# Reata Pharmaceuticals, Inc.

## Protocol 402-C-1504

*Protocol Version 2.0 (October 18, 2016)*

## CATALYST

**A double-blind, randomized, placebo-controlled Phase 3 trial to study the safety, tolerability, and efficacy of bardoxolone methyl in patients with connective tissue disease-associated pulmonary arterial hypertension**

# Statistical Analysis Plan
## Version 2.0

Version 2.0: May 26, 2020
Original: December 19, 2016

# SIGNATURE PAGE

This document has been prepared and/or reviewed by:

[redacted]

[redacted]

5/26/2020 | 4:06:28 PM
Date

[redacted]

5/26/2020 | 4:16:03 PM
Date

This document has been reviewed and accepted by:

[redacted]

[redacted]

5/26/2020 | 1:49:40 PM
Date

[redacted]

5/26/2020 | 4:56:47 PM
Date

# CONTENTS

## TABLES



## FIGURES

# ABBREVIATIONS

| | |
|---|---|
| 6MWT | 6-minute walk test |
| AE | adverse event |
| ALP | alkaline phosphatase |
| ALT | alanine transaminase |
| AST | aspartate transaminase |
| ATS | American Thoracic Society |
| BMI | body mass index |
| BNP | brain natriuretic peptide |
| BUN | blood nitrogen |
| CBC | complete blood count |
| CI | confidence interval |
| CPK | creatine phosphokinase |
| CSR | clinical study report |
| CTD | connective tissue disease |
| CTD-PAH | connective tissue disease-associated pulmonary arterial hypertension |
| [redacted] | [redacted] |
| DSMB | Data Safety Monitoring Board |
| EC | Ethical Committee |
| eCRF | electronic case report form |
| ECG | electrocardiogram |
| eGFR | estimated glomerular filtration rate |
| FC | functional class |
| FCS | fully conditional specification |
| FDA | Food and Drug Administration (US) |
| GFR | glomerular filtration rate |
| GGT | gamma glutamyl transferase |
| Hb | hemoglobin |
| HCG | human chorionic gonadotropin |
| HDPE | high-density polyethylene |
| ICH | International Conference on Harmonization |
| IgG | immunoglobulin G |
| IRB | Institutional Review Board |
| IWRS | Interactive Web Response System |
| ITT | intent-to-treat |
| LDH | lactate dehydrogenase |
| LLD | lower limit of detection |

## ABBREVIATIONS, CONTINUED

| | |
|---|---|
| MCH | mean corpuscular hemoglobin |
| MCHC | mean corpuscular hemoglobin concentration |
| MCV | mean corpuscular volume |
| MDRD | modification of diet in renal disease |
| MedDRA | Medical Dictionary for Regulatory Activities |
| METS | metabolic equivalent tasks |
| MMRM | mixed model repeated measures |
| MNAR | missing not at random |
| NKDEP | National Kidney Disease Education Program |
| PAH | pulmonary arterial hypertension |
| PI | principal investigator |
| PPD | Pharmaceutical Product Development |
| PT | preferred term |
| RBC | red blood cell |
| Reata | Reata Pharmaceuticals, Inc. |
| REB | Research Ethics Board |
| RNA | ribonucleic acid |
| RT-PCR | reverse transcription polymerase chain reaction |
| SAE | serious adverse event |
| SAP | Statistical Analysis Plan |
| [redacted] | [redacted] |
| SD | standard deviation |
| [redacted] | [redacted] |
| SOC | system organ class |
| SRC | Sample Size Re-calculation Committee |
| TEAE | treatment-emergent adverse event |
| ULN | upper limit of normal |
| VCS | visual contrast sensitivity |
| WBC | white blood cell |
| WHO | World Health Organization |

# REVISION HISTORY

| Version | Date | Document Owner | Revision Summary |
|---|---|---|---|
| 1.0 | December 19, 2016 | [redacted] | Initial version. |
| 2.0 | May 26, 2020 | [redacted] | [redacted]<br>Revised definitions of [redacted] clinical improvement<br>Updated rules for rounding when taking averages<br>Removed sensitivity analyses using multiple imputation<br>Added additional analyses of transaminases<br>Added off-treatment safety analyses<br>Added medical history analyses |

**The analysis plan is based on the information from the following document:**

**Protocol Version 2.0, 18 October 2016**

# 1. INTRODUCTION

Study 402-C-1504 (the CATALYST study) is a randomized, double blind, placebo-controlled Phase 3 study designed to compare the safety, tolerability, and efficacy of bardoxolone methyl to placebo in patients with connective tissue disease-associated pulmonary arterial hypertension (CTD-PAH). This statistical analysis plan (SAP) defines the methods and analyses that Reata Pharmaceuticals, Inc. (Reata) will use to analyze data from the CATALYST study.

This SAP complies with the International Conference on Harmonisation (ICH) guidance and relevant Food and Drug Administration (FDA) guidances. It is based on Version 2 of the study protocol dated October 18, 2016. If the protocol is subsequently amended, this SAP may be amended as well. Should the SAP and the protocol be inconsistent with respect to the planned analyses, the language of the SAP is governing. This version of this SAP describes the analyses planned prior to the database lock and supersedes the statistical considerations identified in the study protocol. Reata (the Sponsor) will follow this SAP in analyzing the study data. Unless otherwise specified, these analyses will be summarized in the clinical study report (CSR). Any substantive changes made to the SAP after the database lock will be clearly identified, and any analyses in addition to those specified in the SAP after the database lock will be considered ad hoc. The CSR will describe any deviations from the planned analyses.

# 2. STUDY DESCRIPTION

## 2.1. Study objectives

### 2.1.1. Efficacy

The primary efficacy objective of the CATALYST study is to assess the efficacy of bardoxolone methyl relative to placebo in increasing distance in the six-minute walk test (6MWT) from baseline to Week 24.

### 2.1.2. Safety

The safety objective of the CATALYST study is to assess frequency, intensity, and relationship to study drug of adverse events (AEs) and serious adverse events (SAEs), and change from baseline in the following assessments: physical examinations, vital sign measurements, 12-lead electrocardiograms (ECGs), clinical laboratory measurements, and weight.

## 2.2. Study design

The CATALYST study will randomize approximately 130 patients with CTD-PAH. This number may be increased to approximately 200 patients depending on the results of a blinded sample size re-calculation.

Patients who meet all inclusion criteria and no exclusion criterion and have satisfied screening evaluations will be randomized 1:1 to receive either placebo or bardoxolone methyl administered once daily for 24 weeks. Patients randomized to placebo will remain on placebo throughout the study and undergo sham titration. Patients randomized to bardoxolone methyl will start at 5 mg

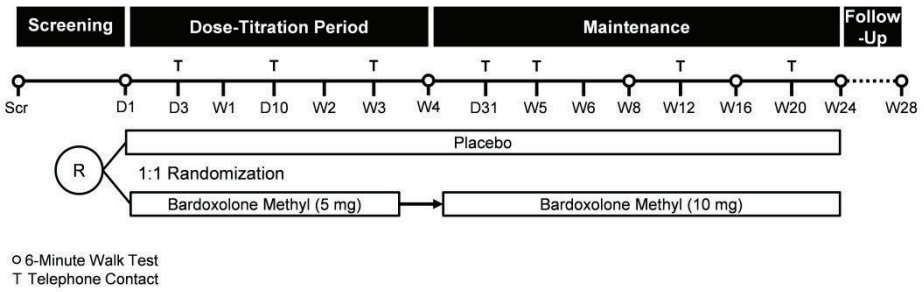
and will dose-escalate to 10 mg at Week 4, unless contraindicated clinically. Dose de-escalation is permitted during the study if indicated clinically.

### 2.2.1. Stratification

Randomization will be stratified by number of background pulmonary arterial hypertension (PAH) therapies into three distinct strata with Stratum 1 consisting of patients with no background PAH therapy, Stratum 2 consisting of patients with one background PAH therapy, and Stratum 3 consisting of patients with two background PAH therapies.

All patients in the study will follow the same visit and assessment schedule, shown in Figure 1. Following randomization, patients will be scheduled to be assessed in person during treatment at Weeks 1, 2, 4, 6, 8, 16, and 24 and by telephone contact on Days 3, 10, 21, 31, 38, 84, and 140. Patients will also be scheduled to be assessed at an in-person follow-up visit at Week 28, four weeks after the end of treatment.

**Figure 1: Schema for study of bardoxolone methyl in patients with CTD-PAH**




The number of administered doses will be calculated based on date of first dose, date of last dose, drug dispensed, and pill counts from drug returned. Patients who administer at least 126 doses (i.e., 75% of the 168 expected doses) will be considered treatment compliant. Patients who complete the study (i.e., complete study visits through Week 28), complete treatment (i.e., complete study medication through Week 24), and are considered treatment compliant will be eligible for a separate open-label extension study.

### 2.2.2. Discontinuation and termination of study patients

Patients may discontinue study drug or withdraw from the study at any time for any reason. Additionally, the sponsor reserves the right to terminate the study, and the investigator may discontinue an individual patient from study drug. Consultation with the medical monitor should occur before discontinuing study drug or withdrawing a patient from the study. The reason for a patient's discontinuation from study drug or study termination will be recorded in the electronic case report form (eCRF).

Patients who discontinue study drug should remain in the study and undergo all scheduled assessments, if possible. Patients who terminate from the study prior to the Week 24 study visit should be brought back to the clinic as soon as possible for early termination assessments (i.e., Week 24 Day Prior, and end-of-treatment visits) as well as a follow-up visit four weeks later. Patients should complete the end-of-treatment procedures, including the 6MWT, prior to any adjustments being made to their PAH treatment regimen.

### 2.2.3. Date of last contact

The last date of contact for each patient is defined as the date of last telephone contact or in-person visit at or before the Week 28 follow-up visit, except in the following cases:

- For those patients who withdraw consent to participate in the study prior to 24 weeks: the date of withdrawal of consent
- For those patients who die prior to 24 weeks: the date of last contact in person or by telephone prior to death

## 2.3. Study drug

The study drug will be supplied in tamper-evident kits containing two 30-cc high-density polyethylene (HDPE) bottles. Each bottle of study drug will contain 30 capsules of 2.5 mg or 5 mg strength bardoxolone methyl or the matching placebo capsules.

## 2.4. Power and sample size

The primary analysis of the efficacy data will be based on the intent-to-treat (ITT) patient population (see Section 3 for a definition of the analysis populations). Primary efficacy is based on an estimate of the change from baseline to Week 24 in the 6MWT. Computation of the statistical power of the primary analysis will be based on the difference between change from baseline in the bardoxolone methyl and placebo groups.

## 2.5. Blinded sample size recalculation

Since there is uncertainty associated with the values used for the computation of the power, the protocol allows recalculation of the sample size when approximately 100 patients have been randomized. Using blinded information, the statisticians responsible for the sample size recalculation will reassess relevant baseline and operational parameters and compute the power using the simulation approach outlined in Appendix A. Based on the results obtained from this recalculation, the study size will be increased, if necessary, to maintain the desired level of power. As no unblinding will occur, the recalculation will have at most a negligible impact on the Type 1 error rate for the final analysis. Appendix B contains details that provide statistical and operational guidelines for conducting the planned recalculation of the sample size.

In the event that the study accrues quickly, so that there is insufficient follow up data for a sample size recalculation, the study size may be increased up to 200 subjects.

## 2.6. Randomization and blinding

Patients will be randomized in a 1:1 ratio to receive bardoxolone methyl or placebo, and randomization will be stratified by number of background PAH medications. Three strata will be used: zero background PAH medications, one background PAH medication, and two background PAH medications.

All patients, investigators, site personnel, laboratories, and Reata personnel with direct involvement in the conduct of the study and their designees will be blinded to treatment assignments and will remain blinded until after the database is locked and access to the randomized treatment codes authorized (Section 6.2). The only individuals with access to treatment assignments will be those individuals who develop and maintain the randomization code, the Interactive Web Response System (IWRS) group, the Data Safety Monitoring Board (DSMB), the external unblinded statistical analysis center providing data to and analyzing data for the DSMB, and safety personnel without direct involvement in the conduct of the study who are assigned to report unblinded data at the occurrence of unexpected SAEs to regulatory authorities as required.

# 3. ANALYSIS POPULATIONS

The study has two analysis populations, which are described below.

## 3.1. Intent-to-treat population

The ITT population consists of all randomized patients categorized by their assigned treatment group regardless of treatment exposure. Analyses of both the primary and secondary efficacy outcomes will use the ITT population. Some sensitivity and exploratory analyses will also use the ITT population.

## 3.2. Safety population

The safety population includes all patients who received at least one dose of randomized study drug. Patients who received at least one dose of bardoxolone methyl will be classified in the

bardoxolone methyl group. Patients who received at least one dose of placebo and no bardoxolone methyl will be classified in the placebo group.

# 4. SOFTWARE

Most statistical analyses will be performed using SAS version 9.3 or later, with SAS program code prepared specifically for the project by qualified statisticians and SAS programmers. Other validated programs (e.g., S-Plus, R, CART) may be used for graphs or for specialized analyses.

# 5. GENERAL CONVENTIONS

The protocol-specified time windows for a visit (Appendix D) and time of a measurement (Section 5.4) determine the visit assignment for each observation.

Continuous data will be summarized using descriptive statistics: number of observations (n), mean, median, standard deviation (SD), minimum, and maximum. Confidence intervals (CIs) will be presented as appropriate. The same number of decimal places as in the observed value will be presented when reporting minimum and maximum; 1 more decimal place than in the observed value will be presented when reporting mean and median; and 2 more decimal places than in the observed value will be presented when reporting SD. When values are averaged in the analysis data (e.g., in the calculation of baseline), the same number of decimal places as in the observed value will be presented.

Frequencies and percentages will summarize categorical data. Presentations of categorical data will generally suppress percentages for items where the count is zero in order to draw attention to nonzero counts.

Confidence intervals will generally be constructed at the 95% level. For binomial variables, exact methods will be employed unless otherwise specified.

Categorical/qualitative data will be presented using frequency counts and percentages. All percentages will be rounded to 1 decimal place, unless otherwise specified. Percentages equal to 100 will be presented as 100% and no percentages will be presented for zero frequencies.

Summary presentations will classify patients by their randomized treatment group for analyses on the ITT population, and actual treatment received for analyses on the safety population, unless otherwise specified. References to treatment group from this point forward in the SAP text indicate randomized treatment group, unless stated otherwise.

## 5.1. Definition of baseline

This section describes which measurements taken prior to first dose of study drug will be used to calculate baseline for each study parameter. No measurement collected after first dose will be used in the calculation of baseline. The collection date and time of the sample will be compared to the date and time of first dose. Should a patient never receive study drug, only measurements collected prior to randomization will be used in the calculation of baseline values.

In general, the baseline value for any laboratory or non-laboratory parameter will be calculated as the mean of data sampled up until and including Day 1, but not occurring after first dose of study drug, except in the following circumstances:

- Whenever specific visit results (i.e., Screening A, B, or Day 1) are indicated, the results will be used regardless of whether the measurements were collected within the protocol-specified window of time (Appendix D), as long as they precede first dose.
- For eligibility purposes, only two 6MWT observations (Screening B and Day 1 preferred to Screening A and Day 1) will be used to determine eligibility (must be within 15% of each other and at least 150 m). Patients not meeting the eligibility criteria are listed.
- As the covariate in the primary efficacy analysis, the screening 6MWT will be either the average of Screening A 6MWT and Screening B 6MWT or the single Screening A 6MWT result if Screening B did not occur.
- For the outcome in primary analysis, the Day 1 6MWT will be subtracted from the 6MWT observation at each followup time to give the change score.
- For discrete efficacy outcomes, [redacted], the latest assessment will be used for analysis purposes.
- If a scheduled measurement prior to first dose is repeated, baseline calculations will use only one result per scheduled measurement (i.e., either the original or repeated). The original value will be used, unless it is missing. The repeated value will only be used if the original value is missing, otherwise the repeated value will be ignored. The CSR will include listings of all repeated measurements collected prior to first dose, indicating whether they were included in the calculation of baseline.
- For repeated laboratory parameters, if both an original and retest result are available in the laboratory dataset, the original result will be used and the second (i.e., 'retested') result will be ignored. Otherwise, if the original result is missing (e.g., because the laboratory detected an error) but a retest result is available, the retested result will be used as the scheduled measurement, and the original result will be ignored.

## 5.2. Definition of scheduled post-baseline measurements

Instead of relying solely on visit labels in the clinical database for post-baseline values, results collected on a date will be attributed to a specific time point by calculating the days relative to randomization. Listings will present all measurements.

Section 5.4 describes how time is calculated for each assessment.

### 5.2.1. Laboratory parameters

Post-baseline laboratory measurements are defined as the scheduled or unscheduled measurements collected closest to (or on) the patient's target visit day (Appendix D), with collection date identified in the central laboratory data. If two assessments are equidistant from a

target study day, the earlier assessment is used. If the visit used for analysis includes two assessments on the same day, the average of the two measurements will be used. As with the baseline retest convention, if both the original and a retest result for a scheduled post-baseline laboratory are available in the central laboratory data, the original result will be used and the second (i.e., 'retested') result will be ignored. Otherwise, if the original result is not available (e.g., central laboratory detects an error) but a retest result is available, the retested result will be used as the scheduled post-baseline measurement.

### 5.2.2. Non-laboratory parameters s

Post-baseline measurements of non-laboratory parameters are defined as the measurements collected on the day closest to the patient's target visit day (Appendix D). If two assessments are equidistant from a target study day, the earlier assessment is used. If the visit used for analysis includes two assessments on the same day, the average of the two measurements will be used. The last clinic visits on and off study drug will be classified according to when they occurred as specified in Section 5.2.3.

### 5.2.3. Last clinic visit and post-treatment results

Patients who terminate from the study before the Week 24 study visit should be brought back to the clinic as soon as possible for early termination assessments as well as a follow-up visit four weeks later. All assessments through Week 24 will be classified according to the time relative to randomization, regardless of the patient's study participation or study drug status. Post treatment assessments will be classified according to the time relative to last dose in Appendix D, for example 4 weeks off-treatment is within 7 to 35 days after last dose of the study.

Patients who discontinue treatment early but continue to be followed until the end of study will have their assessments attributed to visits as specified in Section 5.2.1 and Section 5.2.2.

A listing will indicate when each patient's visits occurred. Appendix D indicates how scheduled results will be classified.

## 5.3. Handling of results of unscheduled tests

As discussed below in Section 5.5, retests will be used for partial 6MWTs stopped for non-medical reasons and repeated at an unscheduled visit.

## 5.4. Definition of time

For visits (or events) that occur on or after randomization, time is calculated in days as:

time = visit (or event) date – date of randomization + 1

For visits (or events) that occur prior to randomization, time is calculated in days as:

time = visit (or event) date – date of randomization

The quantity 'days since first dose' is defined as:

- days since first dose = visit (or event) date – date of first + 1

The quantity 'days since last dose' is defined as:

days since last dose = visit (or event) date - date of last dose For summaries that present distribution of time expressed in weeks and months, weeks will be defined as days divided by seven and months as days divided by 30.4375.

## 5.5. Missing and partial data

Early termination of the 6MWT may be due to one of the six American Thoracic Society (ATS) immediate stopping criteria, low oxygen, or elective non-medical reasons. When a partial 6MWT is caused by non-medical reasons (e.g., a bathroom break), the patient is encouraged to return for an unscheduled visit to redo the 6MWT.

For the analysis of safety variables, only partial dates may be imputed; otherwise, missing data will be treated simply as missing.

The central laboratory uses assays that have lower limits of detection (LLD). All laboratory results below the LLD will be imputed as 1/2 of the LLD. Results reported as greater than a value (i.e., '> value') will be imputed as 1.0 × that value. In the event that a large percentage (> 20% of values) of laboratory values are out of the range of the limits of detection, then a footnote will be added to indicate that the percentage of values falling outside of the limits of detection is greater than 20%.

# 6. DATABASE LOCK AND UNBLINDING

## 6.1. Final database lock

After the database lock and the authorization for unblinding (Section 6.2), the treatment codes will be merged to the analysis datasets. Any change to the clinical database after this time will require written authorization, with explanation, by Reata. In addition, beginning at the time of database lock, an audit trail will be maintained of all versions of the analysis datasets that may result from refinements of the algorithms for derived variables in the course of the analysis.

## 6.2. Authorization for unblinding

After database lock and upon receipt of written authorization from Reata, a blinded study team will receive the actual treatment codes directly from the group maintaining the IWRS. This team will generate top-line results and provide them to Reata along with the analysis files containing the randomized treatment codes. Reata will not have direct access to the randomized treatment codes until they have been provided the top-line results and datasets.

Reata's standard operating procedures dictate how access to the treatment codes will be extended within the company and to investigative sites.

# 7. PATIENT SUMMARIES

## 7.1. Enrollment

Enrollment will be summarized by country and geographical region. A table will present the number and percentages of patients in each stratum by treatment group.

## 7.2. Disposition

The disposition of patients (number randomized, study drug exposure, duration of follow-up, and inclusion in the various study populations) will be tabulated by treatment group in the ITT and Safety populations. Completion of study treatment, completion of study, discontinuation of study drug, discontinuation of study, and reason for discontinuation of study drug treatment and study will be summarized. See Section 8.2 for other summaries of treatment exposure. Time to permanent study drug discontinuation will be summarized. If the treatment end date or study end date are missing, the date will be imputed as the Study Drug Termination (SDT). If the treatment end date or study end date are missing and no reason for treatment or study discontinuation are included in the CRF, the reason for treatment or study discontinuation will be study termination.

## 7.3. Demographic and baseline characteristics

Summaries of demographic and other baseline characteristics will be presented by treatment group for all analysis populations (Section 3). The two treatment groups will be assessed descriptively for comparability of demographic and baseline characteristics for the ITT study population. Demographics and baseline characteristics will be summarized by subgroups.

Demographics and other baseline characteristics include the following:

- age (years) at the date of informed consent, sex, ethnicity, race
- height, weight, body mass index (BMI)
- baseline disease characteristics (e.g., WHO functional class, baseline 6MWT, years since PAH diagnosis)
- connective tissue disease (CTD) etiology (scleroderma, lupus, mixed, and other)
- pulmonary function testing results
- PAH and CTD medications
- geographic region
- baseline eGFR
- baseline eGFR category

  $\geq 90$ (ml/min/1.73m$^2$)

  60 to < 90 (ml/min/1.73m$^2$)

  45 to < 60 (ml/min/1.73m$^2$)

  30 to < 45 (ml/min/1.73m$^2$)

  15 to < 30 (ml/min/1.73m$^{2)}$

  < 15 (ml/min/1.73m$^2$)
- other variables of interest

## 7.4. Medical history

Medical history is summarized by treatment. Medical history is coded using MedDRA (Medical Dictionary for Regulatory Activities) version 19.0. Medical history items are summarized by MedDRA SOC and PT.

# 8. STUDY DRUG AND OTHER MEDICATIONS

## 8.1. Prior and concomitant medications

All concomitant medications collected from screening through the end of the study will be classified by preferred terms according to the WHO Drug Dictionary (version 01MAR2016, type DDE+HD, format B2). Concomitant medications will be summarized by WHO Drug Dictionary classification. Special groups of medications (e.g. PAH, CTD) will be summarized by treatment group, by medical history as appropriate, and dose, if feasible.

## 8.2. Duration of treatment and exposure to study drug

The study drug will be supplied in tamper-evident kits containing two 30-cc high-density polyethylene (HDPE) bottles. Each bottle will utilize foil induction-seal liners and a child-resistant closure. Each bottle of study drug will contain 30 capsules of 2.5 mg or 5 mg strength bardoxolone methyl or the matching placebo capsules. Duration of treatment will be defined for each patient as the number of days the patient received study medication [(date of last dose) – (date of first dose) + 1]. Total number of administered doses will be calculated for each patient using duration of treatment minus missed doses. Missed doses are calculated from 30-count bottles and number of expected doses based on treatment duration. For example, a patient returning 3 capsules in each bottle after 28 days of treatment is assumed to have administered 27 doses and missed 1 dose. Since two bottles are included in each kit, if the number of returned capsules in each bottle is not equal, the bottle with the maximum number of returned capsules will be used for calculating missed doses. Total number of doses dispensed and total dose (mg) dispensed are calculated from total number of kits (bottles) recorded on the Study Drug Dispensation eCRF. Total number of doses received are calculated from information on the eCRF of Study Drug Return and Study Drug Dispensation, as the total number of doses dispensed – total number of doses returned. Study drug compliance (%) is calculated as 100 × total number of doses received / total number of doses dispensed. The proportion of patients who have ≥80% compliance is summarized. To summarize the actual dose-titration, frequency counts and percent of patients on each dose-level (as determined by the dispensed kit strength) will be tabulated for each scheduled dispensing visit.

Standard descriptive statistics will be used to summarize duration of treatment, number of doses administered, number of missed doses, percent treatment compliance, and total exposure. Descriptive statistics will be calculated and tabulated by treatment group and study completion status for the ITT study population, and if warranted, for the other study populations. All dosing information will be listed by treatment group and patient ID. Study drug administration over time will be summarized in a listing showing study drug dispensed at each clinic visit.

# 9. ESTIMATED GLOMERULAR FILTRATION RATE (EGFR)

Estimated glomerular filtration rate (eGFR) value will be calculated at the central laboratory and used in data analyses. The equation used to calculate eGFR for each patient will be the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation:

• eGFR (mL/min/1.73 $m^2$) = $141 \times \min(Scr/\kappa, 1)^{\alpha} \times \max(Scr/\kappa, 1)^{-1.209} \times 0.993^{Age} \times$

1.018 [if female] × 1.159 [if black].

Where Scr is serum creatinine (mg/dL), κ is 0.7 for females or 0.9 for males, and α is -0.329 for females or -0.411 for males, and Age is the age at Screening. Min indicates the minimum of Scr/κ or 1, and max indicates the maximum of Scr/κ or 1.

Missing eGFR values will not be imputed for the analysis of efficacy.

# 10. EFFICACY OUTCOMES

## 10.1. Primary efficacy outcome

The primary efficacy variable is the change from baseline in 6MWT.

## 10.2. Secondary efficacy outcome

The secondary efficacy outcome is the time-to-first persistent clinical improvement event. To declare that clinical improvement has occurred, at least one of the following four criteria must be met:

1. Improvement by at least one WHO functional class coupled with no more than a 15% decrease from baseline in 6MWT.
2. Increase from baseline in 6MWT by at least 10% and stability or improvement in the WHO functional class.
3. Decrease from baseline in creatine kinase (a surrogate biomarker for muscle injury and inflammation) by at least 10% and no worsening in WHO functional class and no more than a 15% decrease from baseline in 6MWT.
4. Improvement in estimated glomerular filtration rate (eGFR) ≥10% of baseline.

The persistence of the change in WHO functional class, 6MWT, eGFR, or creatine kinase must be confirmed by a subsequent assessment that also meets the defined criteria. The confirmatory assessment must occur at least 14 days after the initial assessment, or at the next scheduled assessment. If persistent improvement is confirmed, the date of the event will be considered the initial assessment of improved WHO functional class, 6MWT, eGFR, or creatine kinase.

If confirmatory observation of initial improvement is not available for any reason, such as end of study or death, it may or may not be considered an event as defined by the timing of improvement as it relates to other study events. Below is a list of possible reasons for failure to obtain a confirmatory observation along with the adjustment to the definition of improvement.

These reasons are as follows:

1. Last visit: If the improvement is observed at the last scheduled assessment, the requirement for an additional assessment after 14 days is waived, and the patient is classified as satisfying the criteria for improvement. In the absence of study end date, Study Drug Termination (SDT) is used.
2. Death: If the patient's death occurs before the confirmatory observation, then the patient is classified as not satisfying the criteria for improvement.
3. Study termination: If the patient terminates the study for reasons of "lost to follow-up" or "withdrawal of consent", then the patient is classified as not satisfying the criteria for improvement.
4. Patient starts new PAH drug: If a patient decides to stop study drug and switch to a new PAH drug, then the patient is classified as not satisfying the criteria for improvement.
5. Patient discontinues study drug: If a patient discontinues the study drug due to an AE and improvement is confirmed, then the patient is classified as satisfying the criteria for improvement. If improvement is not confirmed or if the patient has no follow-up measurement, then the patient is classified as not satisfying the criteria for improvement.

## 10.3. Exploratory efficacy outcomes

# 11. STATISTICAL METHODS

## 11.1. Primary efficacy outcome

### 11.1.1. Primary analysis

This section describes the methods planned for the primary analysis, which is based on the primary efficacy outcome, the change from baseline in 6MWT. The analysis will compare the mean change in the bardoxolone methyl group relative to placebo at Week 24. A mixed model will be fit with this outcome using the following PROC MIXED code in SAS, where ID is the patient identifier, SCR_6MWT and DELT_6MWT are the screening values for the 6MWT and change from baseline, respectively, in the primary outcome at each visit (VOBS) by treatment group (GRP), and N_PAH_MEDS is the fixed number of background PAH therapies for a patient. Below is example code used to estimate the treatment effect:

```
proc mixed data = &DAT. method=reml alpha=0.05 covtest;
   class  ID GRP (ref="plc") VOBS (ref=LAST)
     N_PAH_MEDS;
   model DELT_6MWT = SCR_6MWT
                     N_PAH_MEDS
                     GRP
                     VOBS
                     GRP * VOBS
                     VOBS * SCR_6MWT Hb / s ddfm=kr;
   repeated VOBS / subject = ID type = UN r;
   lsmeans  VOBS * GRP / diff om cl e;
run;
```

The model will have the following characteristics:

- The response variable will be the vector of observed change in the primary efficacy outcome from baseline at each study visit where the 6MWT is measured (Week 4, Week 8, Week 16, and Week 24). The outcome will consist of the following four repeated measures: Week 4 6MWT – baseline 6MWT, Week 8 6MWT – baseline 6MWT, Week 16 6MWT – baseline 6MWT, and Week 24 6MWT (averaged over both measurements made at Week 24) – baseline 6MWT. [redacted] 6MWT values collected after the date of clinical worsening will be treated as missing for analysis of 6MWT.
- Repeated post-baseline measurements from each patient will be identified by patient identifier (ID).
- Within-patient correlations will be modeled using an unstructured covariance structure. In the unlikely situation that this model does not converge (i.e., the study has too few observations for the number of parameters estimated), the model will use the heterogeneous Toeplitz structure, which assumes the correlation between two repeated measurements depends solely on their lag in visit number, and that the variance of the outcome may differ over time. If the model still does not converge

with the heterogeneous Toeplitz structure, the model will use a homogeneous Toeplitz structure. Finally, if the model using the homogeneous Toeplitz structure does not converge, the model will use a compound symmetry structure which assumes equal correlation for a patient's measurements, regardless of how far apart in time they were taken.

- The models will include the following covariates:
  - Time defined as a categorical variable (i.e., no restriction is imposed on the trajectory of the mean outcome over time) (VOBS)
  - Treatment group (GRP)
  - N_PAH_MEDS (stratification variable for number of background PAH therapies)
  - Screening 6MWT (SCR_6MWT)
  - Screening 6MWT by time interaction (VOBS * SCR_6MWT)
  - Day 1 hemoglobin (Hb)
  - Time by treatment group interaction (GRP * VOBS)

The MMRM, as defined above, uses the screening value of the 6MWT as a covariate and the baseline value of the 6MWT for the computation of the outcome (the change from baseline). This allows adjustment for the "baseline" value between groups while providing the ability to provide valid estimates of the placebo-adjusted treatment effect.

In the event the MMRM model with an unstructured covariance structure does not converge, the following covariance structures are substituted, in the order listed. Each subsequent covariance structure is used only if each previous covariance structure is used and no previous model converged.

1. Heterogeneous Toeplitz covariance structure (assuming different variances at each time point and that measurements taken closer together in time are more highly correlated than those taken farther apart).
2. Toeplitz covariance structure (assuming measurements taken closer together in time are more highly correlated than those taken farther apart).
3. First order auto-regressive [AR(1)] covariance structure (assuming measurements taken closer together in time are more highly correlated than those taken farther apart, but the correlation is more constrained than the Toeplitz structure).
4. Compound symmetry covariance structure (assuming equal correlation for measurements from a patient, regardless of how far apart in time they were taken). A compound symmetry covariance structure was assumed in the sample size calculation.

Results of MMRM are summarized in tables and line plots. Additionally, line plots of 6MWT change from baseline, labeled by treatment group up to and including Week 24 are generated

## 11.2. Secondary efficacy outcome

The secondary efficacy outcome, the time to clinical improvement, will be analyzed using methods for a censored time-to-event outcome. All analyses will be stratified using number of background PAH medications at baseline to define the strata.

Summary tables will present median time to first clinical improvement and associated 95% CIs for each treatment group as well as hazard ratios and 95% CIs for each comparison of bardoxolone methyl to placebo. Median duration with appropriate 95% CIs, as well as the proportion of patients remaining event-free at times of interest, will be estimated using Kaplan-Meier methods with the product limit estimator computed using PROC LIFETEST in SAS. PROC PHREG in SAS will be used to estimate hazard ratios, mean and standard error time -to-event, and their 95% CIs by Cox regression models.

[redacted]

In the event that the number of events across strata is too small, the strata consisting of zero background PAH medications and one background PAH medication will be collapsed.

[redacted]

Treatment group comparisons will use the Type 3 Chi-square test and unadjusted two-sided p-value associated with the treatment group variable in the Cox proportional hazards model. Time-to-event analyses will use Breslow's method of handling ties in event times [1], which is the default method used in PROC PHREG. The TEST statement in PROC LIFETEST will be used to include the strata and the STRATA statement will be used in PROC PHREG.

Kaplan-Meier plots will supplement these presentations. The number of patients within each treatment group at risk of an event at regular time points will appear at the bottom of these figures.

The secondary outcome is important in understanding the effectiveness of bardoxolone methyl and its ability to improve function.

## 11.3. Exploratory efficacy outcomes

[redacted]

### 11.3.1. Analysis of exploratory change from baseline endpoints

[redacted]

### 11.3.2. [redacted]

[redacted]

[redacted]

### 11.3.3. Analysis of exploratory time-to-event outcomes

[redacted]

### 11.3.4. [redacted]

[redacted]

# 12. SAFETY ANALYSES

## 12.1. Safety variables

Safety variables will include physical examinations, vital sign measurements, ECG, clinical laboratory measurements, concomitant medications, AEs, and SAEs. The safety analyses will include all results collected from randomization through the end of the study. The safety analyses will include only patients in the safety population, with patients classified by whether they received any dose of bardoxolone methyl or not. For laboratory data, vital signs, and ECG data, On-treatment values will be summarized according to the analysis study windows and off-treatment values will be summarized according to the analysis study windows in Appendix D .

## 12.2. Adverse events and serious adverse events

Each treatment emergent AE (TEAE) will be coded using the Medical Dictionary for Regulatory Activities (MedDRA version 19.0). An AE is defined as any untoward medical occurrence in a patient regardless of its causal relationship to study treatment. A TEAE must have an onset or worsen in severity on or after the first dose of study drug and not more than 30 days after the date of the last dose of study drug. AEs with an onset greater than 30 days after the last dose of study drug, and therefore not treatment-emergent AEs (TEAEs), are classified as off-treatment AEs (OTAEs). Analyses described for TEAEs are also performed for OTAEs and tabulated separately.

A SAE is defined as an adverse event that results in death, is life-threatening, requires inpatient hospitalization or prolongation of existing hospitalization, results in persistent or significant incapacity or substantial disruption of the ability to conduct normal life functions, is a congenital anomaly or birth defect in the offspring of a patient taking the study drug, or is an important medical event. Important medical events are defined as events that may jeopardize the patient or result in medical or surgical intervention to prevent one of the outcomes listed above. The term “life-threatening” refers to an event in which the patient was at immediate risk of death at the time of the event.

Adverse events resulting from concurrent illnesses, reactions to concurrent medications, or progression of disease states will be reported. Conditions existing prior to the start of the AE collection period are considered concurrent medical conditions and are not recorded as AEs. If the patient experiences a worsening or complication of a concurrent medical condition, the worsening or complication will be recorded as an AE.

Each AE will be recorded as a single diagnosis. Accompanying signs and symptoms will not be recorded as additional AEs. Signs and symptoms will only be recorded if the diagnosis is not determined. Changes in laboratory tests values or ECG parameters will be recorded if they are determined to be clinically significant; however, abnormal test values or ECG findings will not be reported as AEs if they are the result of pathology for which there is an overall diagnosis.

Elective procedures performed to manage or treat conditions existing prior to the patient’s enrollment in the study are recorded in the patient’s source documents but not in the AE records. If a planned procedure is performed early because the pre-existing condition worsens, the worsening should be documented as an AE.

The investigator must report all SAEs observed or reported by the patient to Reata or designee, regardless of their relationship to the study drug or their clinical significance. SAEs must be followed to resolution or until the investigator deems the event to be chronic or the patient to be stable.

Causality of AE is classified according to the following criteria:

- Unrelated: This relationship suggests that the study drug and the event are not associated.
- Unlikely: This relationship suggests that the association between the study drug and the event is unlikely.
- Possible: This relationship suggests that treatment with study drug might have caused or contributed to the AE. That is, the event follows a reasonable temporal sequence from the time of study drug administration or follows a known response pattern to the study drug, but could have been produced by other factors.
- Probable: This relationship suggests that a reasonable temporal sequence of the event with the study drug administration exists and, based upon the known pharmacological action of the drug, known or previously reported adverse reactions to the drug or class of drugs, or judgment based on the investigator's clinical experience, the association of the event with the study drug administration seems likely.
- Definite: This relationship suggests a definite causal relationship between the study drug administration and the AE, and other conditions (concurrent illness, progression/expression of disease state, or concurrent medication reaction) do not appear to explain the event.

Severity is assessed according to the following criteria

- Mild: Symptoms causing no or minimal interference with usual social and functional activities
- Moderate: Symptoms causing greater than minimal interference with usual social and functional activities
- Severe: Symptoms causing inability to perform usual social and functional activities

The MedDRA dictionary will be used to classify all AEs reported during the study by system organ class (SOC) and preferred term (PT). All summary tables will include counts of patients with treatment-emergent adverse events (TEAEs). TEAEs are defined as those AEs that have an onset time either at or after the start of study drug administration and no more than 30 days after the last dose of study drug is administered. Furthermore, TEAEs may be AEs ongoing at the time of study drug initiation that increase in severity or become closer in relationship to study drug administration during the treatment period. AEs with missing start dates, but with stop dates either overlapping into the treatment period or missing, will be considered TEAEs.

An overall summary of TEAEs will be presented with the number and percentage of patients having a TEAE, a serious TEAE, a TEAE leading to study discontinuation, a TEAE with an outcome of death, a TEAE related to study drug (definitely, probably, or possibly related), or a severe TEAE. The overall incidence of TEAEs will be summarized by SOC and by SOC and PT

using the number and percentage of patients reporting an event and the number of events reported. The incidence of serious TEAEs and TEAEs leading to study discontinuation will be summarized in a similar manner. TEAEs will also be summarized by maximum severity (mild, moderate, or severe) and assessed relationship to study drug with the percentage of patients in each category. A TEAE with missing severity or relationship will be considered severe or related, respectively. If more than one TEAE is recorded for a patient within any SOC or PT, the patient will only be counted once. Serious TEAEs, TEAEs leading to study discontinuation, and TEAEs with an outcome of death will be presented in separate listings.

## 12.3. Laboratory determinations

Actual values and change from baseline values for continuous data from laboratory determinations will be summarized descriptively by received study drug at all scheduled study visits (Sections 5.1 - 5.3, unless otherwise specified). The laboratory tests is listed in Appendix E.

### 12.3.1. Transaminases

Evaluation of Drug-Induced Serious Hepatotoxicity (eDISH) plots are generated for ALT and AST versus TBL

The number and percentage of patients meeting the following thresholds, which are consistent with FDA guidance, are summarized by the maximum dosage received:

**Table 1: Pre-Specified Threshold Levels for Transaminases**

| Lab Parameter | Threshold |
|---|---|
| ALT, AST | ≥ 3 × upper limit of normal (ULN) and < 5 × ULN |
| | ≥ 5 × ULN and < 10 × ULN |
| | ≥ 10 × ULN and < 20 × ULN |
| | ≥ 20 × ULN |
| | ≥ 5 × ULN for more than 2 weeks |
| TBL | ≥ ULN and ≤ 1.5 × ULN |
| | > 1.5 × ULN and ≤ 2 × ULN |
| | > 2 × ULN |
| ALT, /AST, TBL | ALT or AST > 3×ULN and Total bilirubin > 1.5×ULN |
| | ALT or AST > 3×ULN and Total bilirubin > 2×ULN |
| ALT, AST, TBL, INR | > 3 × ULN and (TBL > 2 × ULN or INR > 1.5) |

| Lab Parameter | Threshold |
|---|---|
| ALP | > 1.5 × ULN |

A summary table that includes frequencies and percentages of patients that meet any of the above criteria at any time during the study is provided. A listing of subjects with abnormal ALT, AST, or TBL will also be provided.

### 12.3.2. Line plot and Boxplot presentations

Boxplot figures will present the distribution of laboratory values (both observed and change from baseline) over time by treatment group. Additionally line plots of change from baseline of laboratory values, vital signs and ECG by visit and treatment analysis group will be generated up to and including study week 28.

### 12.3.3. Shift tables

Clinical laboratory test results categorized as in or out of the limits of the normal range will be summarized using "worst case shift tables," that is, by cross-tabulating baseline result categories (high, normal, low) with worst case results during the treatment period. Shift tables will not be limited to results collected at scheduled visits, but will use all available results.

In addition, for ALT, AST and TBL, shift table summarizing (1) baseline to end of treatment, (2) baseline to worst on-treatment, (3) baseline to worst off-treatment, and (4) worst on-treatment to worst off-treatment will be presented.

## 12.4. 12-lead Electrocardiogram

All ECG measurements including ventricular rate, QT interval, Corrected QT interval - Fridericia's formula (QTcF), and the overall interpretation will be displayed in a data listing. Only QTcF will be summarized. The QTcF will be calculated based on the following formula:

- QTcF = QT / cube-root (RR interval)
- RR interval = 60 / heart rate

A summary table of the actual values and change from baseline will be presentated for QTcF by treatment group. In addition, the categories for QTcF of: ≤ 450 msec, > 450 - 480 msec, > 480 - 500 msec, and > 500 msec, and categories of change from baseline for QTcF of: ≥ 30 msec and ≥ 60 msec will be summarized by visit and treatment group.

## 12.5. Vital signs and weight

Actual values and changes from baseline values for vital signs and weight will be summarized descriptively by treatment group at all scheduled study visits. Vital sign measurements will include blood pressure, heart rate, temperature, and respiratory rate.

### 12.5.1. Vital signs

Changes in blood pressure and other vital signs will be summarized by baseline quartile.

#### 12.5.2. Weight

Weight change from baseline will be summarized by baseline BMI category (< 18.5, 18.5 to < 25.0, 25.0 to < 30.0, 30.0 or greater).

# 13. PHARMACOKINETIC ANALYSES

The planned pharmacokinetic analyses are described in a separate document.

# 14. PROTOCOL DEVIATIONS

The Principal Investigator (PI) or designee must document any protocol deviation or violation and report them to the appropriate Institutional Review Board (IRB), Ethical Committee (EC), or Research Ethics Board (REB). The PI must follow the applicable IRB, EC, or REB guidelines.

If there is an immediate hazard to a patient, the PI may deviate from the protocol without prior Sponsor and IRB, EC, or REB approval. Notification to Reata, the IRB, EC, or REB is necessary for the deviation.

## 14.1. Major protocol deviations

Major protocol deviations from entry criteria and treatment compliance will be listed.

## 14.2. Other protocol deviations

Protocol deviations that are not major are termed “other” will not be listed but will be included in the analysis data.

# 15. OTHER TOPICS

## 15.1. Analysis by study site

Because the number of patients at each site is likely to be small, no analyses will be performed by study site.

## 15.2. Interim analysis

A DSMB will be reviewing data from the study to ensure patient safety. The DSMB charter describes the interim reviews of safety. No interim analyses for efficacy are planned.

## 15.3. Subgroups analysis

Safety and efficacy outcomes will be analyzed for the subgroups of interest listed below. The following subgroup analyses will be tabulated using summary statistics:

- Geographic region: US; non US
- Age by median (above and below equal median)
- Sex: female; male

- Race: White; Non-White
- Ethnicity: Non-Hispanic/Latino; Hispanic/Latino
- Number of PAH medications at baseline: 0-1; 2
- CTD type: scleroderma; no scleroderma
- Baseline FC: II; III
- Baseline 6MWT: < 450 m; ≥ 450 m
- On ERA (PAH med): yes; no
- On PD5i (PAH med): yes; no
- Baseline Hb: < 10.5 g/dL; ≥ 10.5 g/dL
- Baseline BMI: < 30; ≥ 30
- End-of-treatment dose: 10 mg; <10 mg; placebo

# 16. ARCHIVING

After finalization of the analysis, the following will be archived:

- SAP
- All computer code used in creating the analysis datasets, in generating the tables and figures, and in analyzing the data
- Tables, listings, and figures as included in the CSR
- SAS datasets

# 17. DEVIATION FROM THE PROTOCOL-SPECIFIED ANALYSIS

The protocol-defined secondary efficacy outcome is the time-to-first persistent clinical improvement event. To declare that clinical improvement has occurred per the language in the protocol, at least one of the following three criteria must be met:

1. Improvement by at least one WHO functional class coupled with no more than a 15% decrease from baseline in 6MWT.
2. Increase from baseline in 6MWT by at least 10% and stability or improvement in the WHO functional class.
3. Decrease from baseline in creatine kinase (a surrogate biomarker for muscle injury and inflammation) by at least 10% and no worsening in WHO functional class and no more than a 15% decrease from baseline in 6MWT.

However, a persistent improvement in estimated glomerular filtration rate (eGFR) ≥10% of baseline was added to this criteria.

## 18. REFERENCES

[1] Klein J, Moeschberger M. *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer-Verlag New York, Inc.; 2003

[2] Pugh M, Buchowski M, Robbins, I et al. Physical Activity Limitation as Measured by Accelerometry in Pulmonary Arterial Hypertension. *CHEST*, 2012; 142:1391-1398.

# APPENDIX A. APPROACH TO ESTIMATING MEANS AND COVARIANCE STRUCTURE USED IN THE ORIGINAL SAMPLE SIZE CALCULATIONS

## A. 1. INTRODUCTION

The primary efficacy analysis will use a MMRM with the change from baseline in 6MWT as the outcome variable (Section 10.1).

## A. 2. TREATMENT EFFECT OF BARDOXOLONE METHYL ON 6MWT

## A. 3. VARIABILITY OF 6MWT IN OTHER PAH STUDIES

**Figure A 2. Observed variance, covariance, and correlation in LARIAT data**

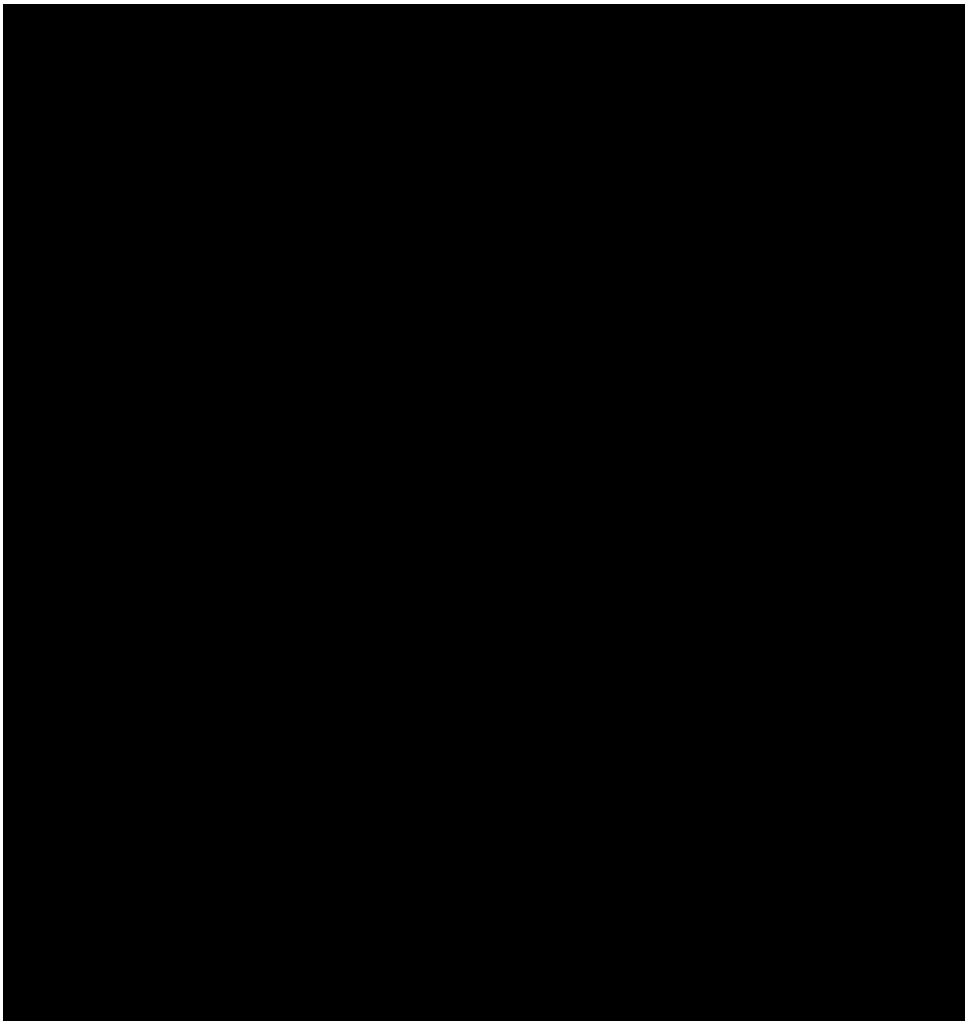


**Table A 1. Variance-covariance and correlation structure for simulation study**

[redacted]

## A. 4. SIMULATION STUDY FOR COMPLETE DATA

[redacted]

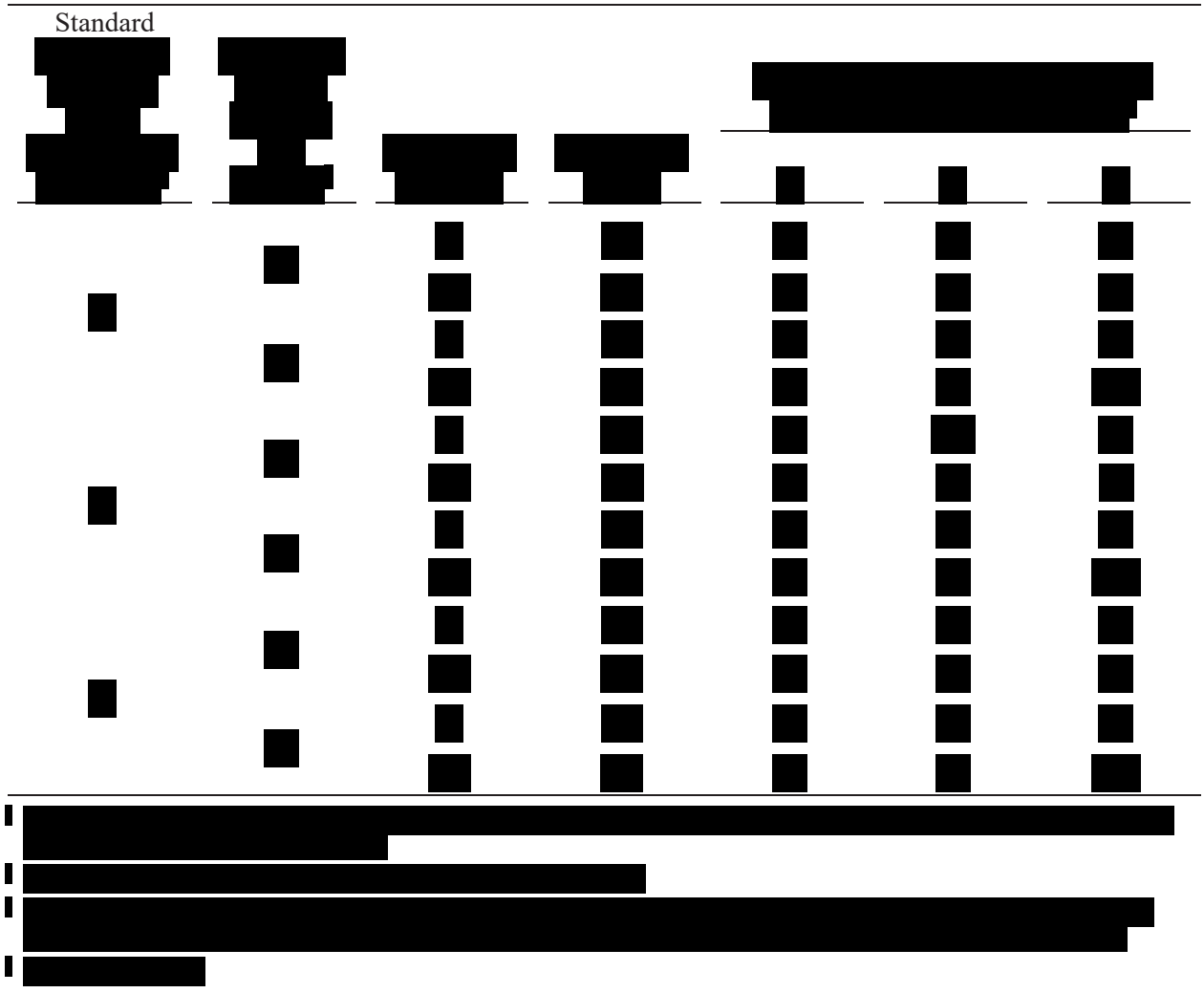
**Table A 2. Power by sample size and variance-covariance of complete data**

[redacted]

## 5. SIMULATION STUDY WITH A MISSING DATA MODEL

[redacted]

**Table A 3. Power by sample size and variance-covariance with a missing data model**



# APPENDIX B. GUIDELINES FOR RECALCULATION OF SAMPLE SIZE

## B. 1. INTRODUCTION

The design of the CATALYST study is based on assumptions, as described in Appendix A of this SAP, regarding the mean trajectory of 6MWT in the treated and placebo groups and the correlation structure of the outcome. As is typical of a Phase 3 randomized clinical trial, these parameters were estimated from data that came from previous studies. Because of the uncertainty of these estimates, the protocol for the CATALYST study allows recalculation of the sample size at an intermediate point in the study. This recalculation will be based on blinded information, that is, data pooled across treatment groups. Because no unblinding will occur, recalculation will have at most a negligible impact on the Type 1 error rate for the final data analysis. See Section 2.6 for further details on blinding procedures.

The recalculation will be performed when approximately 100 patients have been randomized. At this point, relevant baseline and operational parameters will be reassessed and the study size increased if necessary to maintain desired power. A sample size re-calculation committee (SRC), composed of members from Reata and [redacted] will review blinded data at the time of recalculation and make a formal recommendation about study size to Reata.

The remaining sections of this Appendix summarize the current assumptions underlying the sample size and describe the planned recalculation. This appendix will have been finalized and submitted to the FDA as part of the full SAP prior to the sample size recalculation.

## B. 2. STUDY DESIGN ASSUMPTIONS

As described in the protocol and in the main body of the SAP, the CATALYST study will assess efficacy based on the primary outcome, the change from baseline in 6MWT relative to placebo at Week 24. The study design calls for randomizing approximately 130 patients (65 in each group) and for assessing 6MWT at baseline and again at Weeks 4, 8, 16, and 24.

[redacted]

[redacted]

## B. 3. RECALCULATION OF PARAMETERS

When approximately 100 patients have been randomized, a pooled dataset will be constructed with baseline and available follow-up data. Table B 1 details the original assumptions that will be checked at the time of the recalculation. At that time, [redacted] will rerun the sample size simulations based on the updated values for the parameters. Note that all computations will be based on the CATALYST data that is available at the time of the recalculation.

**Table B 1. Tables available at time of the recalculation of parameters needed for power estimates**

Figures will be provided for visual representation of data presented in the source tables.

**Table B 2. Graphical presentations**

| Number | Title | Description |
|---|---|---|
| [redacted] | [redacted] | [redacted] |

In addition to the sample size calculations noted above, the SRC will review data summaries listed in Table B 1 and Table B 2. Using all available data, the SRC will decide whether to recommend increasing the sample size up to 200.

## B. 4. BLINDING PROCEDURES

All measures in place to ensure blinding to treatment assignments, as discussed in the protocol (Section 6.2) and this SAP (Section 2.6), will be maintained through the sample size recalculation. Specifically,

- The only individuals with access to treatment assignments will be the unblinded statistical group [redacted], the DSMB, and the unblinded IWRS group. No additional access to treatment codes will be granted for the sample size recalculation described in this Appendix.
- All members of the SRC as well as all patients, investigators, site personnel, laboratories, and Reata personnel (regardless of their involvement in the conduct of the study) and their designees [redacted] will be blinded to treatment assignments and will remain blinded until after the database is locked and formal access to the randomized treatment codes is authorized by Reata.
- The recalculation will be performed using aggregate data pooled for all patients; no treatment assignments will be provided.
- All available data for the relevant clinical parameters specified in this Appendix will be considered in recalculating sample size.
- SRC recommendations will be based on information limited to the package of data described in this Appendix.

## B. 5. SRC: COMPOSITION AND LOGISTICS

The SRC will perform these evaluations and make recommendations to Reata's senior management. The committee, which will be blinded to treatment allocation, will act by consensus. [redacted], its chair, will serve the SRC as the statistical representative for Reata.

The other members of the SRC will be representatives from Reata who played significant roles in developing the original sample size assumptions: [redacted]

After the recalculation, the SRC will provide a formal recommendation for change or no change with details and rationale to Reata's senior management, who will make the final decision about whether to accept the recommendation. The SRC will use the attached recommendation form (Appendix B. 7) after the recalculation.

Should the SRC and Reata's senior management agree that a sample size increase is warranted after the recalculation, Reata will document the increase in a note to file to the clinical sites and, as needed, to the EC/IRB/REB and the DSMB. If a change in sample size is implemented, a protocol amendment will not be required because the current protocol includes a provision for the sample size recalculation.

## B. 6. CALCULATIONS AND DATA SUMMARIES

[redacted] will provide the SRC with the set of predefined tables and figures presented in Table B 1 and Table B 2.

The SRC will review the materials and provide [redacted] with updated assumptions for the recalculations. [redacted] will detail the updated assumptions, methodology, and document results of the sample size and event recalculations documented in a report provided to the SRC and ultimately to Reata's senior management.

## B. 7. RECOMMENDATION FORM

CATALYST Trial Sample-size Recalculation Committee (SRC) Recommendation Report

Meeting Date: _____

Sample Size Recalculation ☐

Meeting Attendees:

The SRC charged with the review of sample size for the CATALYST Trial reviewed data summaries dated [_____].

As a result, the SRC recommendation is:

☐ To continue the trial unmodified.

☐ To continue the trial and amend the study design as checked:

☐ To increase sample size to [_____] due to:

- ☐ higher variability than expected
- ☐ higher rate of drop-outs than expected
- ☐ lower correlation than expected
- ☐ faster enrollment and limited ability to test assumptions

Comments:

[redacted] (SRC Chair) Date

# APPENDIX C. SCHEDULE OF ASSESSMENTS

The schedule of assessments is available in the protocol.

# APPENDIX D. CLASSIFICATION OF VISITS

**Table D 1. Classification of on treatment visits**

| Timing of assessment (days relative to first dose of study drug[a]) | Target day | Visit associated with assessment |
|---|---|---|
| 1 | 1 | Day 1 [a] |
| 2 to 10 | 7 | Week 1 |
| 11 to 21 | 14 | Week 2 |
| 22 to 35 | 28 | Week 4 |
| 36 to 49 | 42 | Week 6 |
| 50 to 84 | 56 | Week 8 |
| 85 to 139 | 112 | Week 16 |
| 140 to 182 | 168 | Week 24 [a] |
| 7 <= Days after last dose <=35 | 28 | Week 28 [c] |

[a] Day 1 is the day of first dose of study drug.

b. Week 24 values is the average of the two values closest to target date.

[c] Week 28 is only applicable to eGFR change from baseline efficacy analysis

**Table D 2: Analysis Visits for Post-Treatment Safety Analysis**

| Analysis Visit | Label | Target Study Day (Reference Day is Days After Last Dose) | Analysis Window |
|---|---|---|---|
| 40 | 4-weeks – Off Treatment | 28 | 7 ≤ Study Day ≤ 35 |
| 80 | 8-weeks – Off Treatment | 56 | 36 ≤ Study Day ≤ 63 |
| 120 | 12-weeks– Off Treatment | 84 | 64 ≤ Study Day ≤ 91 |
| 240 | 24-weeks– Off Treatment | 168 | 92 ≤ Study Day ≤ 175 |

# APPENDIX E. LIST OF LABORATORY TEST

Blood samples are collected throughout the study for hematology, chemistry, and urinalysis for clinical laboratory evaluation. Test panels include the following lab tests plus additional collected lab assessments based on availability:

**Table E1: List of Laboratory Tests**

| Hematology | Chemistry | Urinalysis |
|---|---|---|
| Hematocrit | Blood urea nitrogen (BUN) | Specific gravity |
| Hemoglobin | Creatinine | Ketones |
| HbA1C | Total bilirubin | pH |
| Red blood cell (RBC) count | Alanine aminotransferase (ALT) | Protein |
| White blood cell (WBC) count | Aspartate aminotransferase (AST) | Blood |
| Neutrophils | Alkaline phosphatase (ALP) | Glucose |
| Bands (if detected) | Ferritin | Urobilinogen |
| Lymphocytes | Sodium | Bilirubin |
| Monocytes | Potassium | |
| Basophils (if detected) | Calcium | |
| Eosinophils (if detected) | Inorganic phosphorus | |
| Absolute platelet count | Magnesium | |
| Mean corpuscular hemoglobin (MCH) | Chloride | |
| Mean corpuscular volume (MCV) | Bicarbonate | |
| Mean corpuscular hemoglobin concentration (MCHC) | Uric acid | |
| Reticulocytes/Erythrocytes | Cholesterol | |
| | Total protein | |
| | Glucose | |
| | Triglycerides | |
| | Albumin | |
| | Creatine phosphokinase (CPK) | |
| | Lactate dehydrogenase (LDH) | |
| | High-density lipoprotein cholesterol (HDL-C) | |
| | Low-density lipoprotein cholesterol (LDL-C) | |
| | Very-low-density lipoprotein cholesterol (VLDL-C) | |
| | Gama-glutamyl transpeptidase (GGT) | |
| | Estimated glomerular filtration rate (eGFR) using the CKD-EPI / Schwartz Bedside Equation | |

# APPENDIX F. IMPUATATION RULES FOR MISSING DATES

*Incomplete diagnosis or treatment date*

- If day is missing, day will be set to 15th of the month.
- If month is missing, month and day will be set to July 1.
- If year is missing, date will be set to missing.

The imputed dates must be logical.

*Adverse Event*

- If onset date is completely missing, onset date is set to date of first dose.
- If year is present and month and day are missing or year and day are present and month is missing:
  - If year = year of first dose, then set month and day to month and day of first dose
  - If year < year of first dose, then set month and day to December 31.
  - If year > year of first dose, then set month and day to January 1.
- If month and year are present and day is missing:
  - If year = year of first dose and
    month = month of first dose then set day to day of first dose date
    month < month of first dose then set day to last day of month
    month > month of first dose then set day to first day of month
  - If year < year of first dose then set day to last day of month
  - If year > year of first dose then set day to first day of month
- For all other cases, set onset date to date of first dose.

*Concomitant Medications*

- If start date is completely missing, start date will not be imputed.
- If year is present and month and day are missing or year and day are present and month is missing, set month and day to January 1.
- If year and month are present and day is missing, set day to 1st day of month.
- If end date is completely missing, end date will not be imputed.
- If year is present and month and day are missing or year and day are present and month is missing, set month and day to December 31.
- If year and month are present and day is missing, set day to last day of the month.