 0 and Week 2. Study treatment information is collected on the ***Rituximab/Placebo Infusion*** CRF page.

Percent compliance for study drug (rituximab or placebo) will be computed for both study visits and overall as:

$$\%\ \text{Compliance} = \frac{\text{Total dose received (mg)}}{\text{Total dose planned (mg)}} \times 100$$

Where:

$$\text{Total dose received (mg)} = \text{concentration of infusion (mg/mL)} \times \text{total volume infused (mL)}$$

For each study arm and overall, percent compliance for study drug will be summarized descriptively by mean, standard deviation, median, minimum and maximum.

Compliance data for study drug will be listed by treatment group, subject, and infusion visit and will include use of pre-infusion medications, duration of infusion, information on incomplete or interrupted infusions, actual and planned dose, as well as concomitant therapy during study drug administration.

7. ENDPOINT EVALUATION

7.1. Overview of Efficacy Analysis Methods

7.1.1. Multicenter Studies

Due to the small sample size and large number of sites needed for this study, study center will not be included as a covariate in the analysis. However, if evidence of a study center effect exists and there are appropriate sample sizes at each site, a sensitivity analysis will be performed including study center in the efficacy analyses. All safety analyses will be based on data pooled across all centers with no adjustment or stratification by center.

7.1.2. Assessment Time Windows

Allowable visit windows for all scheduled visits are provided in Table 7-1.

Table 7-1 Visit Windows

Visit	Window
Screening	-28 Days
Baseline	-7 Days
Day 0	
Week 2	± 4 Days
Weeks 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48	± 7 Days
Quarterly Monitoring	± 14 Days

Unscheduled visits may also occur throughout the study. All data will be included in analyses, regardless of time of assessment.

7.2. Primary Endpoint

7.2.1. Computation of the Primary Endpoint

The primary efficacy endpoint is change in exercise capacity, assessed by the Six Minute Walk Distance (6MWD), from baseline to 24 weeks after treatment initiation. The 6MWD will be performed at Screening and Weeks 4, 12, 24, 36, and 48; additionally, if the 6MWD decreased from the Screening 6MWD by more than 20%, a subject should repeat the 6MWD within 7 days, but not on the same day. A 6MWD may also be performed at an unscheduled visit.

The 6MWD measures the distance a subject is able to walk over a total of six minutes on a hard, flat surface. The goal is for the subject to walk as far as possible in six minutes. The subject is allowed to self-pace and rest as needed as they traverse back and forth along a marked walkway. The total distance walked, in meters, will be recorded for each subject on the *6-Minute Walk Test* CRF page. Information on whether a subject has a musculoskeletal disorder that limits ambulation (yes/no) will also be collected on this CRF page.

If a subject's 6MWD decreased from Screening by more than 20% and a repeat 6MWD was performed, then the better of the two values will be used in the primary analysis. Missing data will not be imputed.

7.2.2. Primary Analysis of the Primary Endpoint

The primary efficacy analysis for this study is designed to test the scientific hypothesis that rituximab has a beneficial clinical impact on SSc-PAH as measured by improvement or stabilization in 6MWD between baseline to 24 weeks in patients treated with standard of care plus rituximab than in patients treated with standard of care plus placebo. Operationally for this study, this scientific hypothesis will be examined by formally testing the following statistical hypothesis:

H_0: The mean change in 6MWD between baseline and 24 weeks does not differ between individuals treated with rituximab and those treated with placebo ($\delta_{rituximab} = \delta_{placebo}$ where δ_x = mean 6MWD at 24 weeks – mean 6MWD at baseline)

versus

H_A: The mean change in 6MWD between baseline and 24 weeks is more beneficial in those individuals treated with rituximab than in those treated with placebo ($\delta_{rituximab} < \delta_{placebo}$ where δ_x = mean 6MWD at 24 weeks – mean 6MWD at baseline)

While this hypothesis is inherently one-sided, a two-sided test at α=0.05 will be used to evaluate the treatment effect at 24 weeks. This primary hypothesis will be tested using a repeated measures random effect model to model the distance walked as a function of treatment, visit week and a treatment by visit week interaction. Additionally, a quadratic term for visit week will be included in the model. A random slope and intercept will be fit for each subject using a separate unstructured covariance matrix for each treatment group. The model will include only 6MWD data up to Week 24. All available data up to Week 24 will be used for subjects who terminate early from the study. All unscheduled visits that occur prior to Week 24 will also be included in the analysis.

Model estimates and standard errors of the distance walked and change from baseline in distance walked will be presented at baseline and Week 24, along with the differences between arms. An appropriate contrast will be used to test the difference (baseline to Week 24) between treatment arms.

Distance walked and change from baseline in distance walked with be shown graphically by treatment arm. Spaghetti plots and summary statistics including 25th percentile, mean, median, and 75th percentile will be plotted for each visit by treatment group. Additionally, mean predicted values from the model will be plotted over time, alongside the actual distance walked values for each study visit.

The primary efficacy analysis will be based on the ITT population.

7.2.3. Sensitivity Analyses of the Primary Endpoint

The following sensitivity analyses will be performed:

1. Primary analyses performed on both PP populations.
2. If a subject's 6MWD decreased from Screening by more than 20% and a repeat 6MWD was performed, then models will be develop using both SMWD values and also using just the second, confirmatory, value in the model.
3. Changes in background PAH medications prior to Week 24 (yes/no) will be added as a covariate to the model.
4. Missing data will not be imputed for the primary endpoint. However, sensitivity analyses may be performed to assess the robustness of the conclusions under different

assumptions. For example, missing data due to death of a subject prior to Week 24 would be handled differently than missing data due to a missed visit. A joint longitudinal/survival model to evaluate trends in clinical assessments under different assumptions about the underlying missing data mechanism may be considered, if appropriate.

5. Demographic and baseline characteristics that differ meaningfully between the two treatment groups (with or without a p-value <0.05) may be included in a covariate-adjusted model, as appropriate.
6. Time-varying covariates, like oxygen use during SMWD, may be added as appropriate.

7.3. Secondary Endpoints

All secondary analyses will be conducted in an exploratory fashion with p-values and confidence intervals presented as descriptive statistics with no adjustments for multiple comparisons. Interval estimates will be generated at the 95% confidence level. Analyses will be conducted on the ITT population and the PP populations. Missing data will not be imputed.

7.3.1. Pulmonary Vascular Resistance

Endpoint: A right heart catheterization (RHC) will be performed at Screening and Week 24. Data for the RHC will be collected on the ***Right Heart Catheterization*** CRF page.

During a RHC a catheter is guided to the right side of the heart and then into the pulmonary artery; blood flow through the heart is observed and is used to measure pressures in a subject's heart and lungs. Pulmonary vascular resistance (PVR) will be collected in Wood Units and method of cardiac output (indirect or estimated Fick, direct Fick, or thermodilution) will be collected as well. Multiple types of PVR (indirect or estimated Fick, direct Fick, or thermodilution) may be collected for a subject. Higher PVR values indicate worse disease status.

PVR will be derived in two ways:

1. First, as noted previously, multiple types of PVR (indirect Fick, direct Fick, or thermodilution) may be collected for a subject. The PVR value used for analysis will be chosen in the following order, as available from the ***Right Heart Catheterization*** CRF page: 1. thermodilution, 2. direct Fick, 3. indirect Fick. Both the screening and Week 24 RHC must use the same method of cardiac output. If the method used for screening and Week 24 do not match, only the baseline measure will be included in model-based analyses.
2. Second, PVR will be derived using the formula:

$$PVR = (mPAP - PCWP)/CO, \quad \text{where}$$

mPAP = mean pulmonary artery pressure (mmHg)

PCWP = mean pulmonary capillary wedge pressure (mmHg)

CO = Cardiac Output (L/min)

Similar to the first method, method of cardiac output will be chosen in the following order, as available: 1. thermodilution, 2. direct Fick, 3. indirect Fick. PCWP can be recorded at end expiration or by a computer-generated mean. If both are recorded, at end expiration will be used.

For each derived PVR variable, change from Screening to Week 24 in the PVR and percent change from Screening to Week 24 in the PVR will be computed for each subject. Missing values will not be imputed.

Analysis: For each derived PVR variable, the following analyses are planned.

1. Descriptive statistics will be presented for PVR, percent change from Screening, and change from Screening in PVR for each treatment group and overall. Spaghetti plots and summary statistics including 25th percentile, mean, median, and 75th percentile will be plotted for each visit by treatment group. Separate summaries will be prepared for all subjects with available data at either time point and for the subset of subjects with measurements at both time points.
2. Additionally, a repeated measures mixed model will be fit with PVR as the outcome and treatment, visit week (actual observed week), and their interaction as predictors. Random intercepts and slopes will be fit for each subject using an unstructured covariance structure for each treatment group. Model estimates and standard errors for PVR and change from baseline in PVR will be presented at baseline and Week 24, along with the differences between arms. An appropriate contrast will be used to test the difference (baseline to Week 24) between treatment arms.

A separate analysis is planned to simultaneously evaluate the all reported PVR (indirect Fick, direct Fick, or thermodilution) measures for a subject without the prioritization rules described above. A multivariable repeated measures mixed model will be fit with PVR as the outcome and treatment, type of PVR (indirect or estimated Fick, direct Fick, or thermodilution), visit week (actual observed week) and interaction terms for type*treatment and type*treatment*week. Random intercepts and slopes will be fit for each subject using an unstructured covariance structure for each PVR type and treatment group. In addition, separate residual variance components will be estimated for each type of PVR.

Note: PVR is measured only at screening and Week 24. Subjects lost prior to Week 24 will contribute only screening data. If subjects are not lost completely at random, findings from these analyses may be biased as they will reflect changes in PVR in the subset of individuals who made it to Week 24.

Analogous analyses will be repeated on the percent change from Screening in PVR at Week 24.

7.3.2. Six Minute Walk Distance

Endpoint: Exercise capacity, as determined by 6MWD, will be assessed as defined in Section 7.2.1, *Computation of the Primary Endpoint*.

Changes in PAH medication use will be recorded on the *Initiation of New PAH Medications and Therapy* CRF page, and will be used to determine whether subjects had changes in their background PAH medications before and/or after the Week 24 visit (yes/no).

Analysis: Analyses analogous to those specified for the primary endpoint will be performed for longitudinal changes in the 6MWD at Week 48. See Section 7.2.2, *Primary Analysis of the Primary Endpoint*, for more details on the analyses. Data for this model will include all available data to Week 48. Additionally, the model will include a covariate to indicate changes in PAH therapeutic agents before and/or after the 24 week time point (yes/no). The interaction between this covariate and treatment will be included if significant (p-value<0.05). Spaghetti plots of the data will also be assessed to define any time trends that may need to be reflected in the model.

Additional models may be developed to assess the effect of other covariates.

7.3.3. Clinical Worsening

Endpoint: Assessment of time to clinical worsening, censored at 48 weeks, will be defined as first occurrence of the following: death, hospitalization for SSc-PAH, lung transplantation, atrial septostomy, addition of other PAH therapeutic agents, or worsening of 6MWD by > 20% and

an increase in functional class. The components of clinical worsening will be assessed as follows.

a. Death: Date of death will be collected on the *Post-Study Follow-Up/Week 104 Vital Status* CRF page.
b. Hospitalization for SSc-PAH: Date of hospitalization for SSc-PAH will be collected on the *Serious Adverse Event* CRF page and determined from SAEs that had seriousness criteria of "Required hospitalization or prolongation of existing hospitalization" and that were marked as being possibly, probably, or definitely related to worsening disease on the *Adverse Events* CRF page. Time to the event will be derived from the date that the AE became serious. If a subject has multiple hospitalizations for SSc-PAH, then the first date will be used.
    i. A second clinical worsening endpoint will be similarly defined except that "hospitalizations for SSc-PAH" will be evaluated by a blinded review panel to exclude those events relating only to a subject's SSc and not PAH.
c. Lung transplantation: Date of lung transplant will be collected on the *Initiation of New PAH Medications and Therapy* CRF page.
d. Atrial septostomy: Date of atrial septostomy will be collected on the *Initiation of New PAH Medications and Therapy* CRF page.
e. Addition of other PAH therapeutic agents: Date of new medications or doses of PAH medications changed for the treatment of PAH will be collected on the *Initiation of New PAH Medications and Therapy* CRF page.
f. Worsening of 6MWD by > 20% and an increase in functional class: The 6MWD and the NYHA Functional Class Assessment will be performed at Screening and Weeks 4, 12, 24, 36, and 48; they may also be performed at an unscheduled visit. Data for the 6MWD will be collected on the *6-Minute Walk Test* CRF page and data on functional class will be collected on the *NYHA Functional Class Assessment* CRF. The first date at which the 6MWD has decreased by >20% from the Screening walk distance and the NYHA Functional Class Assessment has increased a class from Screening will be recorded. If a subject's 6MWD decreased from Screening by more than 20% and a repeat 6MWD was performed, then the better of the two values will be used in the analysis. If a subject is NYHA Class IV at baseline, then only a decrease of 20% in the 6MWD will be required.

Time to clinical worsening will be defined as the first date that meets any of the above criteria and will be calculated in study days as: date of first event – date of randomization. If a subject does not experience a relevant event by Week 48 or if the date of death is after the 48 week follow-up period, then time to clinical worsening will equal the duration of follow-up in the main study for that subject. Appropriate censor variables will be defined to indicate whether the time-to variable represents an event time or a censor time.

Analysis: Time to clinical worsening and type of clinical worsening reached will be summarized by treatment group and overall. Kaplan-Meier curves, including 95% confidence intervals, will be created for each treatment arm and will be compared using the log-rank test. A Cox proportional hazards model may be fit to assess the effect of treatment on time to clinical worsening, if appropriate. Additional covariates may be added to the model as appropriate. Analyses will be run using the definition of hospitalization of SSc-PAH as defined in both point B and B.i above.

7.3.4. Addition of PAH medications

Endpoint: New medications or doses of PAH medications changed for the treatment of PAH will be collected on the *Initiation of New PAH Medications and Therapy* CRF page. Date of the

new or modified medication along with reason for the change will be collected for the first time a PAH medication is modified or added.

Indicator variables for new or modified PAH therapeutic agents will be derived for two time points: prior to Week 24 and after Week 24. Time to the addition or modification of PAH medications will be defined in study days as: date of the first time a PAH medication is modified or added – date of randomization.

Analysis: The number and proportion of subjects who changed or added PAH therapeutic agents prior to Week 24 and after Week 24 will be summarized by treatment group and overall, along with the reason for changes to the background PAH medications. Time (in days) to the addition or modification of PAH medications will also be summarized by treatment group and overall.

Kaplan-Meier curves, including 95% confidence intervals, will be created for each treatment arm and will be compared using the log-rank test.

7.3.5. Quality of Life: SF-36

Endpoint: Change in quality of life as determined by the Short Form Health Survey (SF-36) will be assessed at Baseline and Weeks 24 and 48. Data for the SF-36 will be captured on the *Short Form – 36 Health Survey* CRF page. The SF-36 is a 36-item, patient-reported survey of patient health. Higher scores indicate better outcomes while lower scores indicate more disability.

The following SF-36 scale scores will be calculated: Physical Functioning, Role-Physical, Bodily Pain, General Health, Vitality, Social Functioning, Role-Emotional, and Mental Health. From these scale scores, two summary scores that aggregate measures will be created. The Physical Component Score is comprised of the Physical Functioning Scale, the Role-Physical Scale, the Bodily Pain Scale, and the General Health Scale. The Mental Component Score is comprised of the Vitality Scale, the Social Functioning Scale, the Role-Emotional Scale, and the Mental Health Scale.

Norm-based scores will also be generated for the eight SF-36 scales and the two summary measures. All eight SF-36 scales and the two summary scores will be normalized to have a mean of 50 and a standard deviation of 10 for the 1998 general US population. The advantage of the standardization is that results for one scale can be meaningfully compared with the other scales, and the scores have a direct interpretation in relation to the distribution of scores in the 1998 general US population. Specifically, scores above or below 50 are above or below the average, respectively, for the 1998 general US population. Because the standard deviation is 10 for all 8 scales, each one point difference in scores is equivalent to one-tenth of a standard deviation unit or an effect size of 0.10. See Appendix 14.1, Scoring the SF-36 Version 2, of this SAP for details regarding calculation of these scores.

Change from baseline in the SF-36 Physical Component Summary (PCS) and Mental Component Score (MCS) domains at Weeks 24 and 48 will be computed for each subject. Missing data will not be imputed.

Analysis: Longitudinal changes in the SF-36 at Weeks 24 and 48 will be compared across the two arms using linear mixed models. A repeated measures random effect model with MCS or PCS as the outcome, and treatment, visit week (actual observed week), and their interaction as predictors. Additionally, a quadratic term for visit week may be included in the model, if significant. Random intercepts and slopes will be fit for each subject using an unstructured covariance structure for each treatment group.

Descriptive statistics will be presented for the MCS and PCS and change from baseline in the MCS and PCS for each treatment group and overall by study visit. MCS and PCS and change from baseline in MCS and PCS will be shown graphically by treatment arm. Spaghetti plots and

summary statistics including 25th percentile, mean, median, and 75th percentile will be plotted for each visit by treatment group.

Model estimates and standard errors of the MCS and PCS and change from baseline in the MCS and PCS will be presented at baseline and Weeks 24 and 48, along with the differences between arms. Appropriate contrasts will be used to test the difference (baseline to Week 24 or Week 48) between treatment arms. Mean predicted values from the model will be plotted over time, alongside the actual values for each study visit.

Additional models may be developed to assess the effect of other covariates, as necessary.

7.3.6. Quality of Life: HAQ-DI

Endpoint: Change in quality of life as determined by the Disability Index of the Scleroderma Health Assessment Questionnaire (HAQ-DI) will be completed at Baseline and Weeks 4, 12, 24, 36, and 48. Data for the HAQ-DI will be collected on the *SHAQ (Scleroderma Health Assessment Questionnaire)* CRF page.

Eight domain scores are computed as the maximum value over all variables in the domain. Domains are defined as follows: Dressing/Grooming (Questions 1A, 1B), Arising (Questions 2A, 2B), Eating (Questions 3A, 3B, 3C), Walking (Questions 4A, 4B), Hygiene (Questions 5A, 5B, 5C), Reach (Questions 6A, 6B), Grip (Questions 7A, 7B, 7C), and Activities (Questions 8A, 8B, 8C). Each of the questions are answered on a 4-point scale, where 0 = without any difficulty, 1 = with some difficulty, 2 = with much difficulty, and 3 = unable to do. The total score, HAQ-DI, is then computed as the mean over all the non-missing domain scores. If less than 6 domain scores are available, HAQ-DI is missing. Higher scores indicate more disability. Change from baseline in the HAQ-DI score will also be calculated for each visit. Missing values will not be imputed. Use of devices will not be included in the scoring.

Analysis: Analyses analogous to those specified in Section 7.3.5, *Quality of Life: SF-36*, will be performed for the HAQ-DI.

7.3.7. Digital Ulcers

Endpoint: The total number of digital ulcers will be captured at the Baseline visit, and subsequently the number of new digital ulcers since the subject's last study visit will be captured at the Week 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, and 48 visits and recorded on the *Digital Ulcer Count* CRF page.

The total number of digital ulcers present on the dorsal and palmar surfaces for both the left and right fingers will be captured at the baseline study visit. The number of new digital ulcers since the last study visit (including any ulcers that have appeared and healed since the last study visit) on the dorsal and palmar surfaces for both the left and right fingers will be captured at the post-baseline study visits. A variable for the total number of digital ulcers on both hands will be summed from the number present on the dorsal and palmar surfaces for both the left and right fingers.

Rates (per week) of new digital ulcers for each treatment for each of two periods will be calculated for each subject, the two observations will be defined as:

1) the period including assessments at weeks 4, 8, 12, 16, 20, and 24 (Period=1)
2) the period including assessments at weeks 28, 32, 36, 40, 44, and 48 (Period=2).

For each period, we will define:

1) Total number of new digital ulcers on both hands = sum over all assessments for the period
2) Period Weeks = Sum of weeks between assessments over all intervals in the period (based on actual dates of the visits).
    - the natural log(period weeks) will also be derived

3) PAH Meds = indicator if PAH medications were changed or modified in the treatment period

Analysis: The number of digital ulcers at screening and then the number of new digital ulcers at the follow-up visits will be summarized by treatment group and overall.

Longitudinal rates of the number of new digital ulcers at Weeks 24 and 48 will be compared across the two arms using Poisson marginal models for longitudinal data. Initially, the Poisson model will be used to describe the rate of new digital ulcers (per week) as the outcome, and only treatment, number of digital ulcers at Baseline, and an indicator if the measurement was affected by changed or new PAH therapeutic agents as covariates. The SAS procedure GEE will be used to fit the Poisson model; an unstructured working correlation matrix will be used. Period weeks will be used as the offset variable to represent the different time intervals.

Estimated rates of new digital ulcers (per week) for each treatment group and period will be derived. The rate of new digital ulcers will be compared across treatment groups for each time period. Additionally, within each treatment group, the differences in rates between period 1 and period 2 will be compared, and the differences in the rates between period 1 and period 2 will be compared across treatment arms.

Additional models may be developed to assess the effect of other covariates as appropriate.

7.3.8. Severity of Raynaud's Phenomenon

Endpoint: Change in quality of life as determined by the Scleroderma Health Assessment Questionnaire (SHAQ) will be completed at Baseline and Weeks 4, 12, 24, 36, and 48. Data for the SHAQ will be collected on the *SHAQ (Scleroderma Health Assessment Questionnaire)* CRF page.

Severity of Raynaud's phenomenon will be measured by a VAS scale of the SHAQ. The SHAQ VAS scales include a question asking, "In the past week, how much has your Raynaud's Phenomenon interfered with your activities?" Subjects are asked to place a mark on a 15 cm line, scaled from 0 (does not interfere) to 100 (very severe limitation), to describe the severity of their Raynaud's phenomenon in the past week. The distance from the left edge of the line to the vertical line placed by the subject will be measured in centimeters; VAS scores will be converted to a 0 to 100 scale. Change from baseline in the VAS score will be calculated.

Analysis: Analyses analogous to those specified in Section 7.3.5, *Quality of Life: SF-36*, will be performed for the severity of Raynaud's phenomenon.

7.3.9. DLCO

Endpoint: Pulmonary function tests will be performed at Screening, Week 24, and Week 48; only DLCO will be collected at the Week 24 and 48 visits. Data for DLCO will be collected on the *Pulmonary Function Tests* CRF pages. Lower DLCO values indicate worse disease activity. DLCO will be calculated according to the Global Lung Function Initiative (GLI) all-age reference values and corrected for hemoglobin. The hemoglobin used for adjustment should be done within two weeks of the PFT test and will be recorded on the CRF page. If a hemoglobin measure is not available, the unadjusted value will not be used. The % predicted DLCO is computed as the measured DLCO divided by the computed predicted value and multiplied by 100%. Change from baseline in the % predicted DLCO will be calculated.

Analysis: Analyses analogous to those specified in Section 7.3.5, *Quality of Life: SF-36*, will be performed for DLCO.

7.3.10. Oxygen Saturation

Endpoint: Oxygen saturation will be recorded at Screening, Day 0, Weeks 4, 12, 24, 36, and 48, and will be collected on the *Oxygen Saturation* CRF page. If possible, oxygen saturation

will be measured on room air; if oxygen saturation is <90% on room air, then the flow rate required to maintain saturation ≥90% will be recorded. If a subject is unable to have oxygen saturation measured on room air and supplemental oxygen is required, then the flow rate required to maintain saturation ≥90% will be recorded.

*A*dditionally, oxygen saturation is recorded at Screening and Weeks 4, 12, 24, 36, and 48 on the *6-Minute Walk Test* CRF page, using the "pre walk" values. Oxygen saturation values and whether the value was recorded on room air or supplemental oxygen will be recorded. If oxygen saturation is measured on supplemental oxygen, then the flow rate used will be recorded as well.

Analysis: Oxygen saturation on room air and flow rate required to maintain saturation ≥90% the will be summarized by treatment group and overall by study visit.

7.4. Examination of Subgroups

Secondary analyses of the primary and secondary objectives may be conducted for the following subgroups, including, but not limited to:

- Race
- Ethnicity
- Gender
- Baseline PVR
- Baseline 6MWD
- NYHA Functional Class
- Duration of SSc-PAH disease

Additionally, exploratory analyses of responder subgroups and subgroups defined through the mechanistic and immunological studies may also be conducted.

8. SAFETY EVALUATION

8.1. Overview of Safety Analysis Methods

All safety analyses will be carried out using the safety sample defined in Section 4 unless otherwise noted. Missing safety information will not be imputed. These analyses will not be stratified by site.

Listings will be prepared for applicable safety measurements. All listings will be sorted in order of treatment, subject identifier (ID), and time of assessment (e.g., visit, time, and/or event).}

8.2. Adverse Events

All AEs will be classified by system organ class (SOC) and preferred term, according to a standardized thesaurus (Medical Dictionary for Regulatory Activities [MedDRA] version 13.1). The severity of AEs will be classified using the National Cancer Institute's (NCI's) Common Toxicity Criteria for Adverse Events (CTCAE) version. Each AE is entered on the electronic case report form (eCRF) once at the highest severity. As such, no additional data manipulation is needed to identify events.

Adverse events NCI-CTCAE grade 2 and above will be collected from the time the subject signs the informed consent until he/she initiates study intervention or until he/she is determined to be ineligible to receive study intervention, if the investigator determines that the adverse event is related to a study-mandated procedure, treatment, or change in treatment.

For all participants, adverse events NCI-CTCAE grade 2 and above will be collected from the time of initiation of study intervention (i.e., the administration of the first dose of study drug/study drug) until he/she completes study participation or until 30 days after he/she prematurely withdraws (without withdrawing consent) or is withdrawn from the study. During the quarterly monitoring period of the study, only NCI-CTCAE grade 3 and higher infections are collected on the adverse events CRF page. Grade 3 and higher infections that occur during the quarterly monitoring period will be listed separately and will not be included in data tabulations, unless otherwise specified.

Treatment-emergent AEs will be identified as those with an onset date on or after the first dose of study medication. If the start of the AE in relation to the start of study medication cannot be established (e.g., the start date for the AE is missing), then the AE will be considered treatment-emergent. All data tabulations will be of only treatment-emergent events unless otherwise specified; non-treatment-emergent AEs will be listed.

An overall summary table will be developed to report the number of events and the number and percentage of subjects having at least one event in the following categories:

- AEs
- AEs requiring 24-hour reporting to the sponsor
- AEs indicated as serious
- AEs that lead to study drug discontinuation
- AEs with an outcome of death
- AEs by relationship to a study drug
- AEs reported by maximum severity
- All NCI-CTCAE Grade 3-5 AEs, which are defined as possibly, probably, or definitely related to rituximab
- All NCI-CTCAE Grade 3-5 adverse events

- All infection related AEs

In addition, AEs classified by MedDRA SOC and preferred term will be summarized for each treatment group and overall for each of the following:

- All AEs
- AEs by maximum severity
- AEs by relationship to study drug: relationship to study treatment will be categorized as either treatment related (possibly, probably, or definitely related to study medication) or unrelated
- All infection related AEs

Summary tables will present the total number of events as well as the number and percentage of subjects experiencing the events. If a subject experiences the same AE on multiple occasions, the event will be counted once for each occurrence when reporting the number of AEs. When reporting the number of subjects experiencing the events, a subject will only be counted once if they experience an event within the particular SOC or preferred term. Percentages will be based on the number of subjects in the safety population.

AEs classified by MedDRA system organ class and preferred term will be summarized for each treatment group and overall, and incidence rates for AEs classified by MedDRA SOC and preferred term will be provided for each treatment group and overall. The incidence rate for a SOC or preferred term will be defined as the number of AEs divided by the person-time of exposure aggregated in each treatment group, and an analysis based on a Poisson regression model will be conducted to compare the rate of each type of event in the treatment groups. The proportion of subjects in each study arm experiencing each type of AE will be compared using a Fisher's Exact test.

Separate data listings will be provided for the following subsets:

- Treatment-related AEs
- AEs leading to study drug discontinuation
- Grade 3-5 AEs which are defined as possibly, probably, or definitely related to rituximab
- All Grade 3-5 AEs
- Infections
- Grade 3 and higher infections that occurred during the quarterly monitoring period
- AEs occurring from informed consent and prior to treatment initiation

8.3. Deaths and Serious Adverse Events

8.3.1. Serious Adverse Events

Serious adverse events (SAEs) and events requiring 24-hour reporting to the sponsor will be summarized by MedDRA SOC and preferred term in a manner analogous to what is described in Section 8.3. SAE summary tables by MedDRA SOC and preferred term will be provided by severity and relationship to study therapy as well.

8.3.2. Deaths

Any death that occurs while a subject is active in the main study (up to Week 48) will be collected and reported as a SAE, and any death that occurs after main study completion will only be reported as a SAE if the subject is active in quarterly monitoring for B cells and if the cause of death is determined to be a Grade 3 or higher infection. Additionally, an assessment of vital status approximately two years post randomization will be conducted via telephone or

review of public records or health registries (to which the subject has consented to share this data). Any deaths discovered from this review will be recorded on the *Post-Study Follow-Up/ Week 104 Vital Status* CRF page.

Treatment-related mortality will be defined as death occurring after randomization and before Week 48, and possibly, probably, or definitely resulting from study treatment. Treatment-related mortality can only be assessed if a death was reported as a SAE; relationship to study treatment was not collected for deaths recorded on the *Post-Study Follow-Up/ Week 104 Vital Status* CRF page. As such, treatment-related mortality will be summarized at Week 48.

All-cause mortality, defined as any death occurring after randomization, will be summarized at Week 48 and Week 104.

In addition, time to all-cause mortality will be defined as the number of months post-randomization until death from any cause. If death does not occur during the post-randomization study period (up to Week 104), then time to all-cause mortality will be set equal to Week 104 or the day of the Week 104 vital status telephone call, whichever occurred first. Appropriate censor variables will be defined to indicate whether the time-to variables represent event times or a censor times. Kaplan-Meier survival curves from each treatment arm with be compared using the log-rank test.

Separate displays listing and summarizing death, including time to death and cause of death, will also be created

8.4. Clinical Laboratory Evaluation

Clinical laboratory measurements include serum chemistry and hematology. Subjects randomized prior to March 19, 2013 had serum chemistry and hematology results analyzed by their sites local lab. Data for these labs is collected on the *Hematology* and *Chemistry* CRF pages. Subjects randomized after March 19, 2013 have serum chemistry and hematology labs run by the central lab, ICON. Results will be converted to standardized units where possible.

For numeric laboratory results, descriptive statistics of laboratory values and the change from baseline of laboratory values will be presented for each treatment group and overall. For categorical laboratory results, the number and percentage of subjects reporting each result will be presented for each treatment group and overall.

Laboratory data will be plotted to show patterns over time. Summary statistics including 25th percentile, median, and 75th percentile will be plotted for each visit by treatment group. Lines connecting individual subject results from subjects with grade 2 or higher values will be overlaid on each figure. For lab results that are not gradable, results from subjects with values outside of 2 *upper limit of normal or 0.5*lower limit of normal will be overlaid. Tests with qualitative results (such as “present” or “positive”) will not be plotted. Reference ranges from the central lab will be used.

In addition, abnormal and clinically significant abnormalities will be summarized and listed separately.

8.5. Vital Signs, Physical Findings, and Other Observations Related to Safety

8.5.1. Vital Signs

Descriptive statistics of vital signs results and change from baseline of vital signs will be summarized for each treatment group and overall. Data listings sorted by treatment group, subject, vital sign parameter, and time of assessment will be provided for vital signs measurements. Data will be plotted to show patterns over time. Summary statistics including 25^{th} percentile, median, and 75^{th} percentile will be plotted for each visit by treatment group.

8.5.2. Physical Examinations

Data listings will be provided for physical examination results and sorted by treatment group, subject, body system, and time of assessment.

9. OTHER ANALYSES

9.1. Use of Medications

9.1.1. Prior and Concomitant Medications

Medications will be coded according to the World Health Organization (WHO) Drug Dictionary (version 2010.03). Medications reported on the *Prior/Concomitant Medications* CRF will be categorized for analysis as prior or after study treatment by comparing the medication start and stop dates with the first and last dose of study medication dates. Prior medications will have both the medication start and stop dates prior to the first dose of study medication date. Concomitant medications will have a medication start date on or after the first dose of study medication. Medications taken for the treatment of PAH will be summarizes separately, see Section 9.1.2 below.

The number and percentage of subjects receiving prior and concomitant medications will be presented overall and by medication class. When reporting the number of subjects receiving the medication, a subject will only be counted once if they ever received the medication within the medication class. Percentages will be based on the number of subjects in the analysis population.

9.1.2. Background PAH Medications

Background PAH medications are recorded on the *Background PAH Medications* CRF page at Screening and when the medication is modified or a new medication is added for the treatment of PAH during the course of the study. Medications will be coded according to the World Health Organization (WHO) Drug Dictionary (version 2010.03).

From the time of study entry, subjects are to remain on background PAH medical therapy with either a single agent or a combination of prostanoid, endothelin receptor antagonist, PDE-5 inhibitor, and/or guanylate cyclase stimulators as per the entry criteria. Doses must remain stable through the week 24 primary endpoint visit.

The type and total number of PAH medications that subjects are on at screening will be summarized by treatment group and overall.

See Section 7.3.4, *Addition of PAH medications*, for more information on analyses related to background PAH medications.

10. INTERIM ANALYSES AND DATA MONITORING

The progress of the study will be monitored by the Data and Safety Monitoring Board (DSMB). The Autoimmune DSMB will be chartered to review safety data and to make recommendations regarding continuation, termination, or modification of the study. The DSMB will formally review the safety data at least yearly. The discontinuation of study treatment will also be periodically reported to the DSMB.

In addition, safety data will be reviewed by the DSMB when an event occurs that is of sufficient concern to the National Institute of Allergy and Infectious Diseases (NIAID) medical monitor or protocol chair to warrant review, or when an event occurs that could contribute to a predefined stopping rule specified in the protocol.

Findings will be reported to Institutional Review Boards (IRBs) and health authorities.

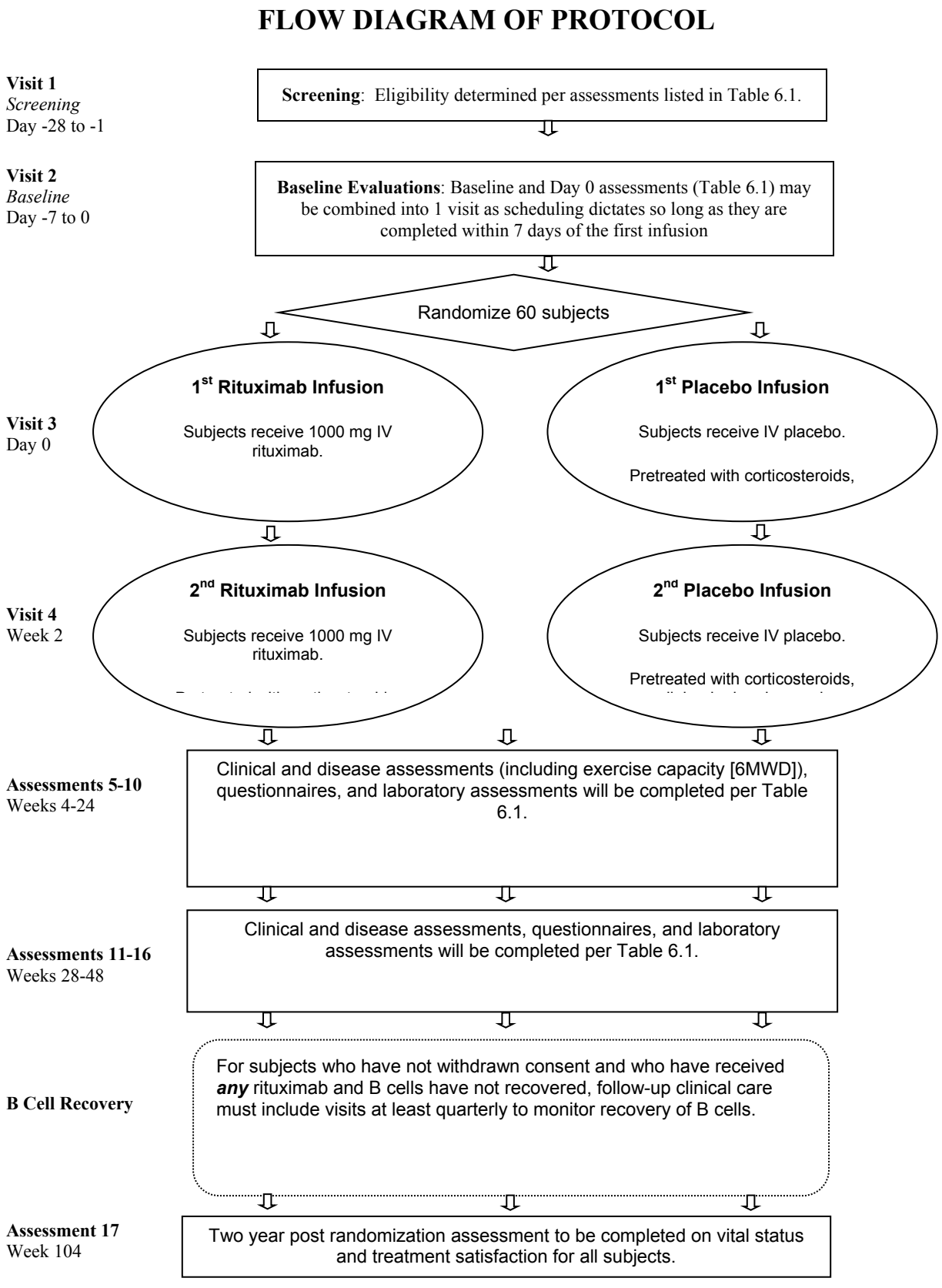
11. CHANGES TO THE ANALYSES PLANNED IN THE PROTOCOL

1. Section 7.2.2: Primary endpoint:
    a. Actual values rather than change from baseline in 6MWD will be used in the model.
    b. Remove baseline measures of 6MWD as covariate
    c. Use visit week rather than visit day
    d. Addition of quadratic term for visit week
    e. Remove limitations from musculoskeletal disease as a covariate. The number of "yes" occurrences is too sparse to inform the analysis.
2. Section 7.3.3: Clinical worsening: Protocol says "worsening of 6MWD by > 20% and a decrease in functional class" vs. corrected "worsening of 6MWD by > 20% and an increase in functional class"
3. Section 7.3.4: Analyses changed from using a logistic regression model to analyzing time to first change or addition of PAH medications.

12. REFERENCES

13. APPENDICES

13.1. Study Flow Chart



13.2. Schedule of Events

Table 6.1, Schedule of Evaluations

Visit Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17		
Description	Screen	Base-line[1]	Day 0[1]	Week 2	Week 4	Week 8	Week 12	Week 16	Week 20	Week 24	Week 28	Week 32	Week 36	Week 40	Week 44	Week 48	Week 104	Quar-terly	Un-sched-uled
Visit window	-28 days	-7 days		±4 days	±7 days	±7 days	±7 days	±7 days	±7 days	±7 days	±7 days	±7 days	±7 days	±7 days	±7 days	±7 days		±14 days	
Type of Visit	Clinic	Clinic	Clinic	Clinic	Clinic	Phone	Clinic	Phone	Phone	Clinic	Phone	Phone	Clinic	Phone	Phone	Clinic	Phone	Clinic	Clinic
Clinical Draw (mL)	21.5	11	2	8	11	0	11	0	0	11	0	0	11	0	0	13	0	3	3
Research Draw (mL)	0	41	0	29	29	0	33	0	0	38	0	0	34	0	0	33	0	24	33
Visit Draw Total (mL)	21.5	52	2	37	40	0	44	0	0	49	0	0	45	0	0	46	0	27	36
General Assessments																			
Informed Consent	X																		
Demographics	X																		
Medical History	X																		
Physical Examination& Vitals[2]	X		X[22]	X	X		X			X			X			X			X
Vaccinations[17]	X																		
NYHA Functional Class Assessment	X				X		X			X			X			X			X[24]
Right Heart Catheterization[20]	X									X									
Pulmonary Function Tests[3]	X									X						X			
Oxygen Saturation[4]	X		X[22]		X		X			X			X			X			X
ECG[5]	X																		
Chest X-Ray	X																		
TB testing[21]	X																		
6MWT [6,18]	X[18]				X		X			X			X			X			X[24]
SF-36		X								X						X			X[24]
SHAQ		X			X		X			X			X			X			X[24]
Digital Ulcer Count		X			X	X	X	X	X	X	X	X	X	X	X	X			X[24]
Concomitant Medications	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X			X
Adverse Event Assessment		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X[23]	X
Post-infusion Phone Assessment[7]			X	X															
Assessment of Vital Status[26]																	X		
Treatment Satisfaction[26]																	X		
Laboratory Assessments																			
Hematology[8]	X	X		X[25]	X		X			X			X			X		X	X[24]
CD19+ B cell count (local)		X																	
Chemistry[9]	X[9]	X[22]			X		X			X			X			X			
Infectious Disease Screen[10]	X																		
Urinalysis	X																		
Pregnancy Test (if applicable)[11]	X		X[22]	X															
CD19+ B Cell Studies[12] (24mL)		X		X	X		X			X			X			X		X	X[24]
Autoantibodies[13] (4mL)		X					X			X						X			X[24]

Serum Igs[14] & IL-1, IL-6,& Other Cytokine Assays[15] (5mL)		X		X	X		X			X			X			X			X[24]
HACA and PK Assays[19] (3-5mL)		X								X			X						
Study Medication																			
Rituximab/Placebo[16]			X	X															

[1] Baseline and Day 0 assessments may be combined into one visit as scheduling dictates so long as they are completed within 7 days of the first rituximab infusion.

[2] Physical Examination: Full PE (including height, weight, vital signs, general appearance, skin, head/eyes/ears/neck/throat, respiratory/chest, cardiovascular, abdominal, neurological, lymph nodes, musculoskeletal/extremities) at Screen, Week 24, and Week 48. Abbreviated PE (including weight, vital signs, skin, respiratory/chest, cardiovascular, abdominal, lower extremity edema) at all other clinic visits.

[3] Pulmonary Function Tests: Height must be measured at time of PFTs. Full PFTs with spirometry, DLCO, and TLC at Screening. DLCO only at Weeks 24 and 48. Sites to send absolute values to Rho to convert for Hgb and altitude adjustments.

[4] Oxygen Saturation: forehead or ear probe must be used for saturations < 95% by finger probe at rest. Measure first at room air, then add/titrate oxygen if needed to maintain O_2 saturation ≥ 90%. The room air oxygen saturation measurement will be conducted at the discretion of the study physician. The measurement should not be performed if subject safety will be compromised.

[5] ECG: If not done with RHC (center-specific).

[6] 6MWT: Subject should complete assessment with same amount of supplemental oxygen used to complete the Screening 6MWT unless safety will be impacted (if the 6MWD decreased from the Screening 6MWD by more than 20%, repeat the 6MWT within 7 days, but not on the same day).

[7] A member of the research staff will contact the subject within 24 hours after discharge from the infusion to assess for adverse events

[8] Hematology: CBC with differential count and platelet count (hematocrit, hemoglobin, platelet count, WBC, bands, eosinophils, lymphocytes, monocytes, and neutrophils). At Baseline, Week 48, and during Quarterly monitoring, CD19+ cells will also be enumerated.

[9] Chemistry: albumin, alkaline phosphatase (ALP), ALT/SGPT, AST/SGOT, BUN, creatinine, potassium, sodium, total and indirect bilirubin, total CO2, total protein, BNP/NT-proBNP. Note: BNP/NT-proBNP is not collected at screening.

[10] Infectious Disease Screen: HIV antibody, hepatitis B surface antibody and surface antigen, HCV antibody with HCV RNA (PCR) if antibody positive.

[11] Pregnancy Test: For women of child-bearing potential a pregnancy test (urine or serum per institutional policy) must be obtained at screening and within 72 hours of each rituximab infusion.

[12] CD19+ B Cell Studies: Blood sent to Stanford Human Immune Processing Center for the Stanford Human Immune Monitoring Center.

[13] Autoantibodies: will include but are not limited to Anti-U1 RNP, anti-U3 RNP, anti-B23, anti-cardiolipin, anti-Th/To, anti-CENP-B, and RF. Frozen serum sent to University of Pittsburgh Immunoassay Core Laboratory.

[14] Serum Immunoglobulins: Quantitative Immunoglobulin Levels including IgG subclasses (frozen serum sent to Stanford Human Immune Processing Center for the Stanford Human Immune Monitoring Center).

[15] IL-1, IL-6, and Other Cytokine Assays: Frozen serum sent to Stanford Human Immune Processing Center for the Stanford Human Immune Monitoring Center.

[16] Rituximab: Per Protocol Section 5, *Treatment of Subjects*.

[17] Vaccinations: All immunizations must be completed at least 4 weeks prior to starting rituximab/placebo. Subject must be vaccinated with the pneumococcal vaccine at least 4 weeks prior to initiation of therapy, unless subject was vaccinated within 5 years of study entry. If vaccination occurred greater than 5 years prior to study entry, the subject must be revaccinated at least 4 weeks prior to initiation of therapy. Depending on the season, influenza vaccine should be administered at least 4 weeks prior to dosing with rituximab/placebo. Otherwise, it should be administered as late as possible in the flu season to still be effective after the last dose of rituximab/placebo.

[18] The 6MWT must be completed within 14 days of the first treatment visit (Day 0).

[19] HACA and PK Assays: HACA and rituximab serum level (PK). Batched and sent to Covance Laboratories.

[20] When completing the repeat RHC, the subject should be on the same amount of supplemental oxygen used during the Screening RHC unless safety will be impacted. The method of calculating cardiac output should be consistent with that used at the Screening RHC.

[21] TB testing (PPD or QuantiFERON®-TB Gold In-Tube Test [QFT-G_IT]): unless performed within 30 days prior to screening and documented as negative in the subject's records or unless subject is known to have a positive test and has documentation of appropriate therapy.

[22] If the Screening physical exam, oxygen saturation, and chemistry panel are performed within 7 days of Day 0, they do not need to be repeated during the Randomization/Baseline or Day 0 visits unless clinically indicated.

[23] Only infectious events ≥ NCI-CTCAE grade 3 will be reported.
[24] Required only if the unscheduled visit is due to an increase in SSc-PAH disease activity; see Section 6.5.8, *Unscheduled Visit* for further details.
[25] Hemoglobin, platelets, lymphocytes, neutrophils, and bands results from a hematology specimen drawn within 24 hours of the infusion must be reviewed prior to initiating the study drug infusion (see Section 5.3.6, *Criteria for Withholding Rituximab Infusion*).
[26] A member of the research staff will obtain verbal consent to conduct an assessment of vital status and treatment satisfaction approximately two years post randomization for each subject. If the subject cannot be reached, a review of public records or health registries should be completed.

14. ATTACHMENTS

14.1. Scoring of the SF-36