ADDENDUM TO STATISTICAL ANALYSIS PLAN

VERSION 1.0

DATED 30-OCT-2020

TO

STATISTICAL ANALYSIS PLAN Version 3.0

Statistical Analysis Plan History 3.0

Version 1.0 – 10 July 2019

Version 2.0 – 15 August 2019

Version 3.0 – 04 November 2019

BASED ON:

Protocol Version 4.0 (April 22, 2019)

Protocol Version 4.0, 4.1, and Protocol 4.2 Addendum (France, Spain, UK Apr 21, 2020; Japan May 12, 2020; US, AUS June 05, 2020)

STUDY DRUG:

RTA 402, BARDOXOLONE METHYL

PROTOCOL NUMBER:

402-C-1603 Phase 3

STUDY TITLE:

A Phase 2/3 Trial of the Efficacy and Safety of Bardoxolone Methyl in Patients with Alport Syndrome ("CARDINAL")

SPONSOR:

Reata Pharmaceuticals, Inc.

5320 Legacy Dr

Plano, Texas 75024

(972) 865-2219

This study is being conducted in compliance with good clinical practice, including the archiving of essential documents.

SIGNATURE PAGE

This document has been prepared and/or reviewed by:

[redacted]

This document has been reviewed and accepted by:

[redacted]

TABLE OF CONTENTS

1. LIST OF ABBREVIATIONS

Table 1: List of Abbreviations

Abbreviation	Term
COVID-19	Coronavirus Disease 2019
eGFR	estimated glomerular filtration rate
FDA	Food and Drug Administration
IP	Investigational Product
SAP	Statistical Analysis Plan
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2

2. RATIONALE FOR ADDENDUM

This addendum describes clarifications to the Statistical Analysis Plan (SAP) version 3.0 as well as additional analyses for Year 2 of CARDINAL Phase 3 to characterize the impact of the COVID-19 pandemic on the trial.

3. CORRECTION AND CLARIFICATION TO THE SAP

3.1. Correction to Covariate Computation

Analyses of efficacy will adjust for the duration of exposure to drug within each year of the study. The fraction of expected exposure duration for Year 2 analyses (i.e., Week 100 and Week 104), calculated as the fraction of expected exposure duration in Year 2, will be included in analyses as a covariate. As outlined in Section 6.5.11 of the SAP, fraction of expected exposure in Year 2 (TRT_EXP2) is calculated using the date of last dose in Year 2. However, any patient who permanently discontinued treatment in Year 1 would not have a Year 2 date of last dose. As a result, the definition has been corrected to use the date of last dose in the study, allowing all randomized patients to have a TRT_EXP2 value and therefore contribute to analyses. Additionally, the length of exposure was extended from 700 to 707 days to account for the COVID-19 protocol modification, which allowed longer dose administration. The computation was also corrected by adding "+1" to the date of last dose. See Table 2 for the corrected definition of TRT_EXP2.

Table 2: Computation of Fraction Received of Expected Two-Year Exposure

SAP definition	Corrected definition
[(date of last dose in second year) – (date of first dose)]/700	[(date of last dose +1) – (date of first dose)]/707

3.2. Clarification of Hierarchical Testing Strategy

Section 8.4 of the SAP specifies that all endpoints tested at Year 1 and Year 2 will use a significance level of 0.025. Within each year, endpoints are tested following a fixed-sequence hierarchical testing strategy. The total significance level (0.05) was split between Year 1 and Year 2 as a strategy to reserve alpha to test Year 2 if the Year 1 testing sequence was not statistically significant. Formal testing in the sequence may proceed so long as statistically significant evidence of benefit continues to be shown. If in the sequence one endpoint does not show statistically significant evidence of benefit, formal statistical testing of subsequent endpoints will not occur.

As a clarification to the SAP, if there is a significant treatment effect for both Year 1 endpoints, then the significance level for Year 1 (0.025) remains available to be carried forward (recycled or passed along) to the Year 2 testing sequence. Thus, if both Year 1 endpoints are significant, the Year 2 testing sequence will be tested using a significance level of 0.05. This approach aligns with the fixed-sequence method in FDA guidance (Multiple Endpoints in Clinical Trials., January 2017).

4. COVID-19 IMPACT

FDA recognizes that the COVID-19 public health emergency may impact the conduct of clinical trials of medical products (FDA Guidance on Conduct of Clinical Trials of Medical Products during COVID-19 Public Health Emergency, September 2020). COVID-19 posed significant challenges to study conduct, which were not unique to the 1603 Phase 3 study including site closures and access restrictions, and quarantines for site personnel and patients creating barriers to scheduling and conducting visits according to the protocol specified schedule. Additionally, the 1603 Phase 3 study population has chronic kidney disease and is therefore categorized as vulnerable and at higher risk for developing serious complications from COVID-19. Specific guidance for these individuals varied over time and by geography, but as a group they were generally advised by health officials to not be in close physical contact with other people (https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/index.html).

This addendum makes the following modifications to the analysis and adds additional analyses to address the impact of COVID-19 on the 1603 Phase 3 study.

4.1. Changes in Study Conduct

Changes to study conduct due to COVID-19 are recorded in the clinical database for each patient. The overall number and percentage of patients impacted by COVID-19 will be summarized and listed. The impact of COVID-19 on study visits (e.g., not done or done outside the window), assessments (e.g., not performed), data collection method (e.g., telephone, video conference, home health, and local lab) will be listed.

4.2. Additional Disposition Categories

Categories for end-of-treatment and end-of-study reasons were added to data capture and disposition summaries:

- Terminated early from the treatment because of COVID-19; and
- Terminated early from the study because of COVID-19.

4.3. Protocol Deviations

Deviations due to the impacts of COVID-19 are documented in the clinical database by the site and the Sponsor. These deviations will be identified in the protocol deviations data and presented in a data listing.

4.4. Changes to Key Efficacy Analyses

COVID-19 created logistical challenges associated with study execution, including possible interruptions in treatment due to quarantine and travel restrictions as well as closure or reduced operating hours of sites/institutions. Many patients enrolled in the 1603 Phase 3 study had to travel long distances requiring full day or overnight travel for in-clinic visits. The National Kidney Foundation advised clinical trial participants to avoid using public transportation when possible (https://www.kidney.org/covid-19). Shipping drug directly to patients and home healthcare visits were added as options to keep patients on treatment per-protocol and overcome limitations associated with travel. High demand for home health nurses during this global pandemic and varying levels of patient comfort with home health nurse interactions during quarantine created continued challenges and limitations with scheduling flexibility.

Key efficacy analyses were modified as follows to address these issues related to COVID-19.

4.4.1. Extended Visit Windows

The Week 100 visit serves to collect the final "on treatment" lab values (including eGFR) from patients. Given the aforementioned logistical challenges due to COVID-19, the protocol was amended to allow up to a two-week extension in the protocol-specified window for the Week 100 visit to increase scheduling flexibility, as needed. As a result, some patients were shipped additional study drug kits so that patients could continue treatment despite their delayed Week 100 visits. Accordingly, some patients were dosed for up to 2 additional weeks and the Week 100 analysis window was adjusted to account for the extended dosing as specified in Table 3.

Table 3: Analysis Window for Week 100 Visit

Patient Treatment Status	Target Study Day[a]	Week 100 Analysis Window	
		Original[b]	Adjusted
Permanently discontinued treatment prior to Week 100	700	659 ≤ study day ≤ 714	No change
Completed treatment at Week 100	700[c]	659 ≤ study day ≤ 714	659 ≤ study day < 728; and Week 100 ≤ 13 days after last dose

[a] Study day is relative to the date of randomization (SAP Section 6.5.2)
[b] Original Week 100 analysis window defined in SAP Section 6.5.11, Table 3
[c] Target study day extended by up to 14 days to account for additional drug dispensed prior to Week 100

Scheduling of the Week 104 visit was adjusted for patients with extended dosing duration to ensure that the Week 104 eGFR value is collected 4 weeks after last dose. The SAP-specified

analysis window of 14 to 35 days with a protocol-specified target of 28 days after the last dose, which was reinforced by the Sponsor, did not allow much flexibility in scheduling. The Week 104 analysis window was extended by removing the upper bound on days after last dose (Table 4) to avoid exclusion of available values collected beyond the original analysis window. If eGFR is assessed more than once within the visit window, the measurement that is closest to 28 days after last dose (i.e., 4-weeks after the last dose as specified in the protocol) is used for the purpose of data analysis.

Table 4: Analysis Window for Week 104 Visit

Target Days After Last Dose[a]	Week 104 Analysis Window	
	Original[b]	**Adjusted**
28	14 ≤ days after last dose ≤ 35	days after last dose ≥ 14

[a] Days after last dose in second-year dosing interval. Patient must have been dispensed treatment in the second year, otherwise Week 104 is considered missing.
[b] Original Week 100 analysis window defined in SAP Section 6.5.4, Table 3

4.4.2. Accounting for Regional Differences

Study 402-C-1603, Phase 3, is a global clinical trial. The regional emergence and impact of COVID-19 is heterogeneous across the various countries contributing participants to the trial. Region (US; non-US) is added as a covariate to the models assessing Week 100 and Week 104 to account for differential COVID-19 impact by region.

4.4.3. No Changes Due to Temporary Interruptions in Study Treatment

A COVID-19 appendix was added to the study protocol to allow for temporary interruptions in study treatment. No changes are required for analyses of Week 100 or Week 104 due to temporary discontinuation and reinitiation of study drug given how the key efficacy analyses were defined:

- Analyses of Week 100 are based on target study day, irrespective of whether or not a patient is receiving study drug.
- Analyses of Week 104 are based on days after final dose following permanent study drug discontinuation in the second year. Analyses of Week 104 data will not include data collected during any intermediate off-treatment period.

4.5. Data Handling for Assessments After Positive COVID-19 Diagnosis

Recent studies have documented that COVID-19 can cause damage to multiple organs, such as the heart, liver, kidneys, and the immune system (Chen, 2020; Wang, 2020). Therefore, the summary tables and analyses do not include efficacy and safety data (clinical laboratory [including eGFR], vital sign, electrocardiogram, visual acuity, and audiology assessments) collected after the start date of adverse event preferred term “coronavirus”. If the number of patients with positive COVID-19 diagnosis is sufficiently large, the data collected for these patients may be summarized together. All data will be presented in by-patient listings.

4.6. Pre-COVID Era and Post-COVID Era Comparison

As an additional analysis, the pre-COVID era which will include all Week 100 and Week 104 visits that were planned on or prior to March 1, 2020 will be compared to the same visits that were planned post March 1, 2020.

4.7. References

Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y, Xia J, Yu T, Zhang X, Zhang L. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet. 2020 Feb 15;395(10223):507-513.

Wang T, Du Z, Zhu F, Cao Z, An Y, Gao Y, Jiang B. Comorbidities and multi-organ injuries in the treatment of COVID-19. Lancet. 2020 Mar 21;395(10228):e52.

STATISTICAL ANALYSIS PLAN

VERSION: 3.0

DATE OF PLAN:

04-NOV-2019

Statistical Analysis Plan History

Version 1.0 – 10 July 2019

Version 2.0 – 15 August 2019

BASED ON:

Protocol Version 4.0 (April 22, 2019)

STUDY DRUG:

RTA 402, BARDOXOLONE METHYL

PROTOCOL NUMBER:

402-C-1603 Phase 3

STUDY TITLE:

A Phase 2/3 Trial of the Efficacy and Safety of Bardoxolone Methyl in Patients with Alport Syndrome ("CARDINAL")

SPONSOR:

Reata Pharmaceuticals, Inc.

2801 Gateway Drive, Suite 150

Irving, Texas 75063

(972) 865-2219

This study is being conducted in compliance with good clinical practice, including the archiving of essential documents.

SIGNATURE PAGE

This document has been prepared and/or reviewed by:

[redacted]

This document has been reviewed and accepted by:

[redacted]

TECHNICAL SUMMARY REPORT (TSR)

Name of Sponsor/Company **Reata Pharmaceuticals**	**Individual Study Table Referring to Part of the Dossier:** **Volume:**	***(For National Authority Use Only)*:**
Name of Finished Product: bardoxolone methyl capsules	**Page:**	
Name of Active Ingredient: bardoxolone methyl		

Title of Study: A Phase 2/3 Trial of the Efficacy and Safety of Bardoxolone Methyl in Patients with Alport Syndrome

Investigators:

[redacted] Gerald Appel; [redacted] Lesley Inker; [redacted] ; Bertrand Knebelmann; [redacted] Kandai Nozu; [redacted] Pablo Pergola; [redacted] Arnold Silva; [redacted] Roser Torra; [redacted] ; Bradley Warady; [redacted]

Study Center(s): Approximately 50 study centers

Studied period (years): 3.5 years	**Phase of development:** 3

Primary and Key Secondary Endpoints:

Phase 3 – Year 1 Endpoints

Primary:

- Change from baseline in estimated glomerular filtration rate (eGFR) at Week 48
- Safety assessments through Week 52

Key Secondary:

- Change from baseline in eGFR, following a 4-week drug treatment withdrawal period, at Week 52

Phase 3 – Year 2 Endpoints

Primary:

- Change from baseline in eGFR at Week 100
- Safety assessments through Week 104

Key Secondary:

- Change from baseline in eGFR, following a 4-week drug treatment withdrawal period, at Week 104

Methodology:

The Phase 3 portion of the 402-C-1603 trial is a double-blind, randomized, placebo-controlled study designed to evaluate the safety and efficacy of bardoxolone methyl up to 30 mg in patients with Alport syndrome. Patients in Phase 3 are randomized 1:1 to receive bardoxolone methyl or placebo. Randomization is stratified by screening urine albumin to creatinine ratio (ACR). Patients randomized to placebo remain on placebo throughout the study, undergoing sham titration.

Adult patients (≥ 18 years of age) receiving bardoxolone methyl start with once-daily dosing at 5 mg. They dose-escalate to 10 mg at Week 2, to 20 mg at Week 4, and then to 30 mg at Week 6 (only if baseline ACR > 300 mg/g) unless contraindicated clinically and approved by the medical monitor. Patients under the age of 18 receiving bardoxolone methyl start dosing at 5 mg every other day during the first week. They begin once-daily dosing with 5 mg during the second week of the study, and then continue with once-daily dosing following the same aforementioned dose-titration scheme based on baseline ACR at Weeks 2, 4, and 6. Dose de-escalation is permitted during the study if indicated clinically; subsequent dose re-escalation is also permitted to meet the dosing objective of the highest tolerated dose.

All patients receive study drug from Day 1 to Week 48. Patients do not receive study drug treatment during a 4-week withdrawal period (between Weeks 48 and 52). Patients re-start study drug treatment at Week 52 at the same dose level that they received prior to the withdrawal period (maintaining the originally randomized treatment group) and continue study drug treatment through Week 100. A follow-up visit occurs at Week 104 (4 weeks after the last dose). The Data Monitoring Committee (DMC) performs quarterly reviews of unblinded data for safety throughout Phase 3.

The Phase 3, Year 1 efficacy is assessed first by the DMC after all patients in the Phase 3 cohort have completed their Week 52 visit. If the Year 1 efficacy endpoints are met, then the Year 1 data will be analyzed by the Reata analysis team to proceed with regulatory filing. Available Year 2 safety data is summarized descriptively as preliminary results at the time of the Year 1 final analysis. The Phase 3, Year 2 efficacy is analyzed after all patients in the Phase 3 cohort have completed their Week 104 visit.

Number of Subjects (planned and analyzed):
Planned: 150

Diagnosis and main criteria for inclusion (see protocol section 8.1):

1. Male and female patients 12 ≤ age ≤ 70 upon study consent;
2. Diagnosis of Alport syndrome by genetic testing (documented mutation in a gene associated with Alport syndrome, including COL4A3, COL4A4, or COL4A5) or histologic assessment using electron microscopy;
3. Screening eGFR (average of Screen A and Screen B eGFR values) ≥ 30 and ≤ 90 mL/min/1.73 m^2. The two eGFR values collected at Screen A and Screen B visits used to determine eligibility must have a percent difference ≤ 25%;
4. Albumin to creatinine ratio (ACR) ≤ 3500 mg/g at Screen B visit. Up to 50% of patients in the Phase 2 cohort and approximately 40% of patients enrolled in the Phase 3 cohort can have ACR of 301 to 3500 mg/g. Once enrollment of these patients is complete, the ACR inclusion criterion is ≤ 300 mg/g;
5. Patients receiving an angiotensin-converting enzyme (ACE) inhibitor and/or an angiotensin II receptor blocker (ARB) should be receiving the maximally tolerated labeled daily dose (MTLDD), for at least 6 weeks prior to the Screen A visit. The dosage of ACE inhibitor and/or ARB should remain the same throughout the remainder of the study (i.e., no change in dosage or medication), and any potential changes should be discussed with the medical monitor. Patients not currently taking an ACE inhibitor and/or ARB because they are not indicated or because of a medical contraindication may be eligible provided the patient has not taken an ACE inhibitor and/or ARB at least 8 weeks prior to the Screen A visit (these patients must be discussed with the medical monitor prior to enrollment);

Test product, dose and mode of administration:
This study uses capsules containing bardoxolone methyl at the 5 and 15 mg strengths. Blinded treatment kits are provided to patients for self-administration orally once daily. Each dose of study drug should be administered at approximately the same time each day, preferably in the morning.

Duration of study:

100 weeks of treatment plus a 4-week follow-up

Reference therapy, dose and mode of administration:

This study uses matching capsules containing placebo. Blinded treatment kits are provided to patients for self-administration orally once daily. Each dose of study drug should be administered at approximately the same time each day, preferably in the morning.

Criteria for evaluation (see protocol section 6.2, 11.0):

Efficacy: eGFR, Patient Global Impression of Change (PGIC), Clinical Global Impression-Improvement (CGI-I), the time-to-first kidney failure outcome event defined as the composite endpoint consisting of 30% decline in eGFR, eGFR < 15 mL/min/1.73 m^2, and end stage kidney disease (ESKD) defined as initiation of maintenance dialysis or kidney transplant.

Safety: Frequency, intensity, and relationship to study drug of AEs and SAEs, and change from baseline in the following assessments: vital sign measurements, 12-lead ECGs, clinical laboratory measurements, weight, visual acuity, and audiology.

Statistical methods:
Primary and secondary analysis of the efficacy data are based on the ITT population, which will include all patients randomized in the Phase 3 portion of the study. Analyses at Weeks 48, 52, 100, and 104 will compare all bardoxolone methyl patients to all placebo patients. Mixed models repeated measures (MMRM) analysis are used to evaluate the treatment effect at Week 48 and 100. ANCOVA is used to evaluate the off-treatment effect at Week 52 and 104. The body of this SAP provides additional detail on analysis of the primary and secondary endpoints, and analyses of exploratory endpoints.

TABLE OF CONTENTS

LIST OF TABLES

1. LIST OF ABBREVIATIONS

Table 1. List of Abbreviations

Abbreviation	Term
μ_{BARD}	bardoxolone methyl group mean
$\mu_{PLACEBO}$	placebo group mean
ARB	angiotensin II receptor blocker
ACE	angiotensin-converting enzyme
ACEi	angiotensin converting enzyme inhibitor
ACR	urine albumin to creatinine ratio
ACR_STRAT	urine albumin to creatinine ratio use for stratification by the randomization list
AE	adverse event
ALT	alanine aminotransferase
ALP	alkaline phosphatase
ANCOVA	analysis of covariance
AR(1)	first order auto-regressive
ARB	angiotensin II receptor blocker
AST	aspartate aminotransferase
BARD	bardoxolone methyl
BCVA	best-corrected visual acuity
BI	bioinformatics
BMI	body mass index
BNP	B-type natriuretic peptide
BUN	blood urea nitrogen
CGI-I	Clinical Global Impression - Improvement
CI	confidence interval
CKD	chronic kidney disease
CKD-EPI	Chronic Kidney Disease Epidemiology Collaboration
CM	concomitant medication
CRF	Case Report Form
CSR	Clinical Study Report
DMC	Data Monitoring Committee
eCRF	electronic case report form
eDISH	evaluation of Drug-Induced Serious Hepatotoxicity
ECG	electrocardiogram
eGFR	estimated glomerular filtration rate
ESKD	end stage kidney disease
FDA	Food and Drug Administration (US)
FMV	first morning void
GFR	glomerular filtration rate
ICH	International Conference on Harmonisation

Abbreviation	Term
ICH E4	International Conference on Harmonisation Tripartite Guideline for Good Clinical Practice E4
ICH E9	International Conference on Harmonisation Tripartite Guideline for Good Clinical Practice E9
IEC	Independent Ethics Committee
IGF-1	insulin-like growth factor-1
INR	international normalized ratio
IRB	Institutional Review Board
ITT	intent-to-treat
IWRS	Interactive Web Response System
Kg	kilogram
LLD	lower limit of detection
logMAR	log of the minimum angle of resolution
LS	least squares
MAR	missing at random
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	mixed model repeated measures
Msec	milliseconds
MTLDD	maximally tolerated labeled daily dose
NDA	new drug application
NT-Pro BNP	N-Terminal Pro-Brain Natriuretic Peptide
OTAE	off-treatment adverse event
PGIC	Patient Global Impression of Change
pH	potential of hydrogen
PK	pharmacokinetic
PP	per-protocol
PT	preferred term
QTc	corrected QT interval
RBC	red blood cell
RBC/HPF	red blood cells per high power field
SAE	serious adverse event
SAP	statistical analysis plan
SCr	serum creatinine
SD	standard deviation
SOC	system organ class
SOP	Standard Operating Procedure
TBL	total bilirubin
TEAE	treatment-emergent adverse event
TLFs	tables, listings, and figures
ULD	upper limit of detection
ULN	upper limit of normal

Abbreviation	Term
US	United States
WOCBP	women of child-bearing potential
XLAS	X-linked Alport syndrome

2. INTRODUCTION

This statistical analysis plan (SAP) describes the planned analyses and data displays to be included in the Clinical Study Report (CSR) for Protocol 402-C-1603 Phase 3.

2.1. Trial Rationale

Conceptually, this trial was designed to test the biological hypothesis that bardoxolone methyl slows disease progression in patients with Alport syndrome, as measured by changes in eGFR. Since on-treatment changes in eGFR may include both reversible, pharmacodynamic effects, as well as irreversible structural effects, analysis of eGFR after withdrawal is used to determine if bardoxolone methyl affects disease progression.

The presence of a sustained eGFR improvement after withdrawal of drug may be indicative of disease-modifying activity. In contrast, a decline in kidney function relative to placebo after drug withdrawal would be unlikely to provide a long-term benefit and could indicate the drug has harmed the kidney.

If the withdrawal analysis indicates that bardoxolone methyl improves kidney function relative to placebo, then this would suggest that the drug is likely to reduce underlying progression of the disease. It would also suggest that the on-treatment eGFR change would likely translate into a long-term renal benefit. Further, in routine clinical practice, physicians are unlikely to withdraw the drug, and an important element of the trial is to characterize the on-treatment eGFR change.

Analytically, the main intent of the trial is to characterize the change from baseline in eGFR following a 4-week withdrawal period. Within the two-year duration, there are two opportunities to test this biological hypothesis and determine if the change from baseline in eGFR is statistically significant relative to placebo. If the off-treatment change from baseline in eGFR is not statistically significant relative to placebo at Year 1, the trial design and analysis plan allow for an additional comparison of the off-treatment effect relative to placebo at Year 2.

Beyond the main intent of the trial, the trial is designed to also determine the following:

- The clinical benefit of bardoxolone methyl while receiving treatment, as measured by the on-treatment assessment of eGFR.
- Whether a treatment effect is observed over a period of up to two years.
- Whether a treatment effect trend is seen across various subgroups, such as stages of disease (i.e., across a spectrum of baseline eGFRs) and the pediatric population.

This document describes the analyses of protocol-specified endpoints to assess the effects of bardoxolone methyl on disease progression in patients with Alport syndrome.

2.2. Background

The original study protocol has undergone three amendments.

Protocol Revision Chronology:		
Protocol version 1	15-Nov-2016	Original

Amendment 1 (Protocol version 2)	03-Aug-2017	• Removed interim analysis for efficacy and potential changes to Phase 3 macroalbuminuria cap • Decreased sample size in Phase 3 portion of study • Global Impressions of Change added as exploratory endpoints
Amendment 2 (Protocol version 3)	30-July-2018	• Added time-to-kidney failure as an exploratory endpoint • Increased cap on patients with macroalbuminuria in Phase 3
Amendment 3 (Protocol version 4)	22-Apr-2019	• Re-grouped objectives and endpoints into first and second year objectives and endpoints • Modified patient discontinuation and termination criteria • Added ESKD follow-up guidance

This SAP was developed in accordance with International Conference on Harmonisation Tripartite Guideline for Good Clinical Practice E9 (ICH E9) guidelines. All decisions regarding the final analysis, as defined in this SAP document, were made prior to Database Lock (unblinding) of the study data. Further information can be found in the protocol. Changes to protocol-specified analyses, including analysis populations, can be found in Section 12.

The SAP is based on:

- Protocol No. 402-C-1603, Version 4, dated April 22, 2019
- ICH guidelines E4 and E9 (Statistical Principles for Clinical Trials)
- Feedback received from FDA throughout clinical development, including the following communications:
  - FDA Meeting Minutes [redacted] for the October 5, 2016 Pre-IND meeting
  - IND Study May Proceed letter [redacted] dated December 15, 2016.
  - Advice/Information Request communication [redacted] dated February 5, 2019.

This SAP describes the study populations, how variables are derived, how missing data are handled, and details concerning the statistical methods to be used to analyze the safety and efficacy data from the Phase 3 portion of Study 402-C-1603. Should the SAP and the protocol be inconsistent with respect to the planned analyses, the language of the SAP is governing.

This SAP was finalized, approved by the Sponsor, and placed on file before database lock. This version of the SAP describes the analyses planned prior to the database lock. Unless otherwise specified, the CSR will summarize these analyses. Any substantive changes made to the SAP

after the database lock will clearly be identified in the CSR, and any analyses in addition to those specified in the SAP prior to the database lock are considered ad hoc.

3. STUDY ENDPOINTS

3.1. Study Endpoints

In patients with Alport syndrome, the study will compare those receiving bardoxolone methyl to those receiving placebo with respect to several endpoints. The timing for analyzing study endpoints is described in Section 5. The strategy for protecting the overall type I error rate of 0.05 for the trial is described in Section 8.4. Primary efficacy analyses are described in Section 8, and additional exploratory efficacy analyses are described in Section 9.

3.1.1. Phase 3 – Year 1 Endpoints

3.1.1.1. Primary Endpoints

- Change from baseline in estimated glomerular filtration rate (eGFR) at Week 48
- Safety assessments through Week 52

3.1.1.2. Key Secondary Endpoint

- Change from baseline in eGFR, following a 4-week drug treatment withdrawal period, at Week 52

3.1.1.3. Exploratory Endpoints

- Percentage of patients with a kidney failure event by Week 48, defined as the composite endpoint consisting of:
  - 30% decline from baseline in eGFR;
  - eGFR < 15 mL/min/1.73 m^2;
  - ESKD (initiation of maintenance dialysis or kidney transplant).
- Percentage of patients with an increase from baseline in eGFR of 30% or more by Week 48
- Percentage of patients with a decrease from baseline in eGFR of 30% or more by Week 48
- Distribution of changes from baseline in eGFR at Week 48
- Distribution of the Patient Global Impression of Change (PGIC) scores at Week 48
- Distribution of the Clinical Global Impression-Improvement (CGI-I) scores at Week 48

3.1.2. Phase 3 – Year 2 Endpoints

3.1.2.1. Primary Endpoints

- Change from baseline in eGFR at Week 100
- Safety assessments through Week 104

3.1.2.2. Key Secondary Endpoint

- Change from baseline in eGFR, following a 4-week drug treatment withdrawal period, at Week 104

3.1.2.3. Exploratory Endpoints

- Time to first kidney failure composite event, defined as the composite endpoint consisting of:
  - 30% decline from baseline in eGFR;
  - eGFR < 15 mL/min/1.73 m^2;
  - ESKD (initiation of maintenance dialysis or kidney transplant).
- Percentage of patients with a kidney failure event by Week 100, defined as the composite endpoint consisting of:
  - 30% decline from baseline in eGFR;
  - eGFR < 15 mL/min/1.73 m^2;
  - ESKD (initiation of maintenance dialysis or kidney transplant).
- Percentage of patients with an increase from baseline in eGFR of 30% or more by Week 100
- Percentage of patients with a decrease from baseline in eGFR of 30% or more by Week 100
- Distribution of changes from baseline in eGFR at Week 100
- Distribution of the Patient Global Impression of Change (PGIC) scores at Week 100
- Distribution of the Clinical Global Impression-Improvement (CGI-I) scores at Week 100

4. STUDY DESIGN

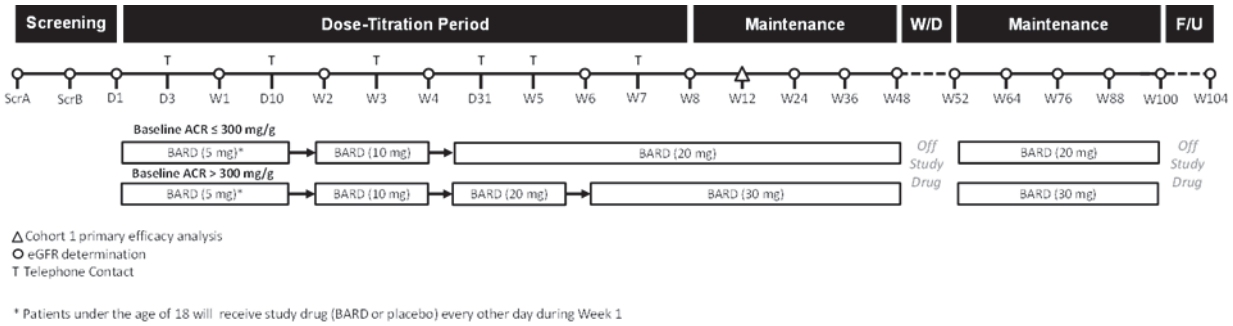
4.1. Summary of Study Design

The Phase 3 portion of the 402-C-1603 trial is a double-blind, randomized, placebo-controlled study designed to evaluate the safety and efficacy of bardoxolone methyl up to 30 mg in patients with Alport syndrome. Patients in Phase 3 are randomized 1:1 to receive bardoxolone methyl or placebo. Randomization is stratified by screening urine albumin to creatinine ratio (ACR). Patients randomized to placebo remain on placebo throughout the study, undergoing sham titration.

Adult patients (≥ 18 years of age) receiving bardoxolone methyl start with once-daily dosing at 5 mg. They dose-escalate to 10 mg at Week 2, to 20 mg at Week 4, and then to 30 mg at Week 6 (only if baseline ACR > 300 mg/g) unless contraindicated clinically and approved by the medical monitor. Patients under the age of 18 receiving bardoxolone methyl start dosing at 5 mg every other day during the first week; they begin once-daily dosing with 5 mg during the second week of the study, and then continue with once-daily dosing following the same aforementioned dose-titration scheme based on baseline ACR at Weeks 2, 4, and 6. Dose de-escalation is permitted during the study if indicated clinically; subsequent dose re-escalation is also permitted to meet the dosing objective of the highest tolerated dose.

All patients receive study drug from Day 1 to Week 48. Patients do not receive study drug treatment during a 4-week withdrawal period (between Weeks 48 and 52). Patients re-start study drug treatment at Week 52 at the same dose level that they received prior to the withdrawal period (maintaining the originally randomized treatment group) and continue study drug treatment through Week 100. A follow-up visit occurs at Week 104 (4 weeks after the last dose). The Data Monitoring Committee (DMC) performs quarterly reviews of unblinded data for safety throughout Phase 3 and will evaluate the initial Year 1 results as described in Section 8.1.

Figure 1. Schema for the Phase 3 Portion of the Study



4.2. Definition of Study Drugs

This study uses capsules containing bardoxolone methyl at the 5 and 15 mg strengths, or the corresponding placebos.

4.3. Sample Size Considerations

The Year 1 and Year 2 endpoints are not independent hypothesis sets. The endpoints are analyzed using a combination of the Bonferroni and fixed-sequence approaches for type 1 error control (Section 8.4).

The protocol specified enrollment of approximately 150 patients in the Phase 3 cohort. Actual enrollment in the Phase 3 cohort was 157. Sample size calculations for each year based on the actual number of patients enrolled are presented below.

4.3.1. Primary Endpoints

With 157 patients enrolled (78 in one group and 79 in the second group), the study will have approximately 80% power to detect a difference between the two treatment groups in change from baseline in eGFR of 3.4 mL/min/1.73 m^2 for the primary endpoint at Week 48 in Year 1. The power calculation for Year 1, which was based on mixed-model repeated measures (MMRM) analysis, assumes the following:

- 9 repeated measurements (Weeks 1, 2, 4, 6, 8, 12, 24, 36, and 48) having compound symmetry covariance structure
- The correlation between observations on the same subject is 0.7
- Two-sided Type I error rate of 0.025
- Standard deviation of change from baseline in eGFR of 8 mL/min/1.73 m^2
- Analyses at Week 48 are based on the intent-to-treat (ITT) population

Analysis of the Year 2 primary endpoint, which also will use the ITT population, will have approximately 80% power to detect a difference of 3.3 mL/min/1.73 m^2 assuming 13 repeated measurements, a correlation of 0.7 between observations, a standard deviation of 8 mL/min/1.73 m^2, and a two-sided type I error rate of 0.025. The primary analysis of efficacy will use an unstructured covariance structure, which is expected to have approximately the same power as the analysis with compound symmetry used for study planning.

4.3.2. Key Secondary Endpoint

With 157 patients enrolled, the study will have a minimum detectable difference (i.e., 50% power) between the two treatment groups in change from baseline in eGFR of approximately 2.9 mL/min/1.73 m^2, and 80% power to detect a difference of 4.0 mL/min/1.73 m^2 for the key secondary endpoint at Year 1 and at Year 2. The power calculation, which was based on a 2-sample t-test as an estimate for the planned ANCOVA analysis, assumes the following:

- Two-sided Type I error rate of 0.025 for Year 1 and 0.025 for Year 2
- Standard deviation of change from baseline in eGFR of 8 mL/min/1.73 m^2
- Multiple imputation is used for missing data
- Analyses of the key secondary endpoint are based on the ITT population

4.4. Randomization

Patients enrolled in the Phase 3 cohort are randomized 1:1 to either bardoxolone methyl (BARD) or placebo. Randomization is stratified by screening ACR (<= 300, > 300 to 1000, and 1000 to 3500). Patients with macroalbuminuria (300 mg/g < ACR ≤ 3500 mg/g) at the Screen B visit do not comprise more than approximately 40% of patients enrolled in the Phase 3 cohort. Randomization is generated using a centralized Interactive Web Response System (IWRS).

4.5. Clinical Assessments

All patients in the study follow the same visit and assessment schedule for the 2 years of total follow-up. Table 2 lists the overall schedule of assessments for the study. Following the first dose of study drug on Day 1, patients are scheduled for in-person assessments during Year 1 of treatment at Weeks 1, 2, 4, 6, 8, 12, 24, 36, and 48, as well as by telephone contact on Days 3, 10, 21, 31, 38, and 45. Patients temporarily discontinue study treatment at Week 48 and are assessed at an in-person Year 1 follow-up visit at Week 52 (4 weeks after the last dose). Patients are dispensed treatment at the end of their Week 52 visit at the same dose level (or matching placebo) as the last dose at Week 48. Patients are assessed for in-person assessments during Year 2 of treatment Weeks 64, 76, 88, and 100. Patients are assessed at an in-person, Year 2 follow-up visit at Week 104 (4 weeks after the last dose).

Table 2. Schedule of Assessments

Assessment	Screen A[a]	Screen B[b]	Day 1[c]	Wk 1 (Phone) Day 3±2	Wk 1 Day 7±3	Wk 2 (Phone) Day 10±2	Wk 2 Day 14±3	Wk 3 (Phone) Day 21±2	Wk 4 Day 28±3	Wk 4 (Phone) Day 31±2	Wk 5 (Phone) Day 38±2	Wk 6 Day 42±3	Wk 7 (Phone) Day 45±2	Wk 8 Day 56±3	Wk 12 Day 84±3
Informed consent	X														
Inclusion/ exclusion	X		X[d]												
Demographics and baseline disease characteristics	X														
Concomitant medications	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Medical history	X														
Height	X		X		X		X		X			X		X	X
Weight in clinic	X		X		X		X		X			X		X	X
Weight at home			X	X	X	X	X	X	X	X	X	X	X	X	X
Dispense weight diary			X				X		X			X		X	X
Collect/review weight diary				X	X	X	X	X	X	X	X	X	X	X	X
ECG	X														
Echocardiogram[e]	X														
Vital sign measurements	X		X		X		X		X			X		X	X
Physical exam	X		X		X		X		X			X		X	X
Pregnancy test for WOCBP[f]	X	X	X						X					X	X
Study drug administration			----------X----------												
Dispense study drug			X				X		X			X		X	X
Collect study drug							X		X			X		X	X
Telephone contact				X		X		X		X	X		X		
Adverse event collection			X[g]	X	X	X	X	X	X	X	X	X	X	X	X
Genetic testing[h]	X														
Clinical chemistry (incl. eGFR)	X	X	X		X		X		X			X		X	X
BNP and NT-proBNP	X		X		X		X		X			X		X	X
IGF-1 and serum ketones	X		X		X		X		X			X		X	X
Hemoglobin A1c	X														X
Hematology	X		X				X		X			X		X	X
Urinalysis and microscopy	X		X				X		X			X		X	X
Urine collection for ACR[i]		X							X					X	X
Visual acuity			X												
Audiology assessment			X												
PGIC															
CGI-I															
Virus serology	X														
PK samples[j]															X

Assessment	Wk 24 Day 168±3	Wk 36 Day 252±3	Wk 48 Day 336±3	Wk 52 Day 364-6	Wk 64 Day 448±5	Wk 76 Day 532±5	Wk 88 Day616±5	Wk 100 or End of Treatment[k] Day 700±3	Wk 104 or Follow-up[k] Day 728-6
Informed consent									
Inclusion/ exclusion									
Demographics and baseline disease characteristics									
Concomitant medications	X	X	X	X	X	X	X	X	X
Medical history									
Height	X	X	X	X	X	X	X	X	X
Weight in clinic	X	X	X	X	X	X	X	X	X
Weight at home	X	X	X	X	X	X	X	X	
Dispense weight diary	X	X	X	X	X	X	X		
Collect/review weight diary	X	X	X	X	X	X	X	X	
ECG			X	X				X	X
Echocardiogram[3]									
Vital sign measurements	X	X	X	X	X	X	X	X	X
Physical exam	X	X	X	X	X	X	X	X	X
Pregnancy test for WOCBP[f]	X	X	X	X	X	X	X	X	X
Study drug administration	---------X---------			----------X----------					
Dispense study drug	X	X		X	X	X	X		
Collect study drug	X	X	X		X	X	X	X	
Telephone contact									
Adverse event collection	X	X	X	X	X	X	X	X	X
Genetic testing[8]									
Clinical chemistry (incl. eGFR)	X	X	X	X	X	X	X	X	X
BNP and NT-proBNP	X	X	X	X	X	X	X	X	X
IGF-1 and serum ketones	X	X	X	X	X	X	X	X	X
Hemoglobin A1c			X					X	
Hematology	X	X	X	X	X	X	X	X	X
Urinalysis and microscopy	X	X	X	X	X	X	X	X	X
Urine collection for ACR[9]	X	X	X	X	X	X	X	X	X
Visual acuity			X					X	
Audiology assessment			X					X	
PGIC	X		X	X		X		X	X
CGI-I	X		X	X		X		X	X
Virus serology									
PK samples[10]									

[a] Total Screening period should not exceed 6 months.
[b] Screen B visit should be no more than 30 days prior to Day 1.
[c] Day 1 is the day of administration of the first dose. **On Day 1, all procedures should be performed before study drug administration.**
[d] Screening eligibility procedures do not need to be repeated on Day 1; however, a review of any changes in eligibility criteria should be evaluated prior to Day 1 procedures, and a urine pregnancy test should be performed for WOCBP.
[e] An echocardiogram performed at the Screen A visit or within 6 months prior to Day 1 may be used to determine eligibility.
[f] A serum pregnancy test is performed at the Screen A visit for WOCBP or at any point in time if a pregnancy is suspected. All other pregnancy assessments are urine pregnancy tests. Additional pregnancy assessments are performed more frequently if required by local law or requested by local regulatory authorities or IRBs/IECs.
[g] AE assessments on Day 1 should be performed following study drug administration.
[h] Patients with definitive diagnosis of Alport syndrome from previous genetic testing will not have genetic testing performed as part of the study, but must provide documentation of genetic diagnosis for eligibility.
[i] Albumin to creatinine ratio is measured by first morning void spot urine collection. Appropriate containers for the collection are provided to the patient at the visit prior to collection.
[j] Patients must be instructed to not take their study drug prior to coming to the clinic for visits when PK samples are collected. Patients must administer the study drug dose in the clinic on PK sample collection visits after the 0 hour PK blood sample is collected. Patients will have blood samples for PK analysis drawn just prior to (0 hour) and after (2 and 4 hours) dose administration.
[k] Patients who terminate from the study prior to the Week 100 study visit should be brought back to the clinic as soon as possible for early termination assessments (i.e., end-of-treatment visit) as well as a follow-up visit 4 weeks later.
Abbreviations: ECG = electrocardiogram, PK = pharmacokinetic, WOCBP = women of child-bearing potential

5. PLANNED ANALYSES

The trial design includes hypothesis tests following one and two years of treatment. All available study data are analyzed (1) when Year 1 data are complete, and (2) when Year 2 data are complete. Details of the database lock plans and blinding are described in the Interim Database Deliverable Plan. The strategy for protecting the overall type I error rate of 0.05 for the primary and key secondary endpoints at Year 1 and Year 2 is described in Section 8.4.

A Data Access Plan describes how the sponsor plans to control access to unblinded data following the Year 1 analysis. In the period between the Year 1 analysis and the Year 2 analysis, all patients, investigators, and site personnel will remain blinded to treatment assignments.

The trial design allows for the opportunity to efficiently submit data for regulatory review following demonstration of efficacy after Year 1 while continuing to collect data through Year 2. A limitation of releasing results of the Year 1 analyses prior to all patients completing Year 2 might include patients' perception of their treatment assignment. Note, however, that calculation of eGFR is an objective measure that is expected to be unaffected by such perceptions, and many patients will be well into their 2nd year of treatment at the time of the release of the Year 1 results. Patients and investigators will be reminded of the importance of their commitment to continue participation in the trial as part of an ongoing messaging campaign that began at the start of the trial. To minimize dropouts the sponsor has a patient retention plan in place, which includes an extended access trial (402-C-1803; EAGLE) with open-label treatment for eligible patients completing the 402-C-1603 (CARDINAL) trial.

5.1. Year 1 Data Complete

Year 1 data are considered complete after the last patient enrolled has completed through Week 52 and the database through Week 52 has been locked. All available data will be analyzed when Year 1 data are complete. An assessment of the Year 1 data by the DMC will occur prior to releasing the randomization to the Reata analysis team. The DMC will assess the final Year 1 primary and key secondary efficacy endpoints for statistical significance prior to Reata being unblinded. This process is designed to prevent Reata from becoming unblinded to Year 1 data in the event that the Year 1 primary and key secondary endpoints do not demonstrate statistically significant evidence of efficacy. The DMC will inform Reata whether or not both the primary and key secondary efficacy endpoints demonstrate statistically significant evidence of efficacy (i.e., $p<0.025$). If both the Year 1 primary and key secondary endpoints are not statistically significant, then Reata will not to unblind until the Year 2 data are complete. However, if both the Year 1 primary and key secondary endpoints are statistically significant, then Reata will unblind and the final analysis of Year 1 data and a preliminary assessment of Year 2 safety data will be performed as described in Section 5.1.1 and Section 5.1.2.

5.1.1. Final Analysis – Year 1

At the time of Year 1 data complete, results from analyses of data through Week 52, including the primary and key secondary efficacy endpoints, are considered final. Because patients will be continuing in the trial, the interim CSR will not include by-patient data listings or patient

narratives to minimize the number of individuals having access to patients' treatment assignments.

5.1.2. Preliminary Analysis – Year 2

At the time of the Year 1 data complete and prior to Year 2 data complete, many patients will have completed visits beyond Week 52. Available safety data for Year 2 visits will be summarized descriptively, including summaries of AEs, vital sign assessments, and laboratory results (including descriptive summaries of eGFR for safety). Year 2 efficacy endpoints will not be assessed with the preliminary Year 2 analysis. The interim CSR will include the preliminary Year 2 safety results.

5.2. Year 2 Data Complete

Year 2 data are considered complete after the last patient enrolled has completed through Week 104 and the database through Week 104 has been locked. Year 2 efficacy endpoints will be assessed only when Year 2 data are complete. The interim CSR will be amended to add Year 2 efficacy results, and to finalize Year 2 safety results. Because the trial is complete when Year 2 data are complete, the final CSR will include by-patient data listings and patient narratives.

6. GENERAL CONSIDERATIONS FOR DATA ANALYSES AND HANDLING

The efficacy and safety analyses use the analysis sets defined in Section 6.3. Patient listings (as appropriate) of all analysis data that support summary tables and/or figures are provided along with their source data. The summary tables do not include measurements from patients excluded from the pre-defined analysis sets or extra measurements (such as unscheduled or repeat assessments) not closest to the target study day unless specified otherwise, but the patient listings do include these data. Missing data are not imputed, unless otherwise specified. In general, patient listings are sorted by patient number and assessment date (time and parameter, as applicable). Any laboratory value (including eGFR), vital sign assessment, or ECG value collected after starting dialysis or after receiving a kidney transplant is considered invalid and will be treated as missing.

6.1. General Summary Table and Individual Subject Data Listing Considerations

Results of statistical analyses are reported using summary tables, listings, and figures (TLFs). All TLFs will use ICH numbering conventions. The strategy for controlling the overall Type I error rate for the Year 1 and Year 2 primary and key secondary efficacy endpoints and other select exploratory endpoints is described in Section 8.4. For endpoints not described in Section 8.4, the reported significance levels are nominal and the following statistical conventions are used:

- Unless otherwise noted, all statistical testing is two-sided and is performed at the 0.05 significance level.
- Tests are declared statistically significant if the calculated p-value is <0.05.

All analyses and summaries are produced using SAS® version 9.3 (or higher).

6.2. Data Presentation Conventions

Unless otherwise specified, descriptive statistics for continuous variables include the number of patients with data (N), mean, standard deviation (SD), median, minimum, and maximum. The same number of decimal places as in the observed value are presented when reporting minimum and maximum; 1 more decimal place than in the observed value is presented when reporting mean and quartiles; and 2 more decimal places than in the observed value is presented when reporting SD.

Categorical data are presented using frequency counts and percentages. All percentages are rounded to 1 decimal place, unless otherwise specified. Percentages equal to 100 are presented as 100% and no percentages are presented for zero frequencies. Where individual variable values are missing, summaries of categorical data are based on reduced denominators (i.e., the denominators include only patients with available data) and the number of missing values is presented. For summaries of AEs and concomitant medications (CM), the percentages are based on the total number of patients in each treatment group.

6.3. Analysis Populations

Analysis populations defined in this section pertain to patients enrolled in the Phase 3 portion of the trial.

6.3.1. ITT Population

The intent-to-treat analysis set is defined as all enrolled patients categorized by their randomized treatment group (whether or not they received study drug). Note that this is the same definition for both the Year 1 and Year 2 analyses.

6.3.2. Per-Protocol (PP) Population – Year 1

A sensitivity analysis exploring the robustness of the primary ITT findings of Year 1 endpoints is based on the patients in the per-protocol, Year 1 population.

The per-protocol (Year 1) population is defined as patients who:

- Received study drug through Week 48; and
- Had no protocol deviation that could potentially affect the efficacy conclusions.

Review of protocol deviations for determining the Year 1 per-protocol population will be performed by a blinded team prior to Year 1 database lock.

6.3.3. Per-Protocol (PP) Population – Year 2

A sensitivity analysis exploring the robustness of the primary ITT findings of Year 2 endpoints is based on the patients in the per-protocol, Year 2 population.

The per-protocol (Year 2) population is defined as patients who:

- Received study drug through Week 100; and
- Had no protocol deviation that could potentially affect the efficacy conclusions.

Review of protocol deviations for determining the Year 2 per-protocol population will be performed by a blinded team prior to the final database lock.

6.3.4. Safety Population

Safety analyses are based on all enrolled patients. The safety population includes all patients who received at least 1 dose of randomized study drug in Year 1. The safety population is used for evaluation of safety variables. 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 dose of bardoxolone methyl will be classified in the placebo group. The safety population is based on this definition for both the Year 1 and Year 2 analyses.

Safety population treatment groups will not be modified after the Year 1 analysis. Patients in the placebo group receiving a dose of bardoxolone methyl in Year 2 are listed.

6.3.5. Japanese Population

The Japanese population includes all patients enrolled at sites located in Japan. A sensitivity analysis exploring safety and efficacy in Japanese patients is performed as appropriate using the

Japanese population. The Japanese population is used for evaluation of efficacy and safety overall and by relevant subgroups to support potential future regulatory submissions to PMDA in Japan.

6.4. Baseline Definition

Baseline values are defined as the last non-missing assessment prior to the first study drug administration, unless otherwise specified below. If the first study drug administration occurs after the date of randomization, the last measurement prior to the first study drug administration is considered the Day 1 measurement for the calculation of baseline. Assessments without time-stamp collected on the same date as the first date of study drug administration are considered to occur before the first dose of study drug administration.

6.4.1. Estimated Glomerular Filtration Rate

Baseline eGFR is defined as the average of screening and Day 1 eGFR measurement, calculated as shown below:

- Screening eGFR = average of the last two eGFR measurements collected prior to the Day 1 eGFR collection
- Day 1 eGFR = the measurement on the date of first study drug administration prior to the first study drug administration
- Baseline eGFR = (0.5 × Day 1 eGFR) + (0.5 × Screening eGFR)

6.4.2. Serum Creatinine

Baseline serum creatinine (SCr) is defined in the same way as baseline eGFR (Section 6.4.1).

6.4.3. Safety Assessments

Baseline for continuous safety assessments (i.e., vital sign assessments, weight, BMI, and laboratory measurements) is defined as the average value of measurements collected up through the date of, but not after, first study drug administration.

6.4.4. Urine Albumin to Creatinine Ratio

Baseline ACR is defined as the last ACR value prior to the first study drug administration. Baseline ACR is summarized using the geometric mean of ACR values.

6.4.5. Natural log(ACR)/eGFR

Baseline of natural log(ACR)/eGFR is defined as the ratio of the natural log of baseline ACR (Section 6.4.4) divided by baseline eGFR (Section 6.4.1).

6.4.6. Baseline Hematuria

Presence of hematuria at baseline is defined as red blood cells per high power field (RBC/HPF) > 2 in two measurements prior to the first dose of study drug. If only one RBC/HPF measurement is available prior to the first dose of study drug, that measurement is used to determine baseline hematuria.

If no RBC/HPF measurements are available prior to the first dose of study drug, urine occult blood dipstick measurements is used to determine baseline hematuria. If the two occult blood measurements prior to the first dose of study drug result is not 'trace' or 'negative' (e.g., result is +1,+2, or +3), then the patient is considered to have baseline hematuria. If no RBC/HPF measurement is available and only one occult blood measurement prior to the first dose of study drug is available, then the patient is considered to have baseline hematuria if the results are not 'trace' or 'negative'.

6.4.7. Baseline Visual Impairment

Patients with normal visual acuity are those with 20/20 vision or better (e.g., 20/15, 20/10, etc.) in both eyes. Any patient whose vision is worse than 20/20 in either eye is considered to have visual impairment.

6.4.8. Baseline Audiology Impairment

Patients with normal hearing are those with pure tone audiometry (PTA) < 25 decibels in both ears. Any patient with > 25 decibel hearing in either ear is considered to have audiology impairment.

6.5. Derived and Transformed Data

6.5.1. Age

Subject's baseline age is defined as the age at consent (Screen A). Subject's age used for each eGFR calculation (Section 6.5.6) is the age on the date of each serum creatinine collection. Calculated age is rounded down to the nearest whole year and expressed as an integer:

- Age (year) = Floor ((visit date – date of birth)/365.25)

6.5.2. Study Day

Study day is the day relative to the date of randomization. Day 1 is defined as the date of randomization. Assessments that occur after randomization but before the first dose of study drug are considered to occur on study Day 1.

Assessments without time-stamp collected on the same date as the first date of study drug administration are considered to occur before the first dose of study drug administration.

For visits (or events) after randomization, day is calculated as:

- Study day = visit (or event) date - date of randomization + 1

For visits (or events) before randomization, day is calculated as:

- Study day = 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 are defined as days divided by seven and months as days divided by 30.4.

6.5.3. Change from Baseline

Change from baseline is calculated using the baseline value (Section 6.4) and the value closest to the target study day, using the rules defined in Section 6.5.4.

6.5.4. Visit Windows

Analysis visits and their windows are defined using derived study day (Section 6.5.2) instead of relying on visit labels in the clinical database because clinical visits may occur outside protocol-specified windows. Study day and days after last dose are calculated using the actual date of each scheduled and unscheduled assessment and compared to the target for each analysis visit as specified in Table 3 and Table 4. They are included in analyses of safety and efficacy as follows:

- Efficacy analyses for endpoints assessed at Week 48 or Week 100 use the analysis windows based on target study day in Table 3, irrespective of whether or not a patient is receiving treatment of study drug efficacy analyses for Week 100 will not include the Week 52 data.
- Efficacy analyses for endpoints assessed at Week 52 or Week 104 use the off-treatment analysis windows based on days after last dose in Table 3.
- Safety analyses use the analysis windows defined in Table 3 so long as the patient is receiving treatment of study drug. Safety follow-up at Week 52 will also be summarized according to Table 3 so long as the patient has not permanently discontinued treatment prior to Week 52. Similarly, safety follow-up at Week 104 is summarized according to Table 3 for patients completing treatment through Week 100. Once a patient permanently discontinues study treatment, safety data are summarized according to time since last dose as defined in Table 4.

If a parameter is assessed or measured more than once within a visit window, the one that is closest to the protocol-scheduled time point (i.e., target) is used for the purposes of data analysis and summary. If two assessments are equidistant from a target, the earlier assessment is used. If the visit used for analysis includes two assessments on the same day, the average of the two measurements is used.

Records from visits not closest to the target study day, and therefore not used in analyses, are presented in by-subject data listings.

Table 3. Analysis Visit Windows

Analysis Visit	Label	Target		Analysis Window
		Study Day[a]	Days After Last Dose	
0	Day 1[b]	1	-	1
1	Week 1	7	-	2 ≤ study day ≤ 10
2	Week 2	14	-	11 ≤ study day ≤ 21

Analysis Visit	Label	Target		Analysis Window
		Study Day[a]	Days After Last Dose	
4	Week 4	28	-	22 ≤ study day ≤ 35
6	Week 6	42	-	36 ≤ study day ≤ 49
8	Week 8	56	-	50 ≤ study day ≤ 70
12	Week 12	84	-	71 ≤ study day ≤ 126
24	Week 24	168	-	127 ≤ study day ≤ 210
36	Week 36	252	-	211 ≤ study day ≤ 294
48	Week 48[c]	336	-	295 ≤ study day ≤ 350
52	Week 52[d]	-	28	14 ≤ Days after last dose[e] ≤ 35
64	Week 64	448	-	407 ≤ study day ≤ 490
76	Week 76	532	-	491 ≤ study day ≤ 574
88	Week 88	616	-	575 ≤ study day ≤ 658
100	Week 100[f]	700	-	659 ≤ study day ≤ 714
104	Week 104[g]	-	28	14 ≤ days after last dose[h] ≤ 35

[a] Study Day is relative to the date of randomization (Section 6.5.2).
[b] Day 1 is the last measurement prior to the first study drug administration.
[c] Week 48 is the end of the first-year dosing interval and should coincide with the last dose in the first year.
[d] Week 52 is the first-year dosing interval 4-week follow-up; must be after the last dose in the first year and before receiving treatment in the second year.
[e] Last dose for analysis visit 52 is the Week 48 dose or date of last dose for permanent study drug discontinuation prior to Week 48.
[f] Week 100 is the end of the second-year dosing interval and should coincide with the last dose in the second year.
[g] Week 104 is the second-year dosing interval 4-week follow-up; patient must have been dispensed treatment in the second year (i.e., Week 52), otherwise Week 104 is considered missing.
[h] Last dose for analysis visit 104 is the Week 100 dose or date of last dose for permanent study drug discontinuation after Week 52 and prior to Week 100.

6.5.4.1. Off-Treatment Visit Windows

Patients who permanently discontinue study drug prior to Week 100 are asked to resume the planned assessments according to the study schedule in the protocol. Off-treatment values for clinical laboratory evaluations (Section 10.5), vital signs (Section 10.6), and electrocardiograms (Section 10.8) are those that occur after the last dose date for patients permanently discontinuing study drug prior to Week 100, and is summarized relative to their last dose of study drug. Off-treatment safety assessments are grouped for analyses according to the strategy in Table 4. Assessments that occur on the date of last dose are considered on treatment.

Table 4. Analysis Visits for Off-Treatment Safety Analysis

Off-Treatment Analysis Visit	Label	Target Study Day (days after last dose)	Analysis Window
0	Last dose (on-treatment)[a]	0	0
4	Off Treatment 4-weeks[b]	28	14 ≤ days after last dose ≤ 35
8	Off Treatment 8-weeks[b]	56	36 ≤ days after last dose ≤ 71
12	Off Treatment 12-weeks[b]	84	72 ≤ days after last dose ≤ 99
24	Off Treatment 24-weeks[b]	168	100 ≤ days after last dose ≤ 252
48	Off Treatment 48-weeks[b]	336	253 ≤ days after last dose ≤ 350

[a] Last dose for patients permanently discontinuing study drug prior to Week 100 Assessments that occur on the date of last dose or less than 14 days after last dose are considered on treatment

[b] The off-treatment values indicate the value closest to the date of last dose. Results are summarized by Year 1, Year 2, and overall.

6.5.5. Laboratory Evaluations

Post-baseline laboratory assessments less than the lower limit of detection (i.e., < LLD) are imputed as LLD/2. If no LLD is available, then the imputed value is the minimum numeric value listed divided by 2 (e.g., < 25 is 25/2=12.5). Laboratory assessments above the ULD are imputed as the ULD. If the lab result is qualitative but presented as > X and X is 10 times greater than the ULN, then the value X is used in the analysis.

6.5.6. Estimated Glomerular Filtration Rate (eGFR)

The eGFR is calculated using the formula below according to patient's age at the date of consent. The formula will not change throughout the study. The Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation is used for adult patients (age at consent at least 18 years):

- eGFR (mL/min/1.73 m^2) = 141 × $\min(S_{cr}/\kappa, 1)^{\alpha}$ × $\max(S_{cr}/\kappa, 1)^{-1.209}$ × 0.993^{Age} × 1.018 [if female] × 1.159 [if black]

The Bedside Schwartz equation is used for pediatric patients (age at consent of 12 to 17):

- eGFR (mL/min/1.73 m^2) = (0.41 × *Height* in cm) / S_{cr}

where S_{cr} 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. Min indicates the minimum of S_{cr}/κ and 1, and max indicates the maximum of S_{cr}/κ and 1. *Age* indicates age at time of serum creatinine collection. *Height* indicates height at time of serum creatinine collection.

6.5.6.1. Prognosis of CKD by eGFR and Albuminuria Category

Prognosis of CKD is determined by the intersection of the following eGFR and albuminuria categories (KDIGO 2012):

Table 5. eGFR Category

eGFR Category	eGFR (mL/min/1.73 m^2)
1	≥ 90
2	60 to < 90
3a	45 to < 60
3b	30 to < 45
4	15 to < 30
5	< 15

Table 6. Albuminuria Category

ACR Category	ACR (mg/g)
1	< 30
2	30 to ≤ 300
3	> 300

6.5.7. Urine Albumin to Creatinine Ratio (ACR)

ACR is provided in the central laboratory database as the ratio of urine albumin to urine creatinine from the first morning void (FMV) urine collection. The ACR value is reported in the central laboratory database as missing when the FMV urine albumin result is < LLD. Urine albumin results < LLD and the corresponding ACR missing values are imputed as follows:

- If urine albumin result = < LLD, then
  - Imputed urine albumin result (mg/dL) = LLD/2
  - Imputed ACR(mg/g) = (imputed urine albumin result in mg/dL) / [(urine creatinine in mg/dL) /1000]

The ACR results are log-transformed for analysis to produce data that are more normally distributed. Any imputed ACR result where ACR=0 is considered to be 0.1 mg/g for purposes of calculating the geometric mean.

6.5.7.1. Baseline Urine Albumin to Creatinine Ratio (ACR) Categorical Status

Baseline ACR status will also be grouped by the following categories using baseline ACR:

- < 30 mg/g
- 30 mg/g < ACR <= 300 mg/g
- 300 mg/g < ACR <= 1000 mg/g
- 1000 mg/g < ACR <= 3500 mg/g
- ACR > 3500 mg/g

6.5.7.2. ACR Category used in Stratification (ACR_STRAT)

The ACR category used for stratification (ACR_STRAT) is determined by the randomization list with the following categories:

- ACR <= 300 mg/g
- 300 mg < ACR <= 1000 mg/g
- 1000 mg/g < ACR <=3500mg/g

6.5.8. Electrocardiogram Fridericia Corrected QT Interval

Electrocardiogram intervals are assessed locally at each site. The following formula is used to calculate the QTcF interval for analysis from QT and RR intervals:

- $QTcF = QT/\sqrt[3]{RR}$

where RR = 60 / (Heart Rate).

6.5.9. Log of The Minimum Angle of Resolution

Best-corrected visual acuity (BCVA) assessments are performed at time points specified in the scheduled assessments in Table 2. The Snellen chart values are used to derive the log of the minimum angle of resolution (logMAR) value as below:

- LogMAR = - log10 (Snellen)

6.5.10. Natural log(ACR)/eGFR

To evaluate ACR after adjusting for filtration rate, ln(ACR)/eGFR (Sections 6.5.7 and 6.5.6) is assessed.

6.5.11. Exposure Duration

Exposure duration is used in analysis of efficacy to adjust analysis of off-treatment eGFR for the exposure duration within each year. Fraction of expected exposure duration is calculated for the first and second years of treatment as follows:

- First year exposure (TRT_EXP1) =

  [(date of last dose in first year) – (date of first dose)]/336

- Second year exposure (TRT_EXP2) =

  [(date of last dose in second year) – (date of first dose)]/700

where 336 is the number of days in the first year (i.e., 48-week) treatment interval, and 700 is the number of days in the second year (i.e., 100-week) treatment interval. Exposure in the second year is simply total exposure and does not subtract the 4-week off-treatment period at the end of the first year.

6.6. Handling of Missing Data

6.6.1. Missing Efficacy Endpoint Data

The primary analysis of efficacy is based on an assumption of missing at random (MAR) (Section 8.5). The tipping point sensitivity analysis (Section 8.6.2.1) is performed to assess how severe departures from MAR must be in order to overturn conclusions from the primary analysis.

6.6.1.1. Missing eGFR for MMRM Analyses

Missing eGFR data are not imputed for the primary MMRM analysis of the primary endpoints (Week 48 and Week 100). A set of sensitivity analyses is included to assess the robustness of conclusions to the MAR assumption (see Section 8.6.2 and Section 8.8.2) for the primary analysis of the primary endpoints (Week 48 and Week 100).

6.6.1.2. Missing eGFR for ANCOVA Analyses

Missing eGFR data are imputed for the primary ANCOVA analysis of the key secondary endpoints (Week 52 and Week 104) using multiple imputation based on an assumption of MAR. A set of sensitivity analyses is included to assess the robustness of conclusions to the MAR assumption (see Section 8.7.2 and Section 8.9.2) for the primary analysis of the key secondary endpoints (Week 52 and Week 104).

6.6.2. Missing Start and Stop Dates for Concomitant Medication

Missing start dates for concomitant medications are not imputed.

Concomitant medications with incomplete end dates are considered concomitant medications if:

- Day and month are missing, and the year is equal to or after the year of the first date of study drug administration;
- Day is missing and the year is after the year of the first date of study drug administration;
- Day is missing and the year is equal to the year of the first date of study drug administration and the month is equal to or after the month of the first date of study drug administration; or
- Year is missing.

6.6.3. Missing Start and Stop Dates for Adverse Events

Treatment-emergent adverse events (TEAEs) are events that either:

- Have a date of onset on or after the date of the first dose and not more than 30 days after the date of the last dose of study drug, or
- Have no recorded date of onset with a stop date after the first dose of study drug, or
- Have no recorded date of onset or stop date.

Adverse events with incomplete start dates are considered after the date of first dose if:

- Day and month are missing and the year is equal to or after the year of the first date of study drug dosing;
- Day is missing and the year is after the year of the first date of study drug dosing;
- Day is missing and the year is equal to the year of the first date of study drug dosing and the month is equal to or after the month of the first date of study drug dosing; or
- Year is missing.

Adverse events with incomplete start dates are considered on or within 30 days of last dose, if:

- Day and month are missing and the year is equal to or before the year of the date of last dose of study drug plus 30 days;
- Day is missing and the year is equal to or before the year of the date of last dose of study drug plus 30 days and month is equal to or before the month of the date of last dose of study drug plus 30 days;
- Year is missing.

7. STUDY POPULATION

7.1. Subjects Disposition

A disposition summary includes the number and percentage of patients in all analysis populations in the following categories:

Disposition - Year 1

- Per-protocol Population – Year 1
- Completed treatment through Week 48
- Discontinued treatment prior to Week 48
  - reason for discontinuing treatment
- Completed study follow-up through Week 52
  - on-treatment
  - discontinued treatment early but completed in-person study visits
  - continued to Year 2 follow-up
- Terminated from the trial before Week 52
- Follow-up by Year 1 study week
  - discontinued treatment
  - follow-up in-clinic
  - follow-up phone

Disposition - Year 2

- Per-protocol Population – Year 2
- Received at least one dose in Year 2
- Completed treatment through Week 100
- Discontinued treatment prior to Week 100
  - reason for discontinuing treatment
- Completed study follow-up through Week 104
  - on-treatment
  - discontinued treatment early but completed study visits
- Terminated from the trial before Week 104
- Follow-up by Year 2 study week

- discontinued treatment
- follow-up in-clinic
- follow-up phone

Patients who completed follow-up are defined as those who were followed by phone or in-person through the Week 104 visit. A listing of disposition is provided for all enrolled patients.

7.2. Screen Failures

Screen failures are not summarized.

7.3. Protocol Deviations

Prior to database lock at Year 1, a blinded team will perform a blinded data review to identify all deviations, including major protocol deviations that could potentially affect the efficacy conclusions of the study, to define the Per-Protocol population (see Section 6.3.2 and Section 6.3.3). A blinded team will repeat the blinded data review process for Year 2 deviations prior to the Year 2 database lock. Protocol deviations for excluded medications, patients who entered the study even though they did not satisfy all entry criteria, patients who received the wrong treatment or incorrect dose, and major protocol deviations are listed.

7.4. Demographic and Baseline Characteristics

Summaries of demographic and other baseline characteristic data are presented by treatment group for all analysis populations. They may also be summarized by select subgroups (including US and non-US) as appropriate (Section 8.3).

The demographic and other baseline characteristics include:

- Baseline Age, Age category (< 18, ≥ 18)
- Sex, Race, Ethnicity
- Weight (kg), Height (cm), BMI (kg/m^2)
- Diastolic and systolic blood pressure (mmHg), Heart rate (bpm)
- Serum creatinine, eGFR, eGFR categorical status (< 60, ≥ 60)
- ACR, ACR categorical status (≤ 300; > 300), ACR Stratum
- Prognosis of CKD by eGFR and Albuminuria categories
- Hematuria (present/absent)
- Angiotensin converting enzyme (ACE)-inhibitor treatment and/or Angiotensin II receptor blocker (ARB) (yes/no)
- Other baseline variables of interest

Alport syndrome related information includes:

- Age at Alport syndrome diagnosis

- Genetic subtype (x-linked, autosomal, unknown)
- Genotype (COL4A3, COL4A4, COL4A5, Unknown)
- Confirmed histologic diagnosis
- Hearing loss (yes/no)
- Visual impairment (yes/no)

7.5. Listing of Subject Inclusion and Exclusion Criteria

A listing of enrolled patients who did not meet inclusion or exclusion criteria is generated.

7.6. Medical History

Medical history is summarized by treatment. Medical history is coded using MedDRA (Medical Dictionary for Regulatory Activities) version 19.1. Medical history items are summarized by MedDRA SOC and PT.

7.6.1. Historical eGFR

To characterize disease progression prior to trial entry, up to 5 years of historical serum creatinine values are collected for each patient as part of their medical history. Collection of historical serum creatinine is subject to availability of data from medical records; therefore, the number of available values will vary by patient and some patients may have no historical values. Historical serum creatinine values are converted to mg/dL and eGFR are calculated for analysis according to the appropriate equation (Section 6.5.6) using the patient age at laboratory collection for adult patients, or patient height at Screening for pediatric patients. Historical serum creatinine values were analyzed at various local laboratories (i.e., not collected as part of central lab data). Therefore, historical serum creatinine and corresponding eGFR values are considered part of a patient's medical history, and distinct from central lab assessments collected as part of the trial.

The eGFR values used in the analysis dataset for the baseline label are the baseline eGFR values for each patient (defined in Section 6.4.1), not the individual Screen A, Screen B, and Day 1 eGFR values.

Historical values are prior to first dose and have a negative Study Day value:

- Study Day = (Date of Historical Lab Collection) – (Date of Screen A)

Summary statistics of available historical eGFR results are based on the following analysis windows:

Table 7. Labels for Summary of Historical Change in eGFR

Analysis Visit	Label	Study Day Windows
-5	Historical Year 5	1825 to 1461 days prior to Screen A
-4	Historical Year 4	1460 to 1096 days prior to Screen A
-3	Historical Year 3	1095 to 731 days prior to Screen A
-2	Historical Year 2	730 to 366 days prior to Screen A
-1	Historical Year 1	365 to 1 days prior to Screen A
0	Baseline	Screen A to Day 1

The summary statistics will include the average eGFR value by patient and analysis time point. For instance, if multiple values are listed in the historical Year 1 results for a given patient, the average of those values is used to calculate the summary statistics.

Summary statistics of eGFR results by historical year (listed in Table 7) will also be summarized.

8. PRIMARY EFFICACY

Analyses of efficacy described in this section are the primary analyses of the Phase 3 efficacy endpoints, and pertain to the ITT population (i.e., all patients randomized in the Phase 3 portion of the study). The trial includes endpoints following one and two years of treatment. Additional details regarding the statistical approach and pseudo SAS code are available in the programming conventions document.

The following abbreviations are used in descriptions of the analysis models:

Table 8. Model Parameter Abbreviations

Abbreviation	Model Term
ACR_STRAT	ACR strata used in randomization
BASE_eGFR	Baseline eGFR
BASE_eGFR x VISIT	the interaction between baseline eGFR and time
TRT	treatment group
TRT_EXP1_	fraction of one-year exposure to treatment (Section 6.5.11)
TRT_EXP2	fraction of two-year exposure to treatment (Section 6.5.11)
TRT x VISIT	the interaction between treatment and time
VISIT	analysis visit used as time

8.1. General Considerations

Analyses are performed for all bardoxolone methyl patients in comparison with all placebo patients. Summary statistics for observed values, change from baseline, and percent change from baseline are presented by randomized treatment group. Missing values are imputed as described in Section 8.5.

8.2. Statement of the Null and Alternate Hypotheses

All primary and key secondary efficacy endpoints compare the change from baseline in eGFR for patients randomized to bardoxolone methyl (μ_{BARD}) to the change from baseline in eGFR for patients randomized to placebo ($\mu_{placebo}$) at the endpoint-specified visit according to the following hypotheses:

- H_0: $(\mu_{BARD}) - (\mu_{placebo}) = 0$ mL/min/1.73 m^2
- H_1: $(\mu_{BARD}) - (\mu_{placebo}) \neq 0$ mL/min/1.73 m^2

Because eGFR decreases with disease progression, an increase in eGFR relative to placebo is considered evidence of benefit.

8.3. Subgroup Analyses

The primary and key secondary efficacy endpoints are analyzed for subgroups of interest to assess homogeneity of the effect size for those subgroups that are sufficiently large to warrant these analyses. Exploratory endpoints are also analyzed for these subgroups.

The following subgroup analyses are tabulated:

- Age (years): < 18; ≥ 18
- Sex: female; male
- Geographic location: US; Non-US
- Ethnicity: Non-Hispanic/Latino; Hispanic/Latino
- Race: White; Non-White
- Baseline BMI (kg/m^2): < 30; ≥ 30
- Baseline eGFR ($mL/min/1.73\ m^2$): ≤ 60; > 60
- Baseline ACR (mg/g): ≤ 300; > 300
- Baseline ACEi/ARB use: Yes; No
- Genetic subtype: X-linked Alport syndrome (XLAS); Non-XLAS

Subgroup analyses will include nominal significance levels for descriptive purposes only. Forest plots are generated for change from baseline eGFR results in each subgroup at Weeks 48, 52, 100, and 104.

8.4. Multiple Comparisons and Multiplicity

The trial includes separate Phase 2 and Phase 3 cohorts, where the cohorts constitute independent sets of patients. Therefore, the Phase 2 analysis has no impact on the Type I error rate for the Phase 3 analysis.

The Phase 3 cohort includes a family of hypothesis tests following the first year of treatment and a second family of hypothesis tests following the second year of treatment. All endpoints tested at Year 1 will use a significance level of 0.025 and those tested at Year 2 will also use a significance level of 0.025. Within each year, endpoints are tested following a fixed-sequence hierarchical testing strategy. Formal testing in the sequence may proceed so long as statistically significant evidence of benefit continues to be shown. At such point in the sequence that one endpoint does not show statistically significant evidence of benefit, formal statistical testing of subsequent endpoints will not occur.

Testing endpoints in the following sequence protects the overall type I error rate of 0.05 for the study:

Year 1 Testing Sequence (all significance levels = 0.025)

1. eGFR at Week 48 (primary at Year 1)
2. eGFR at Week 52 (key secondary at Year 1)

Year 2 Testing Sequence (all significance levels = 0.025)

1. eGFR at Week 100 (primary at Year 2)
2. eGFR at Week 104 (key secondary at Year 2)

3. time-to first kidney failure composite (exploratory at Year 2)
4. distribution of PGIC at Week 100
5. distribution of CGI-I at Week 100

All remaining endpoints are considered "Exploratory" and presented with nominal significance levels for descriptive purposes only.

8.5. Handling of Missing Data

Missing data will not be imputed for analysis of exploratory efficacy endpoints.

8.5.1. Missing eGFR for Sensitivity Analyses

Multiple imputation is the primary approach used to account for missing data in sensitivity analyses of the primary efficacy endpoints (i.e., Week 48 and Week 100) and the key secondary endpoints (i.e., Week 52 and Week 104). Results of sensitivity analyses are summarized in tables and forest plots.

8.5.2. Missing eGFR for MMRM Analyses

The primary analysis of efficacy endpoints is based on the assumption of missing at random (MAR). Missing data are not imputed for the primary analysis of eGFR where the analysis method is defined as MMRM (i.e., Week 48 and Week 100) (Section 6.6.1.1).

8.5.3. Missing eGFR for ANCOVA Analyses

Missing eGFR data are imputed using multiple imputation for the primary analysis of eGFR where the analysis method is defined as ANCOVA (i.e., Week 52 and Week 104) (Section 6.6.1.2). For the primary analysis, multiple imputation is based on the treatment group to which the subject is assigned. A tipping point analysis is used to assess the validity of the MAR assumption.

8.6. Year 1: Primary Efficacy Endpoint

The analyses of the primary and key secondary endpoints are described in Sections 8.6 and 8.7. Additional details and pseudo SAS code are provided in a separate document.

8.6.1. Primary Analysis of the Year 1 Primary Efficacy Endpoint

The Year 1 primary endpoint of this study is the change in eGFR at Week 48 (i.e., the end of the first year of drug exposure). The change from baseline eGFR for patients treated with bardoxolone methyl is compared with placebo at Week 48 using mixed models repeated measures (MMRM) analysis, with baseline eGFR (BASE_eGFR), ACR strata (ACR_STRAT), and fraction of one-year exposure to treatment (TRT_EXP1) as covariates and the following fixed factors: treatment group (TRT), time (i.e., VISIT; analysis visit number), the interaction between treatment and time (TRT x VISIT), and the interaction between baseline eGFR and time (BASE_eGFR x VISIT) as covariates. Inclusion of the fraction of expected one-year exposure to treatment (Section 6.5.11) in the MMRM model as a covariate provides an adjustment for eGFR values collected from early discontinuation patients having less than 1-year of drug exposure.

The trial design separates discontinuation of study drug from discontinuation of study follow-up. Therefore, patients may discontinue early from study drug while continuing follow-up with study visits and assessments. The primary analysis of eGFR at Week 48 uses available eGFR values irrespective of whether or not a patient is receiving treatment. The MMRM analysis uses all eGFR values collected through Week 48 according to analysis visits 0, 1, 2, 4, 6, 8, 12, 24, 36, and 48 (Section 6.5.4). An unstructured covariance matrix is used.

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.

The difference between bardoxolone methyl and placebo in change from baseline of eGFR is estimated along with the 95% confidence interval at Week 48 for the primary analysis of the Year 1 primary endpoint.

8.6.2. Sensitivity Analyses

Sensitivity analyses are included to assess the robustness of conclusions to the primary analysis of the Year 1 primary efficacy endpoint (i.e., change in eGFR at Week 48), and to assess the assumption of MAR. The following sets of sensitivity analyses will be summarized in tables and forest plots: tipping point with multiple imputation, control-based imputation, MMRM fit to the as-treated population, and MMRM fit to the per protocol population. Additional sensitivity analyses may be performed as appropriate.

8.6.2.1. Tipping Point

The following tipping point sensitivity analysis is performed to assess how severe departures from MAR must be in order to overturn conclusions from the primary analysis:

1. The PROC MI procedure in SAS is used to generate 100 datasets satisfying the assumption of monotone missingness, imputing any missing eGFR values using multiple imputation. Imputation is based on the non-missing observations within each treatment group.

2. The PROC MI procedure in SAS is applied to each of the 100 datasets subsetted to the placebo group and all missing values in the placebo group are imputed. These 100 datasets will then be used for the tipping point analysis.
3. For the tipping point analysis with a shift parameter, patients in the bardoxolone methyl group with a missing value will be multiply imputed separately from the placebo-treatment arm and are assigned a shift parameter in the imputation procedure for progressively worse (lower) scores to find the point at which statistical significance is lost. Specifically, if the outcome of the hypothesis test favors bardoxolone methyl over the placebo group (p-value <0.025), then one point is subtracted as a shift parameter in the imputation for the bardoxolone methyl group, while keeping the data for the placebo group unchanged.
4. A mixed model repeated measures (MMRM) model (see Section 8.6.1) is used to analyze each of the 100 completed data sets. PROC MIANALYZE is used to combine these results to obtain the final estimates for the given value of the shift parameter.
5. This "eGFR shifting", as outlined in steps 3 and 4, is repeated with one shift point at a time until the hypothesis test no longer rejects the null hypothesis in favor of the bardoxolone methyl group over the placebo group (i.e., when the p-value becomes greater than 0.025). Additional increments may be used to locate the tipping point. If the result is not quite statistically significant, progressively higher scores are added to find the point at which statistical significance is attained.

8.6.2.2. As-Treated Analysis

The primary analysis of efficacy includes all eGFR values collected during trial follow-up through Week 48, regardless of patient receiving study drug. The as-treated sensitivity analysis to assess the treatment effect while receiving study drug is based on the ITT population, but excludes any eGFR values collected after final dose. The as-treated analysis is performed using the MMRM statistical model defined in Section 8.6.1, with eGFR values collected after final dose treated as missing. Missing data are not imputed for the as-treated sensitivity analysis.

8.6.2.3. Per-Protocol Population Analysis

The per-protocol population analysis is performed for the per-protocol population (Year 1) using the MMRM statistical model defined in Section 8.6.1. Missing data will not be imputed for the per-protocol population sensitivity analysis.

8.7. Year 1: Key Secondary Endpoint

8.7.1. Primary Analysis of Year 1 Key Secondary Endpoint

The Year 1 key secondary endpoint of this study is the change in eGFR at Week 52 (i.e., 4 weeks following the last dose after one year of drug exposure). Analysis of Week 52 assesses the preserved drug benefit (relative to placebo) following withdrawal of treatment; therefore, analysis of Week 52 data does not include the eGFR values collected during treatment.

The change from baseline eGFR at Week 52 for patients treated with bardoxolone methyl is compared to placebo using an analysis of covariance (ANCOVA) model, with BASE_eGFR,

ACR_STRAT, and TRT_EXP1 as covariates, and TRT as the fixed effect. Inclusion of the fraction of expected one-year exposure to treatment (TRT_EXP1; Section 6.5.11) in the ANCOVA model as a covariate provides an adjustment for eGFR values collected from early discontinuation patients having follow-up collected after less than 1-year of drug exposure.

The ANCOVA analysis uses eGFR values collected in the Week 52 analysis window (Section 6.5.4), which includes the follow-up assessment for patients who discontinued treatment early in the first year. Missing Week 52 eGFR data are imputed as defined in Section 8.5.3 using standard imputation.

The difference between bardoxolone methyl and placebo in change from baseline of eGFR is estimated along with the 95% confidence interval at Week 52.

8.7.2. Sensitivity Analyses

Sensitivity analyses are included to assess the robustness of conclusions to the primary analysis of the Year 1 key secondary efficacy endpoint (i.e., change in eGFR Week 52), and to assess the assumption of MAR. The following sets of sensitivity analyses are summarized in tables and forest plots: tipping point with multiple imputation, control-based multiple imputation, and the ANCOVA model fit to the per protocol population. Additional sensitivity analyses may be performed as appropriate.

8.7.2.1. Tipping Point

The following tipping point sensitivity analysis is performed to assess how severe departures from MAR must be in order to overturn conclusions from the primary analysis:

1. The PROC MI procedure in SAS is applied to create 100 datasets with all missing values in the placebo group imputed based on the subjects in the placebo group. These 100 datasets will then be used for the tipping point analysis.
2. For the tipping point analysis with a shift parameter, patients in the bardoxolone methyl group with a missing value will be multiply imputed separately from the placebo-treatment arm and are assigned a shift parameter in the imputation procedure for progressively worse (lower) scores to find the point at which statistical significance is lost. Specifically, if the outcome of the hypothesis test favors bardoxolone methyl over the placebo group (p-value <0.025), then one point is subtracted as a shift parameter in the imputation for the bardoxolone methyl group, whilst the data for the placebo group will not be changed.
3. The ANCOVA model (see Section 8.7.1) is used to analyze each of the 100 completed data sets and PROC MIANALYZE is used to combine these results to obtain the final estimates for the given value of the shift parameter.
4. This “eGFR shifting”, as outlined in steps 2 and 3, is repeated with one shift point at a time until the hypothesis test no longer rejects the null hypothesis in favor of the bardoxolone methyl group over the placebo group (i.e. when the p-value becomes greater than 0.025). Additional increments may be used to locate the tipping point. If the result is not quite statistically significant, progressively higher scores are added to find the point at which statistical significance is attained.

8.7.2.2. Control-Based Multiple Imputation

As a sensitivity analysis, all missing Week 52 eGFR values are imputed with multiple imputation using the Week 52 data from the placebo group. The control-based sensitivity analysis is performed using the ANCOVA statistical model defined in Section 8.7.1.

8.7.2.3. Per-Protocol Population Analysis

The per-protocol population analysis is performed for the per-protocol population (Year 1) using the ANCOVA statistical model defined in Section 8.7.1. Missing data will not be imputed for the per-protocol population sensitivity analysis.

8.8. Year 2: Primary Efficacy Endpoint

The analyses of the primary and key secondary endpoints for the Year 2 analyses are described in Sections 8.8 and 8.9. Additional details and pseudo SAS code are provided in a separate document.

8.8.1. Primary Analysis of the Year 2 Primary Efficacy Endpoint

The Year 2 primary endpoint of this study is the change in eGFR at Week 100 (i.e., the end of the second year of exposure). The change from baseline in eGFR for patients treated with bardoxolone methyl is compared with placebo at Week 100 using the statistical model defined in Section 8.6.1 with the following modifications:

- The analysis uses all eGFR values collected through Week 100 according to analysis visits 0, 1, 2, 4, 6, 8, 12, 24, 36, 48, 64, 76, 88, and 100 (Section 6.5.4).
- TRT_EXP2 is added as a covariate and TRT_EXP1 is removed.

Inclusion of the fraction of expected two-year exposure to treatment (TRT_EXP2; Section 6.5.11) in the MMRM model as a covariate provides an adjustment for eGFR values collected from early discontinuation patients having less than 2-years of drug exposure. All eGFR values collected in the Week 52 analysis window assess the effect following 4-weeks of drug withdrawal and therefore are not included in the MMRM analysis of Week 100.

8.8.2. Sensitivity Analyses

The sensitivity analyses described in Section 8.6.2 (i.e., tipping point, control-based imputation, as-treated, and per-protocol) are performed to assess the robustness of conclusions of the 2-year primary efficacy endpoint analysis of 2-year efficacy. The per-protocol population (Year 2) is used for the per-protocol sensitivity analysis.

8.8.3. Summary of Sensitivity Analyses of the Primary Efficacy Endpoint – Year 2

The following sets of sensitivity analyses are summarized in tables and forest plots: tipping point with multiple imputation, control-based imputation, MMRM fit to the as-treated population, and MMRM fit to the per protocol population.

8.9. Year 2: Key Secondary Endpoint

8.9.1. Primary Analysis of the Year 2 Key Secondary Endpoint

The Year 2 key secondary endpoint of this study is the change in eGFR at Week 104 (i.e., 4-weeks following the last dose after two years of drug exposure). Analysis of Week 104 assesses for retained drug benefit (relative to placebo) following withdrawal of treatment; therefore, the analysis of Week 104 data does not include eGFR values collected during treatment.

The change from baseline eGFR at Week 104 for patients treated with bardoxolone methyl are compared with placebo in a manner similar to the ANCOVA analysis of Week 52 in Year 1 (Section 8.7.1) with the following modification:

- TRT_EXP2 is added as a covariate and TRT_EXP1 is removed

The analysis uses eGFR values collected in the Week 104 analysis window (Section 6.5.4), which includes the follow-up assessment for patients who discontinued treatment early in the second year. Discontinuing treatment in the second year requires that patients received at least one dose of study drug in the second year. Missing Week 104 eGFR data are imputed as described in Section 8.5.3.

The difference between bardoxolone methyl and placebo in change from baseline of eGFR are estimated along with the 95% confidence interval at Week 104.

8.9.2. Sensitivity Analyses

Sensitivity analyses described in Section 8.7.2 are performed to assess the robustness of conclusions of the 2-year key secondary efficacy endpoint. The following sets of sensitivity analyses are summarized in tables and forest plots: tipping point with multiple imputation, control-based multiple imputation, and the ANCOVA model fit to the per protocol population (Year 2).

9. EXPLORATORY EFFICACY

The following data summaries and analyses are presented for the ITT analysis set. Additional sensitivity analyses may be performed as appropriate.

All data values within the appropriate visit windows contribute to the analysis. The exception to this rule applies to any patients who begin receiving dialysis. For them, all their on-dialysis eGFR measurements are considered missing.

9.1. eGFR

9.1.1. Continuous eGFR

Results are presented in tables with nominal significance levels for descriptive purposes only.

9.1.1.1. All Available Data

Changes in eGFR for the bardoxolone methyl group compared to the placebo group are evaluated at Week 48 as the Year 1 primary endpoint, and are evaluated at Weeks 1, 2, 4, 6, 8, 12, 24, and 36 as exploratory analyses using the statistical model defined in Section 8.6.1.

Changes in eGFR for the bardoxolone methyl group compared to the placebo group are evaluated at Week 100 as the Year 2 primary endpoint, and are evaluated at Weeks 64, 76, and 88 as exploratory analyses using the statistical model defined in Section 8.8.1.

Analyses of eGFR using all available data use all eGFR values collected, irrespective of whether or not a patient is receiving treatment.

9.1.1.2. As-Treated Data

To assess the effect over time of receiving bardoxolone methyl treatment, bardoxolone methyl will be compared to placebo using the statistical models defined in Section 8.6.1 (Year 1) to analyze change from baseline in eGFR values collected while receiving treatment at the on-treatment visits for Year 1 (i.e., Weeks 1, 2, 4, 6, 8, 12, 24, 36, 48). Similarly, the statistical models defined in Section 8.8.1 (Year 2) will be used to analyze change from baseline in eGFR values collected while receiving treatment through the on-treatment visits in Year 2 (i.e., 64, 76, 88, and 100).

eGFR values collected 14 days or more after the last dose of study drug will be excluded from the as-treated analysis.

9.1.2. Categorical eGFR Endpoints

A 30% decline or eGFR < 15 is considered confirmed when the threshold is achieved at ≥ 2 visits. Patients who achieved the threshold at their last visit, but did not have a second confirmatory visit, are considered confirmed cases. Similarly, for a 30% or more increase in eGFR, a second confirmatory visit is needed with the exception of the last visit. Patients with a decrease from baseline in eGFR of ≥ 30% at the last on-treatment visit are considered to have a 30% or more decrease from baseline eGFR. In the event a patient has both an increase from baseline in eGFR of 30% or more and a decrease from baseline in eGFR of 30% or more, only the decrease from baseline is considered in composite endpoint analyses. Otherwise, if a patient

has more than one confirmed event, all events are counted in the categorical summaries. Confirmation of kidney failure components is based on all on-treatment eGFR values, irrespective of analysis windows.

The following categorical analyses of eGFR are summarized for the first year of treatment:

- The percentage of bardoxolone methyl versus placebo patients with an increase from baseline in eGFR of 30% or more after 48 weeks of treatment.
- The percentage of bardoxolone methyl versus placebo patients with a decrease from baseline in eGFR of 30% or more after 48 weeks of treatment.
- The distribution of changes from baseline in eGFR in bardoxolone methyl versus placebo patients after 48 weeks of treatment.

The following categorical analyses of eGFR are summarized for the second year of treatment:

- The percentage of bardoxolone methyl versus placebo patients with an increase from baseline in eGFR of 30% or more after 100 weeks of treatment.
- The percentage of bardoxolone methyl versus placebo patients with a decrease from baseline in eGFR of 30% or more after 100 weeks of treatment.

Newcombe's test for the common treatment difference and the corresponding confidence interval adjusting for the randomization strata for each criterion are reported.

9.1.3. Distribution of Changes in eGFR

The proportion of patients at Week 48 and Week 100 with an eGFR change from baseline of 0 mL/min/1.73/m^2 to +10 mL/min/1.73 m^2 at 1-point increments of increased or decreased eGFR (e.g., ≥ 0, $\geq \pm 1$, $\geq \pm 2$, $\geq \pm 3$, $\geq \pm 4$, $\geq \pm 5$, $\geq \pm 6$, $\geq \pm 7$, $\geq \pm 8$, $\geq \pm 9$, $\geq \pm 10$) is summarized by treatment group. The data are summarized similarly for 1-point increments of decreased eGFR to -10 mL/min/1.73 m^2.

A plot of the cumulative distribution of eGFR change from baseline at Week 48 and at Week 100 is generated. A Kruskal-Wallis rank sum test by treatment group of eGFR change from baseline at Week 48 and at Week 100 is used to examine the difference in distributions of eGFR changes between treatment groups at each of these visits.

Waterfall plots of eGFR change from baseline, labeled by treatment group at Week 48 and at Week 100 are generated. A cumulative distribution plot showing the distribution of change from baseline eGFR by treatment group will be generated at Weeks 48, 52, 100, and 104.

Each quartile of change from baseline in eGFR at Week 48, 52, 100, and 104 will be summarized descriptively.

9.2. Patient Global Impression of Change

The Patient Global Impression of Change (PGIC) is a 7-point scale that asks patients to assess how much their illness has improved or worsened relative to a baseline state at the beginning of an intervention (Guy 1976). The PGIC is assessed by completing the following statement "Since I began trial treatment, my overall status is: very much improved (1), much improved (2), minimally improved (3), no change (4), minimally worse (5), much worse (6), very much worse

(7)." All PGIC values collected at each visit using the appropriate analysis windows (Section 6.5.4) are used; the number of patients in each category is summarized at each analysis visit by treatment group.

The analysis of PGIC assesses the distribution of responses in the bardoxolone methyl vs. placebo group at each analysis visit; the distributions are compared using ANCOVA with BASE_eGFR, ACR_STRAT, TRT_EXP1 (or TRT_EXP2 for Year 2), and treatment group as covariates and treatment as the fixed effect.

9.3. Clinical Global Impression - Improvement

The Clinical Global Impression-Improvement (CGI-I) is a 7-point scale that asks the clinician to assess how much the patient's illness has improved or worsened relative to a baseline state at the beginning of an intervention (Guy 1976). The scale is assessed by completing the following statement "Compared to the patient's condition at the start of the trial, this patient's overall status is: very much improved (1), much improved (2), minimally improved (3), no change (4), minimally worse (5), much worse (6), very much worse (7)."

CGI-I is analyzed using the same methods specified for PGIC in Section 9.2.

9.4. Kidney Failure Composite Outcome

9.4.1. Primary Analysis of the Kidney Failure Composite Outcome

The kidney failure composite outcome is defined as reaching one of the following:

- Confirmed 30% decline from baseline in eGFR; or
- Confirmed eGFR < 15 mL/min/1.73 m^2; or
- ESKD (initiation of maintenance dialysis or kidney transplant).

A 30% decline or eGFR < 15 is considered confirmed when the threshold is achieved at ≥ 2 visits. Patients who achieved the threshold at their last visit, and did not have a second confirmatory visit, are considered confirmed. The analysis uses initiation of maintenance dialysis or kidney transplant as reported in the database by the investigator. Patients with a decrease from baseline in eGFR of 30% or more at the last on-treatment visit are considered to have a 30% or more decrease from baseline eGFR. In the event a patient has both an increase from baseline in eGFR of 30% or more and a decrease from baseline in eGFR of 30% or more, only the decrease from baseline is considered in composite endpoint analyses. Confirmation of kidney failure components is based on all on-treatment eGFR values, irrespective of analysis windows. The date of the first eGFR meeting the kidney failure component is used in time-to-event analyses as the date of the event.

For the Year 1 analysis, the number and percentage of patients in each treatment group having a kidney failure composite by Week 52 are summarized using only on-treatment eGFR values. The proportion of patients with a given event is evaluated using Newcombe's test accounting for ACR_STRAT. Similarly, the number and percentage of patients in each treatment group having a kidney failure composite by Week 104 are summarized for the Year 2 analysis.

For the Year 2 endpoint (Section 3.1.2.3), a time-to-event analysis of the kidney failure composite is performed using only on-treatment eGFR values. Missing eGFR data are not

imputed for time-to-event analyses. Hazard ratios and 95% CIs are computed using Cox proportional hazards regression models. Patients who do not experience a confirmed event are censored for time-to-event analysis using the following conventions:

- patients who died from any cause prior to the end of the study are censored at the date of death;
- patients who discontinued the study prematurely, withdrew consent, or were lost to follow-up prior to the end of the study are censored at their last date of contact;
- patients who are alive and still being followed are censored on the study termination (for early termination) or study completion date at Week 104.

9.4.2. Sensitivity Analysis

The primary time-to-event analysis described in Section 9.4.1 is repeated as a sensitivity analysis with all-cause death added to the kidney failure composite.

10. SAFETY AND TOLERABILITY

Safety and tolerability are evaluated by AEs, SAEs, clinical laboratory test results, body weight, vital signs, 12-lead ECG findings, physical examination, visual acuity, and audiology assessment. All analyses of the safety data are performed using the safety analysis set. Descriptive statistics are presented by treatment group assignment in the safety analysis set. No formal statistical testing is performed for safety analyses. Safety is also summarized by the subgroups defined in Section 8.3. On-treatment values are summarized according to the analysis study windows in Section 6.5.4; off-treatment values are summarized according to the analysis study windows in Section 6.5.4.1. Safety data summaries are grouped by treatment year.

Continuous safety parameters (including selected clinical chemistries and vital signs) may be summarized using the methodology described in Section 8.6.1, Section 8.7.1, Section 8.8.1, and Section 8.9.1. Only the visit windows outlined in Table 3 are included in these analyses.

10.1. Adverse Event Preferred Term and Body/Organ System Summary Tables

AEs are summarized by treatment as defined by the safety analysis set. In addition to being summarized by treatment year, AEs are also summarized overall. General considerations for AE summaries and calculations are:

- Multiple events by preferred term (PT) and system organ class (SOC) are counted once only per patient for summaries of AE incidence.
- For summaries of AE incidence by severity, only the most severe event is counted per patient.
- For summaries of AE incidence by relationship, only the most related event is counted per patient.
- An AE with a missing resolution date or incomplete date that is not identified as continuing is assumed to be continuing and no duration is calculated.
- Treatment-emergent adverse events (TEAEs) are summarized separately from off-treatment adverse events (OTAEs).
- For summaries of the number of AE events, all AEs are counted.

AEs are coded using MedDRA (Medical Dictionary for Regulatory Activities) version 19.1. In MedDRA, each verbatim term is mapped to a preferred term and high level term (HLT), which is then mapped to a system organ class. Tables and listings will present data at the SOC and PT level.

Treatment-emergent adverse events (TEAEs) are events that either:

- Have a date of onset on or after the date of the date of first dose and not more than 30 days after the date of the last dose of study drug, or
- Have no recorded date of onset with a stop date after the first dose of study drug, or
- Have no recorded date of onset or stop date

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.
The investigator grades the severity of the AEs as mild, moderate, or severe as defined in the study protocol. The investigators graded association or relatedness to the study medication according to criteria specified in Section 11.4 of the study protocol.
As defined in the protocol, a serious adverse event (SAE) is an adverse event that results in any of the following:

- Death;
- A life-threatening adverse drug experience;
- Inpatient hospitalization or prolongation of existing hospitalization;
- Results in persistent or significant incapacity or substantial disruption of the ability to conduct normal life functions;
- Results in a congenital anomaly or birth defect in an offspring of a patient taking study drug;
- Is an important medical event.

10.1.1. Missing and Partial AE Onset Dates

Rules for handling missing or partial AE Onset Dates are included in Section 6.6.3.

10.1.2. Summaries of Adverse Events

TEAEs and OTAEs are summarized by treatment based on year of AE onset, and overall. For each treatment, SOC, and PT, the number and percentage of patients reporting an event is calculated. In summary tables, SOC is presented alphabetically and events within SOC are presented by decreasing frequency count.

Summary tables (number and percentage of patients and events) of AEs (by SOC and PT) are provided by treatment as follows:

- All treatment-emergent AEs
- All treatment-emergent related AEs (definitely, probably, or possibly related)
- All treatment-emergent AEs by severity
- All treatment-emergent serious adverse events (including deaths)
- All related treatment-emergent serious adverse events (including deaths)
- All treatment-emergent adverse events leading to permanent discontinuation of study drug

Listings are provided showing:

- All AEs
- Serious adverse events (including deaths)

- AEs leading to permanent discontinuation of study drug

10.2. Exposure and Compliance

The duration of study drug exposure is defined as the number of days on treatment from the first dose of study drug until the last dose of study drug (last dose – first dose + 1). Study drug exposure is summarized by descriptive statistics. Summaries include the total dose (mg) received (based on the number of pills returned), study drug compliance, and duration (days) of exposure during the study treatment period. In addition, a summary of the number and percentage of patients by dose (placebo, 5 mg, 10 mg, 20 mg, 30 mg) and by visit is generated. If a patient received more than one dose during a visit window, the duration of the longest dose exposure is used to calculate study drug dose during a given visit window. Summaries of the number and percentage of patients by dose include a subgroup analysis by baseline ACR (ACR ≤ 300 mg/g vs. ACR > 300 mg/g).

Total number of doses dispensed and total dose (mg) dispensed is calculated from total number of kits (bottles) recorded on the Study Drug Dispensation eCRF. Total number of doses received is 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 study days of study participation, excluding the off-treatment periods between Week 48 to Week 52 and Week 100 to Week 104.

10.3. Concomitant Medications

Concomitant medications are coded using the World Health Organization (WHO) drug dictionary (Enhanced version, March 2016, B2 format) for anatomical therapeutic chemical classification (ATC) and preferred drug name. A patient who used multiple medications is counted only once for each ATC and preferred drug name. ATC and preferred drug name within each ATC are sorted alphabetically. Coded concomitant medications are summarized by treatment by WHO ATC class and preferred name. Percentages are based on the number of patients in the safety analysis set. Each summary is ordered by descending order of incidence of ATC class and preferred name within each ATC class.

A concomitant medication is any medication taken at the time of first study treatment or a medication that was started after the start of study drug dosing. Specifically, concomitant medications are medications

- that are continued from screening and continued after the first study drug dosing, or
- that have start dates or stop dates within the treatment period, or
- that have no end date.

Medications with an end date on the date of first study drug administration are not considered concomitant medications.

Patients who take excluded medications (defined in the Protocol Section 9.3.1) during the study are listed.

10.3.1. Missing and Partial Concomitant and Other Medication Start and Stop Dates

Missing and partial concomitant medication start and stop dates are detailed in Section 6.6.2.

10.4. Prior and Concomitant Procedures and Surgeries

Prior and concomitant procedures and surgeries are listed.

10.5. Clinical Laboratory Evaluations

Laboratory data are summarized at baseline and at each time point by treatment.

10.5.1. Continuous Summaries of Laboratory Results

Selected laboratory evaluations and change from baseline are summarized by treatment, laboratory category (hematology, chemistry), test, and study visit using continuous statistics. The estimated glomerular filtration rate (eGFR) results are calculated using formulas described in Section 6.5.6.

The change from baseline in select laboratory evaluations at protocol endpoint time points (i.e., Week 48, 52, 100, and 104) for patients treated with bardoxolone methyl compared to placebo may be performed using the analysis of covariance (ANCOVA) model described in Section 8.7.1 and Section 8.9.1 .

Box plots and line graphs are generated for selected laboratory test results, such as eGFR, ACR, alanine aminotransferase (ALT), aspartate aminotransferase (AST), total bilirubin, BNP, BUN, uric acid, magnesium, creatinine, and creatine kinase (CK). Line graphs include mean ± SE over time for both the observed values and for change from baseline.

Due to the nature of urinalysis parameters, summaries of continuous statistics are not provided unless otherwise specified. The one exception is the urine albumin to creatinine ratio (ACR) from the first morning void urine collection. Qualitative lab results, including urinalysis, are included in the listings, but are not summarized. Laboratory results that are above or below normal limits are flagged in the listings.

Summaries of ACR (Section 6.5.7) will use the geometric mean with 95% confidence intervals instead of the arithmetic mean and will display ACR results in original units of mg/g. Changes from baseline in ACR are reported as the post-baseline/baseline ratios and are summarized by geometric means with 95% confidence intervals at each time point. Summaries of ACR will use the geometric mean instead of the arithmetic mean. Presentations will display ACR results in the original units (mg/g). Changes in ACR are calculated as the ratio of each visit to baseline. Similar ANCOVA analyses as described in this section will be performed for ACR ratio to baseline.

Additionally, ACR is summarized by baseline ACR quartile. Descriptive summaries at each visit include the standard deviation, geometric mean of ACR, as well as the geometric mean of % change from baseline. Summaries of ACR exclude values within the Week 2 and Week 6 visit windows.

10.5.1.1. Natural log(ACR)/eGFR

Mean ratios of natural log(ACR)/eGFR (Section 6.5.10) are summarized at each time point. The ratio of natural log(ACR)/eGFR by analysis visit is summarized by arithmetic means.

10.5.2. Categorical Summaries of Laboratory Results

The number and percentage of patients by laboratory normality and abnormality categories at any time during the study (Normal, Low, High) are summarized by treatment, laboratory category (hematology, chemistry, and urinalysis), and laboratory test. The worst abnormality values are defined as the maximum values while on study treatment, with the exception of magnesium and hemoglobin. The worst abnormal values for magnesium and hemoglobin are defined as the minimum values while on study treatment.

Shift tables summarizing (1) baseline to end of treatment, (2) worst on-treatment to Week 52 (Year 1), and (3) worst on-treatment to Week 104 (Year 2), (4) worst on-treatment to worst off-treatment at Week 52 or Week 104, and (5) baseline to worst on-treatment are presented for lab parameters as appropriate.

An initial set of parameters of specific interest (ALT, AST, ACR) are summarized using shift tables, though additional parameters may be added as appropriate.

In addition, a summary table is provided for number and percentage of patients meeting a following pre-specified threshold level at any time during the study:

- ACR > 3500 mg/g
- Magnesium < 1.3 mEq/L (0.65 mmol/L)
- BNP > 200 pg/mL

10.5.2.1. 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 9. 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

Lab Parameter	Threshold
	ALT or AST > 3×ULN and Total bilirubin > 2×ULN
ALT, AST, TBL, INR	> 3 × ULN and (TBL > 2 × ULN **or** INR > 1.5)
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.

10.6. Vital Signs

Vital signs assessments include systolic blood pressure (SBP, mmHg) and diastolic blood pressure (DBP, mmHg), body temperature (°C), heart rate (HR, bpm), height (m), weight (kg), and BMI (kg/m^2). Vital signs are summarized at baseline and at each time point along with the change from baseline by treatment. Boxplots and line graphs of change from baseline over time for blood pressure are plotted.

10.6.1. Body Weight

Body weight, as collected during each in-office visit, is summarized using descriptive statistics for observed results and change from baseline at time point. In addition, number and percentage of patients experiencing a five pound (2.3 kilogram) or greater increase in weight are summarized by time point. Boxplots and line graphs of change from baseline over time for weight are plotted.

10.7. Physical Examination

Physical exam results are listed.

10.8. Electrocardiogram

Electrocardiogram (ECG) data, such as clinical interpretation of ECGs, heart rate and interval assessments of PR, QRS, and QT are collected on the eCRF. QTcF is calculated for analysis (Section 6.5.8). Descriptive statistics for observed values and change from baseline at each time point are presented for these 12-lead ECG interval assessments. In addition, number and percentage of patients with any abnormal values (i.e., above a pre-specified threshold) at any time during the study are summarized by time point and overall. The pre-specified levels of ECG QTc thresholds are consistent with FDA guidance:

Table 10. Pre-Specified Threshold Levels for ECG Parameters

ECG Parameter	Pre-Specified Level
PR	> 200 msec
QTcF	> 450 and ≤ 480 msec
	> 480 and ≤ 500 msec
	> 500

ECG Parameter	Pre-Specified Level
	Change from baseline: >30 and ≤ 60 msec
	Change from baseline: > 60 msec
Heart rate	< 40 beats/min
	< 100 beats/min

10.9. Visual Acuity

Patients with normal visual acuity are those with 20/20 vision or less (e.g., 20/15. 20/10, etc.) in both eyes. Findings from visual acuity examinations are listed. Observed logMAR value and its change from baseline are summarized descriptively by time point. In addition, shift tables are generated by treatment, visit, and normal versus abnormal visual acuity values.

10.10. Audiology Assessment

Findings from audiology assessments are listed. The average pure-tone hearing threshold for each ear is measured using the automated pure tone test. Patients with normal pure-tone hearing are those with pure tone audiometry (PTA) < 25 decibels in both ears. Any patient with ≥ 25 decibel hearing in either ear are considered to have audiology impairment. The number and percentage of patients with normal or impaired hearing in at least one ear is summarized by time point. In addition, number and percentage of patients within each of the following categories will be summarized at baseline, Week 48, and Week 100:

- PTA ≤ 25 dB
- 25 dB < PTA ≤ 40 dB (mild to moderate hearing loss)
- 40 dB < PTA ≤ 70 dB (moderate to severe hearing loss)
- 70 dB < PTA ≤ 85 dB (severe to profound hearing loss)
- PTA > 85 dB (profound hearing loss)

Descriptive statistics for observed values and change from baseline in hearing threshold for each ear (i.e., right and left) are presented.

10.11. Pregnancy

A listing is provided for serum and urine pregnancy results of all on-study pregnancies.

11. PHARMACOKINETICS

Pharmacokinetic concentration and parameter data will be summarized descriptively. A separate document contains details regarding other planned pharmacokinetic analyses.

12. CHANGES FROM PROTOCOL-SPECIFIED ANALYSES

12.1. Analysis of Key Secondary Endpoints

The protocol defines the primary analysis for the key secondary endpoints as MMRM, using the same model as the primary endpoint analysis applied to the population of patients who completed treatment. This SAP defines ANCOVA as the analysis method for key secondary endpoints using the ITT population. The rationale for this change is that eGFR values following withdrawal of drug are testing a different biological hypothesis than eGFR values while receiving drug. For this reason, it is likely that off-treatment eGFR values (i.e., Week 52 and Week 104) will have different variability and will not correlate in the same way as the on-treatment eGFR repeated measures. To guard against the risk of potential distributional differences for on-treatment as opposed to off-treatment data, the key secondary endpoints are analyzed using ANCOVA.

Because of the above described change in planned analysis method, statistical power for the key secondary endpoints is now calculated from a 2-sample t-test. As a result, the expected minimum detectable effect is changed from 2.2 (from the protocol based on analysis with MMRM and 50% power) to 2.9 (based on analysis with a 2-sample t-test and 50% power).

12.2. Imputation of Missing Data for the Primary Endpoints

This analysis plan specifies that missing data will not be imputed for the primary analysis of the primary endpoints (i.e., Week 48 and Week 100) with MMRM based on the assumption of MAR. (Section 8.5). The protocol stated that missing data for the Week 48 analysis is imputed using Jump to Reference multiple imputation, based on available data collected from patients discontinuing from study treatment but continuing in the study. The rationale for the change in imputation method is the assumption of data being missing at random. A tipping point analysis and control-based imputation are included to assess the MAR assumption as part of a sensitivity analysis.

13. REFERENCES

Guy, W. ECDEU Assessment Manual For Psychopharmacology, DHEW Publication No. ADM 76–338, US Government Printing Office, Washington, DC, USA, 1976.

KDIGO Practice Group. KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. Kidney Int Suppl 2013;3;136-150.