Document Type:	Statistical Analysis Plan
Official Title:	A multicenter, randomized, double-blind, parallel-group, placebo-controlled study to evaluate the efficacy and safety of finerenone on morbidity and mortality in participants with heart failure (NYHA II-IV) and left ventricular ejection fraction = 40% (LVEF = 40%)
NCT Number:	NCT04435626
Document Date:	20 Jun 2024

Title Page

A multicenter, randomized, double-blind, parallel-group, placebo-controlled study to evaluate the efficacy and safety of finerenone on morbidity and mortality in participants with heart failure (NYHA II-IV) and left ventricular ejection fraction ⩾ 40% (LVEF ⩾ 40%)

FINEARTS-HF: **FIN**erenone trial to investigate **E**fficacy and s**A**fety superio**R** to placebo in pa**T**ient**S** with **H**eart **F**ailure

Bayer study drug	Finerenone / BAY 94-8862		
Study purpose:	Efficacy and safety		
Clinical study phase:	III	**Date:**	20 JUN 2024
Study No.:	20103	**Version:**	3.0
Authors:	PPD PPD PPD PPD PPD PPD PPD PPD PPD		

Confidential

The information provided in this document is strictly confidential and is intended solely for the guidance of the clinical investigation. Reproduction or disclosure of this document, whether in part or in full, to parties not associated with the clinical investigation or its use for any other purpose without the prior written consent of the sponsor is not permitted.

Throughout this document, symbols indicating proprietary names (®, TM) are not displayed. Hence, the appearance of product names without these symbols does not imply that these names are not protected.

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

Table of Tables

Table of Figures

Abbreviations

ACEI	Angiotensin-converting enzyme inhibitor
ACM	All-cause mortality
AE	Adverse event
ANCOVA	Analysis of covariance
ARB	Angiotensin receptor blocker
ARNI	Angiotensin receptor neprilysin inhibitor
ATC	Anatomical therapeutic chemical
BMI	Body mass index
BP	Blood pressure
CABG	Coronary artery bypass graft
CEC	Clinical event committee
CI	Confidence interval
CKD	Chronic kidney disease
CKD-EPI	Chronic Kidney Disease Epidemiology Collaboration
CLIPS	Clinical Pharmacology Standards
COPD	Chronic obstructive pulmonary disease
CSR	Clinical study report
CSS	Clinical Summary Score
CoV	Coefficient of Variation
CV	Cardiovascular
CYP3A4	Cytochrome P450 isoenzyme 3A4
DAOH	Days alive and out of hospital
DBP	Diastolic blood pressure
DMC	Data monitoring committee
ECG	Electrocardiogram
eCRF	Electronic case report form
eGFR	Estimated glomerular filtration rate
EoS	End of study
EQ VAS	EuroQol Group visual analogue scale
EQ-5D-5L	EuroQol Group 5-dimension, 5-level questionnaire
FAS	Full analysis set
GCP	Good clinical practice
HF	Heart failure
HHF	Hospitalization for heart failure
HR	Hazard ratio
hs-TNT	High sensitive troponin T
ICF	Informed consent form
ICH	International Council for Harmonisation
ITT	Intention-to-treat
KCCQ	Kansas City Cardiomyopathy Questionnaire
LLOQ	Lower limit of quantification
LVEF	Left ventricular ejection fraction
LWYY	Lin, Wei, Yang and Ying
MedDRA	Medical Dictionary for Regulatory Activities
MLG	MedDRA labeling groupings
MR	Mineralocorticoid receptor
MRA	Mineralocorticoid receptor antagonist
NT-proBNP	N-terminal prohormone B-type natriuretic peptide
NYHA	New York Heart Association
OATP	Organic anion transporting polypeptides
OD	Once daily
OSS	Overall Summary Score
PBMQ	Project-specific Bayer MedDRA Queries
PCI	Percutaneous coronary intervention
PD	Pharmacodynamic

PGIC	Patient Global Impression of Change
PGIS	Patient Global Impression of Severity
PK	Pharmacokinetic
PKS	Pharmacokinetic analysis set
PLS	Physical Limitation Score
PT	Preferred term
PWP	Prentice, Williams and Peterson
RR	Rate ratio
SAE	Serious adverse event
SAF	Safety analysis set
SAP	Statistical analysis plan
SBP	Systolic blood pressure
SD	Standard deviation
SGLT-2	Sodium-glucose transport proteins-2
SMQs	Standardized MedDRA queries
SOC	System organ class
SPA	Special protocol assessment
TEAE	Treatment-emergent adverse event
TIA	Transitory ischemic attack
TLFs	Tables, Listings and Figures
TSS	Total Symptom Score
UACR	Urinary albumin-to-creatinine ratio
ULOQ	Upper limit of quantification
WHO-DD	World Health Organization Drug Dictionary
WLW	Wei, Lin and Weissfeld

1. Introduction

This Statistical Analysis Plan (SAP) is based on the following document:

Clinical Study Protocol Amendment 20103 dated 26 OCT 2022

This SAP describes the statistical analysis of the double-blind placebo-controlled study treatment phase. An independent data monitoring committee (DMC) will be involved in the review of data for safety and efficacy as will be described in the DMC Charter. Blinded adjudication of clinical outcomes will be performed by an independent Clinical Event Committee (CEC), as will be described in the CEC Charter.

2. Study Objectives

Please refer to the study protocol for details on finerenone and on heart failure.

Study 20103 will be the first large-scale, long-term outcome study investigating the efficacy and safety of the non-steroidal mineralocorticoid receptor antagonist (MRA) finerenone on morbidity and mortality in participants with heart failure (New York Heart Association [NYHA] class II-IV) and left ventricular ejection fraction ≥40% (LVEF ≥40%), in comparison to placebo and in addition to standard-of-care therapy for congestion and comorbidities. Primary endpoint includes Cardiovascular (CV) death and total (first and recurrent) heart failure (HF) events (hospitalizations for heart failure [HHF] or urgent HF visits) in HF patients (NYHA class II–IV) and LVEF ≥40%. Secondary endpoints will include: total HF events; improvement in NYHA class from baseline to Month 12; change from baseline to Month 6, 9 and 12 in total symptom score (TSS) of the Kansas City Cardiomyopathy Questionnaire (KCCQ); time to first occurrence of composite renal endpoint: sustained decrease in estimated glomerular filtration rate (eGFR) ≥50% relative to baseline over at least 4 weeks, or sustained eGFR decline <15ml/min/1.73m^2 or initiation of dialysis or renal transplantation; time to all-cause mortality (ACM); and the safety and tolerability of finerenone.

The Phase III program with finerenone in patients with type 2 diabetes and clinical diagnosis of chronic kidney disease (CKD) encompasses 2 placebo-controlled, large-scale, long-term outcome trials: FIDELIO-DKD, the first large-scale, long-term outcome trial that examined whether finerenone can slow the progression of kidney disease and FIGARO-DKD which is examining the effects of finerenone on CV outcomes. The pooled analysis of two complementary trials (FIDELITY-DKD) comprising 13026 patients with a broad spectrum of CKD and type 2 diabetes provides robust evidence of both cardiovascular and kidney protection with finerenone vs. placebo. Across the FIDELITY-DKD population, the relative risk reduction was 14% for the composite cardiovascular outcome of cardiovascular death, non-fatal myocardial infarction, non-fatal stroke, or hospitalization for heart failure and 23% for the composite kidney outcome of kidney failure, a sustained ≥57% decrease in estimated glomerular filtration rate from baseline over ≥4 weeks, or renal death. While MRAs are indicated for the treatment of patients with chronic symptomatic heart failure with reduced ejection fraction, such patients were excluded from the FIDELIO-DKD and FIGARO-DKD studies. The FIDELITY-DKD analysis provides evidence that finerenone use in patients with CKD and type 2 diabetes, a population at high risk of developing heart failure, significantly reduces the risk of developing HHF. In the FIDELITY-DKD analysis, reduction in HHF was

the main driver of the cardiovascular benefit with finerenone, with a relative risk reduction of 22% vs. placebo (p= 0.0030), in a population that excluded patients with chronic symptomatic heart failure with reduced ejection fraction at the run-in visit.An inappropriate release of aldosterone contributes to target organ damage found in HF, myocardial infarction, chronic renal failure, and hypertension. The extensive expression of the mineralocorticoid receptor (MR) in the CV and renal systems, including the heart, endothelial cells, vascular smooth muscle cells, and kidney mesangial cells, provides further evidence for the role of aldosterone in CV and renal injury.

Blockade of the action of aldosterone and potentially other MR ligands such as cortisol has been demonstrated to be of benefit in HF (Pitt et al. 1999, Zannad et al. 2010). Results from a short-term Phase IIb study (ARTS-HF Study 14564) suggest that treatment with finerenone in addition to standard therapy for HF with LVEF ≤40% improves mortality and CV morbidity outcomes; however, long-term outcome conclusive studies examining whether MRAs can prevent CV events are still lacking in this patient population. Study 20103 will be the first study to address these questions in the HF with LVEF ≥40% population.

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

- Demonstrate the superiority of finerenone to placebo in reducing the rate of CV death and total (first and recurrent) HF events (HHF or urgent HF visit) in HF patients (NYHA II–IV) and LVEF ≥40%.

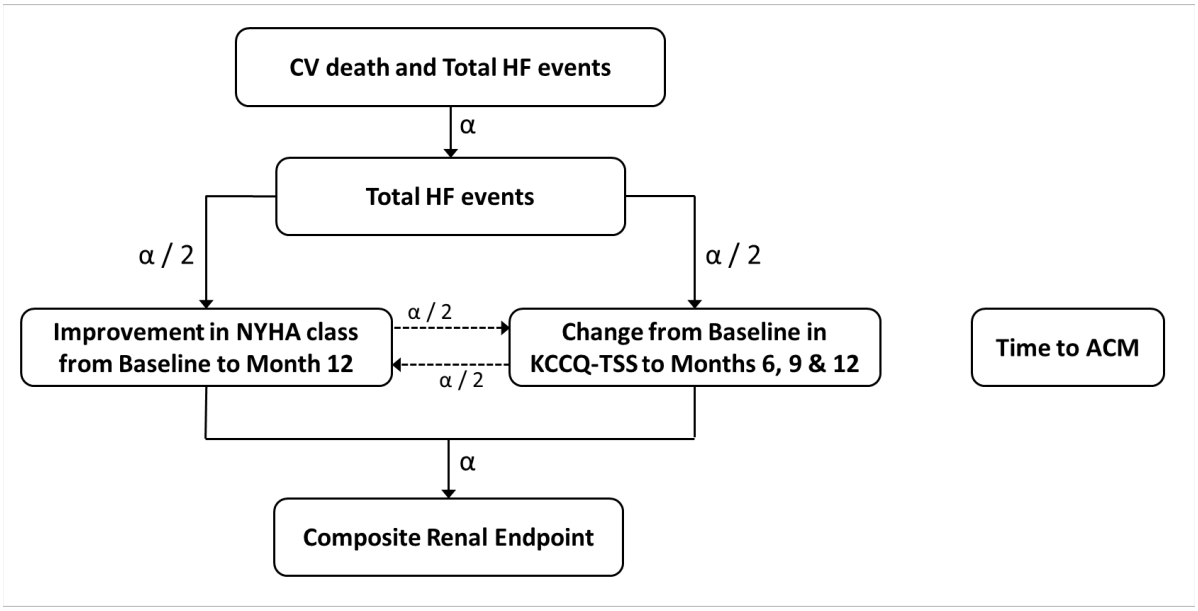
The **secondary objectives** of this study, irrespective of the testing procedure as defined in section 4.7, are to:

- Determine the superiority of finerenone to placebo with regard to each of the following:
    - Reducing the rate of total (first and recurrent) HF events
    - Improvement in NYHA class from Baseline to Month 12
    - Change from baseline to Month 6, 9 and 12 in TSS from KCCQ
    - Time to first occurrence of composite renal endpoint:

      sustained decrease in eGFR ≥50% relative to baseline over at least 4 weeks, or sustained eGFR decline <15ml/min/1.73m^2 or initiation of dialysis or renal transplantation
    - Time to ACM
- Assess the safety and tolerability of finerenone

The testing procedure for the primary and secondary endpoints is presented in Figure 2-1, and described in detail in Section 4.7.

Figure 2-1: Testing procedure for primary and secondary endpoints



In terms of the addendum to International Council for Harmonisation (ICH) E9 (ICH 2019), the five attributes of the primary estimand to address the primary objective of the study are as follows:

a. Population: as described by inclusion/exclusion criteria given in Section 5 of the protocol

b. Variable: Number of unfavorable events including CV death and total (first and recurrent) heart failure events

c. Treatment condition: Finerenone vs. placebo

d. Intercurrent events: There are three important intercurrent events to consider - Treatment discontinuation, CV death and non-CV death. For treatment discontinuation a treatment policy strategy will be applied, i.e. patients will be followed up for events after discontinuing treatment and events and follow-up time after discontinuation of treatment will be included in the analysis. CV death will be counted as both an outcome event as well as a censoring event, so that a combination of a composite and a while alive strategy is used. It is thus assumed that patients could have had further events for HF, if they had not died. This seems appropriate, as including into the model that no further HF events can occur after death, for example by censoring patients at the end of the study, would induce a bias in favor of a treatment group with more early deaths. Non-CV death is assumed to be a censoring event, since the treatment is not assumed to have an effect on these events and interest lies in the treatment effect on composite events while patients are alive.

e. Population-level summary: Ratio of exposure-weighted composite event rates between finerenone and placebo. Exposure-weighted refers to patients being weighted according to their follow-up time in determining the rate.

A pooled analysis of study 20103 with studies FIDELIO-DKD and FIGARO-DKD will be pre-specified and analyses will be described in a separate SAP. Additional analyses for specific scientific questions outside of the scope of the clinical study report (CSR) will be pre-specified in a separate scientific SAP.

3. Study Design

Study 20103 is a randomized, double-blind, parallel-group, placebo-controlled, multicenter, event-driven Phase III study with independently adjudicated clinical outcome assessments. This study will be conducted in patients with HF and LVEF ≥40%. The overall study design is displayed in Figure 3-1.

Participants will be randomized in a 1:1 ratio to either finerenone or placebo. The study is designed to be able to show an effect on the primary endpoint with a power of 90% at an alpha level of 5%. It was originally anticipated that 5500 participants will be randomized and approximately 6900 will be screened (screening failure rate of approximately 20%). A total of approximately 2375 total (first and recurrent) primary composite events are targeted. Due to blinded event rates being lower than those assumed in the sample size calculation, the planned number of randomized participants was increased to approximately 6000. The target number of primary composite events was not changed.

The anticipated duration of the study will be approximately 42 months, with a recruitment period of 24 months. However, as an event-driven study, the actual length of the study will depend on the observed event rates, the participant recruitment rate, and the length of the recruitment period.

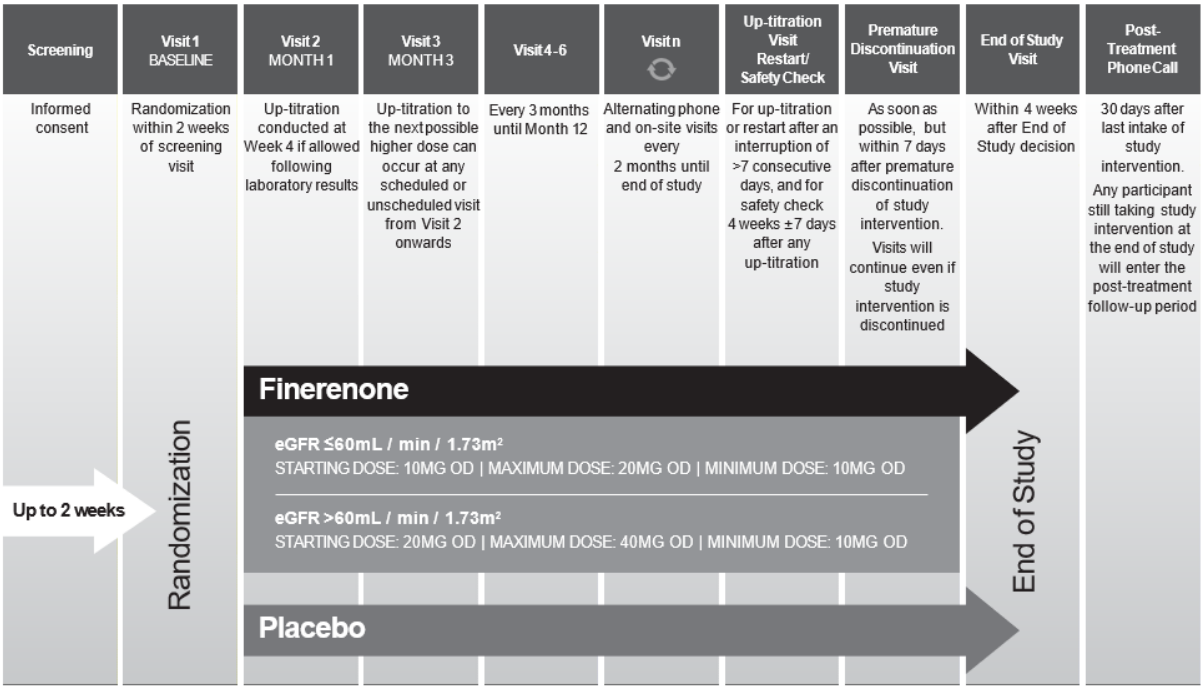
Enrolment in the trial may be capped based on the proportion of patients in certain LVEF categories, in each NYHA class, with/without atrial fibrillation, and by geographic region, among other variables, to ensure recruitment of a representative study population.

The randomization will be stratified by country/region and baseline LVEF (<60%, ≥60%).

Since all randomized participants (excluding those with critical GCP violations) belong to the Full Analysis Set (FAS) on which the efficacy analyses are based, 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 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: Study design



Screening Visit

After providing written informed consent, a Screening Visit to confirm the participant's eligibility will take place prior to randomization. The Screening Visit may take place on the same day as randomization (Visit 1).

Treatment Period

Following a screening period of up to 2 weeks, eligible participants will be randomized in a 1:1 ratio to either finerenone or placebo. Participants with an eGFR ≤60 mL/min/1.73 m² measured at baseline will start with 10 mg once daily (OD) (**dose level 1**) with a maximum maintenance dose of 20 mg OD (**dose level 2**), whereas participants with an eGFR >60mL/min/1.73 m² measured at baseline will start with 20 mg OD (dose level 2) with a maximum maintenance dose of 40 mg OD (**dose level 3**).

There will be at least 2 scheduled visits within the first 3 months from randomization: Visit 2 will take place after 1 month and Visit 3 will take place 3 months after randomization; thereafter, scheduled visits will occur every 3 months until Visit 6 at Month 12. After 1 year from randomization, telephone contact visits will take place at Month 14 and from then onwards every 4 months (i.e. Month 18, Month 22, etc.) alternating with on-site visits (i.e. Month 16, Month 20, etc.) until the end of the study is reached.

Up-titration is expected to occur after 4 weeks ± 7 days of treatment at Visit 2 (Month 1). Ideally, each participant will be on the maximum maintenance dose at this point. In the event of elevated potassium values, participants will be down-titrated to the next lower dose. Down-titrations can be performed at any time after the start of study intervention treatment, at any scheduled or unscheduled visit. At any scheduled or unscheduled visit, the dose of study intervention may be increased to the next possible higher dose, based on serum/plasma potassium level and provided the participant was already on a stable dose for 4 weeks ± 7 days.

Participants will attend an additional unscheduled safety visit 4 weeks ± 7 days after each up-titration; potassium levels and renal function will be monitored at this safety visit. In addition to the protocol-specified visits, participants may be seen at any time throughout the study at the discretion of the investigator.

Any changes in the study intervention dose, including interruption/permanent discontinuation or restart of study intervention, must be recorded in the electronic case report form (eCRF).

It is planned that all randomized participants will remain in the study until either:

a. an instruction is received from the sponsor after the targeted number of primary endpoint events have occurred
or

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

After randomization, study intervention discontinuation does not constitute the participant's withdrawal from the study, and all participants should continue to be followed up. All randomized participants, including any participant who experiences an event considered for the pre-specified primary or secondary endpoints, should continue to receive double-blinded treatment until the study is completed, provided there are no safety grounds for discontinuing treatment.

Post-treatment Follow-up Period

The period between a participant's last intake of study intervention and last visit in the study is referred to as the 'post-treatment follow-up period'.

In case of premature discontinuation of study intervention, participants are expected to continue to attend all protocol-specified study visits, and are expected to perform all scheduled assessments as described in the Premature Discontinuation Schedule of Assessment in the protocol.

Any participant still taking study intervention at the point of end of study will enter the post-treatment follow-up period after stopping study intervention at the End of Study (EoS) Visit. For these participants, this phase will last 30 (+5) days, and will end upon completion of the Post-Treatment Visit (a telephone call visit).

4. General Statistical Considerations

4.1 General Principles

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

The analysis will be based on the Global Standard Tables (Version 4.0 or higher) and the Clinical Pharmacology Standards (CLIPS) (Version 1.2 or higher) where appropriate. Compound Standard Tables will be developed.

The validity of participants for allocation to various analysis sets will be assessed in an ongoing manner in blind review meetings and decisions will be documented in the blind review reports prior to unblinding.

A log-normal distribution is assumed for urinary creatinine and albumin, urinary albumin-to-creatinine ratio (UACR), and N-terminal prohormone B-type natriuretic peptide (NT-proBNP). 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, median, and maximum will be calculated for metric data. The geometric mean, SD and coefficient of variation (CoV) will be provided instead of the arithmetic mean and SD for the variables where log-normal distributions are assumed, as follows:

$$CoV = \sqrt{\exp(SD_{ln}^2) - 1}$$

SD_{ln} being the standard deviation of the log-transformed values. Frequency tables will be generated for categorical data.

The laboratory parameter eGFR will be calculated based on the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) formula (Levey et al., 2009) for all analyses specified in this SAP. For patients recruited in Japan, the Japanese formula adjustment will be applied (Horio et al., 2010).

All participants will be analyzed according to the planned treatment group in FAS per the intention-to-treat (ITT) principle. All participants will be analyzed according to the actual treatment in the Safety Analysis Set (SAF). If a participant 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.

Only adjudicated outcome events will be used for analysis of the primary and secondary efficacy variables (except for sustained decrease in eGFR ≥50% relative to baseline over at least 4 weeks and sustained eGFR decline <15ml/min/1.73m^2 which will be derived from central laboratory measurements and not be adjudicated); Section 6.2.4 specifies how investigator-reported outcomes for these variables will be summarized. Outcome events for exploratory efficacy variables (e.g. non-fatal myocardial infarction, non-fatal stroke, etc.) will not be adjudicated.

The stratified analyses mentioned in this SAP will be conducted in consideration of the randomization stratification factors:

- LVEF: <60%, ≥60%
- Pooled region for stratified analyses: randomization will be stratified by country/region, for the analyses individual countries/regions will be combined into pooled regions as follows:
    - Western Europe and Oceania: Australia, Austria, Denmark, Germany, Israel*, Netherlands, New Zealand, United Kingdom
    - Southwestern Europe: Italy, Portugal, Spain
    - Central Europe: Czechia, Hungary, Poland, Slovakia
    - Southeastern Europe: Bulgaria, Greece, Romania, Turkey
    - Northeastern Europe: Finland, Latvia, Lithuania, Russia, Ukraine
    - Asia: China, Hong Kong, India, Japan, Malaysia, South Korea, Taiwan
    - North America: Canada, United States of America
    - Latin America: Argentina, Brazil, Colombia, Mexico

*Although not geographically located in Western Europe or Oceania, Israel has been included in this pooled region.

In case of issues (e.g. model convergence) with using these predefined pooled regions at the final analyses, the pooled regions may be further combined. Any such changes will be described in the CSR.

All participants will be analyzed according to their correct stratification category. In case of a large number of stratification errors (≥5% of all patients in the FAS), the primary analysis will also be repeated based on the stratification category used in the randomization as a sensitivity analysis.

In case a death cannot be clearly adjudicated as CV-related or non-CV related, it will be adjudicated as undetermined death. Undetermined deaths will be handled as non-CV deaths for the analysis, unless otherwise specified. Where applicable, the combined results of non-CV deaths and undetermined death will be displayed along with results for the individual categories.

4.2 Handling of Dropouts

A participant who has been randomized and discontinues study participation prematurely for any reason, either during study treatment or during post-treatment follow-up, is defined as a 'dropout', even if no study drug has been taken. Dropouts will not be replaced.

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

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

The number of participants 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 "patients still participating in study" and "patients still on study treatment" will be provided.

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

- baseline characteristics
- potential differences between the treatment groups in the proportion of participant 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 eCRF.

General Rules

When appropriate, the following rules will be implemented so as not to exclude participants 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 a participant's final visit will be considered to have started prior to study medication start and ended after stopping 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 earliest and latest possible dates, i.e.:

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

If a participant 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 used.
- In case the range of possible values is further restricted, e.g. because a patient is randomized in the month in which the day is missing or the participant 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.

In case a death date is completely missing, it will be imputed on the basis of the last known contact when the participant was still alive and the first known contact when the participant was dead (e.g. from the participant 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-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.

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

If intensity of the AE is missing, the event will be considered as severe. 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 non-binding interim analysis for futility is planned when approximately 30% (~710) of the required total number of primary endpoint events have been observed. If the observed rate ratio (RR) on the primary endpoint is above 0.95, the trial is planned to be stopped for futility. This gives a probability of approximately 69% to stop under the null hypothesis (i.e. no treatment effect on the composite of HHFs and CV deaths) and leads to a loss in power of less than 1% under the alternative hypothesis of the treatment effect assumed for the sample size determination. No adjustment for this loss in power will be made.

The futility analysis given above is considered to be non-binding, the DMC will be asked to also consider important secondary efficacy endpoints as well as safety in their assessment.

In addition, one formal interim analysis for efficacy is planned when approximately 2/3 (~1580) of the required total number of primary 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 (of the test statistic) in the analysis of the primary efficacy endpoint (two-sided p-value <0.0027) at the interim analysis. In addition, a nominal significant effect on the CV death component should be present (two-sided p-value<0.05) at the interim analysis. Note: The criterion for CV death would not be considered to prove formal statistical significance, as it does not keep the alpha level. It has been added so that the trial is only stopped at the interim if there is at least a certain amount of evidence of a beneficial treatment effect on CV death.

See Section 4.7 for a detailed description of the testing procedure, including an adjustment of the significance level for the interim analysis.

If the study is stopped early, the primary analysis reported in the CSR will consider all events up to the participants' 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. For this sensitivity analysis, censoring dates after the datacut date will be reset to the datacut date.

A detailed plan for the routine DMC safety analyses and the interim analysis will be covered in the DMC charter, the analysis planned to be provided to the DMC will be described in a separate 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. The sponsor will oversee and discuss with the Steering Committee overall blinded event rates to ensure that they are in line with protocol assumptions. If overall event rates are lower than expected, consideration will be given to altering the study design, such as increasing the sample size or extending the study duration without knowledge of any treatment effect.

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, if no assessment from the central laboratory prior to first intake of study drug is available. Otherwise baseline will be missing.

If more than one measurement was planned for a scheduled time point, for example blood pressure (BP) measurements and heart rate, the mean value of the last set of measurements per time point prior to randomization will be used as the baseline value. In case of repeated measurements for pre-treatment visits and Visit 1 (Day 1; baseline), the closest measurement prior to the randomization will be used for analysis instead of the scheduled measurements.

When the Screening and Baseline visits are performed on the same day for a participant, the following assessments (scheduled to be performed at both the Screening and Baseline visits) will only be performed once and the data will appear under Screening:

- NYHA class
- Vital signs
- Local lab potassium and creatinine
- Pregnancy test

Where necessary, these data will be considered in the derivation of baseline values per the rules above.

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 Time Window for Efficacy Events

Events for time to first / recurrent analyses (e.g. primary efficacy endpoint) 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, events will be counted up to the day of the last visit when information on the component is available.

4.5.4 Annual Rate of Recurrent Events

The annual rate of a recurrent event for an individual patient is calculated as:

Annual rate = (Total number of events) / (Follow-up time (days) / 365.25)

4.5.5 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 unless otherwise specified and will be listed only. For example, as described above, local values will be used in the derivation of baseline, if no central measurement is available (cf. Section 4.5.1).

At all visits post-randomization and if not stated otherwise, only the values at scheduled measurements will be used for analysis.

For the derived visit “Any time post baseline” (applicable for efficacy) this will include any measurement after randomization, including unscheduled assessments. For the derived visit

"Any time on treatment" (applicable for efficacy), only assessments on or after study medication start date until 30 days after last study drug administration, including unscheduled assessments, will be considered. For safety, "Any time treatment-emergent" will include measurements on or after study medication start date until 3 days after last study drug administration, including unscheduled assessments.

For values which are < LLOQ (Lower limit of quantification), half the value of the LLOQ will be used for analysis. Differences between two values < LLOQ will be assigned values of 0. Ratios between two 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.

In case of non-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.

4.5.6 Subgroup Analyses

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

Subgroups based on stratification factors

- Pooled region for subgroup analysis (Western Europe, Oceania and Others; Eastern Europe; Asia; North America; Latin America); the subgroup Western Europe, Oceania and Others combines the groups Western Europe and Oceania and Southwestern Europe from the pooled region for stratified analyses. The subgroup Eastern Europe combines the groups Central Europe, Southeastern Europe, and Northeastern Europe from the pooled region for stratified analyses.
- LVEF (<60%, ≥60%)

Key subgroups

- Baseline serum potassium value (≤ 4.5, > 4.5 mmol/L)
- eGFR category at baseline (eGFR <60, ≥60 mL/min/1.73 m²)
- Atrial fibrillation at baseline electrocardiogram (ECG) (present, absent)
- Diabetes Mellitus at baseline (present, absent)
- Index HF event (very recent (≤ 7 days before randomization), recent (> 7 days – ≤ 3 months), > 3 months or no index HF event).

Other subgroups

- Race (white, black, Asian, other)
- Sex (male, female)
- Age (≤ median vs. > median)
- Baseline body mass index (BMI) (< 30 vs. ≥30 kg/m^2)

- Systolic blood pressure (SBP) at baseline (≤ median vs. > median)
- Angiotensin-converting enzyme inhibitor (ACEI), angiotensin receptor blocker (ARB) or angiotensin receptor-neprilysin inhibitor (ARNI) use at baseline (yes, no)
- Treatment with sodium-glucose transport proteins-2 inhibitor (SGLT-2i) (yes, no)
- NYHA functional class at baseline (II, III/IV)
- Baseline NT-proBNP (≤ median vs. > median)
- Baseline UACR (<30 vs. ≥30 mg/g)

For subgroups split by median, the FAS will be taken as reference population for derivation of median.

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

4.6 Blind Review

The results of the final data assessment will be documented in the final list of important deviations, validity findings and assignment to analysis set(s). Any changes to the statistical analysis prompted by the results of the review of study data will be documented in an amendment and, if applicable, in a supplement to this SAP.

4.7 Testing Procedure and Multiplicity Adjustment

If the interim analysis shows clear and consistent benefit in the finerenone treatment group (defined as two-sided p-value <0.0027 for the primary efficacy endpoint and two-sided p-value <0.05 for the CV death component at the formal interim analysis for efficacy) the DMC may recommend early stopping of the study for success (see Section 4.4 for full details).

If the study is stopped early for success: the final analysis for the primary endpoint will be performed at an overall two-sided significance level of 0.0027. Additionally, the secondary endpoints of:

- Total (first and recurrent) HF events
- Improvement in NYHA class from Baseline to Month 12
- Change from baseline to Month 6, 9 and 12 in TSS from KCCQ
- Time to first occurrence of composite renal endpoint

will be formally tested. The testing strategy of the secondary endpoints is as follows:

1. Total HF events will be tested at the 0.0027 two-sided significance level.
2. If the hypothesis of the secondary endpoint total HF events is rejected, the NYHA class and KCCQ endpoints will be tested using the Bonferroni-Holm procedure, i.e. if at least one of the hypotheses of the two endpoints NYHA class and KCCQ can be rejected at the two-sided (0.0027/2) significance level, the remaining of the two endpoints will be tested at the 0.0027 significance level.
3. If the hypotheses for all previous secondary endpoints are rejected, the composite renal endpoint will be tested at the 0.0027 significance level.

If the test for any endpoint produces a non-significant result, the testing of the remaining endpoints further down in the procedure will be performed in an explorative manner only.

Furthermore, the second component of the primary endpoint (CV deaths) will be tested at the 0.0027 two-sided significance level outside of the alpha-preserving procedure for the primary and other secondary efficacy endpoints (Total HF events, NYHA class, KCCQ, renal composite).

If the study is not stopped early for success: a group sequential design with a single interim analysis when 2/3 of the information is available with a stopping rule of two-sided $p < 0.00270$ would require a small adjustment to the alpha level at the final analysis to maintain the overall significance level at 0.05. For an information fraction of 2/3, the adjusted alpha level of 0.04967 applies. If the study is not stopped early for success a p-value of $p < 0.04967$ is therefore required at the final analysis to achieve formal statistical significance.

If the primary hypothesis is rejected, the secondary endpoints of:

- Total (first and recurrent) HF events
- Improvement in NYHA class from Baseline to Month 12
- Change from baseline to Month 6, 9 and 12 in TSS from KCCQ
- Time to first occurrence of composite renal endpoint

will be formally tested with the following test strategy:

1. Total HF events will be tested at the 0.04967 two-sided significance level.
2. If the hypothesis of the secondary endpoint total HF events is rejected, the NYHA class and KCCQ endpoints will be tested using the Bonferroni-Holm procedure, i.e. if at least one of the hypotheses of the two endpoints NYHA class and KCCQ can be rejected at the two-sided (0.04967/2) significance level, the remaining of the two endpoints will be tested at the 0.04967 significance level.
3. If the hypotheses for all previous secondary endopints are rejected, the composite renal endpoint will be tested at the 0.04967 significance level.

If the primary hypothesis is not rejected, these tests will be performed in an explorative manner only; similarly, if the test for any secondary endpoint produces a non-significant result, the testing of the remaining endpoints further down in the procedure will be performed in an explorative manner only.

Furthermore, if the primary hypothesis is rejected then the second component of the primary endpoint (CV deaths) will also be tested at the 0.04967 two-sided significance level outside of the alpha-preserving procedure for the primary and other secondary efficacy endpoints (total HF events, NYHA class, KCCQ, renal composite).

Regardless of whether the study is stopped early for success: as a hard endpoint and objective indicator of benefit-risk, time to ACM will be tested at a two-sided significance level of 0.05, after the rejection of the primary hypothesis. Testing of time to ACM will thus be done outside of the alpha-preserving procedure for the primary and other secondary efficacy variables (total HF events, NYHA class, KCCQ, renal composite).

5. Analysis Sets

5.1 Assignment of Analysis Sets

Final decisions regarding the assignment of participants to analysis sets will be made during the review of study data and documented in the final list of important deviations, validity findings and assignment to analysis set(s) (see Section 4.6).

For purposes of analysis, the following populations are defined:

Population	Description
Enrolled	All participants who sign the informed consent form (ICF)
Randomly assigned to study intervention	All participants randomly assigned to study intervention
Full analysis set (FAS)	All randomized participants. Participants will be analyzed according to the intervention they were randomized to. The only potential reasons for exclusion would be a clearly erroneously randomization, or major good clinical practice (GCP) violations, for example, a suspicion of fraud.
Safety analysis set (SAF)	All participants in the FAS who take at least 1 dose of study intervention. Participants will be analyzed according to the intervention they actually received.
Pharmacokinetic analysis set (PKS)	All finerenone-treated participants (with the exception of participants excluded on the grounds of critical GCP violations) with at least 1 valid finerenone plasma concentration and without validity findings which would interfere with the evaluation of the pharmacokinetic (PK) data.
Listing-only set	All participants enrolled who were not randomized or were excluded from the FAS. Their data is provided as individual participant data listings and will not be included in any statistical analyses.

6. Statistical Methodology

6.1 Population Characteristics

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

6.1.1 Disposition

The number of participants enrolled, randomized and valid for the FAS and SAF will be summarized overall and by treatment group, country/region and study site. The number of participants 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 participants with important deviations will be presented

overall, by country/region 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 (continuous and categorized by 40-<65, 65-<75, 75-<85, ≥85), sex, race, pooled region (for stratified analysis and for subgroup analysis), body weight (continuous and categorized by <60, 60-<90, ≥90 kg), body height, BMI (continuous and categorized by <30 vs. ≥30 as well as by <18.5, 18.5 to <25, 25 to <30, 30 to <35, ≥35 kg/m^2), hip and waist circumference and waist-hip ratio, smoking history (never, former, current smoker) and alcohol consumption.

Other baseline characteristics include baseline

- LVEF (continuous and categorized by <60% vs. ≥60% as well as <50, 50-<60, ≥60%)
- NYHA Class (II, III, IV)
- index HF event (randomized during/at HF event, very recent (≤ 7 days from randomization), recent (>7 days - ≤ 3 months), >3 months, no index; as well as ≤ 7 days from randomization, >7 days - ≤ 3 months, >3 months or no index event)
- type of (latest) index HF event with respective timing (hospitalization for heart failure, urgent HF visit, no index event)
- serum potassium (continuous and categorized by ≤4.5 mmol/L vs. >4.5 mmol/L)
- eGFR (calculated by CKD-EPI formula, Japanese formula adjustment made for participants recruited in Japan; continuous and categorized by <60 vs. ≥60 mL/min/1.73 m² as well as <45, 45 to <60, 60 to <90, ≥90 mL/min/1.73 m²),
- serum creatinine (continuous)
- SBP (continuous and categorized by <90, 90-<130, 130-<160, ≥160 mmHg)
- diastolic blood pressure [DBP] (continuous)
- heart rate (continuous)
- NT-proBNP (continuous)
- UACR (continuous and categorized by <30 vs. ≥30 as well as <30, 30-<300, ≥300 mg/g)
- history of LVEF<40% (yes [improved], no). For participants with a history of LVEF <40%, prior LVEF values will be summarized.
- Atrial fibrillation at baseline per ECG
- Diabetes Mellitus at baseline
- ACEI, ARB or ARNI use at baseline
- SGLT-2i use at baseline.

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 the SAF if ≥5% of randomized patients do not take at least one dose of study intervention (i.e. are in the FAS but not the SAF).

As stated in Section 4.2, demographics and other baseline characteristics will also be presented separately for participants prematurely discontinuing the study and for participants permanently discontinuing study treatment.

6.1.3 Medical History

Medical history will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). Medical history will be presented for each MedDRA Primary System Organ Class (SOC) and Preferred Term (PT) by treatment group and overall in a summary table. Additional medical history terms by the following Standard MedDRA Queries (SMQs), Project-specific Bayer MedDRA Queries (PBMQs), Bayer MedDRA Labeling Groupings (MLGs) or selected PTs will also be presented:

- Hyperlipidemia (MLG)
- Hypertension (MLG)
- Type 2 diabetes mellitus (PT)
- Atrial fibrillation/flutter (PBMQ)
- Ischemic Stroke/Transitory Ischemic Attack (TIA) (PBMQ)
- Myocardial Infarction (MLG)
- Coronary Artery Disease (PBMQ)
- Peripheral Arterial Occlusive Disease (PBMQ)
- Cardiac Failure (MLG)
- Coronary Artery Bypass Graft (CABG) (PBMQ)
- Chronic Obstructive Pulmonary Disease (COPD) (PT)
- Percutaneous coronary intervention (PCI) (PBMQ)
- COVID-19 (SMQ narrow)
- Hepatic cirrhosis (PT)
- Sleep Apnea Syndrome (PT)
- Chronic Kidney Disease (CKD) (PT)

The medical history tables will be repeated for SAF.

6.1.4 Concomitant Medications

Concomitant medications will be coded using the World Health Organization Drug Dictionary (WHO-DD). The number of participants who took at least one concomitant medication and the number of participants who took at least one medication that started before administration of study drug will be presented by treatment group and overall using anatomical therapeutic chemical (ATC) classes and subclasses.

These tables will be repeated summarizing the number of participants with medications in the Standard drug groups of interest. In addition, the number of participants who took at least one concomitant medication that started after start of study drug and the number of participants who took at least one medication ongoing at baseline (i.e. starting before or on the day of randomization and ending at least one day after the day of randomization) will be presented by treatment group and overall. Standard drug groups of interest are:

- ACEIs and ARBs
- ARNIs
- Beta-blockers
- Loop diuretics
- Thiazide diuretics
- Digoxin
- Nitrates
- Potassium supplements
- Potassium lowering agents (including binders)
- Alpha blocking agents
- Calcium channel blockers
- Centrally acting antihypertensives
- Strong, unclassified, moderate, weak cytochrome P450 isoenzyme 3A4 (CYP3A4) inhibitors
- Strong, unclassified, moderate, weak CYP3A4 inducers
- Aspirin
- Statins
- MRAs
- Organic anion transporting polypeptides (OATP) substrates

Anti-diabetic drugs

- Insulin and analogues
- SGLT-2 inhibitors
- Other anti-diabetic drugs (Dipeptidyl Peptidase 4 inhibitors or Glucagon-like peptide-1 agonists or Biguanides or Sulfonylureas or Alpha glucosidase inhibitors or Metiglinides or Thiazolidinediones.)

A participant will be counted only once within each ATC class / subclass or Standard drug group, respectively.

A listing will be provided including all medications classified as a strong, unclassified, moderate, or weak CYP3A4 inhibitor according to the drug groupings together with the respective classification information.

For potassium lowering agents, ACEIs, ARBs, diuretics and SGLT-2 inhibitors shift tables for changes of use for baseline vs. any time treatment-emergent will be provided.

The number of participants with MRA use during follow-up will be given by substance.

6.1.5 Treatment and Study Duration, Extent of Exposure and Compliance

The analyses described in this section will be repeated for the SAF if ≥5% of randomized patients do not take at least one dose of study intervention (i.e. are in the FAS but not the SAF). 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. The total duration in patient-years will be provided. In addition, treatment duration will be categorized to ≤ 1 month, 1 – 3 months, >3 – 6 months, >6 – 12 months, and then further six-monthly intervals, and presented with the corresponding number and percentage of participants. Cumulative treatment duration will be categorized to at least one dose, at least 1 month, at least 3 months, then further 3 monthly intervals. A table will be presented with the absolute and relative frequencies of participants still on study medication at each visit. Kaplan-Meier plots for "patients still on study treatment" will be provided, as also described in Section 4.2.

The above analyses will be repeated for study duration, from the day of randomization to the EoS 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. The table will be repeated by participants starting on 10 mg and 20 mg, respectively.

The number and percentage of participants on each dose level (blinded) will be summarized by visit and treatment group, overall and differentiated by participants starting on 10 mg or 20 mg. The overall titration status, regardless of actual or sham up-titration, will be summarized with absolute and relative frequencies per treatment group, overall and differentiated by patients starting on 10 mg or 20 mg. In addition, the number of patients with study drug down-titrated or temporarily interrupted (dose recorded as 0 mg) as well as associated reason will be summarized with absolute and relative frequencies per treatment group, overall and differentiated by participants starting on 10 mg and 20 mg, respectively.

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

100 * Number of tablets taken / Number of planned tablets.

The number of planned tablets will be calculated as follows:

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

For participants who withdraw prematurely from the study drug, compliance will be calculated up to the time of last dose.

The overall compliance will be summarized descriptively by treatment group and overall. In addition, percentage compliance will be categorized into three 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

Events that could potentially fulfill the criteria for primary efficacy variables during the study will be evaluated by the CEC. Definitions of individual endpoints (e.g. CV death) will be provided in the Endpoint Manual.

6.2.1.1 Primary Efficacy Variable: Primary Analysis

The primary Estimand as defined in Section 2 will be used for the primary analysis. The primary endpoint is the composite of CV death and total (first and recurrent) HF events (HHF or urgent HF visit) in HF patients. The primary analysis of this endpoint will be performed in the FAS using the planned treatment group, in line with the ITT principle.

Participants without an event of the primary composite endpoint at the time of analysis will be censored at the date of their last contact or date of non-CV death.

The number and incidence rate of primary endpoint events and censoring events (i.e. non-CV death) will be summarized per treatment group, both overall and per individual event category. 95% CIs of the incidence rates will be derived based on a Poisson model with robust variance estimator.

The primary analysis of the primary composite endpoint will be based on a stratified Andersen-Gill model (Andersen, 1982) including treatment group as fixed effect and including pooled region for stratified analyses and baseline LVEF (<60%, ≥60%) as stratification factors. Robust standard errors (sandwich estimator) will be used to account for correlations of event times within a participant. As shown by Lin et al. 2000, the Andersen-Gill model with robust standard errors can be interpreted as a proportional rates model. After the authors of the paper, the model is also referred to as Lin, Wei, Yang and Ying (LWYY) model. Let θ be the RR for the finerenone versus placebo group. In order to evaluate whether finerenone is superior to placebo in reducing the rate of the composite event of CV death and total HF events the following null hypothesis will be tested using the model above (see Section 4.7 for details regarding the nominal significance level):

$$H_0: \theta = 1 \text{ versus } H_1: \theta \neq 1,$$

where θ < 1 represents a treatment benefit of finerenone over placebo.

A point estimate of the RR together with a 95% CI will be presented alongside the point estimate and hazard ratio for the censoring event of non-CV death, calculated using a stratified Cox proportional hazards model.

The primary analysis method has been investigated with extensive simulation studies and it has been confirmed that it keeps the alpha level and has good operating characteristics across a range of plausible scenarios. A small adjustment will be made to the nominal significance level and the critical value at the final analysis to take into account the interim analysis (see Section 4.7 for details). No adjustment to the sample size calculation is done for this.

The SAS code below illustrates the program for the Anderson-Gill model:

```
PROC PHREG DATA=primary COVS(aggregate);
  MODEL (time_start, time_rec)*status(0)=treat/ties=efron rl;
  ID patid;
```

```
  STRATA {stratum};
RUN;
```

primary is the input dataset, time_start is the previous event stop time and time_rec is the current event stop time, the censoring variable status (0 for censored and 1 for event) should take the value 1 if the last event is a CV death and 0 if it is censored for a non-CV death or at the end of study for the given patient; patid is the participant ID and treat is the treatment group identifier.

If a participant experiences an HF event and subsequently dies for a cardiovascular reason, this will be considered as two separate events for the primary analysis unless the participant dies on the same calendar day as the HF event (both events would still be considered for the analyses of the separate components). If a participant is hospitalized for HF shortly after an urgent HF visit, this will be considered as two separate events for the primary analysis unless they occur on the same calendar day.

Additionally, plots and summaries of the mean cumulative function for the primary endpoint (Nelsen-Aalen estimate) will be presented by treatment group.

6.2.1.2 Primary Efficacy Variable: Sensitivity Analysis

As a sensitivity analysis, the number of primary composite events will also be analyzed using a negative binomial regression model including stratification factors and treatment group as covariates and log follow-up time as an offset parameter.

As a sensitivity analysis, plots and summaries of the mean cumulative function for the primary endpoint will be derived based on a competing-risk approach (Ghosh and Lin, 2000) and cumulative incidence function for the competing event of non-CV death (Aalen and Johansen, 1978) will be presented by treatment group. The following SAS Code illustrates the program for the mean cumulative function for the primary endpoint:

```
PROC PHREG DATA=primary;
  MODEL (time_start, time_rec)*event(0)= / eventcode=1;
  STRATA treat;
  ID patid;
  BASELINE OUT=mcfdat CIF=cif LOWERCIF=lcif UPPERCIF=ucif /
           SEED=999;
RUN;

DATA mcfdat;
  SET mcfdat;
  mcf = -log(1-cif)
  lmcf = -log(1-lcif)
```

```
  ucif = -log(1-ucif)
RUN;
```

`primary` is the input dataset, `time_start` is the previous event stop time and `time_rec` is the current event stop time, the event variable `event` (0 for censored, 1 for primary event, and 2 for competing event) and `patid` is the participant ID and `treat` is the treatment group identifier.

6.2.1.3 Primary Efficacy Variable: Supportive Analysis of CV death component

As part of the primary analysis, a separate estimate of the treatment effect for CV death as one of the components of the primary endpoint will be obtained. The second component of total HF events will be a secondary endpoint and analyzed as described in Section 6.2.2.1.

The main cause-specific treatment effect estimate for CV death will be derived from a stratified Cox proportional hazards model for time to CV death and the main p-value from a stratified log-rank test. A cause-specific treatment effect estimate for the censoring event of non-CV death will also be calculated using a stratified Cox proportional hazards model and presented with associated 95% confidence interval. The cumulative incidence functions for time to CV death and time to non-CV death will also be calculated using Aalen-Johansen estimates.

Note that the study is not powered to show an effect on CV death alone. While this is the case, a sufficient number of deaths are expected so that an excess risk in mortality can be excluded. Under the assumptions of the sample size determination, approximately 535 CV deaths and approximately 775 all-cause deaths are expected to occur in the study. Even though no formal statistical tests for exclusion of an increased risk will be performed, these expected event counts would result in a relatively high power to exclude increased hazard ratios (HRs) on ACM. Table 6–1 provides the respective power values to exclude HRs above 1.15 and 1.25 under different assumed values for the true HR on CV death and assuming no treatment effect on non-CV deaths (HR_{NonCVD}=1.0). Similar to the primary endpoint, a treatment policy strategy is used for treatment discontinuation. With exclusion of a certain HR value it is meant that the upper limit of a 95% CI is below the value.

Table 6–1: Power to exclude increased HR on ACM under different assumed treatment effects on CV death

True HR_{CVD}	Exclude HR_{ACM}>1.15	Exclude HR_{ACM}>1.25
0.8	94%	>99%
0.9	78%	97%
1.0	52%	88%

6.2.1.4 Primary Efficacy Variable: Supportive Analysis of Time-to first event of Composite Endpoint

As supportive analysis, stratified Cox proportional hazard regression analysis will be performed for the time to first composite of HF event or CV death and a plot of Aalen-Johansen estimates of the cumulative incidence function will be provided.

6.2.1.5 Primary Efficacy Variable: Other Supportive Analyses

A supportive analysis of the primary endpoint will exclude urgent HF visits and consider only CV deaths and HHFs as events. Also, an additional analysis of the primary endpoint will restrict CV deaths to HF-related events and thus will consider HF events and CV deaths due to HF. These analyses will both be performed for total (first and recurrent) events and for first events only.

A total-time approach considering times from randomization to the onset of first, second, third composite event using a Wei, Lin, and Weissfeld (WLW, 1989) model will be applied. This model enables analysis of the cumulative effect on the primary endpoint from randomization (i.e. the effect on second event includes the effect on the first, and the effect on third event includes the effects on the first and second). The corresponding individual HRs with 95% CIs comparing treatment groups on the first, second, and third event will be presented.

In addition, a conditional gap-time model according to Prentice, Williams and Peterson (PWP, 1981) will be applied to obtain HR estimates with 95% CIs for the time from first to second and from second to third event (note that this gives a non-randomized comparison). Both models will employ robust standard errors and include the stratification factors and treatment group as fixed effects. Both WLW and PWP approaches are known to have limitations and hence are strictly only performed as supplemental analyses since they aim at describing different aspects of recurrent events.

An "on-treatment" analysis will be performed, including only events occurring up to 30 days after treatment discontinuation. This analysis will be performed in the SAF instead of the FAS.

In addition, table and figure of risk ratios and respective confidence intervals will be provided for the primary efficacy endpoint with patients being censored sequentially at each study day similar to figure 3 in Packer et al. (2021). The first day where the upper CI of the RR is below 1 and stays below for the remainder will be marked.

The primary analysis for the primary endpoint will also be repeated for the "Total HF events and ACM" endpoint.

In addition, the primary analysis will also be repeated where patients are included with only up to a maximum of 4 composite events, to examine the impact of patients with a large number of events. For this analysis, patients who experienced 4 or more events will be censored at the time of their 4^{th} event.

An additional analysis of the primary endpoint will include a time-dependent covariate for SGLT-2 inhibitor use.

6.2.2 Analysis of Secondary Efficacy Variables

6.2.2.1 Secondary Efficacy Variables: Primary Analysis

Secondary efficacy variables are the following:

- Total (first and recurrent) HF events
- Improvement in NYHA class from Baseline to Month 12
- Change from baseline to Month 6, 9 and 12 in TSS of the KCCQ

- Time to first occurrence of composite renal endpoint: sustained decrease in eGFR ≥50% relative to baseline over at least 4 weeks, or sustained eGFR decline <15ml/min/1.73m^2 or initiation of dialysis or renal transplantation
- Time to ACM

See Section 4.7 for details regarding the testing hierarchy for secondary endpoints and nominal significance levels.

Total HF events will be analyzed in a similar fashion to the primary endpoint, using an LWYY model including treatment group as fixed effect and including pooled region for stratified analyses and baseline LVEF (<60%, ≥60%) as stratification factors. A treatment policy strategy will be applied for treatment discontinuation, i.e., all events and follow-up time will be included in the analysis; all-cause death will be a censoring event for a while-alive approach.

The percentage of participants with improvement in NYHA class from Baseline to Month 12 will be analyzed with a logistic regression model including factors for treatment group and stratification levels. A patient is considered as having improved in NYHA class, if the NYHA class at Month 12 (Visit 6) is at least one category improved compared to the baseline visit. A composite strategy will be applied to those cases, where no measurement at Visit 6 is available due to stop of treatment prior to Visit 6. That means these patients are considered to have not improved in NYHA class. Participants who are still in the treatment period at Visit 6, but have NYHA assessment missing, will be imputed by taking the mean of the last available measurement prior to Visit 6 and the first measurement thereafter. Participants with no further value available after Visit 6 or whose only available information is from EoS visit will be imputed as non-responders. Participants with missing baseline value will be excluded from the analysis. Odds ratio and two-sided 95% confidence intervals will be provided for the comparison of finerenone vs. placebo treatment group. In addition, change from baseline in NYHA class will be summarized descriptively using shift tables, presented by visit and any time post-baseline. These tables will present the number of participants with the class at a certain visit by their respective baseline class.

The absolute change from baseline including measurements up to Month 12 of the KCCQ TSS will be analyzed by a repeated measures mixed model including the factors treatment group, baseline, visit, baseline-by-visit interaction, and factors for the stratification levels. For each treatment group a separate covariance pattern will be estimated based on an unstructured covariance to adjust for within participant variance.

In case the model does not converge, different covariance patterns will be used, in the following order, until convergence is met: separate unstructured covariance patterns, separate Toeplitz covariance patterns, separate Autoregressive(1) covariance patterns, and finally separate Compound Symmetry covariance patterns will be used.

Differences between the finerenone and the placebo treatment groups will be calculated with two-sided 95% CIs. The comparison assumes a common treatment effect across Month 6, 9 and 12 and will be considered primary. This analysis will investigate the effect on the KCCQ TSS while patients are alive and irrespective of any permanent treatment discontinuation. This means that all observed values will be included in the analysis.

The primary analysis of the secondary time-to first event variables (i.e. composite renal endpoint and ACM) will be done with a stratified log-rank test for testing and a stratified Cox proportional hazards model for obtaining a point estimate with 95% CIs. The Cox proportional hazards model will be stratified according to the stratification factors and include

treatment group as fixed effect. For the composite renal endpoint, the cause-specific point estimate with 95% confidence interval for the censoring event of death will also be presented. Cumulative incidence function plots and summaries (calculated using Aalen-Johansen estimates) will be produced for the composite renal endpoint and the censoring event of death. In addition, components of the composite renal endpoint will be analyzed. Furthermore, Aalen-Johansen plots will be displayed for the ACM endpoint.

Only central laboratory measurements before initiation of dialysis or renal transplantation will be considered for the definition of the individual eGFR-based components of the renal endpoint. At the up-titration, restart and safety check visits, only a local laboratory measurement is obtained. These values will be checked for a potential eGFR event and, in case of decline, the investigators will be advised to retest eGFR centrally in an unscheduled visit; determination of an eGFR event will then be based on this value. If an initial decrease in eGFR occurs on the EoS visit, there will be another confirmatory measurement taken at least 4 weeks later to confirm the initial decrease. The individual components "Sustained decrease of eGFR ≥ 50% from baseline over at least 4 weeks" as well as "Sustained eGFR decline <15ml/min/1.73m^2" will be programmatically derived. Only in the event that the eGFR decrease was confirmed by at least one additional eGFR measurement taken at least 4 weeks later, it will be considered as a sustained decrease and counted for the renal endpoint. The confirmatory additional eGFR measurements will typically be taken during an unscheduled visit. The date used for the analysis will be the date of the initial sample exceeding the threshold. If there was no confirmatory assessment, events will only be counted for the renal endpoint when the patient died after the initial decrease or the patient went on renal replacement therapy such as dialysis or transplantation prior to their scheduled confirmatory assessment. If there is an intermediate measurement that does not confirm the initial decrease, the event will not be counted for the renal endpoint.

eGFR events will be counted from the day of randomization until the EoS visit (or EoS notification + 28 days if EoS visit is missing). Participants will be censored at the earliest of the date of their last visit when a central eGFR measurement is available or EoS visit date (or EoS notification + 28 days if EoS visit is missing). If no post-baseline eGFR measurement is available, participants will be censored at day 1 (randomization date).

The other two components of the renal composite endpoint, i.e. initiation of dialysis or renal transplantation, will be adjudicated. To account for events of initiation of dialysis or renal transplantation after the last eGFR is recorded at a clinic visit, such events will be included in the efficacy analysis of the composite renal endpoint if they occur in the period up to the next planned clinic visit (+ 1 month) or EoS visit date (or EoS notification + 28 days if EoS visit is missing) or death date or last contact date. If no post-baseline eGFR measurement is available, the next planned visit is the date of Visit 2 (Month 1). Censoring will be applied at the last eGFR date (or day 1 if last eGFR date is missing) or at the earliest of EoS visit date (or EoS notification + 28 days if EoS visit is missing) or death date or last contact date, if these occur earlier than the next eGFR date + 1 month.

Randomized participants without an event of the composite renal endpoint at the time of analysis will be censored at the latest censoring date of their individual components.

6.2.2.2 Secondary Efficacy Variables: Sensitivity Analysis

As a sensitivity analysis for the total HF events endpoint, a joint frailty model (Rogers et al., 2016) with constant baseline hazard for CV death and constant baseline intensity for HF events will be fitted including effects for treatment group, pooled region for stratified analyses

and baseline LVEF (<60%, ≥60%). A gamma frailty distribution will be assumed. This model gives a treatment effect on total HF events which is adjusted for a potential treatment effect on CV death. An effect on CV death might otherwise dilute the effect seen on the hospitalizations, i.e. an effective treatment will prevent CV deaths especially in the more severely ill participants, which then potentially results in many hospitalizations.

The joint frailty model can sometimes have convergence issues (Toenges & Jahn-Eimermacher, 2020); additionally, estimates obtained from the model have sometimes been observed as unstable (e.g., large differences with changes in starting values and/or ordering of covariates in the model). Should such problems be encountered, a simpler model including only treatment group as a fixed effect will be used instead. The joint frailty model additionally produces an estimate for CV death; however, this will be considered only supportive for the analysis of this component and instead the main analysis for the CV death component is described under the primary efficacy variable in Section 6.2.1.2.

6.2.2.3 Secondary Efficacy Variables: Supportive Analysis

As supportive analysis for the total HF events endpoint, stratified Cox proportional hazard regression analyses will be performed for the following endpoints and plots of Aalen-Johansen estimates of the cumulative incidence functions will be provided:

- Time to first HF event
- Time to first HHF
- Time to first urgent HF visit

The additional analyses of the secondary time-to-first event endpoints will include an "on-treatment analysis". For the renal composite endpoint, events will only be counted if they occur within 5 months after the last visit with complete information on all components of the composite primary endpoint. A 5-month time window is used as visits are 4-monthly and in order to allow for late attendance by an additional 1 month. The proportional hazards assumption will be investigated by plotting smoothed Schoenfeld residuals. For the renal endpoint, a time-to-first event analysis will be done separately for each of the components.

A supportive analysis of the KCCQ TSS will apply a worst-case imputation for death which means that if a patient dies, a worst score of 0 for the TSS will be imputed for all subsequent visits after the patient's death (i.e., composite strategy). Treatment effects at Month 6, 9 and 12 will also be investigated individually by adding a treatment-by-visit interaction into the model.

A responder analysis for the KCCQ TSS will also be performed, defining patients with an increase of ≥5 points from baseline to Month 12 (or, for those with a baseline score of >95, a score of >95 at Month 12 without decline from baseline) as a responder. All observed values will be included irrespective of any permanent treatment discontinuation. In case of missing data, a patient's last available post-baseline score prior to Month 12 will be used (i.e. while-alive strategy) unless the patient died before Month 12 in which case they will be imputed as a non-responder (i.e. composite strategy). Responder status will be analysed using a logistic regression model including treatment, baseline TSS and stratification factors as covariates; the odds ratio and associated 95% CI will be reported. This analysis will be repeated for cut-offs of ≥10 points increase from baseline to Month 12 (or maintaining a score of >90 from baseline to Month 12 without decrease from baseline) and ≥20 points increase (or maintaining a score of >80 without decrease from baseline). These cut-offs correspond to small (≥5), moderate (≥10) and large (≥20) clinically meaningful improvements (Spertus et al, 2005). A

further analysis will define those responders who do not experience a ≥5 points decrease from baseline (or, for those with a baseline score of <5, a score of ≥5 at Month 12). This is equivalent to not experiencing a small deterioration. The number and percentage of patients who are responders or non-responders per each of the above criteria will be presented at Months 6, 9 and 12. This will include a breakdown of the criteria met for response (e.g. increase from baseline of ≥5, >95 at baseline and post-baseline visit) or non-response (e.g. change from baseline of <5, >95 at baseline and ≤95 at post-baseline visit, missing score at post-baseline visit).

A second responder analysis for the KCCQ TSS will use the thresholds derived from the anchor-based analyses with the Patient Global Impression of Severity (PGIS) and Patient Global Impression of Change (PGIC) for a clinically meaningful within-patient change in KCCQ TSS at month 6, 9 and 12, respectively, which have been performed separately on blinded data (cf. Section 6.2.3.1). The derived thresholds are

- Minimally important within-patient improvement: 9.09
- Moderate within-patient improvement: 19.85

These thresholds will be used for each timepoint.. All observed values will be included irrespective of any permanent treatment discontinuation. In case of missing data, a patient's last available post-baseline score prior to Month 6, 9 or 12, respectively, will be used unless the patient died before the respective scheduled visit in which case they will be imputed as a non-responder. Responder status will be analysed using a logistic regression model at each timepoint including treatment, baseline TSS and stratification factors as covariates; the odds ratio and associated 95% CI will be reported.

Empirical cumulative density functions will be plotted for Months 6, 9 and 12, with change from baseline in KCCQ TSS (+100 to -100, ordered from greatest possible improvement to greatest possible worsening) on the x-axis and proportion of participants achieving this change or greater on the y-axis. Separate curves will be presented for each treatment group.

6.2.3 Analysis of Further Exploratory Efficacy Variables

Other exploratory efficacy variables will be as follows:

- Time to first CV hospitalization
- Total number of CV hospitalizations
- Time to first all-cause hospitalization
- Total number of all-cause hospitalizations
- Time to first occurrence of the following composite endpoint: CV death or non-fatal CV event (i.e. non-fatal myocardial infarction, non-fatal stroke, or HHF)
- Time to first occurrence of the following composite endpoint: sustained decrease in eGFR ≥ 57% relative to baseline over at least 4 weeks, or sustained eGFR decline < 15 ml/min/1.73m^2 or initiation of dialysis or renal transplantation
- Change in eGFR from baseline
- Mean rate of change in eGFR slope and its subcomponents acute and chronic slope
- Change in UACR from baseline

- Days alive and out of hospital (DAOH)
- Time to new onset of atrial fibrillation
- Change in health-related quality of life summary scores from baseline measured by KCCQ and EuroQol Group 5-dimension 5-level questionnaire (EQ-5D-5L)

Exploratory time-to-event variables will be analyzed using the stratified log-rank test and the stratified Cox proportional hazards model. Plots of Aalen-Johansen estimates of the cumulative incidence function will be provided.

The total number of CV hospitalizations will be analyzed using an LWYY model, similarly to the primary efficacy endpoint, and will be summarized descriptively by treatment group together with the annual rate of CV hospitalizations. These summaries and analyses will be repeated for all-cause hospitalizations.

The absolute change from baseline in eGFR at each visit until Visit 10 (Month 24) will be analyzed by a repeated measures mixed model with the factors treatment group, baseline eGFR, visit, treatment-by-visit interaction, baseline-by-visit interaction, and factors for the stratification levels (pooled region for stratified analyses and LVEF). Differences between the finerenone and placebo treatment groups at each visit will be calculated, and corresponding two-sided 95% CIs will be computed. For each treatment group a separate covariance pattern will be estimated based on an unstructured covariance to adjust for the within participant variance. Change in logarithmized UACR from baseline will be analyzed in an identical fashion. Results will be back-transformed to the original scale so that ratios will be displayed in table outputs.

Frequency tables will be generated for the number and percentage of patients with a relative decrease in eGFR of ≥25%, ≥30%, ≥40%, ≥50% and ≥57% from baseline. The analysis will be performed for each visit and for any time post-baseline.

The rate of change of eGFR will be compared between the finerenone and the placebo group by estimating the total eGFR slope using available assessments from baseline to planned end of the treatment period. It is assumed that changes in the mean response can be expressed in terms of a linear trend, and the treatment effect can be expressed in terms of the difference in slope between finerenone and placebo. For the analysis of the total slope, the serial change in eGFR will be modeled using a two-slope linear spline mixed-effects model in which a fixed change point will be defined to separate acute and chronic eGFR slope at Month 3 (Section 2.1 of Vonesh et al. 2019). In addition to fixed effects for the treatment, time (continuous) and treatment by time interaction, the model will include fixed effects for the stratification factors and random effects for the intercept, acute slope (baseline to Month 3), and chronic slope (Month 3 to planned end of treatment period). An unstructured covariance will be used to model the between-participant errors. Within-participant errors are assumed to be homogenous. Linear contrasts will be constructed to estimate the acute, chronic, and total slope in eGFR. LS means and differences of the acute, chronic and total eGFR slope for finerenone and placebo group will be provided with 95% confidence intervals (and corresponding p-values for the differences).

DAOH will be summarized descriptively by treatment group; the number and percentage of DAOH with respect to total potential follow-up time will be provided alongside the number and percentage of days dead and days in hospital, including breakdown into type of death. These analyses will be performed overall and separately by the stratification factors (pooled region for stratified analyses and LVEF).

DAOH will be analyzed by an ANCOVA model including potential follow-up time, treatment group, and stratification factors as fixed effects. Potential follow-up time is defined as the time from randomization up to end of study or lost to follow-up or withdrawal date, in case the patient did not complete the study.

DAOH will be analyzed once considering the total potential follow-up time and once considering only the first year of follow-up.

For the KCCQ, 3 further summary scores (physical limitation score [PLS], clinical summary score [CSS] and overall summary score [OSS]) will be derived. For the KCCQ PLS, CSS and OSS, the absolute change from baseline including measurements up to month 12 of the KCCQ TSS will be analyzed by a repeated measures mixed model including the factors treatment group, baseline, visit, baseline-by-visit interaction, and factors for the stratification levels. Differences between the finerenone and the placebo treatment groups will be calculated with two-sided 95% CIs. In addition, descriptive statistics will be presented by visit and treatment group: number of observations, number of missing values, minimum, first quartile, mean, standard deviation, median, third quartile, and maximum, including the changes from baseline. The analyses for TSS are described in Section 6.2.2.

For the EQ-5D-5L, summary scores will be calculated from the 5 dimensions according to the scoring instructions from UK and the US (refer to the EQ-5D-5L User Guide (EuroQoL Group 2013) and to the EQ-5D Value Sets (Szende et al. 2007). The values and the changes from baseline of the summary scores and the EuroQol Group visual analogue scale (EQ VAS) will be summarized by treatment group and visit using the same descriptive statistics as for KCCQ.

6.2.3.1 Patient Global Impression of Change (PGIC) and Severity (PGIS)

A sub-population of approximately 1200 participants is being asked the following questions at baseline (PGIS only) and at Visit 4 (Month 6), Visit 5 (Month 9) and Visit 6 (Month 12):

- PGIC: the participant is asked to assess the degree of change in their HF symptoms compared to the start of the treatment using the following response options: much better, better, a little better, the same, a little worse, worse or much worse
- PGIS: the patient is asked to assess the current severity of their HF symptoms due to HF using the following response options: no symptoms, mild, moderate, severe or very severe

These questions will be used as an anchor to provide an estimate of clinically meaningful change in the KCCQ TSS. Details of the analysishave been described in a separate SAP. The analysis has been conducted on a blinded dataset and will be reported separately from the CSR.

6.2.4 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 aEs. Only adjudicated outcome events will be used for the analysis of the primary composite endpoint. Adjudication of the secondary renal endpoint will be restricted to cases of initiation of dialysis or renal transplant. No adjudication will be done for events only included in an exploratory efficacy endpoint (e.g. non-fatal myocardial infarction) and therefore the investigator-reported events will be used in the analysis of those endpoints. An overall summary of all outcome events will be generated by treatment group.

The number of participants 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 and outcome events by maximum intensity will be summarized by treatment group using MedDRA terms grouped by Primary SOC and PT.

The incidence rate of outcome events per 100 patient-years will also be provided by treatment group using MedDRA terms grouped by Primary SOC and PT. 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 participant in case no such event is recorded.

Outcome events will be summarized separately in the CSR Section 8.2 tables.

6.3 Pharmacokinetics/pharmacodynamics

6.3.1 Pharmacokinetics

The finerenone plasma concentration versus time data collected at various study 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 Pharmacokinetic (PK) methods, followed by PK / Pharmacodynamic (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

Analysis of the pharmacodynamics parameters (e.g. blood pressure, heart rate, laboratory values) will be described in detail in a separate SAP.

6.4 Safety

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

6.4.1 Adverse Events

AEs will be coded using the latest version of MedDRA available prior to database freeze. A listing will be provided linking the original investigator terms and the coded terms. AEs will be presented grouped by SOCs and PTS.

AEs that occurred or worsened after the first dose of study drug and up to 3 days after the last dose of study drug will be considered as treatment-emergent AEs (TEAEs).

To comply with local regulatory requirements in Japan and India, certain cardiovascular disease-related outcome events will also be documented as (S)AEs in Japan and India. These will be included in the outcome event tables (see Section 6.2), and to avoid double-counting of such events, they will not be included in the adverse event summary tables or listings. Separate listings will be generated for all AEs excluded from the AE analysis due to double reporting in Japan and India, respectively.

An overall summary of all AEs, pre-treatment AEs, post-treatment AEs occurring more than 3 days after stop of study drug and TEAEs will be generated by treatment group. TEAEs and treatment-emergent SAEs will be summarized by subgroups as defined in section 4.5.6.

The number of participants with TEAEs, post-treatment AEs occurring more than 3 days after stop of study drug, treatment-emergent serious adverse events (SAEs), treatment-emergent study drug-related AEs, treatment-emergent study drug-related SAEs, TEAEs and treatment-emergent SAEs resulting in discontinuation of study drug, treatment-emergent study drug-related AEs resulting in discontinuation of study drug, treatment-emergent non-serious AEs, TEAEs and treatment-emergent SAEs by maximum intensity, drug-related TEAEs by maximum intensity, TEAEs and treatment-emergent SAE by worst outcome and drug-related TEAEs by worst outcome will be summarized by treatment group using MedDRA terms grouped by Primary SOC and PT.

Hyperkalemia and worsening of renal function are considered events of special safety interest. Hyperkalemia will be defined by MLG 'Hyperkalemia' and worsening of renal function will be defined by

- List of PTs: 'Acute kidney injury', 'Blood creatinine increased', 'Glomerular filtration rate decreased', 'Postrenal failure', 'Prerenal failure', 'Renal failure' and 'Renal impairment'
- SMQ 'Acute renal failure' (narrow search)
- SMQ 'Acute renal failure' (broad search).

An overall summary for each definition of worsening of renal function as well as for hyperkalemia will provide the number of participants once for all events and once for all treatment-emergent events by treatment group.

In addition, further safety variables listed in Section 9.4.2 of the protocol are:

- Number of participants with hospitalization for hyperkalemia
- Number of participants permanently discontinuing study intervention due to hyperkalemia
- Number of participants with hospitalization for worsening of renal function
- Number of participants permanently discontinuing study intervention due to worsening of renal function

These will be summarized for treatment-emergent events by treatment group using frequency counts and grouped by Primary SOC and PT.

Cumulative incidences based on Aalen-Johansen estimates and accounting for mortality as competing risk will be provided for the time to first treatment-emergent hyperkalemia event. For this analysis, the person-time at risk for a single participant is the number of days from first intake of study intervention until the event of interest or until the minimum of (date of death, last exposure to treatment + 3 days). Since the number of particpants with hospitalization or permanently discontinuing study intervention due to hyperkalemia is very low, no cumulative incidences will be displayed for these variables.

For hyperkalemia AEs, an additional sensitivity analysis will be performed where events are defined as treatment-emergent, if the AE started or worsened after the first dose of study drug up to 3 days after any temporary or permanent interruption of study drug.

The incidence of TEAEs and treatment-emergent SAEs per 100 person-years will also be provided by treatment group using MedDRA terms grouped by Primary SOC and PT. This analysis will consider the first AE per Primary SOC or PT for a participant. Per SOC or PT, the incidence per 100 person-years will be derived as

100 * (number of participants with TE(S)AE) / (sum of time at risk).

The time at risk per patient is defined as time from first dose of study drug to last dose of study drug + 3 days (treatment-emergent), or death, if earlier, for those patients without a respective AE. For patients with AE, it is the time from first dose of study drug to AE start date. In case the AE start date is (partially) missing, the earliest possible date will be imputed; i.e. first day of a month, first month of a year, restricted to date of first study drug intake.

In case of events with different intensity within a participant, the maximum reported intensity will be used. If intensity 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 participant, the event will be considered as related to study drug. If the drug relationship is missing, the event will be considered as being related to the study drug. Deaths and SAEs will be listed separately.

Any AEs/SAEs related to study procedure recorded after signing of informed consent but prior to randomization will be tabulated separately.

6.4.2 Laboratory Parameters

Generally, only central laboratory measurements will be used for analyses. The only exception is for hematology, where a few countries have the possibility to use local labs in exceptional circumstances (e.g. logistical challenges due to global pandemic, natural disaster, or regional crisis). These local measurements will also be used for the analysis.

The number of participants with treatment-emergent (after the first dose of study drug and up to 3 days after last dose 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, NT-proBNP, high sensitive troponin T (hs-TNT), clinical chemistry and urinalysis. Geometric statistics and ratios to baseline will be presented for urinary creatinine, albumin and UACR and NT-proBNP instead of arithmetic statistics with changes from baseline. For eGFR the relative change will be displayed in addition to the absolute change from baseline. Graphical displays will be provided for change from baseline and ratio to baseline by visit, respectively. Proportion of patients with available and missing measurements will also be displayed.

Summary statistics for serum potassium, eGFR, NT-proBNP 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 participants with safety events as described below by treatment group, visit and for any time treatment-emergent (including unscheduled assessments) and up to 3 days after last study drug administration. This will also be performed by stratification factors. The summaries will be performed for the number of participants with:

- Absolute value of serum potassium >5.0 mmol/L, >5.5 mmol/L (hyperkalemia), >6.0 mmol/L (severe hyperkalemia) and >7.0 mmol/L
- Relative decrease from baseline in eGFR of ≥25%, ≥30%, ≥40%, ≥50% and ≥57%, also sustained decrease over 4 weeks
- Absolute value of eGFR < 30 ml/min/1.73m²
- Increase from baseline in serum creatinine >0.3 mg/dL and >0.5 mg/dL.

The percentage of participants with the respective events (non-stratified) at any time post-baseline (including unscheduled assessments) and within 3 days after last study drug administration will be compared between the finerenone and placebo group by applying separate explorative χ^2 tests with continuity correction. If the expected number of participants in at least 1 cell of the 2x2 contingency table is <5 (Agresti 2005), Fisher's exact test will be applied instead of the χ^2 test. Estimates and two-sided 95% CIs will be provided for each treatment group and the treatment differences. Clopper-Pearson CIs will be calculated for each treatment group, while for treatment differences the exact unconditional confidence limits will be calculated.

6.4.3 Other Additional Safety Variables

6.4.3.1 Vital Signs

At the corresponding visits, 2 BP and 1 pulse measurements of vital signs parameters will be taken. Averages of non-missing values of these two BP measurements will be calculated and used for the statistical analysis. If only one 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. Changes from baseline will also be displayed graphically. The analysis will be repeated for SBP stratified by baseline SBP ≥90 to <130 mmHg, 130 to <160 mmHg and ≥160 mmHg.

6.4.3.2 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.

6.5 COVID-19 and Related Issues

It is expected that the COVID-19 pandemic – ongoing at the start of the trial (14-Sep-2020) – will have some impact on this trial. Every effort will be made to capture the effect of COVID-19 on the study conduct. All adjudicated outcome events will additionally be adjudicated for relationship to COVID-19 (yes, possibly, no). COVID-19 related study disruptions comprise missing visits or procedures, study drug interruptions or permanent discontinuations, AEs related to COVID-19, and (other) protocol deviations and will be analysed as follows:

- All patients affected by study disruption related to the COVID-19 pandemic will be listed together with site information and the type of study disruption(s)
- Number of patients affected, number of missed visits, number and type of protocol deviations will be tabulated
- COVID-19 pandemic related reasons (participant decision, physician decision, or logistical reasons) for premature discontinuation of study epochs (e.g., screen failure, discontinuation of study drug, etc.) and changes in study treatment (e.g., dose titration, interruption, discontinuation, etc.) will be included in the relevant summaries
- Coded terms for COVID-19 will be included in medical history and adverse event summaries

Supportive analyses to evaluate the impact of the COVID-19 pandemic on the primary analysis will be conducted for the primary and secondary efficacy endpoints.

The primary analysis of the primary and secondary endpoints mostly follows the treatment policy strategy as described in ICH E9 (ICH 2019), in which participants are analyzed as they were intended to be treated and all relevant data are used regardless of any previous intercurrent events like COVID-19 infections or study drug interruptions which could have an influence on the occurrence of the event of interest. Only the following intercurrent events are handled with different approaches:

Endpoint	**Intercurrent events and strategies**
Primary endpoint (CV death and total HF events)	Non-CV death is treated as a censoring event for a while-alive strategy. Additionally, CV death is counted as both an outcome event as well as a censoring event, hence a combination of composite and while-alive strategy
Total HF events	Non-CV death is treated as a censoring event for a while-alive strategy. Additionally, the effect on total HF events is adjusted for a potential treatment effect on CV death
Improvement in NYHA class from Baseline to Month 12	A composite strategy is applied when no measurement is available at Month 12 (due to death, or other reasons)
Change from baseline to Month 6, 9 and 12 in TSS from KCCQ	All observed values up to death for any cause are included, i.e. a while-alive strategy is used
Time to first occurrence of renal composite endpoint	All-cause death is treated as a censoring event for a while-alive strategy. If a participant had a decrease or decline in eGFR but died before a confirmatory eGFR measurement could be taken, then the event is counted for the analysis
Time to ACM	Not applicable

This general approach is still considered valid despite the presence of the COVID-19 pandemic, considering the following factors:

- COVID-19 related study disruptions are expected to be equally likely to occur in either treatment group, therefore not creating bias for the treatment effect
- Information on most adjudicated endpoint events can still be collected even when physical visits cannot take place although incidences or reporting of some adjudicated events might change, e.g. as fewer participants might be hospitalized and have a higher risk of death

To quantify the impact of the COVID-19 pandemic, the following supplemental analyses will be performed:

- Primary endpoint (CV death and total HF events)
    - To address the hypothetical scenario in which participants did not experience any study disruption due to the COVID-19 pandemic, three separate analyses

will be conducted using the LWYY model described in Section 6.2.1.1 where participants are censored at

a) Date of permanent discontinuation of study treatment due to COVID-19

b) Date of first COVID-19 adverse event (SMQ narrow)

c) Date of first direct or indirect COVID-19 study disruption

This analysis will be performed once including all primary efficacy events and once excluding those events adjudicated as related to COVID-19. For the latter analysis, the affected participants will be censored at the first occurrence of a COVID-19 related event

Additionally, plots and summaries of the mean cumulative function for the primary endpoint (Nelsen-Aalen estimate) will be presented by treatment group.

- To investigate the effect of COVID-19 related study disruption on the results of the primary analysis, a time-dependent covariate capturing whether a participant is affected by COVID-19 at the respective time together with its interaction with treatment group will be included in the LWYY model described in Section 6.2.1.1. Separate models will be used for the three COVID-19 related categories described above (permanent discontinuation of study treatment, AEs, and study disruption). Events occurring before and after participants are affected by COVID-19 will also be summarized by treatment group
- Primary efficacy events (CV death and HF events) and the competing event of non-CV death will be summarized by treatment group and relationship to COVID-19 as adjudicated by the CEC, defined as follows:
  - Yes (positive testing, typical clinical trajectory)
  - Possibly (inconclusive or absent testing, typical clinical trajectory)
  - No (testing negative or not done, not suspected)
- If ≥5% of the total number of primary efficacy events are adjudicated as "Yes" or "Possibly" related to COVID-19, a further analysis using the LWYY model will be performed excluding such events; the affected participants will be censored at the first occurrence of a COVID-19 related event
- Primary efficacy events occurring during interruption of study treatment (due to COVID-19, or due to other reasons) up to 30 days after any temporary or permanent stop of study treatment will be summarized by treatment group

- Total HF events
  - Summaries of HF events occurring before and after participants are affected by COVID-19, and during interruption of study treatment, will be included as part of the summaries for the primary endpoint described above
- Improvement in NYHA class from Baseline to Month 12
  - Reasons for a participant being classified as "not improved" at Month 12 under the composite strategy will be summarized, including:

  - NYHA measured at Month 12 and not improved from baseline
  - NYHA not measured at Month 12 due to COVID-19 related reason
  - NYHA not measured at Month 12 due to other reason
- Change from baseline to Month 6, 9 and 12 in TSS from KCCQ
  - No additional summaries or analyses are proposed. This endpoint will be analyzed by a repeated measures mixed model which is valid under the missing at random assumption, and COVID-19 related study disruptions are expected to be equally likely to occur in either treatment group and therefore not related to study treatment
- Time to first occurrence of renal composite endpoint
  - Events occurring before and after COVID-19 related study disruption, and during study treatment interruptions due to COVID-19 or other reasons, will be summarized by treatment group
- Time to ACM
  - To address the hypothetical scenario in which participants were not affected by the COVID-19 pandemic, the stratified log-rank test and stratified Cox proportional hazard model as described in Section 6.2.2.1 will be repeated excluding any death adjudicated as "Yes" or "Possibly" related to COVID-19; these participants will instead be censored at the date of death. Aalen-Johansen plots for non-COVID and COVID related deaths will also be displayed

6.6 Regional crisis between Russia and Ukraine and Related Issues

During the conduct of the FINEARTS-HF study, the conflict between Russia and the Ukraine escalated in February 2022 and, as of finalization of this amendment of the SAP, the regional crisis is ongoing in Ukraine. Every effort will be made to capture the effect of this regional crisis on study conduct.

Crisis-related study disruptions comprise missing visits or procedures, study drug interruptions or permanent discontinuations, AEs related to the regional crisis, and (other) protocol deviations and will be summarized as follows:

- All patients affected by study disruption related to the regional crisis will be listed together with site information and the type of study disruption(s)
- Number of patients affected, number of missed visits, number and type of protocol deviations will be tabulated
- Crisis -related reasons (participant decision, physician decision, or logistical reasons) for premature discontinuation of study epochs (e.g., screen failure, discontinuation of study drug, etc.) and changes in the study treatment (e.g., dose titration, interruption or discontinuation of study drug, etc.) will be included in the relevant summaries
- AEs related to the regional crisis will be included in the relevant summaries; if possible, these will be identified and highlighted in the CSR

See Section 6.5 (COVID-19 and Related Issues) or individual analysis sections for a description of the intercurrent event strategies for the primary and secondary endpoints.

Similarly to the impact of COVID-19, the general approach for these endpoints is still considered valid despite the presence of the regional crisis, considering the following factors:

- Crisis-related study disruptions are expected to be equally likely to occur in either treatment group, therefore not creating bias for the treatment effect
- Information on most adjudicated endpoint events can still be collected even when physical visits cannot take place although incidences or reporting of some adjudicated events might change

Despite this, it is recognized that issues arising from the regional crisis (including but not limited to missing data, data of compromised quality and crisis-triggered intercurrent events) may have a substantial impact on individual data points, patients, sites and/or countries. Therefore, the situation will be monitored on an ongoing basis and a final decision will be made prior to unblinding as to whether specific data-handling rules are required for individual data points, patients, sites and/or countries. This decision will be documented in an SAP amendment or – if that is not possible – a note to file, and all rules will be described and justified in the CSR. Additional summaries and sensitivity analyses may also be required, and these will be reported in the CSR.

For example, due to substantial impact of the regional crisis on a specific site, a decision may be made to censor all study participants from that site at a specific date (e.g. 24-Feb-2022, recognized as the date that the conflict escalated).

Regardless of any decision to implement data-handling rules for data affected by the regional crisis, the following supplemental analyses will be performed to quantify the impact of the regional crisis:

- Primary endpoint (CV death and total HF events)
    - To address the hypothetical scenario in which Ukraine participants did not experience any study disruption due to the regional crisis, two separate analyses will be conducted using the LWYY model described in Section 6.2.1.1 where Ukraine participants are censored at
        a) Date of first direct or indirect crisis-related study disruption (missed assessment, missed visit, study drug interruption or permanent discontinuation due to the regional crisis)
        b) 24-Feb-2022, recognized as the date that the conflict escalated (date of the television broadcast “On conducting a special military operation” by Russian president Vladimir Putin)

      Additionally, plots and summaries of the mean cumulative function for the primary endpoint (Nelsen-Aalen estimate) will be presented by treatment group.
    - To investigate the effect of regional crisis related study disruption on the results of the primary analysis, a time-dependent covariate capturing whether a Ukraine participant is affected by the regional crisis the respective time together with its interaction with treatment group will be included in the LWYY model described in Section 6.2.1.1 Separate models will be used for the two dates described above. Events occurring before and after participants

are affected by the regional crisis will also be summarized by treatment group (separately for FAS and for Ukraine participants in the FAS only)
    - Primary efficacy events occurring during interruption of study treatment (due to the regional crisis, or due to other reasons) up to 30 days after any temporary interruption or permanent discontinuation of study treatment will be summarized by treatment group (separately for FAS and for Ukraine participants in the FAS only)
- Total HF events
    - Summaries of HF events occurring before and after participants are affected by the regional crisis, and during interruption of study treatment, will be included as part of the summaries for the primary endpoint described above
- Improvement in NYHA class from Baseline to Month 12
    - Reasons for a participant being classified as "not improved" at Month 12 under the composite strategy will be summarized, including:
        - NYHA measured at Month 12 and not improved from baseline
        - NYHA not measured at Month 12 due to crisis-related reason
        - NYHA not measured at Month 12 due to other reason
- Change from baseline to Month 6, 9 and 12 in TSS from KCCQ
    - No additional summaries or analyses are proposed. This endpoint will be analyzed by a repeated measures mixed model which is valid under the missing at random assumption, and crisis-related study disruptions are expected to be equally likely to occur in either treatment group and therefore not related to study treatment
- Time to first occurrence of renal composite endpoint
    - Events occurring before and after crisis-related study disruption, and during study treatment interruptions due to regional crisis or other reasons, will be summarized by treatment group (separately for FAS and for Ukraine participants in the FAS only)
- Time to ACM
    - Events occurring before and after crisis-related study disruption, and during study treatment interruptions due to regional crisis or other reasons, will be summarized by treatment group (separately for FAS and for Ukraine participants in the FAS only)

7. Document history and changes in the planned statistical analysis

- SAP version 0.4 dated 16 DEC 2019: unsigned draft version for special protocol assessment (SPA) submission
- SAP version 1.0 dated 04 SEP 2020
- SAP version 2.0 dated 24 FEB 2023
- SAP version 3.0 dated 20 JUN 2024

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

Description of the finerenone program in DKD is updated.

Primary and secondary objectives of the trial are moved from Section 3 to Section 2. Description of the primary Estimand is moved from Section 6.2.1.1 to Section 2. A figure of the testing procedure is included.

Total (first and recurrent) HF events is elevated from supportive analysis of the primary endpoint to a new secondary endpoint. Supportive analyses of the primary endpoint concerning only HF events are moved accordingly to become supportive analyses of the new secondary endpoint. This is based on changes made to the secondary endpoints in protocol V3.0

Improvement in NYHA class from Baseline to Month 12 is elevated from exploratory endpoint to new secondary endpoint. A logistic regression analysis is newly specified for the endpoint accordingly in accordance with protocol V3.0.

The composite renal secondary endpoint is changed to include sustained decrease in eGFR ≥50% relative to baseline instead of ≥40% in accordance with protocol V3.0.

Additional separate SAPs are mentioned for specific analyses (pooled analyses, scientific SAP).

Increase in sample size from 5500 randomized to 6000 randomized is described.

Handling of death adjudicated as undetermined death is clarified for the analyses.

Clarification of use of local laboratory values for analysis.

Clarification of definition of on-treatment analysis for efficacy vs. treatment-emergent analysis for safety is aligned throughout the document.

Subgroups are revised (pooled region; baseline eGFR; index HF event; age; BMI; SBP; ACEI/ARB/ARNI), additional subgroup is included (Baseline UACR) and some ‘other’ subgroups are removed (weight; MRA, beta-blocker, diuretic, anti-diabetics, potassium, CYP3A4 or Entresto use; history of CAD; Baseline hs-TNT).

The testing procedure and multiplicity adjustment is adjusted to reflect the two new secondary endpoints in accordance with protocol V3.0.

Demography and baseline characteristics are revised to ensure consistency with subgroups yet keeping baseline characteristics that are not any more used for subgroup analyses.

Inclusion of additional specific medical history terms (COVID-19, Liver cirrhosis, Sleep Apnea Syndrome, Chronic Kidney Disease) and clarification of derivations (via e.g. PT, MLG, etc).

Remove concomitant BCRP substrate use from concomitant medications of interest.

Additional tables for exposure and titration by starting dose have been specified.

Analysis strategies for the primary endpoint and secondary endpoints are revised to include the handling of censoring events. This concerns the primary endpoint, the supportive analysis of CV death, as well as secondary and exploratory time-to event endpoints.

Summary of incidence rate of primary endpoint events has been included.

Handling of multiple primary endpoint events (i.e. ‘7-day’ rule) has been revised to a ‘same calendar day’ rule. The supportive analysis on the ‘7-day’ rule has been removed accordingly.

Ghosh and Lin competing risk approach for the mean cumulative function has been moved to sensitivity analysis and respective example SAS Code for the mean cumulative function has been included.

Sensitivity and supportive analyses for the primary efficacy variable have been structured in separate sections (Section 6.2.1.2: Sensitivity and Section 6.2.1.3 to Section 6.2.1.5: Supportive). For the secondary endpoint of total HF events, a separate section with sensitivity analysis is included as well.

An additional sensitivity analysis with LVEF as continuous variable has been specified for the primary efficacy endpoint.

Further supportive analyses for the primary efficacy endpoint have been included (restriction to HF-related CV death; sequential censoring; time-dependent covariate of SGLT2-inhibitor use).

The analysis of the secondary endpoint of total HF events is changed from a joint frailty model (as specified in protocol V3.0) to the LWYY model. A joint frailty model with constant baseline hazard for CV death and constant baseline intensity for HF events was instead included as sensitivity analysis. This change in analysis strategy is based on potential convergence issues and unstable estimates of the joint frailty model.

Additional model specification is included for the repeated measures model of the change in KCCQ TSS.

An additional sensitivity analysis with LVEF as continuous variable has been specified for the total HF endpoint.

Analyses with joint frailty models as supportive analyses of recurrent HHFs and recurrent urgent HF events have been removed.

A responder analysis based on thresholds of the anchor-based analyses with the PGIS/PGIC are included for KCCQ TSS. Empirical cumulative density functions are included for change from baseline in KCCQ TSS.

Time to first occurrence of a new composite renal endpoint including sustained decrease in eGFR ≥57% relative to baseline, mean rate of change in eGFR slope, and change in UACR from baseline are included as new exploratory endpoints and respective analyses of these endpoints are specified.

Additional model specification is included for the repeated measures model of the change in eGFR and UACR.

Exchanging analysis of AEs by SMQs with MLGs.

Handling of SAE and outcome events for Japan and India have been clarified.

Inclusion of analyses for hyperkalemia and acute renal failure.

Clarification of derivation of incidences per 100 person-years.

Handling of local hematology measurements due to local protocol amendment has been clarified.

Revision of cutoffs for special safety parameters.

Listing further safety variables in line with protocol.

Inclusion of further analyses for COVID-19 related issues.

Inclusion of analyses due to regional crisis between Russia and Ukraine.

A new section on document history and changes is included as this is required to document respective changes from the first version.

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

Removed reference to older protocol versions.

Clarification of log-normality assumption for specific parameters added.

Specification for the derivation of median for subgroup splits included.

Updated definition of SAF – clarifying excludion of participants with GCP violation.

Included definition of listing-only participants.

Removed ‘ethnicity’ from demography as not being collected in the study. Included a further category of <90 for SBP categorization, due to respective observed values.

Updated PT Term from ‘Liver cirrhosis’ to ‘Hepatic cirrhosis’.

Restricted some analyses on concomitant medications to medications in the standard drug groups of interest.

Separated CYP3A4 inhibitors and inducers.

Removed analysis of mean daily dose of MRAs due to unclean data.

Included total treatment duration in patient-years.

Removed analyses with continuous covariate of baseline LVEF as well as analysis with time by treatment interaction for primary and secondary endpoints.

Updated definition of ‘on-treatment’ in general and for renal composite endpoint specifically.

Included information on handling time after the forth event for analysis of the first 4 composite events.

Included further imputation rules for secondary endpoint ‘Improvement in NYHA class'.

Rules for usage of covariance patterns in case of non-convergence included for MMRM analysis of KCCQ TSS.

Updated the information on censoring for the renal endpoint and specifically for the different components.

Clarified intercurrent event strategies for KCCQ responder analyses.

Included thresholds for clinically meaningful change derived from separate anchor analysis.

Clarified analysis of UACR used log-transformation.

Removed breakdown of hospitalizations for DAOH.

Removed MedDRA Labelling Groupings analyses for Adverse events. Included further analyses for AEs by SOC and PT. Included analyses of AEs by subgroups.

The section 6.4.3.3 Further Safety Variables has been integrated into Advere Events Section 6.4.1 with further details included on the definition of Hyperkalemia and Worsening renal function events.

Included display of cumulative incidences for TE hyperkalemia events.

Updated list of lab parameters to be analyzed with geometric statistic (i.e. assuming log-normal data). Included graphical displays for lab parameters.

Included graphical displays for vital signs.

Minor wording updates throughout the document (e.g. changed ‘conflict’ to ‘regional crisis’; ‘subject’ to ‘participant’) and correction of typos.

8. References

Agresti, A. (2007): *An Introduction to Categorical Data Analysis*, Chapter 2, Second Edition, New York: John Wiley & Sons.

Andersen, P. K. A. G., R.D (1982). Cox's Regression Model Counting Process: a Large Sample Study. *Annals of Statistics,* 10**,** 1100-1120.

EuroQoL_Group T. EQ-5D-5L user guide version 2.0: EuroQoL Group, 2013. http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/UserGuide_EQ-5D-5L_v20_October_2013.pdf

International Council for Harmonization (2019): Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. ICH E9 (R1).

Ghosh D and Lin DY (2000) Nonparametric analysis of recurrent events and death. Biometrics; 56:554–562.

Green CP, Porter CB, Bresnahan DR, Spertus JA. Development and evaluation of the Kansas City Cardiomyopathy Questionnaire: a new health status measure for heart failure. J Am Coll Cardiol. 2000 Apr;35(5):1245-55

Horio M, Imai E, Yasuda Y, Watanabe T, Matsuo S. Modification of the CKD epidemiology collaboration (CKD-EPI) equation for Japanese: accuracy and use for population estimates. Am J Kidney Dis. 2010 Jul;56(1):32-8.

Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, 3rd, Feldman HI, et al. A new equation to estimate glomerular filtration rate. Ann Intern Med. 2009;150(9):604-12.

Lin DY, Wei LJ, Yang I and Ying, Z (2000) Semiparametric regression for the mean and rate functions of recurrent events. J R Stat Soc; B, 62:711–730.

Liu L, Huang X. The use of Gaussian quadrature for estimation in frailty proportional hazards models - ABSTRACT. Stat Med. 2008 Jun 30;27(14):2665-83.

Packer M, Butler J, Zannad F, Filippatos G, Ferreira JP, Pocock SJ, Carson P, et al. Effect of empagliflozin on worsening heart failure events in patients with heart failure and preserved ejection fraction (EMPEROR-Preserved Trial). Circulation 2021;144:1284-1294.

Pitt B, Zannad F, Remme WJ, Cody R, Castaigne A, Perez A, et al. The effect of spironolactone on morbidity and mortality in patients with severe heart failure. Randomized Aldactone Evaluation Study Investigators. N Engl J Med. 1999 Sep 2;341(10):709-17.

Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. Biometrika, Volume 68, Issue 2, August 1981, 373–379.

Rogers JK, Yaroshinsky A, Pocock SJ, Stokar D, Pogoda J. Analysis of recurrent events with an associated informative dropout time: Application of the joint frailty model. Stat Med. 2016 Jun 15;35(13):2195-205.

Spertus J, Peterson E, Conard MW, Heidenreich PA, Krumholz HM, Jones P, McCullough PA, Pina I, Tooley J, Weintraub WS, et al; Cardiovascular Outcomes Research Consortium. Monitoring clinical changes in patients with heart failure: a comparison of methods. Am Heart J. 2005;150:707–715.

Szende A, Devlin N, editors. EQ-5D value sets: inventory, comparative review and user guide. Dordrecht, Netherlands: Springer. 2007.

Toenges G, Jahn-Eimermacher A. Computational issues in fitting joint frailty models for recurrent events with an associated terminal event. Comput Methods Programs Biomed. 2020 May; 188:105259.

Wei LJ, Lin DY, Weissfeld L (1989): Regression analysis of multivariate incomplete failure time data by modeling marginal distributions, J Am Stat Assoc;84,1064-1072.

Zannad F, McMurray JJ, Drexler H, Krum H, van Veldhuisen DJ, Swedberg K, et al. Rationale and design of the Eplerenone in Mild Patients Hospitalization And SurvIval Study in Heart Failure (EMPHASIS-HF). Eur J Heart Fail. 2010 Jun;12(6):617-22.

9. Appendix

9.1 The Kansas City Cardiomyopathy Questionnaire Scoring Instructions

There are 10 summary scores within the KCCQ, which are calculated as follows:

1. Physical Limitation

- Code responses to each of Questions 1a-f as follows:

  Extremely limited = 1
  Quite a bit limited = 2
  Moderately limited = 3
  Slightly limited = 4
  Not at all limited = 5
  Limited for other reasons or did not do = <*missing value*>

- If at least three of Questions 1a-f are not

  missing, then compute

  Physical Limitation Score = 100*[(mean of Questions 1a-f actually answered)

  – 1]/4 *(see footnote at end of this document for explanation of meaning of*

  "actually answered")

2. Symptom Stability

- Code the response to Question 2 as follows:

  Much worse = 1
  Slightly worse = 2
  Not changed = 3
  Slightly better = 4
  Much better = 5
  I've had no symptoms over the last 2 weeks = 3

- If Question 2 is not missing, then compute

  Stability Score = 100*[(Question 2) – 1]/4

3. Symptom Frequency

- Code responses to Questions 3, 5, 7 and 9 as follows:

  <u>Question 3</u>
  Every morning = 1
  3 or more times a week but not every day = 2
  1-2 times a week = 3

Less than once a week = 4
Never over the past 2 weeks = 5

Questions 5 and 7
All of the time = 1
Several times a day = 2
At least once a day = 3
3 or more times a week but not every day = 4
1-2 times a week = 5
Less than once a week = 6
Never over the past 2 weeks = 7

Question 9
Every night = 1
3 or more times a week but not every day = 2
1-2 times a week = 3
Less than once a week = 4
Never over the past 2 weeks = 5

- If at least two of Questions 3, 5, 7 and 9 are not missing, then compute:

  S3 = [(Question 3) – 1]/4
  S5 = [(Question 5) – 1]/6
  S7 = [(Question 7) – 1]/6
  S9 = [(Question 9) – 1]/4

  Symptom Frequency Score = 100*(mean of S3, S5, S7 and S9)

4. Symptom Burden

- Code responses to each of Questions 4, 6 and 8 as follows:

  Extremely bothersome = 1
  Quite a bit bothersome = 2
  Moderately bothersome = 3
  Slightly bothersome = 4
  Not at all bothersome = 5
  I've had no swelling/fatigue/shortness of breath = 5

- If at least one of Questions 4, 6 and 8 is not missing, then compute

  Symptom Burden Score = 100*[(mean of Questions 4, 6 and 8 actually answered) – 1]/4

5. Total Symptom Score

= mean of the following available summary scores: Symptom Frequency Score

Symptom Burden Score

6. Self-Efficacy

- Code responses to Questions 10 and 11 as follows:

Question 10
Not at all sure = 1
Not very sure = 2
Somewhat sure = 3
Mostly sure = 4
Completely sure = 5

Question 11
Do not understand at all = 1
Do not understand very well = 2
Somewhat understand = 3
Mostly understand = 4
Completely understand = 5

- If at least one of Questions 10 and 11 is not missing, then compute Self-Efficacy Score = 100*[(mean of Questions 10 and 11 actually answered) – 1]/4

7. Quality of Life

- Code responses to Questions 12, 13 and 14 as follows:

Question 12
It has extremely limited my enjoyment of life = 1
It has limited my enjoyment of life quite a bit = 2
It has moderately limited my enjoyment of life = 3
It has slightly limited my enjoyment of life = 4
It has not limited my enjoyment of life at all = 5

Question 13
Not at all satisfied = 1
Mostly dissatisfied = 2
Somewhat satisfied = 3
Mostly satisfied = 4
Completely satisfied = 5

Question 14
I felt that way all of the time = 1
I felt that way most of the time = 2
I occasionally felt that way = 3
I rarely felt that way = 4
I never felt that way = 5

- If at least one of Questions 12, 13 and 14 is not missing, then compute

Quality of Life Score = 100*[(mean of Questions 12, 13 and 14 actually answered) – 1]/4

8. Social Limitation

- Code responses to each of Questions 15a-d as follows:

Severely limited = 1
Limited quite a bit = 2
Moderately limited = 3
Slightly limited = 4
Did not limit at all = 5
Does not apply or did not do for other reasons = <*missing value*>

- If at least two of Questions 15a-d are not missing, then compute

Social Limitation Score = 100*[(mean of Questions 15a-d actually answered) – 1]/4

9. Overall Summary Score

= mean of the following available summary scores: Physical Limitation Score
Total Symptom Score Quality of Life Score Social Limitation Score

10. Clinical Summary Score

= mean of the following available summary scores: Physical Limitation Score
Total Symptom Score

Note: references to "**means of questions actually answered**" imply the following.

f If there are *n* questions in a scale, and the participant must answer *m* to score the scale, but the participant answers only *n-i*, where *n-i* >= *m*, calculate the **mean of those questions** as
(sum of the responses to those *n-i* questions) / (*n-i*) **not**
(sum of the responses to those *n-i* questions) / *n*