Document Type:	Statistical Analysis Plan
Official Title:	A randomized, double-blind, placebo-controlled, parallel-group, multicenter, event-driven Phase III study to investigate the efficacy and safety of finerenone, in addition to standard of care, on the progression of kidney disease in subjects with type 2 diabetes mellitus and the clinical diagnosis of diabetic kidney disease
NCT Number:	NCT02540993
Document Date:	14-FEB-2020

Statistical Analysis Plan for study 16244 (FIDELIO-DKD) in finerenone

Phase III study in diabetic kidney disease

Bayer study drug	BAY 94-8862 / Finerenone		
Study purpose:	Efficacy and safety		
Clinical study phase:	III	**Date:**	14 FEB 2020
Study No.:	16244	**Version:**	4.0
Authors:	PPD PPD PPD		

This Statistical Analysis Plan is produced on a word-processing system and bears no signatures.

The approval of the Statistical Analysis Plan is documented in a separate Signature Document.

Table of Contents

Abbreviations

ACEI	Angiotensin-converting-enzyme inhibitor
AE	Adverse event
AP	Asia Pacific
ARB	Angiotensin-receptor blocker
ATC	Anatomical Therapeutic Chemical
BMI	Body mass index
BNP	B-type natriuretic peptide
BOCF	Baseline observation carried forward
CEC	Clinical event committee
CHF	Chronic heart failure
CI	Confidence interval
CKD	Chronic kidney disease
CKD-EPI	Chronic Kidney Disease Epidemiology Collaboration
CLIPS	Clinical Pharmacology Standards
CSR	Clinical Study Report
CV	Cardiovascular
CV	Coefficient of Variation
CVD	Cardiovascular disease
CYP3A4	Cytochrome P450, Family 3, Subfamily A, Polypeptide 4
DAOH	Days alive and out of hospital
DKD	Diabetic kidney disease
DM	Diabetes mellitus
DMC	Data monitoring committee
DPP-4	Dipeptidyl peptidase 4
eCRF	Electronic case report form
eGFR	Estimated glomerular filtration rate
EOS	End-of-study
EQ-5D-5L	EuroQol Group 5 dimension, 5 level
ESRD	End-stage renal disease
FAS	Full analysis set
GCP	Good Clinical Practice
GFR	Glomerular filtration rate
GLP-1	Glucagon-like peptide-1
HbA1c	Glycated hemoglobin
HEOR	Health Economics, Outcomes & Reimbursement
HF	Heart failure
HFrEF	Heart failure with reduced ejection fraction
HRQoL	Health-related quality of life
ICH	International Committee on Harmonization
ITT	Intent-to-treat
IxRS	Interactive voice/web response system
KDQOL	Kidney Disease Quality of Life
LLOQ	Lower limit of quantification
LOCF	Last observation carried forward
LOS	Listing only set
LVSD	Left ventricular systolic dysfunction
MAR	Missing at random
MedDRA	Medical Dictionary for Regulatory Activities
MCS	Mental component summary
MMRM	Mixed model repeated measures
MR	Mineralocorticoid receptor
MRA	Mineralocorticoid receptor antagonist
NT-proBNP	N-terminal prohormone B-type natriuretic peptide
NYHA	New York Heart Association
N/A	Not applicable
OD	Once daily

PCS	Physical component summary
PD	Premature discontinuation
PDD	Protocol deviation document
PK	Pharmacokinetic
PKS	Pharmacokinetic analysis set
PPS	Per-protocol set
PRO	Patient-reported outcome
PT	Post-treatment
RAS	Renin-angiotensin system
SAE	Serious adverse event
SAF	Safety analysis set
SAP	Statistical Analysis Plan
SD	Standard deviation
SF-12	Short form 12
SGLT	Sodium-glucose transport proteins
SMQs	Standardized MedDRA Queries
SoC	Standard of care
T2DM	Type 2 diabetes mellitus
TEAE	Treatment-emergent adverse event
TLF	Tables, Listings and Figures
UACR	Urinary albumin-to-creatinine ratio
UAE	Urinary albumin excretion
ULOQ	Upper limit of quantification
WHO-DD	World Health Organization Drug Dictionary

1. Introduction

This is a Statistical Analysis Plan (SAP) for study 16244 (FIDELIO-DKD) in finerenone.

It is based on the following document(s):

Integrated Clinical Study Protocol 16244 (DKD) version 3.0 dated 26 FEB 2019

Amendment 1 (local, Japan), dated 18 AUG 2015

Assessment Criteria Identification Requirement Document for the phase III DKD studies 16244 and 17530 version 4.0 dated 27 AUG 2019

DMC Charter version 3.0 dated 11 OCT 2016

CEC Charter version 2.0 dated 23 NOV 2016

This SAP describes the statistical analysis of the period from informed consent to randomization, the double-blind placebo-controlled study treatment and post-treatment follow-up phase (expected duration 3 years or more) and the 4 week period up to the post-treatment visit (if applicable). One formal interim analysis is currently planned. An independent data monitoring committee (DMC) will be involved in the review of data for safety and efficacy as described in the DMC Charter. Blinded adjudication of clinical outcomes will be performed by an independent Clinical Event Committee (CEC), as described in the CEC Charter.

Background

Finerenone (BAY 94-8862) is a next-generation, oral, selective, non-steroidal MRA. In animal models, finerenone reduced cardiac and renal hypertrophy, plasma prohormone of brain natriuretic peptide (BNP) and proteinuria more efficiently than in those treated with the steroidal mineralocorticoid receptor antagonist (MRA) eplerenone, when comparing equi-natriuretic doses. Finerenone's tissue distribution pattern in rats was found to differ from the steroidal MRAs, i.e. spironolactone and eplerenone, which showed a higher accumulation of the drug equivalent concentration in kidney than in heart tissue, in contrast to finerenone which was found to be equally distributed in both the kidney and heart tissue. Steroidal MRAs are known to interfere with the steroid hormone receptor, which can cause sexual side effects such as gynecomastia in men. However, finerenone is a non-steroidal and selective MRA in vitro, without any detectable affinity for the related androgen receptor; sexual side effects are therefore not expected to occur with finerenone at therapeutic dose levels. These characteristics may translate to a drug with a positive benefit-risk ratio in patients with DKD.

Treatment with renin-angiotensin system (RAS) inhibitors has been of particular interest given the ability of these drugs to reduce the rate of progression of renal disease, independently and in addition to their antihypertensive effects. Therefore, angiotensin-converting-enzyme inhibitors (ACEIs) and angiotensin-receptor blockers (ARBs) can be considered as standard of care (SoC) therapy in patients with DKD. During the run-in period of the phase III DKD studies, the subject's SoC therapy including treatment with ACEIs or ARBs will be optimized.

Finerenone has been investigated in phase II for two indications that are planned to proceed into phase III: diabetic kidney disease (DKD) and chronic heart failure (CHF).

Study 14563 was a phase IIa study, where finerenone was studied in individuals with stable HFrEF and mild (estimated glomerular filtration rate [eGFR] 60 to < 90 mL/min/1.73 m2) -to-moderate (eGFR 30–60 mL/min/ 1.73 m^2) chronic kidney disease (CKD).

Study 16243 was a Phase IIb multicentre, randomized, double blind, placebo-controlled, parallel-group study investigating the safety and efficacy of several oral doses of finerenone in patients with type 2 diabetes mellitus (T2DM) and a clinical diagnosis of DKD, based on either high albuminuria (urinary albumin-to-creatinine ratio [UACR] 30 to 299 mg/g) or very high albuminuria (UACR 300 to 3000 mg/g), and with an eGFR of 30 mL/min/1.73 m^2 or more.

Study 14564 was a Phase IIb randomized, adaptive, double-blind, double-dummy, comparator-controlled, parallel-group, multi-center study, that investigated the safety and efficacy of finerenone in subjects with worsening CHF and left ventricular systolic dysfunction (LVSD) and either T2DM with/without CKD, or moderate CKD alone.

Diabetic kidney disease

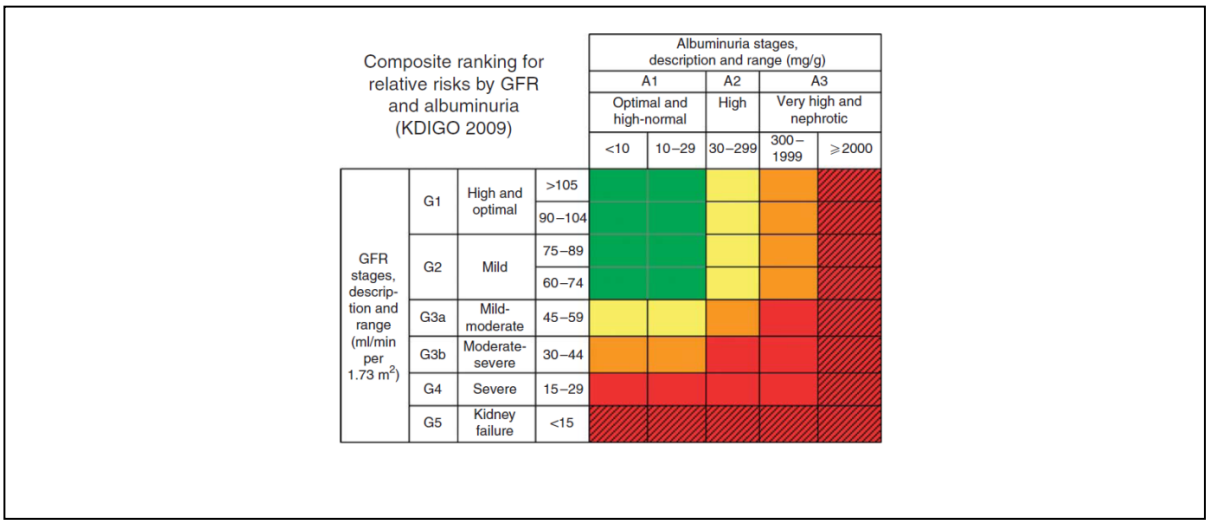
DKD is the most frequent cause of end-stage renal disease (ESRD) in western countries. In addition, the risk of cardiovascular (CV) disease and death increases in patients with DKD with decreasing glomerular filtration rate (GFR) and increasing albuminuria levels.

As the T2DM population rapidly grows throughout the world within the next years, and with it the DKD population, there is an increasing need for new therapeutic agents that effectively target underlying disease mechanisms and slow or halt the progression of kidney disease, whilst also addressing the high CV morbidity and mortality in this population.

DKD is a clinical syndrome affecting individuals with diabetes that is characterized by albuminuria on at least 2 occasions separated by 3 to 6 months. DKD is usually accompanied by hypertension, progressive rise in proteinuria, and decline in renal function. DKD is categorized based on the level of urinary albumin excretion (UAE), either high albuminuria (UACR 30 to <300 mg/g) or very high albuminuria (UACR ≥300 mg/g).

Increased UAE and decreased GFR are both associated with a worse prognosis (Figure 1–1), with an increase in all-cause and CV mortality. Additionally, increased UAE and decreased GFR are also risk factors for ESRD, acute kidney injury and progressive CKD, independent of each other and of other cardiovascular disease (CVD) risk factors in general as well as in high-risk populations.

Figure 1–1 Prognosis of CKD by GFR and Albuminuria Categories



In study 16244, finerenone will be investigated in patients at high risk to develop end-stage renal disease and will therefore focus on renal endpoints. In study 17530, patients at high risk to develop CVD will be enrolled, and here the primary endpoint will be cardiovascular events.

2. Study Objectives

The **primary objective** of this study is to:

- Demonstrate whether, in addition to standard of care (SoC), finerenone is superior to placebo in delaying the progression of kidney disease, as measured by the composite endpoint of time to first occurrence of kidney failure, a sustained decrease in eGFR of ≥40% from baseline over at least 4 weeks, or renal death.

The **secondary objectives** of this study are to determine whether, in addition to SoC, finerenone compared to placebo:

- Delays the time to first occurrence of the following composite endpoint: cardiovascular (CV) mortality and morbidity in subjects with T2DM and the clinical diagnosis of DKD
- Delays the time to all-cause mortality
- Delays the time to all-cause hospitalization
- Reduces UACR from baseline to Month 4
- Delays the time to first occurrence of the following composite endpoint: onset of kidney failure, a sustained decrease of eGFR ≥ 57% from baseline over at least 4 weeks, or renal death.

3. Study Design

The primary and secondary objectives in this phase III study are mainly based on time to specified events. The key objectives are to delay the time to the first event of a composite or single type of event.

This is a randomized, double-blind, placebo-controlled, parallel-group, multicenter, and event-driven study.

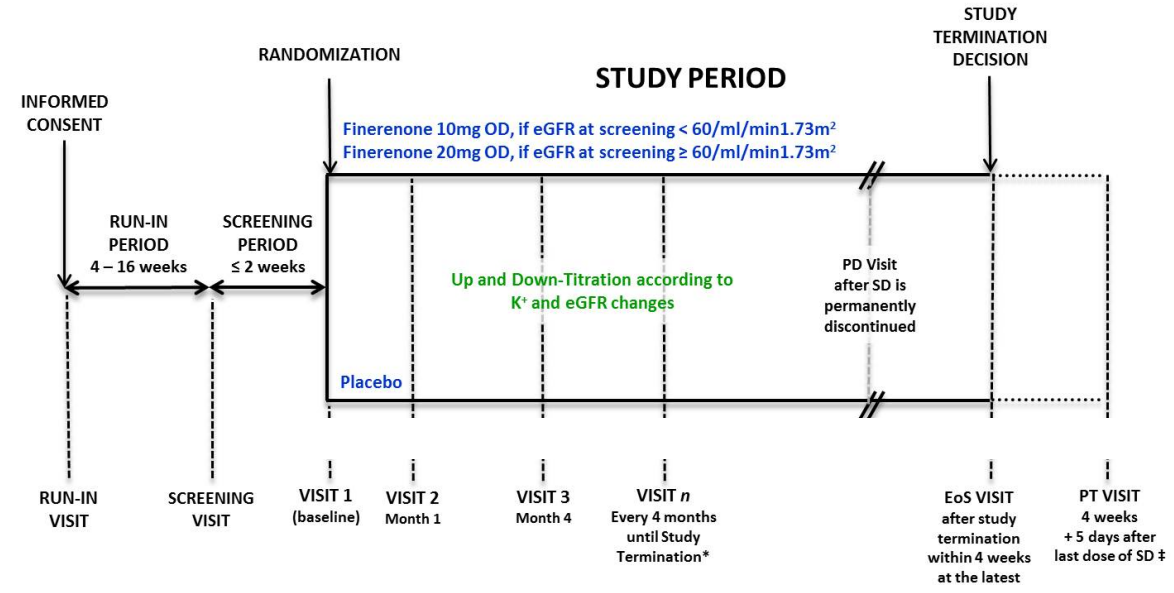
A total of 1068 primary efficacy endpoint events will have a minimum 90% power to demonstrate superiority of finerenone to placebo using a logrank test at a two sided significance level of 3.3333%, assuming a 20% relative risk reduction, i.e. a true hazard ratio of 0.80 (the hazard ratio that will be observed in the study will be different, i.e. closer to 1 due to treatment discontinuations). The study will stop when at least 1068 primary efficacy endpoints accrue across both treatment arms.

With an assumed study duration of 44 months (duration of recruitment period: 33 months, equal recruitment pattern during the accrual period, maximum treatment period of the last subject recruited: 11 months), the planned total number of subjects to be randomized is estimated to be 4690 subjects, assuming an annual placebo event rate of 12%, a common annual lost to follow-up rate of 0.7% in both treatment groups, an annual finerenone discontinuation rate of 5% and assuming that placebo discontinuation will not change the hazard. 4800 subjects were originally planned to be randomized taking a certain ramp-up during recruitment into account. Assuming a screening failure rate of 50%, 9600 subjects were planned to be screened. Due to the observation of a slower-than-assumed event rate during the study conduct phase, the required number of randomized subjects was increased to approximately 5800.

As all randomized patients (except for cases with critical GCP violations) will belong to the Full Analysis Set (FAS) and be part of the efficacy analyses, it is important to avoid randomization of non-eligible patients into the study.

The general study design as applied to this study is shown in Figure 3–1. There is a run-in period and screening period, a double-blind treatment period and a safety follow-up period. Patients prematurely terminating from the study and up to the primary study completion will be asked to attend scheduled visits to collect efficacy data.

Figure 3–1 Overall study design



* Scheduled visits will continue even if treatment with SD is discontinued

‡ For all subjects still on treatment with SD at the EoS Visit

K+ = Blood Potassium
SD = Study Drug
OD = once daily
PD = Premature Discontinuation
EoS = End of Study
PT = Post-treatment

Run-in Period (up to 16 weeks)

Subjects with written informed consent who complete the Run-in Visit and meet all eligibility criteria will be enrolled into a mandatory Run-in Period, the purpose of which is to ensure that the subject's SoC therapy including treatment with ACEIs or ARBs is optimized and that all inclusion and exclusion criteria are met at the Screening Visit. The Run in Period will last for a minimum of 4 weeks and a maximum of 16 weeks.

In the absence of documentation of diagnosis of high or very high albuminuria, the subject may still be enrolled into the study as albuminuria will be measured at the Run in Visit; in this case, the Run in Period must last for a minimum of 12 weeks and a maximum of 16 weeks.

Screening Period (up to 2 weeks)

At the end of the run-in period, a screening visit to confirm the subject's eligibility will take place within ≤14 days prior to the planned randomization. At this visit, it will be assessed whether the subject still meets all the inclusion and none of the exclusion criteria.

For those subjects without a prior documented diagnosis of high or very high albuminuria, the Screening Visit must be performed at least 12 weeks after the Run-in Visit, and albuminuria should then be re-evaluated. If the subject still suffers from high or very high albuminuria whilst on SoC treatment, and fulfills all other eligibility criteria, she/he can be randomized into the study.

Treatment period

Eligible subjects will be randomized to receive once daily (OD) oral doses of finerenone (10 mg and 20 mg) or placebo in addition to their SoC therapy.

There will be up to 4 planned visits (including randomization at Visit 1) in the first 4 months, thereafter visits will take place every 4 months until the end of the study. Study drug dose can be up-titrated from Visit 2 (Month 1) onwards or down-titrated at any point (even between scheduled visits). Subjects may be seen at any time throughout the study, in addition to scheduled visits, at the discretion of the investigator.

All randomized patients, including any patient who has experienced a health event considered for the pre-specified primary or secondary endpoints, should continue to receive study drug until the trial is completed provided there are no safety grounds for discontinuing treatment.

If the study drug is temporarily discontinued, it should be re-introduced as soon as medically justified in the opinion of the investigator. Any changes in the study drug dose, including interruption/permanent discontinuation or restart of study drug, must be recorded in the electronic case report form (eCRF).

It is planned that all randomized subjects will remain in the trial until either:

a. an instruction is received from the sponsor after the projected number of primary endpoints is positively adjudicated, or

b. the trial is terminated prematurely at the recommendation of the independent DMC.

Discontinuation of study drug (for any reason) does not constitute the subject's withdrawal from the study, except if the investigator believes that it is in the best interest of the subject or if the subject withdraws consent.

Follow-up period

The period between the subject's last intake of study drug and last visit in the study is referred to as the "Follow-up Period". If a subject withdraws from study drug permanently but does not withdraw from the study, this would apply to the period between the Premature Discontinuation (PD) Visit, which should take place as soon as possible following permanent discontinuation of study drug, and the End-of-Study (EOS) Visit. In the case where a subject discontinues treatment at the EOS Visit, the Follow up Period applies to the period between the EOS visit and Post-treatment (PT) visit.

During this period, subjects are expected to continue to attend all protocol specified study visits, and should be encouraged to perform all assessments (with the exception of local laboratory assessments) as described in the visit schedule.

All subjects who withdraw consent can be followed up for vital status if they do not sign the *'Declaration of Objection'* form. In addition, vital status can be obtained by the investigator from publicly available data sources. The collection of vital status must be obtained within the timelines provided by Bayer.

End of study

The end of the study as a whole will be reached as soon as the last visit of the last subject of the study has been reached in all centers in all participating countries.

After notification of study termination, for all subjects still participating in the study, an EOS Visit should be scheduled as soon as possible (but within 4 weeks at the latest) to determine whether the subject had an event for inclusion in the primary or secondary endpoints.

Subjects who are still on treatment will discontinue study drug treatment at the EOS Visit and must perform the PT Visit (4 weeks + 5 days) after their last dose of study drug.

Subjects who are no longer taking study drug must also be contacted as soon as possible after issue of the notification of study termination and asked to attend the EOS visit.

4. General Statistical Considerations

4.1 General Principles

The statistical evaluation will be performed by using the software package SAS release 9.4 (SAS Institute Inc., Cary, NC, USA).

The analysis will be based on the Global Standard Tables (Version 3.0 or higher) and the Clinical Pharmacology Standards (CLIPS) (Version 1.2 or higher) where appropriate.

The validity of subjects for allocation to various analysis sets will be assessed in an ongoing manner in several validity review meetings and decisions will be documented in the validity review reports prior to unblinding (potential conditional validity for per-protocol set [PPS] and pharmacokinetic analysis set [PKS]). This SAP may be updated based on the results of the validity review meetings.

Only adjudicated outcome events will be used for all efficacy analyses except for that specified in section 6.2.3.6.

A log-normal distribution is assumed for serum creatinine and UACR. For all other metric variables, a normal distribution is assumed. The distributional assumptions will be

investigated and if necessary, nonparametric methods or transformation of the data will be considered.

All variables will be analyzed by descriptive statistical methods. The number of data available, mean, standard deviation (SD), minimum, lower quartile, median, upper quartile, and maximum will be calculated for metric data. The geometric mean and SD will be provided instead of the arithmetic mean and SD for the variables where log-normal distributions are assumed. Frequency tables will be generated for categorical data.

The laboratory parameter eGFR will be calculated based on the CKD-EPI formula for all analyses specified in this SAP.

All subjects will be analyzed according to the planned treatment group in FAS (the intent-to-treat or ITT principle). All subjects will be analyzed according to the actual treatment in the safety analysis set (SAF). If a subject receives both treatments due to a bottle error, the treatment actually received for the majority of the time in the study will be used in SAF.

The stratified analyses mentioned in this SAP will be conducted in consideration of the randomization stratification factors region, eGFR category and type of albuminuria unless specified otherwise, see also section 4.5.4. The strata variables type of albuminuria and eGFR category used in the statistical analysis will be based on the screening UACR and eGFR assessments and corrected via the interactive voice/web response system (IxRS) if necessary. All subjects will be analyzed according to their correct stratification category. If screening values are not available, the run-in values will be used. If a subject's actual UACR / eGFR combination results in a stratification group that does not exist in the study (e.g. in 16244, very high albuminuria and eGFR ≥75 mL/min/1.73m^2) then the stratification group the investigator originally specified in IxRS will be kept.
In case of stratification errors, the primary analysis will also be repeated based on the stratification category used in the randomization as a sensitivity analysis.

4.2 Handling of Dropouts

A subject who has been randomized and discontinues study participation prematurely for any reason is defined as a "dropout", even if no study drug has been taken. Dropouts will not be replaced.

Data from subjects who prematurely terminated the study will be used to the maximum extent possible.

The number of subjects discontinuing the epochs, together with the primary reason for discontinuation, will be summarized as described in section 6.1.1.

The number of subjects who prematurely discontinue the study and/or study treatment for any reason, as well as the reasons for premature discontinuation of study and/or study treatment, will be reported. Kaplan-Meier plots for "Time to end of study" and "Time to end of study treatment" will be provided.

All dropouts and subjects prematurely discontinuing study treatment will be evaluated with respect to

- baseline characteristics
- potential differences between the treatment groups in the proportion of patient withdrawals or in the timing of withdrawals

- the reasons for premature discontinuation of study and/or study treatment.

4.3 Handling of Missing Data

All missing or partial data will be presented in the subject data listing as they are recorded on the Case Report Form (CRF).

General rules

When appropriate, the following rules will be implemented so as not to exclude subjects or observations from statistical analyses due to missing or incomplete data.

Concomitant medications with missing start and stop date but flagged as being ongoing at end of study will be considered to have started prior to study medication start and end after stop of study medication. The start and end reference period will be imputed as "before" for the medication start and as "during/after" for the medication end.

In case of (partially) missing dates for interruptions or permanent stop of study medication intake, a 'worst-case' approach will be applied to impute the start and end dates of study medication intake as the minimum and maximal possible dates, i.e.:

- first month of the year, or first day of the month for a partially missing start date, and
- last month of the year, or last day of the month for a partially missing end date.

If a subject died earlier than the imputed worst study medication end date, the death date will be taken as the study medication end date. However, if these imputations lead to a temporal overlap between different exposure date records, the imputed dates will be adjusted so that no overlap exists and the time on the higher dose is maximized. The date of first exposure to treatment is not expected to be missing as the patients are instructed by the investigator to take their first dose of study drug directly at Visit 1, but in the very rare case that this date is not recorded, it will be imputed according to the rules outlined above for missing start dates, but not earlier than the randomization date.

When only partial dates are available for clinical events in the efficacy analysis, a median imputation rule will be used:

- For example if the day is missing and the month is July, then day 16 is chosen.
- If the number of potential values is even, the lower of the 2 middle numbers is taken. For example, if the day is missing and the month is June, then day 15 is imputed. The same rule applies if the day and month are missing, e.g. if the year is 2017 and the day and month are missing, 2nd July is taken.
- In case the range of possible values is further restricted, e.g. because a patient died in the month in which the day is missing, the median in the restricted set of possible values is calculated. For example, if the clinical event occurred in June 2017 and the respective patient died on 11th June 2017, 6th June 2017 is imputed as the date of the clinical event.
  The same principle also applies to partially missing death dates, e.g. if only the year of death is available, but there is a last contact in the given year, then the date of death is imputed as the median of this date and the end of the given year.

In case a death date is completely missing, it will be imputed on the basis of the last known contact when the subject was still alive and the first known contact when the subject was dead

(e.g. from the subject health status follow-up page) as the median of these two dates. As above, if the number of potential values is even, the lower of the two middle numbers is taken.

In case both a non-fatal clinical event and death have partially missing dates, then death takes precedence and will be imputed first according to the rules outlined above. This also applies for non-renal and non-CV death.

However, given the importance of an accurate determination of the adjudicated event date in relation to randomization date for the time to event analysis, we would expect a minimal number of such missing dates.

For the analyses of the secondary endpoint UACR at Month 4, the closest post-baseline set of measurements to day 120 (+/-30days) will be taken as Month 4. If there is no set of measurements within this time frame, the subject will be excluded from the primary analysis of UACR at Month 4.

A worst case approach will be applied for determining whether an AE with partially missing dates is treatment-emergent or not, i.e. if it is possible that the adverse event start date is within a period of study drug intake +3 days, then the adverse event is considered treatment-emergent.

If intensity of the AE is missing, the event will be considered as severe. If the same event is reported as both unrelated and related to the study drug within a subject, the event will be reported as related to study drug. If the drug relationship is missing, the event will be considered as being related to the study drug.

4.4 Interim Analyses and Data Monitoring

One formal interim analysis is planned when 2/3 of the required total number of primary efficacy endpoint events have been observed.

If the interim analysis shows clear and consistent benefit in the finerenone treatment group, the DMC may recommend early study termination. The Haybittle-Peto rule will be used to guide the decision regarding early stopping of the study for success: a reduction of 3 standard deviations in the analysis of the primary and the key secondary efficacy endpoint at the interim analysis (two-sided p-value <0.00270).

For a detailed description of the testing procedure, including an adjustment of the significance level for the interim analysis, please see section 4.7.

In case of an early stop of the study for success, the final analysis reported in the CSR will consider all events up to the subjects' respective EoS visits as would have been the case if the study had not stopped early. A sensitivity analysis will take all events up to the interim analysis datacut date into consideration and will be based on the same final clean database as done for the final analysis reported in the CSR (instead of the interim analysis database). For this sensitivity analysis, censoring dates after the datacut date will be reset to the datacut date.

For a lack of efficacy, a non-binding futility approach will be utilized at the time of the planned interim analysis. If the conditional probability of rejecting the null hypothesis for the primary comparison, given the assumed and current event rates, falls to an unacceptably low level (as specified in the DMC Charter), the DMC may consider recommending early termination of the study.

A detailed plan for the routine DMC safety analyses and the interim analysis is covered in the DMC charter, the analysis planned to be provided to the DMC is described in a separate DMC SAP with Tables, Listings and Figures (TLFs) attached to the DMC charter. The DMC will review the data in an unblinded manner, both for the routine safety tables and the interim analysis. There are no predefined stopping conditions for the ongoing safety monitoring of this trial. The statistical analysis for the DMC meetings will be performed by an independent statistical analysis center.

4.5 Data Rules

General data rules are described in this section, further data rules for specific parameters or analyses are specified in the respective subsections of section 6.

4.5.1 Baseline values

Baseline values will be defined as the last non-missing measurement before or on the day of randomization. If the last observation available prior to randomization is the measurement from the Screening Visit, this would be used as the baseline value. This also includes assessments from a local laboratory, in case that prior to randomization, no assessment from the central laboratory is available. Otherwise baseline will be missing. The measurement from the Run-in Visit will not be used as a baseline value unless otherwise specified.

If more than one measurement was planned for a scheduled visit, for example blood pressure measurements and heart rate, the mean value of these measurements per time point prior to randomization will be used as the baseline value.

4.5.2 Change from baseline

Change from baseline will in general be displayed as absolute change from baseline defined as the difference to baseline, i.e.:

Absolute change = Post baseline value – baseline value.

Some parameters will be additionally analyzed as relative change defined as

Relative change = 100 * [(post baseline value – baseline value) / baseline value].

For specific analyses, the relative decrease of a variable will be analyzed instead of the relative change. The relative decrease is equivalent to the negative of the relative change and defined as

Relative decrease = 100 * [(baseline value – post baseline value) / baseline value].

4.5.3 Other data handling

Only the data provided by the central laboratory will be used for analysis, values from local laboratories will not be used in the statistical analysis and will be listed only unless there are no values from the central laboratory available for baseline.

In case of retests for Run-in and Screening visits, the closest measurement prior to randomization will be used for analysis instead of the scheduled measurements. At all visits starting at Visit 1 and if not stated otherwise, only the values at scheduled measurements will be used for analysis. In the unexpected case that more values than planned are available for a scheduled visit, only the first valid value will be used.

For the derived visit "Any time post-baseline" (applicable for efficacy) this will include any measurement after date of randomization, including unscheduled assessments. For the derived visit "Any treatment-emergent" (applicable for safety), only assessments on or after study medication start date until 3 days after the date of any temporary or permanent interruption of study drug, including unscheduled assessments, will be considered.

For those laboratory values which are <LLOQ (Lower limit of quantification), half the value of the LLOQ will be used for analysis. Differences between 2 values <LLOQ will be assigned values of 0. Ratios between 2 values <LLOQ will be assigned a value of 1. For values which are >ULOQ (Upper limit of quantification), the ULOQ will be used for analysis. This rule does not apply to the finerenone plasma concentration data.

In case of log-normally distributed data, descriptive statistics other than minimum, maximum and median will only be calculated if at least 2/3 of the individual data were measured and were above the lower limit of quantification. In tables showing descriptive statistics, where values below LLOQ are included, these descriptive statistics will be marked.

Urinalysis is assessed 3 times per visit. UACR will be determined 3 times at each visit from first morning void urine samples collected on 3 consecutive days. For the analyses of UACR, the 3 measurements at one visit will be combined provided that at least 2 measurements are available. First, the coefficient of variation (CV) will be calculated for the 3 values as follows considering the log-normal distribution for UACR [7]:

$$CV = \sqrt{exp\left(SD_{ln}{}^{2}\right) - 1},$$

where SD_{ln} is the standard deviation of the 3 log-transformed UACR values. If the coefficient of variation exceeds 25%, the median from the measurements will be used for the analyses. The median in case of an even number of values will be defined as the geometric mean from the 2 middle values. If the coefficient of variation is 25% at the most, the geometric mean will be used. If 2 or 3 scheduled assessments are missing or invalid and unscheduled assessments have been performed, then the set of unscheduled measurements closest to the planned visit will be used to determine the UACR at the respective visit if performed within 30 days before or after the planned time point for all visits from Visit 3 (Month 4) onwards (with the following exception: only samples with 2/3 of values no later than 30 days (for "last on-treatment" analysis) or 3 days (for "last treatment-emergent" analysis) after last study drug intake will be considered). For Visit 1, 2 out of 3 samples have to be on or before the day of randomization and have to lead to non-missing values in order for the set of samples to be valid for baseline calculation; otherwise (also including the case when there is one valid sample at Visit 1), the set of screening samples will be considered for baseline.

In the case of ESRD, the UACR and eGFR values after the onset date of ESRD will be excluded from all efficacy analyses, but included in the safety analyses.

For the stratification group type of albuminuria at screening (high or very high), the classification supplied from the central lab via IxRS will be used. The rule that has been applied by the central lab for this classification is, if there are 3 values, the median value should be taken; if 2 values, then the arithmetic mean value, regardless of degree of variation between the values. The lab reports from the run-in and screening visits will include the type of albuminuria, to assist the investigators in reducing stratification and randomization errors.

4.5.4 Subgroup analyses

Exploratory subgroup analysis will be performed for the primary and secondary efficacy variables. The subgroup analyses will include the randomization stratification factors. The list of key subgroups (in addition to the stratification factors) and other subgroups analysed is specified below. Analysis will include descriptive statistics, graphical display of estimated treatment effects with 95%-CIs in a forest plot and a statistical test for interaction.

Stratification factors:

- Region (North America, Latin America, Europe, Asia, Others)
- eGFR category at screening (eGFR 25 to <45, 45 to <60 and ≥60 mL/min/1.73m^2)
- Type of albuminuria at screening (high albuminuria, very high albuminuria)

Persistent high albuminuria is defined as UACR of ≥30 mg/g (≥3.4 mg/mmol) but <300 mg/g (<33.9 mg/mmol) in 2 out of 3 first morning void samples. Persistent very high albuminuria is defined as UACR of ≥300 mg/g (≥33.9 mg/mmol) in 2 out of 3 first morning void samples.

Key subgroups (in addition to stratification factors)

- History of CVD (present, absent)
- Sex (male, female)
- Race (white, black, Asian, other)
- Age at run-in visit (<65, ≥ 65 years)
- eGFR category at baseline (eGFR <25, 25 to <45, 45 to <60 and ≥60 mL/min/1.73m^2)
- Type of albuminuria at baseline (normalbuminuria (UACR <30 mg/g), high albuminuria, very high albuminuria)
- Baseline serum potassium value (≤ median and > median in the FAS)
- UACR at baseline (≤ median and > median in the FAS)
- Systolic blood pressure at baseline (≤ median and > median in the FAS)
- Baseline BMI (<30, ≥30 kg/m^2)
- Hemoglobin A1C (≤ 7.5% / > 7.5%)
- SGLT-2 inhibitors treatment at baseline (yes, no)
- GLP-1 agonists treatment at baseline (yes, no)

Other subgroups

- Baseline serum potassium value (≤4.5, >4.5 mmol/L)
- Baseline serum potassium (by quartiles in the FAS: ≤ Q1, >Q1 and ≤Q2, >Q2 and ≤Q3, >Q3)
- Baseline serum potassium value (< 4.8 mmol/L, ≥4.8 to 5.0 mmol/L, > 5.0 mmol/L)
- Baseline hemoglobin A1C (by quartiles in the FAS: ≤ Q1, >Q1 and ≤Q2, >Q2 and ≤Q3, >Q3)
- Baseline C-reactive protein (by quartiles in the FAS: ≤ Q1, >Q1 and ≤Q2, >Q2 and ≤Q3, >Q3)

- Systolic blood pressure at baseline (<130 mmHg, 130 to <160 mmHg, and ⩾160 mmHg)
- Age at run-in visit (18 to 44 years, 45 to 64 years, 65 to 74 years, 75 years and over)
- Ethnicity (hispanic or latino, not hispanic or latino, not reported)
- Baseline BMI (<20, 20 to <25, 25 to <30, 30 to <35, ≥35 kg/m2)
- Baseline weight (<60, 60 to <90, ≥90 kg)
- eGFR at baseline 25 to < 45 ml/min/1.73m² and baseline serum potassium value >4.5 mmol/L (yes, no)
- ACEI at baseline (yes, no)
- ARB at baseline (yes, no)
- Beta-blocker at baseline (yes, no)
- Diuretic at baseline (yes, no)
- Statins at baseline (yes, no)
- Other anti-diabetic treatment at baseline: insulin and analogues (yes, no); DPP-4 inhibitors (yes, no); biguanides (yes, no); sulfonylureas (yes, no); alpha glucosidase inhibitors (yes, no); meglitinides (yes, no); thiazolidinediones (yes, no)
- Potassium supplementation at baseline (yes, no)
- Potassium lowering agents (including binders) at baseline (yes, no)
- Potency of concomitant CYP3A4 inhibitor medication at baseline (strong, weak, moderate, unclassified, none)
- Potency of concomitant CYP3A4 inducer medication at baseline (strong, weak, moderate, unclassified, none)
- Baseline waist circumference (normal [men <94cm, women<80cm], increased [men 94-102cm, women 80-88cm], substantially increased [men >102cm, women >88cm])

If the result for a subgroup cannot be calculated due to a small sample size or number of events, then this subgroup will be dropped.

Individual country analyses, e.g. for Japan, required for regulatory purposes, will be included in a country-specific study SAP.

For the safety variables

- Number of subjects with hospitalization for hyperkalemia
- Number of subjects discontinuing study drug permanently due to hyperkalemia
- Number of subjects with hospitalization for worsening of renal function
- Number of subjects discontinuing study drug permanently due to worsening of renal function

Exploratory subgroup analyses will be performed in the following selected subgroups:

- eGFR category at baseline (eGFR 25 to <45, 45 to <60 and ≥60 mL/min/1.73m2)

- Baseline serum potassium value (≤4.5, >4.5 mmol/L)
- Anti-diabetic treatment at baseline: DPP-4 inhibitors (yes, no); GLP-1 agonists (yes, no); SGLT-2 inhibitors (yes, no)
- Potency of concomitant CYP3A4 inhibitor medication at baseline (strong, weak, moderate, unclassified, none)

4.6 Validity Review

The results of the validity review meetings will be documented in the validity review reports and may comprise decisions and details relevant for statistical evaluation. Any changes to the statistical analysis prompted by the results of the validity review meetings will be documented in an amendment and, if applicable, in a supplement to this SAP.

4.7 Testing procedure and multiplicity adjustment

If the decision will be to stop the study early for success (see section 4.4 for a description of the stopping rules), the final primary analyses, both for the primary and secondary efficacy endpoints, will be performed at a two-sided significance level of 0.270%. Only if both the primary and the key secondary efficacy endpoint achieve this significance level, the corresponding null hypotheses can be rejected and the remaining secondary endpoints will also be tested at a two-sided level of 0.270% according to the following hierarchy:

- Time to all-cause mortality
- Time to all-cause hospitalization
- Change in UACR from baseline to Month 4
- Time to first occurrence of the following renal composite endpoint: onset of kidney failure, a sustained decrease in eGFR of ≥ 57% from baseline over at least 4 weeks or renal death.

If the decision will be not to stop the study early for success, the final primary analyses will be performed as follows: A group sequential design with a single interim analysis when 2/3 of the information is available with a stopping rule of two-sided $p < 0.00270$ requires a small adjustment to the final analysis that is slightly lower than $p = 0.05$ to maintain the overall significance level at 5%. The weighted Bonferroni-Holm procedure will be used for the hierarchical testing of the primary and secondary efficacy endpoints with the following adjusted alpha levels, which apply for an information fraction of 2/3:

- If the primary renal composite endpoint achieves statistical significance at a two-sided p value ≤ 0.03282695, the secondary CV endpoint will be tested at the two sided 0.04967388 level.
- Alternatively, if the secondary CV endpoint achieves statistical significance at a two sided p value ≤ 0.01576184, the primary renal composite endpoint will be tested at the two-sided 0.04967388 level.
- Only if both the renal and CV endpoints achieve formal statistical significance, the remaining secondary endpoints will be tested at a two-sided level of 0.04967388 according to the hierarchy above.

According to Tang and Geller (1999) [9] this procedure controls the overall familywise error rate at 5%.

If the testing strategy stops at one point due to a non-significant result, the testing of the remaining secondary efficacy variables will be performed in an explorative manner only.

5. Analysis Sets

Subjects will be excluded from all below-mentioned analysis sets if they are related to critical GCP violations.

5.1 Safety analysis set (SAF)

All randomized subjects who have taken at least 1 dose of study drug (with the exception described in Section 5).

5.2 Full analysis set (FAS)

All randomized subjects (with the exception described in Section 5).

5.3 Per-protocol analysis set (PPS)

All subjects of the FAS without any of the following validity findings:

- The subject is younger than 18 years (the lower age limit may be higher if legally required in the participating country)
- Subject without type 2 diabetes mellitus
- Subject with no clinical diagnosis of DKD (no persistent high/very high albuminuria at run-in or screening visit (UACR<30mg/g (3.4 mg/mmol)).
- Subject with clinical diagnosis of DKD and persistent high albuminuria, but eGFR < 25 mL/min/1.73 m² or ≥75 mL/min/1.73 m² or missing at run-in or screening, or for patients who switched study after screening only at the screening visit
- Subject with clinical diagnosis of DKD and persistent very high albuminuria, but eGFR < 25 mL/min/1.73 m² or ≥60 mL/min/1.73 m² or missing at run-in or screening visit, or for patients who switched study after screening only at the screening visit
- Subject with clinical diagnosis of DKD and persistent high albuminuria, but no history of diabetic retinopathy in the medical history
- UACR >5000 mg/g (565 mg/mmol) or missing at the run-in visit or screening visit
- Dialysis for acute renal failure within 12 weeks prior to the run-in visit
- Subject was not treated with any ACEI or ARB starting from run-in visit
- Participation in another clinical study or treatment with another investigational product 30 days prior to randomization except for the run-in and screening of study 17530
- Known significant non-diabetic renal disease, including clinically relevant renal artery stenosis
- Investigational drug other than the study drug is used concomitantly or within 30 days from stop of study drug

- An ACE inhibitor and an ARB OR more than 1 ACE inhibitor OR more than 1 ARB were used in combination during study drug treatment for more than 30 days
- Eplerenone, spironolactone, any renin inhibitor, or potassium sparing diuretic were used during study drug treatment for more than 30 days
- Potent CYP3A4 inducers were taken for more than 30 days during study drug treatment
- Wrong study medication (incorrect treatment (placebo or verum)) was dispensed to subject
- The subject is not compliant (<80%)
- The subject is not compliant (>120%)
- Subject was randomized but never received any study medication
- Visit 1 was done on the same day as screening and/or run-in visit

5.4 Pharmacokinetic analysis set (PKS)

All finerenone-treated subjects (with the exception described in Section 5) with at least 1 valid finerenone plasma concentration and without validity findings which would interfere with the evaluation of the pharmacokinetic (PK) data.

5.5 Listing only set (LOS)

All other subjects enrolled who were not randomized or were excluded from FAS will be classified as LOS. Their data will be presented in the individual subject data listings but will not be included in any statistical analysis.

5.6 Assignment of analysis sets

The allocation of each subject to analysis sets will be documented before the database lock (conditional validity for PPS and PKS).

Final decisions regarding the assignment of subjects to analysis sets will be made during the validity review meetings and documented in the validity review reports.

If a subject has no signed informed consent or subject is a minor, then all data should be deleted from all databases and such a subject cannot be considered valid for any analysis set.

6. Statistical Methodology

6.1 Population characteristics

Population characteristic analyses, except for subject disposition, will be performed for the FAS, if not stated otherwise.

6.1.1 Disposition

The number of subjects enrolled, randomized and valid for the SAF, FAS, PPS and PKS will be summarized overall and by treatment group, region, country and study site. The number of subjects discontinuing each epoch, together with the primary reason for discontinuation will be presented by treatment group (post-randomization epochs only) and overall in separate tables. In addition, the number of subjects with important deviations and validity findings will

be presented overall and by country for each treatment group, and in total. The frequencies of each important deviation and validity finding will be presented by treatment group and in total.

6.1.2 Demography and other baseline characteristics

Demography includes age, age group (as defined in the subgroup analysis), sex, race, ethnicity, region (North America, Europe, Asia, Latin America and others; see section 9.3), body weight, body height, BMI, hip and waist circumference, waist-hip ratio, smoking history (never, former, current smoker) and alcohol consumption (abstinent, light, moderate, heavy). Other baseline characteristics include baseline UACR, serum potassium, categories for serum potassium (≤4.5 mmol/L and >4.5 mmol/L), eGFR (calculated by CKD-EPI formula), serum creatinine, HbA1c, values for vital signs parameters (i.e. systolic blood pressure, diastolic blood pressure and heart rate) and additional subgroup categories as described in section 4.5.4.

All demographic data and baseline characteristics will be tabulated by treatment group and overall. The demographic and other baseline characteristics table will also be presented, separated by each level of the stratification factors.

The non-stratified demographic and other baseline characteristics table will be repeated for all other analysis sets.

As stated in section 4.2, demographics and other baseline characteristics will also be presented separately for subjects discontinuing the study (dropouts) and for subjects discontinuing treatment with study drug.

6.1.3 Medical History

Medical history will be coded using the MedDRA dictionary. Medical history will be presented for each MedDRA Primary System Organ Class and Preferred Term by treatment group and overall in a summary table. Additional medical history terms by the following Project specific Bayer MedDRA Queries (PBMQs), Bayer MedDRA Labeling Groupings (MLGs) or selected Preferred Terms will also be presented, including the following:

- Hyperlipidemia (MLG)
- Hypertension (MLG)
- Diabetic Retinopathy (Preferred Term)
- Diabetic Neuropathy (PBMQ)
- Atrial fibrillation/flutter (PBMQ)
- Ischaemic Stroke/Transitory Ischemic Attack (TIA) (PBMQ)
- Myocardial Infarction (Preferred Term)
- Coronary Artery Disease (Preferred Term)
- Myocardial Ischaemia and Angina pectoris (MLG)
- Peripheral Artery Disease (Preferred Term)
- Cardiac Failure (MLG)
- Coronary Artery Bypass Graft (CABG) (PBMQ)

- Percutaneous coronary intervention (PCI) (PBMQ)
- Carotid endarterectomy (Preferred Term)
- Periodontal disease (Preferred Term)

The medical history tables will be repeated for SAF.

6.1.4 Prior and Concomitant Medications

Concomitant medications will be coded using the WHO Drug Dictionary (WHO-DD). The number of subjects who took at least one concomitant medication, the number of subjects who took at least one medication that started and ended before administration of study drug, the number of subjects who took at least one concomitant medication that started after start of study drug and the number of subjects who took at least one medication ongoing at baseline (i.e. starting before or on the day randomization and ending at least one day after the day of randomization) will be presented by treatment group and overall using ATC classes and subclasses.

The concomitant medication tables will be repeated for SAF.

The concomitant medication tables will be repeated summarizing the number of subjects with medications in the Standard or Bayer Drug Groupings of interest:

- ACEIs
- ARBs
- Beta-blockers
- Diuretics
- Loop diuretics
- Thiazide diuretics
- Potassium-sparing diuretics
- Potassium supplements
- Potassium lowering agents (including binders)
- Alpha blocking agents
- Calcium channel blockers
- Centrally acting antihypertensives
- Strong CYP3A4 inhibitors
- Moderate CYP3A4 inhibitors
- Weak CYP3A4 inhibitors
- Unclassified CYP3A4 inhibitors
- Strong CYP3A4 inducers
- Moderate CYP3A4 inducers
- Weak CYP3A4 inducers
- Unclassified CYP3A4 inducers

- Oral anticoagulants
- NSAIDS (excluding acetylsalicylic acid)
- Acetylsalycylic acid and its salts
- Statins
- Erythropoetin stimulating agents
- ARNIs
- Platelet aggregation inhibitors (excluding heparin)

Anti-diabetic drugs

- Insulin and analogues
- Dipeptidyl Peptidase 4 inhibitors
- Glucagon-like peptide-1 agonists
- SGLT-2 inhibitors
- Biguanides
- Sulfonylureas
- Alpha glucosidase inhibitors
- Meglitinides
- Thiazolidinediones.

A subject will be counted only once within each ATC class / subclass or Bayer drug grouping, respectively.

A supporting listing will be provided including all medications classified as a weak, moderate or strong CYP3A4 inhibitors and inducers according to the protocol together with the respective classification information.

The percentage of patients who are on maximum labeled dose of ACEIs and ARBs (yes / no) 4 weeks before the Screening Visit, and reasons why not on maximum labeled dose will be presented by treatment group and overall. In addition, the percentage of patients at baseline above maximum, at maximum, between minimum and maximum, at minimum and below minimum labeled dose, using the dose ranges in Table 8-1 of the protocol, will be presented by treatment group and overall.

For potassium lowering agents, ACEIs, ARBs and diuretics, shift tables for changes of use for baseline vs. any time post-baseline will be provided.

6.1.5 Treatment duration, extent of exposure and compliance

The analyses described in this section will be repeated for the SAF and PPS if they differ in sample size from the FAS. All tables and figures regarding treatment duration, extent of exposure and compliance will be presented by treatment group and overall (unless otherwise stated).

Treatment duration, defined as time from start of study drug to permanent stop of study drug (in months) will be summarized using descriptive statistics by treatment group and overall. In addition, treatment duration will be categorized to ≤1 month, 1 - 4 months, then 4 monthly

intervals, and presented with the corresponding number and percentage of subjects. Cumulative treatment duration will be categorized to at least one dose, at least 1 month, at least 4 months, then further 4 monthly intervals. Cumulative treatment exposure over the study in person-years will be given. A table will be presented with the absolute and relative frequencies of subjects still on study medication at each visit. Kaplan-Meier plots for "Time to end of study treatment" will be provided by treatment group, as also described in section 4.2.

The above analyses will be repeated for study duration, from the day of randomization to the End of Study visit.

The extent of exposure to study drug (total amount of intake in grams) and the average daily dose in mg during treatment will be summarized using descriptive statistics by treatment group, also by eGFR category at screening.

The overall titration status, regardless of actual or sham up-titration, will be summarized with absolute and relative frequencies per treatment group, differentiated by patients starting on 10 mg or 20 mg:

- For subject starting on 10 mg: never up-titrated, up-titrated once, and up-titrated more than once.
- For subjects starting on 20 mg: never down-titrated, down-titrated once, and down-titrated more than once.

Furthermore, the dose status will be summarized with absolute and relative frequencies per treatment group for each visit with regard to the question "Was 20 mg study drug dispensed at visit?" and the reason why 20 mg was not dispensed. The number of subjects not dispensed with 20 mg as well as associated reason will also be presented overall, i.e. on the patient level.

In addition, the number of patients with study drug down-titrated to 10 mg or temporarily interrupted (dose recorded as 0 mg) as well as associated reason will be summarized with absolute and relative frequencies per treatment group.

The compliance (as percentage) will be calculated as follows:

100 * Number of taken tablets / Number of planned tablets

The number of planned tablets is calculated as follows:

(Days from randomization to last intake of study drug + 1) * Number of planned tablets per day.

All tablets, including the dummy placebo tablets, will be counted. For subjects who withdraw prematurely from the study drug, compliance will be calculated up to the time of last dose.

The compliance will be summarized descriptively by treatment group and overall. In addition, percent of compliance will be categorized into 3 groups, less than 80%, 80 to 120% and greater than 120%, and the categories will be summarized by treatment group and overall.

6.2 Efficacy

6.2.1 Analysis of primary efficacy variable

6.2.1.1 Primary efficacy variable: primary analysis

The primary efficacy variable is the time to first occurrence of the composite endpoint of onset of kidney failure, a sustained decrease of eGFR ≥ 40% from baseline over at least 4 weeks, or renal death. All events classifying for the primary endpoint have to be adjudicated positively by an independent adjudication committee.

The primary analysis of the primary efficacy variable will be performed in the FAS.

Events for inclusion in the primary analysis will be counted from the day of randomization (planned at Visit 1) onwards until the EOS visit following the study termination decision, or until the date of EOS notification + 4 weeks, if the EOS visit has not been performed. In the event of premature discontinuation from the study with no subsequent follow-up information, renal events will be counted up to the day of the last visit when complete information on all components of the composite renal endpoint is available with an additional rule for events occurring after this visit. To account for events of kidney failure or renal death after the last eGFR is recorded at a clinic visit, such events will be included in the efficacy analysis if they occur in the period up to one day before the next planned clinic visit. This period would be a maximum of 5 months, given the next scheduled clinic visit should take place 4 months later, plus a time window for late attendance of one month. In addition, should the subject have a non-renal death in this period then the date of death will be used as the censoring date. Randomized subjects without an event of the renal composite endpoint at the time of analysis will be censored at the date of their last visit when complete information on all components of the composite renal endpoint is available, up to and including the EOS visit (should this visit satisfy this rule), or date of non-renal death using a time window of 5 months as above if a subsequent clinic visit had been planned. Subjects without any information about the primary composite endpoint after baseline will be censored at Day 1.

The weighted Bonferroni-Holm procedure which will be used for the hierarchical testing of the primary and secondary efficacy endpoints, is described in section 4.7. In order to evaluate whether finerenone is superior to placebo in prolonging the time to the first event of the primary composite endpoint, the following null hypothesis will be tested using the logrank test at the two-sided significance level given above:

$$H_0: \lambda_{finerenone,k}(t) = \lambda_{placebo,k}(t) \text{ for all time points } t \geq 0 \text{ and each stratum } k$$

The alternative hypothesis will be:

$$H_1: \lambda_{finerenone,k}(t) \neq \lambda_{placebo,k}(t) \text{ for at least one time point } t \geq 0 \text{ and at least one stratum } k,$$

where $\lambda_{finerenone,k}$ denotes the hazard rate of the finerenone treatment group in stratum k and $\lambda_{placebo,k}$ denotes the hazard rate of the placebo treatment group in stratum k. In this context, the hazard rate is defined as the instantaneous risk of a primary composite event at a given time point provided that no such event has yet taken place. The test statistic of the stratified logrank test compares the observed number of events with the number of events expected under the null hypothesis H_0 in each treatment group per stratum.

The following decision rule to test the null hypothesis will be applied. According to the size of this study and the expected number of events, it is justified to assume under H_0 a sufficiently close approximation of the logrank test (see reference 4) to the normal

distribution. If the logrank test statistic (i.e. the standardized difference of the observed number of events and expected number of events stratified by the stratification factors region, type of albuminuria and eGFR category) is larger than the critical quantile from the normal distribution, the null hypothesis will be rejected in favor of the alternative hypothesis.

The nominal significance levels and the critical values at the final analysis will be adjusted to account for the interim analysis, see section 4.7 for more details.

In order to provide a point estimate of the hazard ratio and a corresponding two-sided 95% confidence interval, a stratified Cox proportional hazard regression model will be used. The stratification variables will be included in the STRATA statement in SAS PROC PHREG, which will assume proportional hazards for each level of one particular stratification factor.

Kaplan Meier curves will be provided for the cumulative incidence risk of outcome events by treatment groups.

To derive the logrank test statistic and the corresponding p-value, SAS program code corresponding to the following will be used:

```
PROC LIFETEST DATA = <dataset> ALPHA=0.05 METHOD=KM;
  TIME ttevalue * ttecnsr(1);
  STRATA {stratumn} / GROUP=trtgrpn  TEST=(logrank);
RUN;

/*
where
dataset  = name of sub-dataset including all ITT subjects
randomized
trtgrpn  = variable coding randomized treatment group
ttevalue = time to first occurrence of primary efficacy
outcome event
ttecnsr  = censoring index (1 = right-censored, 0 = event)
stratumn = variables for stratification factors
*/
```

For the Cox proportional hazard regression SAS program code corresponding to the following will be used:

```
PROC PHREG DATA = <dataset>;
  MODEL ttevalue * ttecnsr(1) = trtgrpn / RL TIES=EFRON
ALPHA=0.05;
  STRATA {stratumn};
RUN;
```

6.2.1.2 Primary efficacy variable: supportive analysis

The primary analysis of the primary efficacy variable will be repeated in the PPS as a supportive analysis. For the PPS analysis, the censoring rules for the primary analysis

specified in Section 6.2.1.1 will be further restricted to events occurring up to 30 days after last study drug intake.

Accordingly, subjects with censoring or event date more than 30 days after last study drug intake in the primary analysis will be censored at the date of last study drug intake +30 days.

The following supportive analyses of the primary efficacy variable will be performed in the FAS and PPS unless specified otherwise.

The originally planned presentation of absolute and relative frequency of subjects with an event of the primary composite endpoint will be superseeded by the Kaplan-Meier estimates described below.

Kaplan-Meier estimates of the time to the first event of the primary composite endpoint (including 95% confidence interval, calculated based on a normal approximation and using the Greenwood formula for the standard deviation [4]), overall survival rates at appropriate time points and Kaplan-Meier curves will be presented for each treatment group. In this context, "survival" means no event of the respective primary efficacy endpoint. These analyses will be presented in total and for each level of the stratification factors.

For the primary composite endpoint and the individual events, the incidence rate per 100 patient-years will be presented by treatment group, which is defined as the number of subjects with an event divided by the time a patient is at risk for the respective event, i.e. either the time from randomization until the first event or the censoring time, multiplied by 365.25*100. Kaplan-Meier plots, logrank tests and Cox proportional hazards regression for the time to first event of the individual components will also be provided as part of exploratory efficacy (see section 6.2.3).

An '*on-treatment*' analysis will also be performed, repeating the inferential statistics (stratified logrank test and stratified Cox regression model) on the FAS. For the on-treatment analysis, as described for the PPS analysis above, the censoring rules for the primary analysis will be adapted by further restricting to events occurring up to 30 days after last study drug intake.

The logrank test as well as the Cox regression model will also be performed without including stratification factors. The plausibility of the proportional hazards assumption will be assessed by visually examining the plot of the log of the negative log of Kaplan-Meier estimates of the survival function versus log time for evidence of non-parallelism, and by including a treatment by log-transformed time interaction into the Cox model. For the latter, the SAS code is adapted as follows:

```
PROC PHREG DATA = <dataset>;
  MODEL ttevalue * ttecnsr(1) = trtgrpn trtltime / RL
TIES=EFRON ALPHA=0.05;
  STRATA {stratum};
  trtltime = trtgrpn*log(ttevalue);
RUN;
```

The significance of the interaction will be tested at the 5% type I error level. If the interaction is significant and there is strong evidence of non-proportionality from the plots, time-dependent hazard ratios will be estimated with the model that includes the interaction term.

The censoring mechanism of subjects without an event of the primary composite endpoint at the time of analysis is assumed to be non-informative for the primary efficacy analysis. This also implies that the primary analysis is valid under a missing at random (MAR) assumption. Sensitivity analyses will be performed, assessing the impact of a potential informative censoring of subjects lost to follow-up or with withdrawn consent. Non-renal deaths will be treated as censored observations as in the primary analysis.

A summary of the amount of missing data by treatment arm will be given, including number of subjects who have incomplete follow-up on the respective endpoint.

Robustness with respect to missing data will be investigated with a tipping point analysis on the FAS (based on the ITT approach). The analysis will:

- Estimate the hazard at time of loss to follow-up, using a Weibull survival model fitted to the population of subjects who prematurely discontinued study treatment and adjusted for treatment group and stratification factors.
- Inflate the hazard only in the finerenone group by a known factor; assumes non-informative censoring in the control group.
- Impute events to the end of the study 1000 times, calculate the hazard ratios with the stratified Cox model specified in section 6.2.1.1 and combine them using standard multiple imputation combining rules.
- Increase the inflation factor until upper limit of two-sided confidence interval for the hazard ratio crosses 1.0; this will be the “tipping point”. The actual confidence level of the above mentioned confidence interval will depend on whether the study is stopped for overwhelming efficacy at the interim analysis or not, and on the p-value of the primary analysis of the key secondary endpoint. The confidence intervals will be as follows:
  - At the interim analysis: Use two-sided 99.73% confidence interval
  - At the final analysis:
    - p-value of primary analysis of key secondary endpoint < 0.01576184: Use two-sided 95.032612% confidence interval
    - p-value of primary analysis of key secondary endpoint ≥ 0.01576184: Use two-sided 96.717305% confidence interval

The tipping point will show how much higher the event risk after drop-out would need to be in the finerenone group so that statistical significance is lost. Since there is no particular reason to assume that the event rate should be higher, a large tipping point will indicate robustness of the study results. Technical details of the model are provided in section 9.1.

6.2.2 Analysis of secondary efficacy variables

6.2.2.1 Secondary efficacy variables: primary analysis

Secondary efficacy variables are as follows:

- Time to first occurrence of the following CV composite endpoint: cardiovascular (CV) death or non-fatal CV events (i.e. myocardial infarction, stroke, or hospitalization for HF)
- Time to all-cause mortality

- Time to all-cause hospitalization
- Change in UACR from baseline to Month 4
- Time to first occurrence of the following renal composite endpoint: onset of kidney failure, a sustained decrease in eGFR of ≥ 57% from baseline over at least 4 weeks or renal death.

The primary analysis of the secondary efficacy variables will be conducted in the FAS.

See section 4.7 for the testing hierarchy for secondary endpoints and for multiplicity adjustment of significance levels.

The primary analysis of the secondary time-to-event endpoints will be conducted analogously to the primary analysis of the primary composite endpoint, with modifications to the censoring rules as the events will differ. In the event of premature termination from the study with no subsequent follow-up information, subjects without any information about the composite endpoint after baseline will be censored at Day 1, and the following censoring rules will also apply:

- CV events will be counted up to the day of withdrawal from the study or non-CV death. Randomized subjects without an event of the CV composite endpoint at the time of analysis will be censored at the date of their last contact up to and including the EOS visit or date of non-CV death.
- Hospitalization events will be counted up to the day of withdrawal from the study or death. Randomized subjects without a hospitalization event at the time of analysis will be censored at the date of their last contact up to and including the EOS visit or date of death.
- All-cause mortality will be counted up to the last date vital status could be obtained. Randomized subjects without such an event at the time of analysis will be censored at the last date vital status could be obtained up to and including the EOS visit.

An ANCOVA model will be fitted to the logarithmized ratios of UACR at Month 4 to UACR at baseline including the factors treatment group, the stratification factors region, type of albuminuria and eGFR category and the logarithmized baseline UACR as covariate. Corresponding two-sided 95% confidence intervals will be computed. As the baseline UACR defines the type of albuminuria, the log-transformed baseline UACR will be nested as covariate in type of albuminuria.

6.2.2.2 Secondary efficacy variables: supportive analysis

The primary analysis of the secondary efficacy time-to-event endpoints will be repeated in the PPS as a supportive analysis. For the PPS analysis, the censoring rules for the secondary analysis specified in Section 6.2.2.1 will be further restricted to events occurring up to 30 days after last study drug intake.

Accordingly, subjects with censoring or event date more than 30 days after last study drug intake in the primary analysis will be censored at the date of last study drug intake +30 days.

The supportive analyses of the secondary time-to-event endpoints will correspond to those done for the primary composite endpoint in the FAS and PPS (see section 6.2.1.2), except for the ‘on-treatment’ and tipping point analysis which will only be performed for the key secondary endpoint based on the FAS.

UACR at Month 4 in the PPS will be analyzed using an ANCOVA as described above. In addition, the following sensitivity analyses accounting for missing data at Month 4 will be conducted:

- On-treatment LOCF approach, only including data up to 30 days after stop of study medication (FAS and PPS).
- Baseline observation carried forward analysis (BOCF), imputing the baseline value for missing data and thus including all subjects with missing Month 4 data with a value of 1 for the ratio of UACR baseline to Month 4 (FAS).
- Multiple imputation analysis (FAS): For reproducibility, the SAS seed number for creating the random numbers will be set to the study (16244). Further details are given in section 9.2.

For all PPS analyses of UACR at month 4, only data up to 30 days after stop of treatment will be considered, all other data will be set to missing before employing the above-mentioned imputation techniques.

6.2.3 Analysis of other exploratory efficacy variables

Other exploratory efficacy variables will be as follows:

- Time to onset of kidney failure
- Time to onset of ESRD
- Time to onset of eGFR decrease to less than 15 mL/min/1.73 m² sustained over at least 4 weeks
- Time to onset of eGFR decrease of ≥ 40 % sustained over at least 4 weeks
- Time to onset of eGFR decrease of ≥ 57% sustained over at least 4 weeks
- Time to CV death
- Time to non-CV, non-renal death
- Time to first CV hospitalization (either hospitalization for HF, other cardiovascular hospitalization or adjudicated CV event associated with hospitalization)
  - Causes for other cardiovascular hospitalizations include unstable angina, arrhythmia, and peripheral arterial disease.
  - CV events associated with hospitalization include (if not already considered as "other cardiovascular hospitalization") CV death, new onset of atrial fibrillation/flutter, non-fatal myocardial infarction, non-fatal stroke and transient ischemic attack.
- Time to first hospitalization for HF
- Time to first non-fatal stroke
- Time to first non-fatal myocardial infarction
- Time to first occurrence of the composite endpoint of CV death, non-fatal stroke or non-fatal myocardial infarction
- Time to first occurrence of the composite endpoint of CV death or hospitalization for HF

- Time to first occurrence of the composite endpoint of CV death, kidney failure, eGFR decrease of ≥ 57% sustained over at least 4 weeks or renal death
- Number of subjects with new diagnosis of atrial fibrillation or atrial flutter
- Number of subjects with new diagnosis of HF
- Change in UACR from baseline
- Change in eGFR from baseline
- Regression from very high to high albuminuria and high albuminuria to normalbuminuria accompanied by a decrease in UACR of at least 30% from baseline
- Number of subjects with UACR decrease of at least 30% from baseline at any time post-baseline
- Number of subjects with UACR decrease of at least 50% from baseline at any time post-baseline
- Change in QoL summary scores measured by the following health-related quality of life (HRQoL) questionnaires: Kidney Disease Quality of Life (KDQOL 36) and EuroQol Group 5 dimension, 5 level questionnaire (EQ 5D 5L)

All these other efficacy variables will be analyzed in FAS and PPS (as applicable) in an explorative manner. For all PPS analyses, only data points and events up to 30 days after stop of study drug will be considered.

6.2.3.1 Components of primary and secondary composite endpoints

Consistent with the methods for the primary and secondary efficacy variables, components or composites of the components of the primary and secondary time-to-event endpoints (as included in the list of other exploratory efficacy variables above) will be analyzed as exploratory time-to-event variables using the stratified logrank test stratified by the stratification factors. A point estimate of the hazard ratio and a corresponding two-sided 95% confidence interval will be calculated using a stratified Cox proportional hazard regression model. Kaplan Meier curves will be provided for the cumulative proportions of events by treatment groups.

6.2.3.2 UACR and albuminuria

UACR during the study will be summarized descriptively by treatment group and visit including ratios to baseline. These analyses will be performed overall and separated by the stratification factors (region, type of albuminuria at screening and eGFR category at screening).

The log-transformed ratio of UACR to baseline at each visit up to the last visit with a reasonable number of subjects with UACR measurements will be analyzed by a mixed model repeated measures (MMRM) with the factors treatment group, visit, treatment-by-visit interaction, factors for the stratification levels, log-transformed baseline value as covariate nested within type of albuminuria and log-transformed baseline value by visit interaction. Pairwise ratios between the finerenone and the placebo treatment group will be calculated and corresponding two-sided 95% confidence intervals will be computed.

An on-treatment analysis including measurements restricted to up to 30 days after last intake of study drug will be performed using the same model as above.The SAS procedure PROC

MIXED will be used estimating covariance patterns within subjects to adjust for the within subject variance. For each treatment group a separate covariance pattern will be estimated based on an unstructured covariance. In case of convergency issues, an alternative covariance matrix will be used, with Toeplitz matrix as first choice alternative.

Frequency tables will be generated for the number of subjects with a relative decrease and increase in UACR of ≥30%, ≥40% and ≥50% from baseline UACR. The analysis will be performed for each visit and for any time post baseline. The analysis will also be performed stratified for each level of the stratification factors.

A shift table will be provided displaying the number of subjects who changed from baseline to each visit from very high albuminuria to high albuminuria, from very high albuminuria to normalbuminuria (UACR <30mg/g), from high albuminuria to normalbuminuria (UACR <30mg/g), from high albuminuria to very high albuminuria and from normalbuminuria (UACR <30mg/g) to high and very high albuminuria by treatment group and overall. The albuminuria category changes will only be considered as shifts if they are accompanied by a UACR decrease or increase of at least 30% from baseline to each visit.

The additional categorical UACR efficacy variables will be summarized for presence or absence of the event using logistic regression with the factors treatment group and stratification levels. Pairwise differences between the finerenone and the placebo treatment group will be calculated and corresponding two-sided 95% confidence intervals will be computed.

6.2.3.3 Decrease in eGFR

eGFR will be summarized descriptively by treatment group and visit including absolute and relative changes to baseline. These analyses will be performed overall and separated by the stratification factors.

The absolute change of eGFR to baseline at each visit until the last visit with a reasonable number of subjects with eGFR measurements will be analyzed by a mixed model with the factors treatment group, visit, treatment-by-visit interaction, factors for the stratification levels, baseline value as covariate (nested within eGFR category) and baseline by visit interaction. Pairwise differences between the finerenone and the placebo treatment group will be calculated and corresponding two-sided 95% confidence intervals will be computed.

An on-treatment analysis including measurements restricted to up to 30 days after last intake of study drug will be performed using the same model as above.

The SAS Procedure PROC MIXED will be used estimating covariance patterns within subjects to adjust for the within subject variance. For each treatment group a separate covariance pattern will be estimated based on an unstructured covariance.

Frequency tables will be generated for the number of subjects with a relative decrease in eGFR of ≥30, ≥40%, ≥50% and ≥57% from baseline eGFR. The analysis will be performed for each visit and for any time post baseline.

For each patient with 1 or more post-baseline eGFR measurements, the annualized change in eGFR will be calculated by fitting the patient's eGFR assessments into a linear regression model with time in years as the independent variable. The derived annualized change will be analyzed using an ANCOVA model including baseline, treatment group and stratification factors as fixed-effects.

The following additional analysis of acute and chronic eGFR slopes will be performed.

- Change in eGFR from baseline (Visit 1) to the PD or EOS Visit
- Acute eGFR changes from baseline to Month 4
- Change in eGFR from Month 4 to the PD or EOS Visit
- Correlation analyses between short term eGFR changes between baseline and Month 4 after randomization, and long term eGFR decline (chronic slope).

An ANCOVA model will be fitted to the eGFR efficacy variables including the factors treatment group, the stratification factors and the baseline eGFR as covariate. Corresponding two-sided 95% confidence intervals will be computed. In addition, correlation analyses between short term eGFR changes between baseline and Month 4 after randomization, and long term eGFR decline (chronic slope) will be performed.

6.2.3.4 Health-related quality of life

The KDQOL-36 will be used. The items of the KDQOL 36 are grouped as follows:

- Items 1-12: Physical Component Summary (PCS) / SF-12 physical health composite on physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, mental health
- Items 1-12: Mental Component Summary (MCS) / SF-12 mental health composite on physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, mental health (including a different weighting than the PCS)
- Items 13-16: Burden of Kidney Disease; interference with daily life, time to deal with kidney disease, frustration, feeling like a burden
- Items 17-27: Symptoms / Problems; general health, activity limits, ability to accomplish desired tasks, depression/anxiety, energy level, social activities (item 28 concerning dialysis will be only collected in patients on dialysis)
- Items 29-36: Effects of Kidney Disease; impact of fluid & diet limits, ability to work around the house and to travel, feeling depending on medical team, stress or worries, sex life, personal appearance

Domain scores for each of the 5 domains (Physical Component Summary, Mental Component Summary, Burden of Kidney Disease, Symptoms / Problems, and Effects of Kidney Disease) will be calculated according to the KDQOL scoring instruction (see reference 1). The KDQOL-36 domain scores will be presented by visit and treatment group by means of number of observations, number of missing values, minimum, first quartile, mean, SD, median, third quartile, and maximum, including the changes from baseline. In addition to the analysis of the KDQOL-36 for the overall population, this will be repeated for patients with or without ESRD and/or dialysis at any point during the study.

The absolute change of the PRO to baseline at each visit up to the last visit with a reasonable number of subjects with PRO measurementswill be analyzed by a mixed model with the factors treatment group, visit, treatment by visit interaction, factors for the stratification levels, baseline value as covariate and baseline by visit interaction. Pairwise differences between the finerenone and the placebo treatment group will be calculated and corresponding two-sided 95% confidence intervals will be computed.

The SAS Procedure PROC MIXED will be used estimating covariance patterns within subjects to adjust for the within subject variance. For each treatment group a separate covariance pattern will be estimated based on an unstructured covariance.

For the EQ-5D, summary scores will be calculated out of the 5 dimensions according to the scoring instructions from Europe (UK as a representative country within Europe) and the US (see reference 2) and to the EQ-5D Value Sets (see reference 3). The values and the changes from baseline of the summary scores and the EQ Visual Analogue scale (VAS) will be summarized by treatment group and visit using the same descriptive statistics as for KDQOL.

The EQ 5D 5L will be analyzed similarly to the KDQOL-36.

6.2.3.5 New diagnosis of atrial fibrillation, new diagnosis of heart failure and regression of albuminuria

These other exploratory efficacy variables will be summarized for presence or absence of the event using logistic regression with the factors treatment group and stratification levels (type of albuminuria, region and eGFR category). Pairwise differences between the finerenone and the placebo treatment group will be calculated and corresponding two-sided 95% confidence intervals will be computed.

6.2.3.6 Outcome events reported by the investigators

Outcome events using the investigator-reported terms will be summarized by treatment group, using tables analogous to those for adverse events. No statistical testing of these events will be performed, as the adjudicated outcome events are components of the primary, secondary and exploratory efficacy variables.

An overall summary of all outcome events will be generated by treatment group.

The number of subjects with all outcome events, outcome events from randomization up to 30 days after stop of study medication, post-treatment outcome events occurring more than 30 days after stop of study drug or after the EOS Visit, outcome events by maximum intensity and outcome events by worst outcome will be summarized by treatment group using MedDRA terms grouped by Primary System Organ Class and Preferred Term.

The incidence rate of outcome events per 100 patient-years will also be provided by treatment group using MedDRA terms grouped by Primary System Organ Class and Preferred Term. The time under risk for the incidence rates is defined as the time from randomization until the first onset of the event or the last date of contact with the subject in case no such event is recorded.

Outcome events will be listed separately in the section 14.2 tables.

6.3 Pharmacokinetics/pharmacodynamics

6.3.1 Pharmacokinetics

The plasma concentration versus time data of Visit 3 (Month 4) and at subsequent yearly visits will be evaluated descriptively, separated by dose and visit. Plots will be prepared of all individual plasma concentrations vs. actual relative study times (time of sample collection after time of study drug administration).

Evaluation of the concentration data will be performed using Population PK methods, followed by PK/PD analyses. These analyses will be described in a separate Analysis Plan outside of this document and will be reported separately.

6.3.2 Pharmacodynamics

Not applicable.

6.4 Safety

All analyses on safety and tolerability data will be performed in the SAF.

6.4.1 Adverse events

Adverse events (AEs) will be coded using the Medical Dictionary for Regulatory Activities (MedDRA, latest version available prior to data base freeze). A listing will be provided linking the original investigator terms and the coded terms. AEs will also be presented grouped by SMQs.

AEs that started or worsened after the first dose of study drug up to 3 days after any temporary or permanent interruption of study drug will be considered as treatment-emergent AEs (TEAEs).

An overall summary of all AEs and TEAEs will be generated by treatment group.

The number of subjects with TEAEs, pre-randomization AEs, post-treatment AEs occurring more than 3 days after any temporary or permanent stop of study drug, treatment-emergent SAEs, treatment-emergent study drug-related AEs, treatment-emergent study drug-related SAEs, treatment-emergent SAEs and TEAEs causing permanent discontinuation of study drug, treatment-emergent non-serious AEs, non-serious AEs, TEAEs by maximum intensity, treatment-emergent SAEs by maximum intensity, drug-related TEAEs by maximum intensity, TEAEs by worst outcome and treatment-emergent SAEs by worst outcome will be summarized by treatment group using MedDRA terms grouped by Primary System Organ Class and Preferred Term.

To comply with local regulatory requirements in Japan, some cardiovascular disease-related outcome events will also be documented as (S)AEs in Japan. These will be included in the outcome event tables (see section 6.2.3.6), and to avoid double-counting of such events, they will not be included in the adverse event summary tables or listings. A separate listing will be generated for all AEs excluded from the AE analysis due to double reporting in Japan.

In case of events with different intensity within a subject, the maximum reported intensity will be used. If intensity is missing, the event will be considered as severe. If the same event is considered as both unrelated and related to the study drug within a subject, the event will be reported as related to study drug. If the drug relationship is missing, the event will be considered as being related to the study drug.

These further safety variables are listed in section 10.3.4 of the protocol:

- Number of subjects with hospitalization for hyperkalemia
- Number of subjects discontinuing study drug permanently due to hyperkalemia
- Number of subjects with hospitalization for worsening of renal function

- Number of subjects discontinuing study drug permanently due to worsening of renal function

These will be summarized by treatment group using frequency counts. For the number of subjects with these events, the summary will be provided once for all events and once for all treatment-emergent events.

Separate tables summarizing TEAEs, study drug-related TEAEs, and SAEs that occurred in more than 5% of the subjects will be provided.

Deaths, SAEs and AEs leading to permanent study drug discontinuation will be listed separately.

6.4.2 Laboratory parameters

The number of subjects with treatment-emergent (until 3 days after any temporary or permanent interruption of study drug) abnormal laboratory values above or below the normal range will be tabulated by the laboratory parameter and treatment group.

Summary statistics including changes to baseline will be calculated by treatment group and visit for all quantitative laboratory parameters, e.g. for hematology, HbA1c, clinical chemistry and urinalysis. Geometric statistics and ratios to baseline will be presented for creatinine and UACR instead of arithmetic statistics with changes from baseline. For eGFR the relative change will be displayed in addition to the absolute change from baseline.

Summary statistics for serum potassium, eGFR and serum creatinine will also be repeated by treatment group and visit separately for each level of the stratification factors.

The following special safety parameters will be further assessed by displaying the number of subjects with treatment-emergent safety events. This will also be performed by visit and by stratification factors. The summaries will be performed for the number of subjects with:

- Absolute value of serum potassium >5.5 mmol/L and >6.0 mmol/L,
- Relative decrease from baseline in eGFR of ≥30%, ≥40%, ≥50% and ≥57%
- Absolute value of eGFR <30 mL/min/1.73m².

The percentage of subjects with the respective treatment-emergent events (non-stratified) will be compared between the finerenone and the placebo or active control group by applying separate explorative χ^2 tests with continuity correction. If the expected number of subjects in at least 1 cell of the 2x2 contingency table is less than 5 (see reference 5), Fisher's exact test will be applied instead of the χ^2 test. Estimates and two-sided 95% confidence intervals will be provided for each treatment group and the treatment differences. Clopper Pearson confidence intervals will be calculated for each treatment group, while for treatment differences, due to the computational burden of exact unconditional confidence limits, Wald asymptotic confidence limits will be calculated.

6.4.3 Vital Signs

The analysis of vital signs includes blood pressure and pulse measurements. Averages of non-missing values will be calculated and used for the statistical analysis. If only 1 of the planned measurements is available, this value will be used.

Vital signs values will be summarized by treatment group and visit using descriptive statistics including absolute changes from baseline. The analysis will be repeated for SBP stratified by

baseline SBP <130 mmHg, 130 to <160 mmHg and ≥160 mmHg and SBP above and below median.

6.4.4 Weight and BMI

The values and the changes from baseline will be summarized by treatment group and visit using descriptive statistics for weight and BMI.

7. Document history and changes in the planned statistical analysis

SAP version 0.1 dated 14 DEC 2015 submitted for internal review

SAP version 0.2 dated 29 JAN 2016 submitted to the FDA

SAP version 0.3 dated 08 JUL 2016 submitted for internal review

Approved SAP version 1.0 dated 03 AUG 2016

Approved SAP version 2.0 dated 07 JUN 2019

Approved SAP version 3.0 dated 12 SEP 2019

Approved SAP version 4.0 dated 14 FEB 2020

7.1 Overview of Changes to SAP – from version 1.0 to version 2.0

Editorial, administrative, and typographical corrections were made that do not affect the content of the overall integrated SAP. These changes are not described in this section.

The SAP, Version 1.0, dated 03 AUG 2016, was amended with the changes resulting from global CSP amendments, the latest being global amendment 4, forming integrated CSP Version 3.0, dated 26 FEB 2019. SAP modifications resulting from the integrated CSP Version 3.0 are primarily:

- administrative, typographical and consistency-related corrections,
- definition of treatment-emergent adverse events and laboratory parameters to also consider temporary interruptions of study drug as well as modification of special safety variables for hyperkalemia and worsening of renal function to only consider permanent discontinuations of study drug, and
- replacement of the date of first intake of study drug with the randomization date as the baseline date for all analyses (except those explicitly requiring start of study drug intake as a reference).

In addition, the following changes were introduced in SAP Version 2.0, dated 07 June 2019.

Modification 1: Update on sample size

Rationale: Due to the lower-than-expected event rate, the sample size of the study was increased by approximately 1000 subjects. As this has no impact on the original planning assumptions, the protocol text remained unchanged, but for descriptive purposes the expected sample size was added in the SAP

Sections affected:

- Section 3:Study Design

Modification 2: Exclusion from analysis sets due to critical GCP violations

Rationale: After detection of issues with investigator fraud and duplicate subjects, it was decided and agreed with the FDA to exclude the affected subject records from all analysis sets (except for the listings only set) as this data could not have been trusted.

Sections affected:

- Section 3: Study Design
- Section 5: Analysis Sets

Modification 3: Types of events used for efficacy analyses

Rationale: It was clarified that only events adjudicated by an independent adjudication committee will be used for all efficacy analyses (except for those explicitly mentioning investigator-reported events).

Sections affected:

- Section 4.1: General Principles
- Section 6.2.1.1 Primary efficacy variable: primary analysis

Modification 4: Stratification groups

Rationale: It was clarified that if the corrected stratification group (due to incorrect allocation at randomization) does not exist in the study, the stratification group the investigator originally specified in the IxRS will be kept. This is in line with the randomization and ensures that no strata with few patients are opened up which would result in potential non-convergence of the stratified Cox model, for example.

Sections affected:

- Section 4.1: General Principles

Modification 5: Handling of missing data for study drug exposure, clinical events, UACR at Month 4 and adverse events

Rationale: Texts were added to describe the handling of missing data for study drug exposure (conservative approach maximizing time on higher dose, but avoiding overlaps) and for clinical events (elaboration on median imputation rule). The handling of missing start dates for adverse events was simplified and aligned with the definition of treatment-emergence (see change related to Integrated CSP Version 3.0 above). For the analysis of UACR at Month 4, it was clarified that the closest set of measurements in the timeframe of Month 4, i.e. on days 120 (+/-30days), is used, regardless of whether it is labeled as the month 4 visit to also consider unscheduled visits or other visits in case these should be closer to the relevant date.

Sections affected:

- Section 4.3: Handling of Missing Data

Modification 6: Primary analysis when stopping at the interim analysis

Rationale: The significance level which would remain for the final analysis after conducting the interim analysis was made more precise and specified as 4.967388%. Furthermore, it was clarified that in case of stopping at the interim analysis, the primary analysis reported in the CSR will take all events up to the subjects' respective EoS visits into account as would be the case for the final analysis if the study does not stop prematurely.

Sections affected:

- Section 4.4 Interim Analysis and Data Monitoring

Modification 7: Derivation of the coefficient of variation, window for on-treatment/PPS analysis and baseline flag for UACR

Rationale: The formula for the coefficient of variation was added. The time window to consider unscheduled measurements in case a scheduled measurement is missing at a given visit for an efficacy analysis and to consider for the on-treatment analysis was set to 30 days in line with the time window for the on-treatment analysis/PPS analysis for clinical events. The 30 days window also applies to investigator-reported clinical events. However, a 3 days window is still used for any safety analyses of UACR. In addition, the derivation of baseline UACR was clarified to require a set of two non-missing values on or before day of randomization; otherwise the screening set of UACR measurements would be used.

Sections affected:

- Section 4.5.3 Other data handling
- Section 6.2.3.6 Outcome events reported by the investigators

Modification 8: eGFR and UACR values after onset of ESRD for efficacy analyses

Rationale: It was clarified that UACR and eGFR values after onset of ESRD would be excluded from all efficacy analyses because they would be strongly related to the event of ESRD and thus would not be meaningful in themselves.

Sections affected:

- Section 4.5.3 Other data handling

Modification 9: Addition of subgroup factors and analyses for medical history and concomitant medications

Rationale: Additional subgroup factors (UACR at baseline below and above median and potency of concomitant CYP3A4 inducer medication) were added for reasons of completeness. A new analysis was added for concomitant medications ongoing at baseline, i.e., at randomization, which was not adequately captured before as the other analyses for concomitant medication are in relation to the dates of first intake of study drug, and which is also relevant considering that there are several subgroup factors for concomitant medications

at baseline. Furthermore, analyses for concomitant medications and medical history will also be repeated on the SAF because it is certain that the SAF will differ from the FAS.

Sections affected:

- Section 4.5.4: Subgroup Analyses
- Section 6.1.3: Medical History
- Section 6.1.4: Prior and Concomitant Medications

Modification 10: Titration status

Rationale: The analyses for titration related to different questions on the eCRF were restructured and split up into three different analyses/tables for easier interpretability.

Sections affected:

- Section 6.1.5 Treatment duration, extent of exposure and compliance

Modification 11: Description of primary analyses for the primary and secondary efficacy time-to-event variables

Rationale: The description of the hypotheses was made more accurate by adequately taking into account the stratification factors. At the same time, survival rates were exchanged with hazard rates to create an obvious link from the hypothesis to the hazard ratio which is given as the primary output of the Cox model. However, this exchange has no impact on the content, as survival rates and hazard rates can be translated into each other and the hypotheses of equality of survival rates and of equality of the hazard rates (per stratum) are thus equivalent. Furthermore, the SAS code for the Cox model was corrected to the stratified Cox model specified in the text (by means of adding the stratification factors to the STRATA statement).

Sections affected:

- Section 6.2.1.1: Primary efficacy variable: primary analysis

Modification 12: Modifications and additions to the sensitivity analyses for the primary and secondary efficacy variables

Rationale: The analysis of number of (recurrent) events per 100 patient-years was replaced by an analysis of incidence rate, no longer considering recurrent events because this is not meaningful for the clinical events in this study as most of them usually do not happen more than once. An on-treatment analysis was added for the FAS in line with the protocol. Further supportive analyses were added to check the proportional hazards assumption for the Cox model (plot of the log of the negative of Kaplan-Meier estimates versus log time and Cox model including treatment by log-transformed time as an interaction). More details were given for the tipping point analysis, and it was clarified that this analysis and the newly added on-treatment analysis will only be done for the primary and key secondary renal endpoint based on the FAS.

Sections affected:

- Section 6.2.1.2: Primary efficacy variable: supportive analysis
- Section 6.2.2.2: Secondary efficacy variables: supportive analysis
- Section 9.1: Appendix A: Technical details missing data imputation model for time to event

Modification 13: Censoring rules for on-treatment/PPS analysis of key renal endpoint

Rationale: Censoring rules for the key secondary renal composite endpoint were clarified to consider both the time window of 30 days after treatment discontinuation as usual for this type of analysis in this SAP and the 5 months after the last visit with information of all components of the renal composite endpoint as done for the FAS analysis of this endpoint.

Sections affected:

- Section 6.2.2.2: Secondary efficacy variables: supportive analysis

Modification 14: Modification of list of further efficacy time-to-event endpoints

Rationale: Further exploratory time-to-event endpoints were added for the sake of completeness to adequately depict all the single components as well as reasonable sub-composites of the previously mentioned primary and secondary composite endpoints.

Sections affected:

- Section 6.2.3.1: Exploratory time to event variables

Modification 15: Deletion of analysis for relationship between a change in UACR from baseline to month 4 and the incidence of cardiovascular and renal events

Rationale: This analysis was removed because it was not deemed appropriate to compare the relationship between two post-baseline events in the described manner as the possibility of confounding with the factor treatment cannot be excluded.

Sections affected:

- Section 6.2.3.2: UACR and albuminuria

7.2 Overview of Changes to SAP – from version 2.0 to version 3.0

Major changes compared to version 2.0:

- Full specification of testing procedure and multiplicity adjustment, moved to new section 4.7.
- Deeper specification of imputation rules for missing event dates (section 4.3)
- Correction of data handling for UACR measurements (section 4.5.3)
- Definition of subgroup analyses for safety events (section 4.5.4)
- Alignment of cut-off values for baseline UACR/blood pressure categories

- Drop of name of investigator in disposition tables (section 6.1.1)
- Deletion of information from the combined treatment and follow-up epoch (section 6.1.1)
- Correction of ACEIs/ARBs baseline analysis (section 6.1.4)
- Amount of missing data will be summarized (section 6.2.1.2)
- Superseding frequency tables at certain time points by more powerful KM estimates (section 6.2.1.2)
- Specification of UACR analysis (section 6.2.2.1)
- Specification of analysis for components of primary and secondary endpoints (section 6.2.3.2)
- Specification of censoring rules for exploratory efficacy variables (section 6.2.3.6)
- Clarification about analysis of certain safety parameters (Hyperkalemia, worsening of renal function) (section 6.4.1)

7.3 Overview of Changes to SAP – from version 3.0 to version 4.0

Major changes compared to version 3.0:

- Clarifiaction on imputation rules for partially missing death dates (section 4.3)
- Additional subgroups added, “key” and “important” subgroups re-ordered (section 4.5.4)
- Platelet aggregation inhibitors (excluding heparin) added to the list of concomitant medications (section 6.1.4)
- Clarification that time-to-event analysis on-treatment and for PPS will be censored at the end-of-study visit (sections 6.2.1.2 and 6.2.2.2)
- Deletion of one endpoint and addition of three endpoints to the list of other exploratory efficacy variables (section 6.2.3)
- Specification of Toeplitz as first choice alternative covariance matrix in case of convergency issues in PROC MIXED (section 6.2.3.2)
- Deletion of analyses for the iohexol substudy, since the results of the substudies will be reported separately, and not in the clinical study report (section 6.2.3.3)
- Change from time-to-event analyses to logistic regression for other exploratory efficacy analyses (New diagnosis of atrial fibrillation, new diagnosis of heart failure and regression of albuminuria) (section 6.2.3.5)
- A separate listing will be generated for all AEs excluded from the AE analysis due to double reporting in Japan (section 6.4.1)
- For the other safety events, summary will be provided once for all events and once for all treatment-emergent events (section 6.4.1)
- Alignment of use of “treatment-emergent” (for safety) and “on-treatment” (for efficacy) analyses (e.g., section 6.4.2)

- Wald confidence limits for treatment differences in certain laboratory parameters will be calculated, due to the computational burden of exact unconditional confidence limits (section 6.4.2)

8. References

1. http://www.rand.org/health/surveys_tools/kdqol.html (6th September 2013).
2. EQ-5D-5L User Guide, Version 2.0, EuroQoL Group, 2013.
3. Szende, A., Oppe, M., Devlin, N. (2007): EQ-5D Value Sets: Inventory, Comparative Review and User Guide. Chapter 4. Springer.
4. Collett, D. (2003): Modelling Survival Data in Medical Research, Chapter 2. Chapman and Hall.
5. Agresti, A. (2007): An Introduction to Categorical Data Analysis, Chapter 2, Second Edition, New York: John Wiley & Sons.
6. Little, R. J., et al. (2016): The treatment of missing data in a large cardiovascular clinical outcomes study. Clinical Trials, 13(3): 344-351.
7. Koopmans, L. H., Owen, D. B., Rosenblatt, J. I. (1964): Confidence intervals for the coefficient of variation for the normal and log normal distributions. Biometrika. 51: 25–32.
8. Carpenter, J. R., Kenward, M. G. (2007): Missing data in randomized controlled trials - a practical guide. National Health Service Coordinating Centre for Research Methodology: Birmingham.
9. Tang, D. I., Geller, N. L. (1999): Closed testing procedures for group sequential clinical trials with multiple endpoints. Biometrics, 55(4): 1188-1192.

9. Appendix

9.1 Appendix A Technical details missing data imputation model for time to event

To investigate the impact of missing data under a missing not at random (MNAR) assumption a methodology similar to pattern-mixture-models will be employed as described by Little et al. (2016) [6]. In this approach, a Weibull survival model will be fitted to all subjects who prematurely discontinued study treatment in order to estimate the individual hazard rates for patients with withdrawn consent or loss to follow-up, but without a previous endpoint event. Based on these individual hazard estimates at the time of censoring (possibly penalized through inflation), event times for patients alive but censored before the global cut-off date (notification of study termination + 4 weeks) will be imputed.

Details:

The SAS LIFEREG procedure will be used where the natural logarithm of the survival time $w = log(t)$ is modeled. The survival function has the form

$$S(w) = \exp\left(- \exp\left(\frac{w - \mu}{\sigma} \right) \right),$$

where μ is the location parameter (intercept) and σ is the scale parameter in the SAS output.

The observed hazard rates (scale and shape parameters) for individual subjects are obtained from a Weibull model fitted to all subjects who prematurely discontinued study treatment, adjusted for treatment group (coded 0=placebo, 1=finerenone) and the stratification factors. The Weibull probability density function is re-parametrized to have the following form

$$g(t|\, \lambda, \beta) = (\beta/\lambda)\ (t/\lambda)^{\beta-1}\ \exp(-(t/\lambda)^{\beta}),$$

where $\lambda = \exp(\mu)$ is the new scale parameter = $\exp(\mu)$ and $\beta = 1/\sigma$ is the re-parametrized shape parameter.

With this re-parametrization, the hazard function is given by

$$h(t|\, \lambda, \beta) = (\beta/\lambda)\ (t/\lambda)^{\beta-1}$$

and the survival function can be written as

$$S_2(t) = \exp(-(t/\lambda)^{\beta}).$$

For subjects with consent withdrawn or who are lost to follow-up and who did not experience a primary efficacy endpoint, random variables will be simulated using the conditional time to event distribution after the last contact date given that no event was observed before. If this randomly generated variable has a value less than the elapsed time between the last contact date and the global cut-off date, i.e. the date when all subjects should have performed their EOS visit (notification of study termination + 4 weeks), the subject will be counted as having observed an event at the last contact date plus the random variable. Otherwise, the subject is re-adjusted to be censored at the global cut-off date.

These imputed events will be added to the observed events in the study and the primary efficacy variable will be re-analyzed. This process will be repeated 1000 times. For the final analysis, the hazard ratio with 95.032612% or 96.717305% confidence intervals (the actual confidence level will depend on the p-value of the primary analysis of the key secondary endpoint and is adjusted for the interim analysis, for further details see Section 6.2.1.2) will be estimated for each imputed data set and then combined using standard multiple imputation combining rules.

Details:

When the survival time T follows a Weibull distribution with survival function S_2, then the conditional distribution function of time T after the last contact date/censoring date (LC) is given for $x \geq 0$ by:

$$y = \Pr(T \leq x + LC | T > LC) = 1 - \Pr(T > x + LC | T > LC)$$
$$= 1 - \frac{\Pr(T > x + LC)}{\Pr(T > LC)} = 1 - \frac{S_2(x + LC)}{S_2(LC)}$$

With the estimated shape parameter β and the scale parameter λ from the underlying Weibull model, this results in:

$$y = 1 - \exp\left(-\left(\frac{x + LC}{\lambda}\right)^{\beta}\right) \Big/ \exp\left(-\left(\frac{LC}{\lambda}\right)^{\beta}\right)$$
$$\equiv \log(1 - y) = -\left(\frac{x+LC}{\lambda}\right)^{\beta} + \left(\frac{LC}{\lambda}\right)^{\beta}$$
$$\equiv x = \lambda \times \left(\left(\frac{LC}{\lambda}\right)^{\beta} - \log(1 - y)\right)^{1/\beta} - LC$$

If y is replaced with a random variable with uniform distribution between 0 and 1, the random variable x can be generated using the inverse transformation technique from

$$\lambda \times \left(\left(\frac{LC}{\lambda}\right)^{\beta} - \log(1 - rand('uniform')\right)^{1/\beta} - LC.$$

In addition, the estimated hazard rates for the finerenone group will be inflated by a known factor *k*, with 1000 simulations performed at each rate of inflation. Inflation will be done until the upper limit of the confidence interval of the combined estimate of the hazard ratio crosses 1.

Details:

The imputation above is repeated with different inflation factors for the subjects in the finerenone group. The individual hazard rates is multiplied with *k*, leading to

$$k \times (\beta/\lambda)\ (t/\lambda)^{\beta-1}$$

with new scale parameter $s = \lambda/(k^{1/\beta})$.

9.2 Appendix B Technical details missing data imputation model for UACR

The evaluation of the secondary efficacy endpoint UACR is the ratio of change from baseline to Month 4 in UACR. In case of a missing value at Month 4 (on or off treatment) the subject will be excluded from this analysis. This strategy leads to unbiased estimates of the treatment effect only if the missing data are "missing completely at random" (MCAR), i.e. the missingness is independent of both observed and unobserved outcomes. Although this condition might hold approximately, they are unlikely to hold exactly. In accordance with the EMA "Guideline on missing data in confirmatory clinical trials" other ways of handling missing data will be investigated.

The frequency, proportion and the reasons for dropouts will be tabulated within the main analysis. In addition the frequency and proportion for time of dropout as well as for the overall pattern of missing observations will be given. To investigate whether missingness seems to be "missing at random", the mean UACR at baseline will be summarized for those with and without a post-baseline UACR. These analyses will be done for all patients grouped together, and stratified by treatment. For the ratio of change to Month 4 analysis, as UACR is only recorded at baseline and then at Month 4, the pattern of UACR over time up to Month 4 cannot be systematically investigated further. However, if there are sufficient dropouts with UACR at Month 4, the ratio of change to from baseline to Month 4 will be summarized by those still on study treatment and those having terminated study treatment.

Missing not at random (MNAR) means that missingness depends both on observed and unobserved outcomes. This requires an explicit model for the patient's statistical behavior after drop-out. A fixed increase above the group average for patients dropping out can be assumed, where the pattern and size of the increase is varied across several sensitivity analyses. This is sometimes called the delta method of sensitivity analysis in a pattern mixture framework (see reference 8, Chapter 6).

The following scenarios of no penalty and fixed penalties after drop-out will be investigated, where different values are assigned to the treatment groups:

Penalties after drop-out (ratio baseline : Month 4), based on an expected ratio from baseline to Month 4 on finerenone of 0.70.

- Placebo 0 / finerenone 0
- Placebo 0 / finerenone +0.15
- Placebo 0 / finerenone +0.30
- Placebo +0.15 / finerenone +0.15
- Placebo +0.15 / finerenone +0.30
- Placebo +0.30 / finerenone +0.30
- Tipping point analysis: Placebo 0 / finerenone +X, X is the value where the p-value of the difference between placebo and finerenone becomes 0.05 (two-sided).

To properly account for the incompleteness of the data, multiple imputation will be used to draw sets of completed data that will then be modified according to the scenarios given above. Multiple imputation will be done using SAS PROC MI using the following generic code, where stratification group are coded as dummy variables using the first factor in each stratification group as the respective reference groups:

```
PROC MI data=<dataset> seed=<study number> nimpute=50
out=<output dataset>;
     BY trtgrpn;
     MCMC niter=500 nbiter=500;
     VAR logUACRbase logUACRmonth4 <stratumn>
RUN;
```

After modifying the completed data sets according to the scenarios, the ANCOVA of the main analysis will be performed at Month 4 for each completed data set. The results are them combined using SAS PROC MIANALYZE.

For each scenario, the ratio of treatment difference at Month 4 from baseline will be given with a 95%-confidence interval and a p-value. In addition, the minimum and maximum of observed treatment differences over the imputed data sets will be presented.

9.3 Appendix C Region grouping

The following regions have been defined as used for stratification:

- North America,
- Latin America,
- Europe,
- Asia,
- Others

Region grouping is based on the location of a country to that continent, which in most cases is unambiguous. These rules have been used for the following countries:

- Israel and Saudi Arabia are geographically located in Asia, and have been allocated to that region.
- Turkey and Russia are located in both Europe and Asia; they have been allocated to region Europe.
- Mexico is allocated to region Latin America.
- Puerto Rico is in the Caribbean and an unincorporated US territory; it is therefore linked to the US and included in North America.
- Others includes countries in Africa and Australasia: South Africa, Australia and New Zealand.

The IxRS stratification scheme has Italy, China and Japan as their own individual stratification groups, due to the ioxehol substudy (Italy) and regulatory requirements to demonstrate consistency with the main study (China, Japan). For region groupings, Italy will be included with Europe, China and Japan with Asia.