NOVARTIS

Clinical Development

LCZ696D

Clinical Trial Protocol CLCZ696D2301 / NCT01920711

A multicenter, randomized, double-blind, parallel group, active-controlled study to evaluate the efficacy and safety of LCZ696 compared to valsartan, on morbidity and mortality in heart failure patients (NYHA Class II-IV) with preserved ejection fraction

RAP Module 3 – Detailed Statistical Methodology

Author: [redacted]

Document type: RAP Documentation

Document status: Amendment 3

Release date: May 28, 2019

Number of pages: 58

Table of contents

Document History – Changes compared to previous version of RAP module 3 (Final Draft version 1.0, October 16, 2013).

Version	Date of change	Page	Changes
Final version 2.0 (Amendment 1 to Final Draft 1.0, dated Oct 16, 2013)	9/15/2016	7	(Analysis sets) Several items in exclusions from analysis sets due to PDs have been changed due to the changes in the PD list and the reconsideration of the analysis sets.
		10	(Subgroup analyses) Subgroup defined by the randomization baseline eGFR values and NYHA classifications are added; Definition of regions has been changed according to the countries involved in the study.
		12	Baseline for biomarkers has been corrected to visit 101/102. MMSE has been added to the list to be collected at randomization baseline. Analysis cut-off date has been updated for the new sample size.
		10	South Africa and Israel have been moved to ***Asia/Pacific and other*** under the new definition of region. Slovenia is added to Central Europe. Countries not participating the study are removed.
		15, 16, & 21	The ACEi intolerant subgroup summary for background and demographic characteristics has been removed. Baseline LVEF <= median subgroup summary has been added.
		17 to 19	The exposure duration calculation formulas for the dose level 2 in sub-run-in phase have been corrected. The dose level 3 was erroneously included in the calculation for the run-in and has been removed. More exposure duration analyses for dose down- and up-titration are added.
		20	Formula of "Daily dose for each patient" are updated to reflect BID.
		24 to 26	Secondary analysis (based on the protocol amendment 2): (1) The extended composite endpoint variable (CV death, total HF hospitalizations, total non-fatal strokes, and total non-fatal MIs) and time to new onset Atrial fib are moved to the exploratory analysis from the secondary analysis; KCCQ clinical summary score at month 8 and time to 1st composite renal impairment endpoint have been added to the secondary analysis. In total we still have four secondary endpoints. Exploratory analysis: (2) The extended composite endpoint variable is added; (3) Number of worsening HF events or CV death per-subject has been replaced by number of total worsening HF events and CV death per-subject. The definition of worsening HF event is given; (4) Number of days alive and out of hospital (DAOOH) during the randomized, double-blind study period and percentage of DAOOH over potential follow-up time is added;

Version	Date of change	Page	Changes
			(5) Time to 1st occurrence of composite renal impairment event is removed; (6) Time to new onset AF is moved here; (7) Number of ICU stays is added; (8) Cardiac related biomarkers are added to the analysis list; (9) Change from baseline in MMSE summary score at year 2 (week 96) is added; (10) Indicator of presence of AF >5 minutes (for the AF substudy) is removed; (11)AF burden measured by the total AF duration over total monitoring time is removed (12)Echocardiographic parameters (in a subset of patients) are removed.
		27	The 0.02499 in "The primary hypothesis will be tested at a one-sided significance level of 0.02499 adjusted for the interim analysis." has been corrected to 0.024. Nelsen-Aalen estimate has been moved to Page 30.
		27	In the primary analysis model, "ties = breslow" has been changed to "ties = efron" as requested by the FDA. The general baseline rate and hazard functions are changed to constant baseline rate and hazard in the joint shared frailty model, to simplify the modeling based on the experience from PARADIGM-HF. Piece-wise constant baseline rate and hazard functions will be used for the sensitivity analysis. The BIC criterion will be used to select the number of pieces. The marginal time-to-CV death analysis has been added to use Cox's regression analysis. The following sentence "The inference on the total HF hospitalizations can only be made when both the composite endpoint and the total HF hospitalizations itself showed statistically positive results" has been removed in order to reduce the confusion about how to make the primary inference.
		31	A new supportive analysis of performing the LWYY model on HF hospitalizations alone has been added. A pooled-weighted effect size from this analysis and the Cox's regression analysis on CV death has also been added.
		32	R codes for multi-state model have been modified. On-treatment LWYY analysis is added as supportive analysis to the primary efficacy.
		33 to 35	Due to the change in the secondary objectives, the endpoints and the testing scheme in the multiple hypotheses testing procedure for the secondary efficacy analysis have been changed accordingly.
		35 to 36	The analysis specifications for the KCCQ CSS (mean change and responder analyses) at Month 8 have been added. The analysis for composite renal impairment has been added. The analyses for the extended composite endpoint and new onset AF have been removed from secondary analyses. Missing NYHA class change due to death will be imputed as worsened.

Version	Date of change	Page	Changes
		37 to 43	For the exploratory analyses, the following analyses are updated according to protocol amendment (PA) 2: 1. Time to new onset AF; 2. Time-to-recurrent-composite event of CV death, total HF hospitalizations, total strokes, and total MIs; 3. Analyses for changes in health-related quality of life change (KCCQ scores and EQ-5D); 4. Analysis for rate of change in eGFR; 5. Analysis of biomarkers; 6. Analysis of days alive out of hospital; 7. Analysis of MMSE; 8. Analysis of count data. The details for the analysis for MMSE are added. The subgroup analysis for MMSE is also specified.
		39	Analysis related to pharmacokinetics parameters are added.
		45	In Table 10-2: 1. Hyperkalemia related AE by baseline hyperkalemia status is removed. 2. Symptomatic hypotension related are removed. 3. Cutoff for SBP is updated to 100 from 90 mmHg. 4. Risks deleted: gastric lesions, stimulation of lipolysis, QT-prolongation. 5. Risks added: embryo-fetal toxicity/lethality, neonatal/infantile toxicity, statin drug-drug interaction.
		47	Analysis for cognitive impairment has been explicitly listed in the analysis for change in cognitive function.
		50 to 51	For interim analysis, the following modifications have been made according to the protocol amendments (PA) 2 and 3. 1. Move the interim efficacy analysis to occur after approximately two-thirds of the target number of primary composite events (i.e., ~1231 primary composite events) rather than half as previously defined; (PA 2) 2. Eliminate the futility analysis; (PA 2) 3. The statistical stopping rules for superiority of LCZ696 over valsartan is modified from one-sided p-value of <0.0001 for the primary endpoint to one-sided p-value of <0.001 for both the primary endpoint and CV death at the interim efficacy analysis. (PA 3)

Version	Date of change	Page	Changes
		51, 52	According to the protocol amendment 3, the total sample size has been readjusted from 4300 to 4600 to increase the statistical power from 81% to close to 85% to detect a 25% reduction in recurrent heart failure hospitalizations. The sample size re-estimation was based on a recent analysis on recurrent heart failure hospitalization in the PARADIGM-HF, which showed that LCZ696 resulted in approximately a 25% reduction in recurrent heart failure hospitalizations relative to enalapril. The target number of primary events is also increased to 1847, which corresponds to conducting the interim efficacy analysis when ~1231 primary composite events have been confirmed by adjudication. As described in Table 13-1, a 25% reduction in recurrent heart failure hospitalizations is expected to correspond to an overall 19% reduction in the primary endpoint (CV deaths and total recurrent heart failure hospitalization).
		53, 54	For the secondary endpoints, the powers are also updated according to the new sample size.
		54	Reference of Andersen and Gill (1982) is added.
Amendment of final version 2.0 (Amendment 2 to Final Draft 1.0, dated Oct 16, 2013)	8/1/2018	10	Slovenia is added to Central Europe. Countries not participating the study are removed.
		20	Formula of “Daily dose for each patient” are updated to reflect BID.
		32	On-treatment LWYY analysis is added as supportive analysis to the primary efficacy.
		39	Analyses related to pharmacokinetics parameters are added.
		45	In Table 10-2: 1. Hyperkalemia related AE by baseline hyperkalemia status is removed. 2. Symptomatic hypotension related are removed. 3. Cutoff for SBP is updated to 100 from 90 mmHg. 4. Risks deleted: gastric lesions, stimulation of lipolysis, QT-prolongation. 5. Risks added: embryo-fetal toxicity/lethality, neonatal/infantile toxicity, statin drug-drug interaction.

Version	Date of change	Page	Changes
Amendment 3 (Amendment 3 to Final Draft 1.0, dated Oct 16, 2013)	5/28/2019	10	(Table 2-1) Protocol deviations leading to exclusion from analysis sets are updated. Removed: 1. History of severe pulmonary disease (exclusion criteria); 2. LVEF ≥45% not measured by echocardiogram (echo) during the screening epoch, or within 6 months prior to Visit 1 (inclusion criteria); 3. Lack of evidence for left atrial (LA) enlargement or left ventricular hypertrophy (LVH) as defined by the protocol (inclusion criteria). Added: 1. Most recent LVEF measured <45 by ECHO, or LVEF % not measured; 2. NYHA class I Visit 1.
		11	(Subgroup analyses) Subgroup of atrial fibrillation based on history prior to randomization is added in addition to the existing atrial fibrillation based on ECG at randomization.
		14	Analysis cut-off date information is added.
		32	Analysis of time to first primary composite event (CV death or hospitalization for HF) was added as supportive analysis to the primary efficacy objective. Analysis includes: 1. Cox proportional hazards model; 2. Additionally, Log-rank test will be performed.
		34	CV death component in primary LWYY analysis is replaced with the following as supportive analyses to the primary efficacy objective: 1. Death due to CV or unknown cause; 2. All-cause death. Time-dependent systolic blood pressure (SBP) covariate adjusted LWYY and Cox analyses are added as supportive analyses to the primary efficacy objective.
		39	Component analysis (joint frailty model) for item 2 is removed.
		41	Pharmacokinetics related analyses will be presented in a separate report.
		42	The subgroup analysis for MMSE is specified with more interaction terms.
		48	In Table 10-2: 1. Risks deleted: change in bone growth/bone mineral density. 2. Risks updated or added: neonatal/infantile toxicity through exposure from breast milk, anaphylaxis/anaphylactoid reactions, malignancy.
		55	Table 13-1: one typo is corrected.
		#	Besides the above noted changes, some minor edits to the text are also made to correct some words or order of texts.

Statistical and analytical plans

1 Introduction

Data will be analyzed according to the data analysis section 10 of the study protocol which is available in Appendix 16.1.1 of the CSR. Important information is given in the following sections and details are provided, as applicable, in Appendix 16.1.9 of the CSR.

2 Analysis set

The following analysis populations will be defined for statistical analysis:

- **Screened set (SCR)** – All patients who signed the informed consent. The *screened set* includes only *unique screened patients*, i.e., in the case of *re-screened patients* only the chronologically last screening data is counted.
- **Enrolled set (ENR) (run-in phase)** – All patients who received at least one dose of run-in study drug.
- **Valsartan run-in set (VRS)** – All patients who are in the ENR and received at least one dose of run-in valsartan.
- **LCZ696 run-in set (LRS)** – All patients who are in the ENR and received at least one dose of run-in LCZ696.
- **Randomized set (RAN)** – All patients who received a randomization number, regardless of receiving trial medication.
- **Full analysis set (FAS)** – All randomized patients. This is the primary efficacy population applied in efficacy analyses for all efficacy endpoints. Following the intent-to-treat principle, patients are analyzed according to the treatment they have been assigned to at randomization. However, patients who did not qualify for randomization but were inadvertently randomized into the study and did not receive study medication may be excluded from FAS. Further exclusions from the FAS may only be justified in exceptional circumstances (e.g., *serious* GCP violations). These potential incidences imply the possibility that the FAS may be smaller than the RAN.
- **Per protocol set (PPS)** – A subset of the FAS, consists of all randomized patients in FAS who received at least one dose of study drug during the double blind period of the study and have no major protocol deviations. Major protocol deviations are those affecting the primary endpoint analyses, which will be identified prior to the final database lock.
- **Safety set (SAF)** – All randomized patients who received at least one dose of study drug during the double blind period of the study. Patients will be analyzed according to treatment received. Treatment received will be considered identical to the randomized treatment if the patient has received at least one dose of the randomized treatment.

The following Table 2-1 contains a sample of current study specific important protocol deviations that lead to exclusion from analysis sets in the study. The final list may be different from this and will be signed before DBL.

Table 2-1 Specification of protocol deviations which lead to exclusion from analysis sets

Category	Description of Protocol Deviation	Inclusion / Exclusion in analyses	Identification method	Person responsible for identification
General	GCP issue present at site	Excluded from PPS,FAS	Manual	Field monitor (FM)
	Blind broken locally	Excluded from PPS	Manual	FM
	Informed consent missing	Excluded from PPS, FAS, SAF, RAN	Manual	FM
Study treatment	Patient randomized but no study medication taken.	Excluded from SAF	Programmable	TDM
	Patient was mis-randomized (patient randomized in error and no DB study medication taken).	Excluded from PPS, FAS, SAF	Manual	FM
Exclusion criteria	History of dilated cardiomyopathy, including peripartum cardiomyopathy, chemotherapy induced cardiomyopathy, or viral myocarditis.	Excluded from PPS	Manual	FM
	Heart transplant or ventricular assistance device (VAD) or intent to transplant (on transplant list) or implant a VAD	Excluded from PPS	Manual	FM
Inclusion criteria	NYHA class I at Visit 1	Excluded from PPS	Programmable	TDM
	Most recent LVEF measured <45 by ECHO, or LVEF % measured by non-ECHO methods	Excluded from PPS	Manual	FM
	Lack of 1) HF hospitalization (defined as HF as the major reason for hospitalization or treatment for HF lasting ≥12 hours and including intravenous [IV] diuretics at a healthcare facility) within 9 months prior to Visit 1, OR 2) lack of elevated NT-	Excluded from PPS	Manual	FM

Category	Description of Protocol Deviation	Inclusion / Exclusion in analyses	Identification method	Person responsible for identification
	proBNP > 300 pg/ml for patient not in AF on Visit 1 ECG or > 900 pg/ ml for patient in AF on Visit 1 ECG			

3 Subgroup definitions and subgroup analyses

3.1 Subgroups

Subgroups will be formed for the analyses for the FAS or SAF depending on the parameters under consideration to explore the consistency of treatment effects and safety profiling between the subgroups and the overall population.

In general, subgroups will be defined based on baseline information. In this study, since we have a run-in phase to test patients' tolerability to the study drugs before they can enter the double-blind phase, we have defined two baselines (Section 4): the run-in baseline and the double-blind baseline. Subgroups may be formed using one of these baselines according to their analysis purposes.

In Table 3-1, we have listed all subgroups defined for this study and the ways to derive them. In different analysis, such as efficacy analysis or safety analysis, a different subset of them will be selected. Also note that only important parameters or variables in these analyses may have subgroup analyses. The details about the parameters having subgroup analyses will be presented in the corresponding sections below. Also, additional subgroups may be formed later for regional or country-wise analyses per request for exploratory purposes.

Table 3-1 Specification of subgroups

Subgroup	Method of Derivation	Disposition /Background & demo. /Exposures	Efficacy (Primary and secondary)	Health economics (KCCQ and EQ-5D only)	Safety
Age groups: <65, ≥65; <75, ≥75 years	Screening	x	x	x	x
Gender	Screening	x	x		x
Race (Caucasian, Black, Asian, Other)	Screening	x	x		x
Region *	Derived (pooled countries or country), using	x	x	x	x

	Screening data				
Diabetic at baseline (yes/no)	Randomization (Derived)		x	x	
Baseline LVEF ≤ median,>median	Screening	x	x		
Atrial Fibrillation based on ECG at baseline: yes/no	Randomization		x	x	
Atrial Fibrialtion based on prior history at baseline: yes/no	Randomization		x	x	
Baseline NT-proBNP: ≤ median and > median	Screening		x	x	
SBP ≤ median,>median	Screening		x		
Use of Aldosterone antagonist at baseline: yes/no	Randomization (Derived)		x	x	
ACEi intolerant: yes/no	Screening		x		
Baseline eGFR (<60 vs ≥60 mL/min/1.73 m^2)	Randomization (Derived)		x		
NYHA classification at baseline I/II vs III/IV	Randomization		x		
UK				x	

* ***North America*****:** USA, Canada

Latin America (including Central America)**:** Argentina, Brazil, Colombia, Guatemala, Mexico, Peru

Western Europe**:** Austria, Belgium, Denmark, Finland, France, Germany, Greece, Italy, Netherlands, Norway, Spain, Sweden, Switzerland, UK

Central Europe**:** Bulgaria, Croatia, Czech Republic, Hungary, Poland, Rep. of Slovakia, Romania, Russia, Serbia, Turkey, Slovenia

Asia/Pacific and other**:** Israel, South Africa, Australia, China, India, Japan, Rep of Korea, Philippines, Singapore, Taiwan.

3.2 Subgroup analyses

Subgroup analyses will be performed depending on the analysis purposes. Subgroups selected and types of analyses are specified in Table 3-1 and will be detailed later in the sequel sections. In principle, there will be no adjustment for multiple comparisons for subgroup analyses.

4 Definitions of baseline, post baseline, analysis cut-off, and close-out

Single-blind run-in phase (Treatment run-in epoch)

The single-blind run-in phase (or run-in phase) is defined as the period between the start of the study drug (in general it should be Visit 101 or Visit 102) and the time prior to randomization (Visit 201). The run-in phase will be divided further into the valsartan run-in phase and the LCZ696 run-in phase. The start of the valsartan phase is the date the patient receives the run-in dose of valsartan. In principle, the day before a patient receives the first dose of LCZ696 (any dose level) will be the end of the valsartan run-in phase and the day the patient receives the first dose of LCZ696 will be the start day for the LCZ696 run-in phase for that patient. If the date of the start day of receiving LCZ696 is missing, then the date of Visit 103, which is the protocol specified run-in drug switching day, will be used as the start date for the LCZ696 phase. If the date of the last dose of valsartan run-in medication falls on the same day of the first dose of LCZ696 run-in medication, this day will be considered the end of the valsartan run-in phase and the following day will be considered as the start of the LCZ696 run-in phase. The LCZ696 run-in phase ends at the time when the patient receives the randomization drug. During these two run-in phases, patients will be on the run-in medications in a single blind manner.

Randomized treatment phase

The randomized treatment phase begins at the time of randomization and ends with the last study drug intake or the death of the patient, whichever comes earlier. During the randomized treatment phase, patients will return for scheduled clinic visits. For all related safety analyses randomized treatment starts with the first intake of randomized, double-blind study drug. Temporary discontinuation of the study drug will not be counted as randomized treatment phase discontinuation.

Post-randomized treatment phase

The post-randomized treatment phase (usually after unscheduled, permanent study drug discontinuation) begins after last study drug intake + 1 day and ends on the date last seen (or vital status confirmed by indirect contact).

Double-blind (DB) phase (Randomized treatment epoch)

The double-blind phase is the combination of the randomized treatment phase and the post-randomized treatment phase.

Baseline for run-in phase

The treatment run-in baseline is defined as the last available measurement prior to or at time of Visit 101 or Visit 102, whichever is earlier. For some parameters, values may be only collected at screening (Visit 1). For them the treatment run-in baseline will be at screening.

Baseline for double blind phase

The baseline for the double blind phase is defined depending on the parameter considered. In general, the baseline for the double blind phase is defined as the last available measurement during the LCZ696 *run-in* phase and randomization (generally Visit 199 or Visit 201). For the parameters that are not designated to have measurements collected during the LCZ696 run-in or randomization, the run-in baseline may be used as the baseline for the double blind phase. The following parameters will have the double blind phase baseline at randomization visit:

NYHA class, HF signs and symptoms, vital signs, waist/hip circumference, eGFR, safety laboratory values, endpoints, 1st morning urine (sub-study patients only), lab assessments, patient global assessment (it is assumed that patients will assess their health status comparing to that at randomization), KCCQ scores and EQ-5D assessments and MMSE.

The following parameters will have their double blind phase baseline measured only at screening visit (Visit 1), i.e., prior or at Visit 101 or 102:

Height, (HF, diabetes, liver, and other) medical histories, demography, echocardiogram, and inclusion/ exclusion criteria.

NT-proBNP will be collected in all patients at Visits 1, 101/102 (whichever occurs first), 103, 199/201, 203 and 205. Additional biological markers will be collected in a subset of patients. For these biomarkers, the baseline will be defined as values at Visit 101 or 102, whichever occurs first.

Unscheduled visit

Only for the analysis of safety laboratory evaluation will unscheduled measurements be taken into account. For efficacy evaluations, measurements from unscheduled visits will generally not be used, unless specifically specified.

Analysis cut-off date (Study endpoint for analysis)

The trial was designed to be terminated when 4600 patients have experienced a total of about 1847 confirmed primary composite events of heart failure hospitalization and CV death. If the assumptions of the study design were kept unchanged, the analysis cut-off date will be a predicted date when either the targeted 4600 patients have experienced 1847 primary composite events during a period of 2.42 years of enrollment plus 2.17 years of minimum follow-up or a study termination date has been decided based on other study termination decisions. Analysis cut-off date is determined to be Apr 30, 2019 based on prediction analysis.

Final visit

The final visit will be the last time a patient's study record has been taken. For a patient who is still alive and in the study the final visit will be his/her last clinical visit. For a patient who died or is determined to be lost to follow-up before the analysis cut-off date, the final visit will be his/her death/lost-to-follow-up day. For a patient who is alive but no longer has regularly scheduled clinic visits and whose study records only can be obtained by telephone or indirect contact, the final visit will be the day the study records being taken.

Close-out

Close-out is a final study period (about two-to-three months) consisting of arranging patients' final visits and completing study conduct, starting from the date (close-out date) when all the study sites are informed to schedule patients' final visits for those still alive. The close-out process will start before the analysis cut-off date. The final visit for each patient may occur prior to or immediately after the analysis cut-off date, depending on the timing the final visits can be scheduled in each site.

5 Missing date handling

Visit date missing

In any analysis or evaluation, if the visit date(s) is used but is missing, then the date(s) calculated based on the planned date(s) in the schedule specified in the protocol should be used to impute the missing date(s).

Event date missing or partially missing

If the date of an event is not known or is incomplete, the imputation rules are:

a) If the day of the event is unknown, then the 15th day of this month will be imputed for a missing day;

b) If only the month is unknown, then July will be used for imputation of the missing;

c) If only the year of the event is known, then the 1st of July will be imputed for a missing day and month;

d) The above rules are only for general case. If there is additional information available for the missing date, then the information should be used and the imputation of missing date should be treated differently. For example, if an event occurs between two visits and its date is missing, then the date in the middle of these visits may be used.

Medication stop date missing

If medication stop date is unknown or is incomplete, the imputation rules are:

a) If only the day field of the drug stop is missing, then the missing date is imputed by using the 15th of the month;

b) If year and month are missing, then use the next scheduled visit date (using the protocol specified visit schedule) from the previous last non-missing visit date to replace the missing drug stop date;

c) If the drug stop date is completely missing, then:

    a. If patient had fatal AEs (identified as either start or end date is equal to the date of death and the AE is flagged as an SAE), handling rules are (in the specified order):

        i. AE end date is not missing: use the AE end date to replace the missing drug stop date;

ii. AE end date is completely missing but AE onset date not missing: use the AE onset date to replace the missing drug stop date;

iii. AE end date is partially missing (only day field is missing): use Novartis standard procedure to impute the AE end date, and then use the imputed AE end date to replace the missing drug stop date;

iv. AE end date is completely missing and AE onset date is partial missing (missing the date field only): impute the AE onset date using Novartis standard procedure, and then use the imputed AE onset date to replace the missing drug stop date;

v. If both AE onset and end dates are completely missing, then use the last previous non-missing visit date plus 35 days to replace the missing drug stop date.

b. If patients had no fatal AEs, handling rules are the same with the case where year and month are missing.

6 Subject disposition, background and demographic characteristics, medical history

6.1 Subject disposition

The number and percentage of patients successfully screened will be presented. In addition, the reasons for screen failures will be provided. For patients who are screened more than once, the information from the last screen will be used in the summary. The analysis set is the SCR.

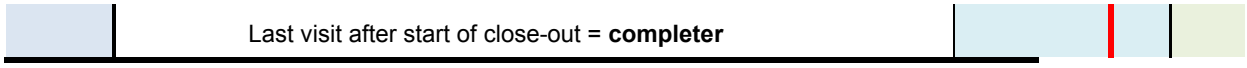
The number and percentage of subjects enrolled, completed, and failed in run-in will be summarized for the valsartan run-in and LCZ696 run-in separately. The reasons for run-in failures will be provided. The analysis set is the ENR. The same information for the valsartan run-in phase (VRS) and the LCZ696 run-in phase (LRS) will also be summarized.

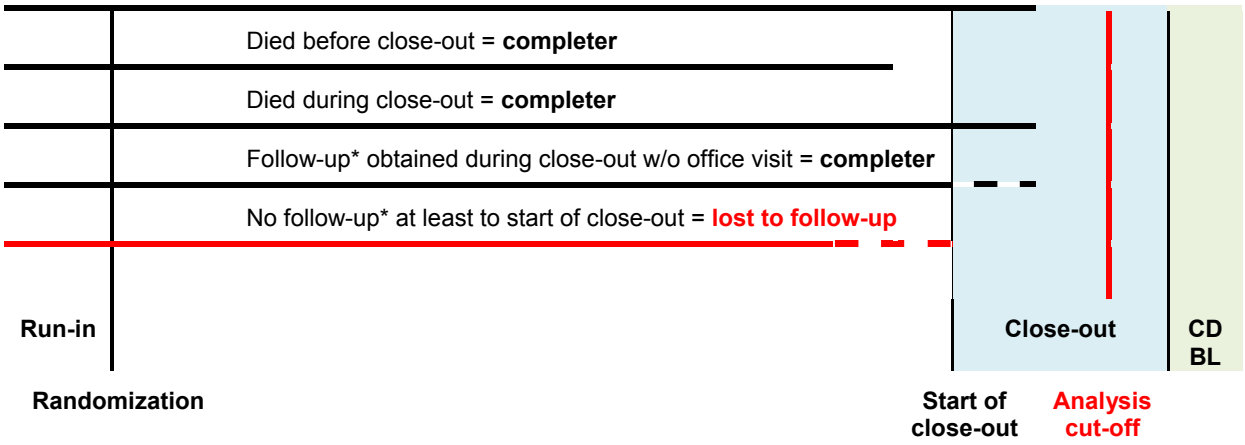
The number of subjects randomized (RAN) and included in the full analysis set (FAS) will be presented by treatment group. The number and percentage as well as the reasons that subjects had been excluded from the RAN will be summarized by treatment group. The number and percentage of subjects in the FAS who completed the study ("Study completer" see Figure 6-1), who discontinued the study and the reasons for discontinuation will be presented for each treatment group. The number and percentage of subjects with protocol deviations as well as the criteria leading to exclusion from analysis sets will be presented in separate tables for the FAS.

Study completer is defined as the one died or having vital status available after the close-out date.

Furthermore, the number of subjects enrolled and randomized per region and per country will be presented descriptively for the ENR and the FAS, respectively.

Figure 6-1 Definition of study completer





*: Follow-up means patient's vital/medical status information. It may be obtained through indirect contact.

As described in Section 3, the disposition analysis will be repeated for the following subgroups:

- Age group <65, ≥65 years
- Age group <75, ≥75 years
- Gender (male, female)
- Race (Caucasian, Black, Asian, Other)
- Region
- Baseline LVEF <=median, > median

6.2 Background and demographic characteristics

For the run-in phase, summary statistics will be provided by total number of patients based on the ENR for background and demographics, disease characteristics, and cardiovascular risk factors for the run-in phase baseline, including the following parameters: age, age group (<65 years vs. ≥65 years; <75 years vs. ≥75 years), sex, race, ethnicity, weight, height, body mass index (BMI) (mean summary and frequencies in ≤30, >30 and ≤35, and > 35), category of prior CHF medication, prior HF hospitalization, etiology of HF (ischemic vs. non-ischemic), NYHA class, NT-proBNP, BNP, history of MI, history diabetes, history of hypertension, device use, eGFR (mean summary and frequencies in < 60 and ≥ 60), LVEF (mean summary and frequencies in < 50 and ≥ 50), vital signs, and echocardiography parameters. BMI will be calculated as weight (kg) / height^2 (m^2) from the collected height and weight at Visit 1 (Screening Visit).

Similarly, for the double blind phase, summary statistics will be provided by treatment group and total based on the FAS for the above mentioned parameters plus: category of other select cardiovascular medications, duration of heart failure, medical history (including history of MI, diabetes mellitus, hypertension, atrial fibrillation/flutter, stroke). The summary will be based on the data collected at the double blind phase baseline, if available. Biomarkers will use their baseline at Visit 101/102.

Continuous variables will be summarized using n, mean, standard deviation, median, minimum, the first quartile, the third quartile, and maximum. For in-text tables, minimum and maximum will not be presented. Categorical variables will be summarized using frequency and percentage.

As described in Section 3, the background and demographic characteristics analysis will be repeated for the following subgroups and for the FAS:

- Age group <65, ≥65 years
- Age group <75, ≥75 years
- Gender (male, female)
- Race (Caucasian, Black, Asian, Other)
- Region
- Baseline LVEF <=median, > median

For baseline comparability analysis, categorical variables will be analyzed by Cochran-Mantel-Haenszel Chi-square test (Fleiss 1981) and continuous variables will be analyzed by F-test. The results of these analyses will be presented in Appendix 16.1.9 only. For categorical variables, the number and percentage of subjects with missing data will be provided but those subjects who are included in the missing category will not be in the denominator for the Cochran-Mantel-Haenszel Chi-square test.

Note that these tests of comparability are performed for descriptive purposes only, and will not serve as a basis for determining entry of explanatory variables into the respective models. However, when these tests yield statistically significant results referring to a p-value less than 0.05, they can be used as supportive information in interpreting the statistical analyses performed on the primary and secondary efficacy variables.

6.3 Medical history

Any condition entered on the relevant medical history / current medical conditions CRF will be coded using the most updated version of MedDRA dictionary. Medical history includes heart failure history and cardiovascular disease history, and other medical history in this study, which are collected at Visit 1 (Screening visit). The number and percentage of subjects with each medical condition will be provided by treatment group and system of organ class for the ENR set and for the FAS set.

7 Study medication

7.1 Run-in phase

7.1.1 Duration of treatment exposure

The duration of the run-in study treatment exposure will be summarized (n, mean, standard deviation, minimum, Q1, median, Q3 and maximum) by treatment in a sequential manner: the valsartan run-in phase (VRS) followed by the LCZ696 run-in phase (LRS) and (frequency and

percentage of patients) by exposure duration category in each phase. The exposure duration categories are defined for each phase as:

For the valsartan run-in phase:

- 0 days
- 0 - 1 week
- 1 - 2 weeks
- >2 weeks

For the LCZ696 run-in phase:

- 0 days
- 0 - 1 week
- 1 - 2 weeks
- 2 – 4 weeks
- > 4 weeks

The valsartan run-in phase, as defined in Section 4, starts at the date when the patient receives the run-in valsartan and ends at the day prior to receiving LCZ696 or when the patient discontinued, whichever comes first. The LCZ696 run-in starts at the date when the patient receives the run-in LCZ696 and ends at the day prior to randomization or when the patient discontinued, whichever comes first.

The mean duration of *each* run-in treatment phase (valsartan and LCZ696) will be computed based on each sub run-in set for overall duration and for each up-titration dose levels 1 and 2 (40 mg BID, 80 mg BID, and 80 mg matching placebo BID for valsartan; and 50 mg matching placebo BID, 100 mg matching placebo BID, and 100 mg BID LCZ696, respectively) using the following algorithm:

- Overall valsartan run-in: valsartan run-in treatment duration (days) = min(date the patient died in run-in, date failed in valsartan run-in, date received the first run-in LCZ696 drug – 1) – date received the first run-in valsartan + 1;
- Overall LCZ696 run-in: LCZ696 run-in treatment duration (days) = min(date the patient died after receiving the first run-in LCZ696, date failed in LCZ696 run-in, date randomized –1) – date received the first run-in LCZ696 + 1.

In each sub run-in phase:

- Run-in dose level 1 for valsartan: Run-in dose level 1 valsartan treatment duration (days) = min(treatment start date of valsartan run-in dose level 2 medication – 1, date run-in failure date, date died) – run-in valsartan dose level 1 medication date + 1;
- Run-in dose level 2 for valsartan: Run-in dose level 2 valsartan treatment duration (days) = min(treatment start date of LCZ696 run-in dose level 2 medication – 1, date run-in failure date, date died) – run-in valsartan dose level 2 medication date + 1;

- Run-in dose level 2 for LCZ696: Run-in dose level 2 LCZ696 treatment duration (days) = min(randomization date – 1, date failed run-in, date died) – run-in LCZ696 dose level 2 medication date + 1.

Time from the first study dose to its first top dose will also be summarized for valsartan and LCZ696, if necessary.

The above specified analysis will be repeated for the SAF to investigate the run-in drug administration profile for patients who eventually enter the double blind phase.

7.1.2 Dose level

Doses and dose levels in the run-in phase are:

Table 7-1 Study drug dispensed for the treatment run-in epoch by study visit

Study visit	Dose level	LCZ696	Valsartan
101[a]	1	50 mg matching placebo b.i.d.	40 mg b.i.d.
102	2	100 mg matching placebo b.i.d.	80 mg b.i.d.
103	2	100 mg b.i.d.	80 mg matching placebo b.i.d.

Number and percentage of patients will be summarized by dose level (or visit) and treatment.

This analysis will be repeated for the SAF.

7.2 Double-blind phase

7.2.1 Duration of treatment exposure

The duration of the treatment exposure for a patient, regardless of temporary interruptions of usage of the study drug, is defined as

- date of last study drug intake – first study drug date (usually the randomization date) + 1.
- Duration of the treatment exposure will be summarized by treatment group descriptively (i.e. n, mean, standard deviation, min, Q1, median, Q3, max) and the number (percentage) of patients will be summarized by exposure duration category:
- < 2 week
- 2 to < 8 weeks
- 8 weeks to < 4 months
- 4 months to < 1 year
- 1 to < 2 years
- 2 to < 3 years
- 3 to < 4 years
- 4 to < 5 years
- > 5 years.

Overall patient-years on-treatment and average patient-years on-treatment will be computed based on the duration of the treatment exposure as follow: Overall patient years

$$= \sum_{i=1}^{n} duration of\ treatment exposure_i / 365.25,$$

where *n* is the total number of patients in the SAF. Average patient-years on-treatment = overall patient-years on-treatment/*n*. They will be summarized by treatment group.

The durations on each dose level, time from randomization to the first dose down-titration, and time from the first dose down-titration to the start of the next up-titration for patients with dose down- and up-titration will also be summarized by treatment group.

Duration of total exposure to study drug (excluding interruptions) will be computed as

- date of last study drug intake – first study drug date (usually day after randomization date) + 1 – *number of days of treatment interruption*.

The duration of total exposure to study drug will be summarized by treatment group (mean, standard deviation, median, minimum and maximum) as well as frequencies per duration category defined below.

- < 2 week
- 2 to < 8 weeks
- 8 weeks to < 4 months
- 4 months to < 1 year
- 1 to < 2 years
- 2 to < 3 years
- 3 to < 4 years
- 4 to < 5 years
- > 5 years

Percentage exposure to study drug for each patient is defined as the ratio of the total exposure to study drug divided by the treatment exposure x 100. The percentage exposure to study drug will be summarized by treatment group (mean, standard deviation, median, minimum and maximum).

The time from randomization to the permanent discontinuation of study drug, excluding treatment discontinuation caused by death, will be summarized by Kaplan-Meier curve by treatment group.

7.2.2 Dose levels

Doses and dose levels in the double blind phase are:

Table 7-2 Study drug dose levels during randomized treatment epoch

Dose level	LCZ696 Treatment Arm	Valsartan Treatment Arm
3	200 mg b.i.d.	160 mg b.i.d.
2	100 mg b.i.d.	80 mg b.i.d.
1	50 mg b.i.d.	40 mg b.i.d.

Average dose and dose level will be summarized by visit and overall study and treatment group (mean, standard deviation, median, minimum and maximum). Average dose will be calculated as, in each treatment group, at a given visit:

Sum of reported doses (including zero dose for interruption) / Total number of patients.

The formula for the average dose level is similar. The overall average (daily) dose will be calculated as:

1. Daily dose for each patient = $\frac{\sum_{i=0}^{3}(no.of\ days\ on\ dose\ level\ i)\times(dose\ level\ i\times 2)}{\sum_{i=0}^{3}(no.of\ days\ on\ dose\ level\ i)}$
2. Average daily dose for each patient over patients in each treatment group.

For the average overall dose level, the formula is similar, replacing "dose level i×2" " by i.

Frequency and percentage of patients at each dose level will be summarized by visit and treatment group.

The last recorded treatment on the drug administration form will be presented with the number and percentage of patients on each dose level, including off-treatment. The average dose related to the last recorded treatment will be also be summarized by treatment group, for both cases of including all patients in the SAF and excluding off-treatment patients.

7.2.3 Other exposure analyses

The number and percentage of patients with down titrations will be summarized together with the study specific reasons for down titration. This summary will be repeated for down titrations to no treatment.

The number and percentages of treatment interruption episodes and the reasons for permanent treatment discontinuation and treatment interruption episodes (any and >14 days) will be summarized by treatment group.

7.2.4 Treatment exposure in subgroups

The above treatment exposure analyses in Section 7.2.1 will be repeated for the following subgroups:

- Age group <65, ≥65 years
- Age group <75, ≥75 years
- Gender (male, female)
- Race (Caucasian, Black, Asian, Other)
- Region

- Baseline LVEF <=, > median

7.3 Study exposure

As defined before (Section 4) the double-blind study phase = the on-randomized treatment phase + the post-randomized-treatment phase. The study exposure here means the exposure for the double blind study phase. The summary of the exposure for the study phase will be computed for the FAS.

The duration of the on-treatment phase exposure (duration of study drug exposure, regardless of temporary discontinuation of study drug) will be computed for the FAS using the following formula:

- Date of last study drug intake – randomization date + 1.

The duration of the post-treatment phase will be computed for patients in the FAS who complete the last scheduled visit of the double-blind phase or have the vital status available using the following algorithm:

- Duration of the post-treatment phase (days) = date last seen (or vital status confirmed by indirect contact) - date of the visit at which the study medications was permanently stopped + 1

The overall duration of the double blind study phase will be computed for patients in the FAS.

- Duration of the double blind study phase (days) = date last seen (or vital status confirmed by indirect contact) – randomization date + 1

The durations of the on-treatment, post-treatment phase, and overall double blind study phase will be summarized by treatment group (n, mean, standard deviation, median, minimum Q1, Q3, and maximum) and the number and percentage of patients, in terms of the duration of the overall double blind study phase, will also be summarized by exposure duration category.

The study exposure duration categories are defined as

- < 2 week
- 2 to < 8 weeks
- 8 weeks to < 4 months
- 4 months to < 1 year
- 1 to < 2 years
- 2 to < 3 years
- 3 to < 4 years
- 4 to < 5 years
- > 5 years

8 Concomitant medication

Medications will be identified using Novartis Drug and Therapy Dictionary, NovDTD which is a modified Novartis internal version of Q3 2004 WHO Drug Reference List (DRL) including Anatomical Therapeutic Chemical (ATC) code.

For the run-in phase, the run-in concomitant medication information will be summarized based on the ENR.

Prior and run-in concomitant medications taken at any time between the screening visit and the randomization visit will be summarized in sequential format by treatment of valsartan for the VRS and LCZ696 for the LRS based on the latest version of dictionary. Medications will be presented in alphabetical order, by ATC codes and grouped by *anatomical main group* (the 1st level of the ATC codes). Tables will show the overall number and percentage of patients receiving at least one drug of a particular ATC code and at least one drug in a particular anatomical main group.

Prior medications are defined as drugs taken prior to first dose of run-in study medication. Any medication given at least once between the day of first dose of run-in study medication and the last day prior to randomization visit will be a **run-in concomitant medication**, including those which were started pre-screening and continued into the run-in phase Prior or run-in concomitant medication will be identified based on recorded or imputed start and end dates of taking medication. The rules for imputing incomplete (start and end) dates are described in Section 5.

The concomitant medication information for the double blind phase will be summarized based on the SAF.

Concomitant medications taken at any time since the randomization visit will be summarized by treatment group based on the latest version of the dictionary. As before, medications will be presented in alphabetical order, by ATC codes and grouped by anatomical main group (the 1st level of the ATC codes). Tables will also show the overall number and percentage of subjects receiving at least one drug of a particular ATC code and at least one drug in a particular anatomical main group.

Any medication given at least once between the day of first dose of randomized study medication and the last day of study visit will be a double-blind concomitant medication, including those which were started pre-randomization visit and continued into the double blind treatment phase.

The following important classes of concomitant heart failure and CV medications administered during run-in and during double-blind phase will be summarized separately in a similar way as described above.

- ARBs
- ACE inhibitors
- Diuretics
- Beta blockers
- Aldosterone antagonists
- Cardiac glycosides (Digoxin/digitalis glycoside)

- Calcium antagonists
- Other vasodilators
- Anticoagulants
- Antiarrhythmic agents
- Aspirin
- Other antiplatelet agents
- Statins
- Nitrates
- Other lipid lowering agents

The classes of medications will be defined in a separate EXCEL sheet with ATC preferred term and WHO drug code. This EXCEL sheet will be stored in CREDI at the RAP level after the content is agreed to by the Global Program Medical Director (GPMD) and prior to CDBL

9 Efficacy evaluation

Efficacy evaluation will focus on the double blind phase and be based on the FAS. Endpoint events collected during run-in will not be utilized in the comparison of the efficacy endpoint.

The endpoint events related to the primary or secondary objectives collected during run-in will be simply summarized with adjustment for treatment exposure duration using the annualized event rates, according to the designed sequential sub-run-in periods. In case a baseline is needed for the evaluation, the run-in phase baseline will be referred to.

9.1 Event counting for efficacy

All events occurring prior to or at the analysis cut-off date will be included in the efficacy analyses.

Only adjudicated and confirmed events will be counted in the primary and secondary analyses. Event information received after the analysis cut-off date will not be included in the primary and secondary efficacy analyses, but they will be included in the investigator reported event analysis.

For a patient who did not complete the study, i.e. who was lost to follow-up or withdrew consent, if the vital status for the patient can be retrieved from public information (or vital status can be confirmed by indirect contact), the last known alive date defined for the analysis will be the last alive date obtained from the retrieved information.

9.2 Variables

The variables defined in this section are only for the double blind period. The baseline referred to in the variable definitions below is always the double blind phase baseline, unless otherwise indicated. For variables related to changes at Month 8, Month 8 refers to Week 32.

Primary variable:

- Cumulative number of primary composite events, i.e., the composite events of (total) HF hospitalizations and CV death, for a given subject, over time.

Secondary variables:

- Change from baseline in KCCQ clinical summary score at Month 8;
- Change from baseline to Month 8 NYHA class;
- Time from randomization to first occurrence of composite renal endpoint event, defined as either:
  - Renal death, or
  - a 50% decline in estimated glomerular filtration rate (eGFR) relative to baseline, or
  - reaching end stage renal disease (ESRD);
- Time to all-cause mortality.

Exploratory variables:

- Cumulative number of events of the extended composite endpoint of CV death, total HF hospitalizations, total non-fatal strokes, and total non-fatal MIs, for a given subject, over time;
- Change in clinical composite assessment (NYHA, global patient assessment, and clinical events defined as CV death and HF hospitalization) at Month 8;
- Patient global assessment at Month 8;
- Changes from baseline in health-related QoL (assessed by the total score, clinical summary score, and individual scores of the sub-domains from the KCCQ and assessments of the EQ-5D for health status);
- Number of HF events per-subject;
- Number of total worsening HF events and CV death per-subject. A subject will be defined as having a CV death or worsening HF event when the subject has:
  1. CV death or
  2. A hospitalization for HF or
  3. Received intravenous (IV) decongestive therapy (IV diuretics, IV neseritide or other natriuretic peptide, IV inotropes, and IV nitroglycerin [NTG]), and does not result in formal inpatient hospital admission, regardless of the setting (i.e. in an emergency room (ER) setting, in the physician's office, an outpatient treatment facility, etc.);
- Number of all-cause hospitalizations per-subject and number of cause specific hospitalizations per-subject;
- Number of days alive and out of hospital (DAOOH) at Month 12;

- Number of days alive and out of hospital (DAOOH) during the randomized, double-blind study period and percentage of DAOOH over potential follow-up time (defined as date of end of study minus date of randomization plus 1);
- Rate of change in eGFR (eGFR slope);
- Time from randomization to new onset diabetes mellitus (NODM);
- Time to new onset of atrial fibrillation (AF);
- Number of days staying in intensive care unit (ICU), number of stays in intensive care unit (ICU), number of re-hospitalizations for HF, and number of ER visits for HF;
- Indicator of 30 day HF rehospitalization (after a prior HF hospitalization during double blind phase);
- Number of rehospitalizations within 30 days after discharge;
- Time between HF hospital readmissions;
- Changes in pre-specified biomarkers (e.g., cardiac, vascular, renal, collagen, metabolism, and inflammatory biomarkers) from baseline to predefined time-points (in a subset of patients);
- Variables to characterize the PK of valsartan, AHU377, and LBQ657 at steady-state in patients receiving LCZ696 using population modeling and/or non-compartmental based methods;
- Change from baseline in Mini-Mental State Examination (MMSE) summary score at year 2 (week 96).

Among the variables above, some of them or their components need to be adjudicated to confirm the reported types and occurrences based on the unified criteria for analyses. The adjudication of these events is based on an adjudication process through an independent Endpoint Adjudication Committee. The following events will be assessed by the Endpoint Adjudication Committee during the study:

- All death events;
- Unplanned hospitalization for HF;
- MI;
- Stroke;
- Renal dysfunction (including occurrence of end stage renal disease);
- Others (new onset atrial fibrillation (NOAF), NODM)

In the analysis, as specified before, only adjudicated events will be utilized if the events included in the analysis are required to be adjudicated. However, in the sensitivity analysis, the investigator reported events may also be used.

As specified before, all events and endpoints occurred or measured prior to or at the specified analysis time-points (such as Month 8, Month 12, etc.), including the analysis cut-off date, will be counted, unless otherwise specified.

Censoring variables:

Number of events and time-to-event variable are usually subject to censoring. The censoring can be caused by withdrawal of consent, lost to follow-up, other competing events, or administrative reasons (interim or analysis cutoff date). Below are the rules for deriving censoring variables in the analyses (primary, secondary, exploratory):

- For all-cause death, the censoring will occur at minimum (date of withdrawal of consent, last known alive date, analysis cutoff date)
- For CV death, the censoring will occur at minimum(date of withdrawal of consent, last known alive date, analysis cutoff date, date when patient died for non-CV causes)
- For a non-mortality and non-composite endpoint, the censoring will occur at minimum(date of withdrawal of consent, date of lost to follow-up, analysis cutoff date, date when patient died)
- For a composite endpoint, the censoring will occur at the earliest censoring date among the components.

In some additional analyses, one may be interested in all events occurring before major protocol deviations. In such cases, the events occurred after having major protocol deviations need to be censored. The censoring variables are:

- Minimum (the censoring time defined for the primary, secondary, exploratory analyses, date of the first major protocol deviation).

9.3 Statistical hypothesis, model, and method of analysis

9.3.1 Primary analysis

All adjudication confirmed primary events occurred between randomization and the analysis cut-off date will be included in the primary analysis.

9.3.1.1 Analysis for primary endpoint

The primary efficacy endpoint of the study consists of the times to recurrent hospitalization due to heart failure and death time due to CV reasons during the patient's follow-up. The comparisons between two treatment groups will be made using statistical procedures which deal with such multiple event time observations. The semi-parametric proportional rates model (abbreviated as LWYY model) (Lin et al 2000) will be utilized for quantifying the treatment difference.

Specifically, let $\lambda_{ij}(t, x_{ij})$ be the individual rate of primary composite events for subject i in region j, given that the patient has not died from a CV reason at time t. It is dependent on the time from randomization (t) and treatment group (x_{ij}). Let $x_{ij} = 1$ if the subject is in the LCZ696 group and $x_{ij} = 0$ if the subject is in the valsartan group.

Under the proportional rates model, the individual rate function for the composite endpoint of CV death and total HF hospitalizations is assumed to be, $\lambda_{ij}(t, x_{ij}) = Y_{ij}(t)\lambda_{0j}(t)\exp(\beta_0 x_{ij})$, where $Y_{ij}(t) = 1$ if subject i in region j is at risk for HF hospitalization or CV death at time t and $Y_{ij}(t) = 0$ if subject i in region j is censored or died from a CV reason at time $t-$ and $\lambda_{0j}(t)$ is the baseline rate function for the event in region j. ($Y_{ij}(t) = I(\min(T_{ij}, C_{ij}) \geq t)$, where T_{ij} and C_{ij} are CV death time and censoring time for subject i in region j. $I(.)$ is an indication function.)

The primary hypothesis to be tested is, H_{10}: $\beta_0 \geq 0$ versus H_{1a}: $\beta_0 < 0$, where $\exp(\beta_0)$ is the relative risk or rate ratio (RR) of total hospitalizations for HF and CV death in the LCZ696 group relative to the valsartan group given the patient has not died from a CV reason at time t, which is assumed to be constant over time and across regions.

The primary hypothesis could be equivalently written as:

H_{10}: Rate ratio LCZ696/valsartan ≥ 1 versus H_{1a}: Rate ratio LCZ696/valsartan < 1,

A rate ratio < 1 indicates an effect in favor of LCZ696.

Note that $\exp(\beta_0)$ can also be considered as a ratio of two mean cumulative frequencies of a subject having hospitalizations for HF and CV death when the rate ratio is constant over time and CV mortality is balanced between treatment groups.

The primary hypothesis will be tested at a one-sided significance level of 0.024 adjusted for the interim analysis. The rate ratio and its 95% confidence interval will be estimated from the above proportional rates model through maximization of a partial likelihood score function. The resulting estimate of $\exp(\beta_0)$ is identical to the one described by Anderson and Gill (Anderson and Gill 1982), but unlike Anderson-Gill, a robust variance estimator (sandwich estimator) is used to account for the dependency of within subject events. Note that having CV death is not considered as a censoring variable, but as a primary endpoint event and a conditional factor in this analysis. The "conditional factor" means that the rate of the composite endpoint at time t is evaluated conditional on the patient being at risk of CV death at time t. Time to non-CV death will be considered as a censoring variable. Any censoring due to non-CV death is assumed to be non-dependent in the analysis.

The above presented analysis method based on LWYY provides a treatment comparison based on primary endpoint rate ratio conditional on not having died from a CV reason. A marginal interpretation of the estimates requires that any censoring of the primary composite events will be non-informative and may be challenging if there is an imbalance in CV mortality.

The LWYY analysis can be carried out using the following SAS code:

```
proc phreg data=primary_data covs(aggregate);
      model (time_start, time_rec)*status_rec(0)=treat/ties=efron rl;
      id sid1a;
run;
```

where `primary_data` 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_rec` (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 any non-CV death reasons for the given patient; `sid1a` is the subject ID.

As a supportive part of the primary analysis, the two components in the composite endpoint (total HF hospitalizations and CV mortality) will be analyzed separately to quantify the respective treatment effects and check the consistency between the composite and the components.

For the analysis of total HF hospitalizations component, occurrence of CV death can be regarded as semi-competing risk (informative censoring) and may introduce a bias in the treatment effect estimate for HF hospitalizations (dilution of effect size if the drug has a positive effect on both components). In order to address this concern and to account for the correlation between the two components, the joint modeling (frailty model) approach (Cowling et al 2006) will be used for the component analyses. The joint model is specified as follows.

$$\lambda_{ij}^{H}(t) = Y_{ij}(t)\lambda_0^{H}\exp(\beta_0 x_{ij})\nu_{ij}$$

$$\lambda_{ij}^{T}(t) = Y_{ij}(t)\lambda_0^{T}\exp(\gamma_0 x_{ij})\nu_{ij}^{\alpha}$$

$$\nu_{ij} \sim f(.)$$

where $\lambda_{ij}^{H}(t)$ is the rate function of the total hospitalizations for HF for subject *i* in region *j* and λ_0^{H} is its constant baseline and $\lambda_{ij}^{H}(t)$ has a similar format as the model described above for the primary endpoint except for the multiplier ν_{ij} that is a subject-specific random effect factor (frailty); and $\lambda_{ij}^{T}(t)$ is the hazard function for the component of CV mortality for subject *i* in region *j* and λ_0^{T} is its constant baseline and $\lambda_{ij}^{T}(t)$ is a frailty proportional hazards model. The parameter α measures the association between $\lambda_{ij}^{H}(t)$ and $\lambda_{ij}^{T}(t)$. The frailty distribution $f(\cdot)$ is taken to be the Gamma distribution with shape = θ and scale = $1/\theta$. The parameters in the joint model, including the unknown power for the frailty, will be estimated through maximization of the corresponding likelihood. The model assumes that the correlation between total HF hospitalizations and CV death for each patient is completely explained by x_{ij} and the frailty and the two analysis variables corresponding to the two components are independent given the patient's frailty and x_{ij}.

For the analysis of CV death, it will be analyzed using Cox's proportional hazards model with a fixed treatment group factor and stratified by region. The analysis in the joint frailty model for this component, which is actually subject specific, will be considered as supportive for the analysis for CV death whenever an inference is needed for it.

In all the proposed analyses, it is acknowledged that the study will not be powered to achieve statistically significant results for the CV death.

The proposed joint frailty model assumes a constant baseline intensity and constant baseline hazard model for the recurrent HF hospitalization and CV death components, respectively. The marginal model for the recurrent hospitalizations for heart failure of this joint frailty model is a negative binomial model.

As a sensitivity analysis, the constant baseline intensity function λ_0^{H} and constant baseline hazard function λ_0^{τ} will be replaced by piece-wise constant functions (i.e. multiple-step

functions). The number of steps in the step functions will be selected based on the Bayesian Information Criterion (BIC). The partitions of these constant step intervals can be obtained based on the equally-divided percentiles (Liu and Huang 2008).

The joint model will be estimated using SAS PROC NLMIXED. The below SAS code illustrates the program for a piece-wise constant intensity and hazard baseline function with 10 steps:

```
proc nlmixed data=data_joint qpoints=5 corr;
parms / data=inpar;
bounds r01 r02 r03 r04 r05 r06 r07 r08 r09 r10 h01 h02 h03 h04 h05 h06
h07 h08 h09 h10 alpha >=0;

/* g is Gamma distributed with mean 1 and variance 1/alpha*/
p=cdf('NORMAL', nu);
if p > .999999 then p=.999999;
g2=quantile('GAMMA', p, alpha);
g=g2 * 1/alpha;
/* baseline hazard and cum baseline hazard, recurrent events */
base_haz_r=r01*event_r1+r02*event_r2+r03*event_r3+r04* event_r4 +
r05*event_r5+r06*event_r6+r07*event_r7+r08*event_r8
+r09 * event_r9 + r10 * event_r10;
cum_base_haz_r=r01*dur_r1+r02*dur_r2+r03*dur_r3+r04*dur_r4 +
r05*dur_r5+r06*dur_r6+r07*dur_r7+r08*dur_r8+r09*dur_r9 +
r10 * dur_r10;
/* baseline hazard and cumulative baseline hazard for death */
base_haz_d=h01*event_d1+h02*event_d2+h03*event_d3+h04*event_d4 +
h05*event_d5+h06*event_d6+h07*event_d7+h08*event_d8+
h09 * event_d9 + h10 * event_d10;
cum_base_haz_d=h01 * dur_d1 + h02 * dur_d2 + h03 * dur_d3 +
h04 * dur_d4 + h05 * dur_d5 + h06 * dur_d6 + h07 * dur_d7 +
h08* dur_d8 +h09 * dur_d9 + h10 * dur_d10;
mu1= beta1 * trt + log(g);/* for recurrent event */
mu2= alpha1 * trt + gamma * log(g); /* for death event */
loglik1=-exp(mu1) * cum_base_haz_r;
loglik2=-exp(mu2) * cum_base_haz_d;
/*log likelihood for recurrent event */
if status_rec = 1 then loglik=log(base_haz_r) + mu1 ;
/*log likelihood for death */
if status_rec = 2 then loglik=loglik1 +log(base_haz_d)+mu2+loglik2;
/*log likelihood for censoring */
if status_rec = 0 then loglik=loglik1 + loglik2;
model time_stop ~ general(loglik);
random nu ~ normal(0, 1) subject = sid1a;
ods output ParameterEstimates=jm.est FitStatistics=jm.fit;
run;
```

where `data_joint` is the input dataset, `jm.est` and `jm.fit` for the parameter estimates and model fitted statistics. The dataset `inpar` contains the initial guesses for `r01-r10, h01-h10` and the other model parameters. The input dataset `data_joint` should include the values of event indicators `event_r1- event_r10` and `event_d1- event_d10` and durations `dur_r1- dur_r10` and `dur_d1- dur_d10`.

A non-parametric estimate of the primary endpoint event rates (Nelson-Aalen estimates) will be provided. Non-parametric estimates of conditional HF hospitalization rates (Nelson-Aalen estimates), given the patient has not died due to a CV reason, over time (Cook and Lawless 1997) and of unconditional HF hospitalization rates over time allowing for death as terminal event will be provided as well (Ghosh and Lin 2000).

9.3.1.2 Supportive analyses for primary efficacy

The supportive analyses for the primary analysis results will be performed in two ways:

1) Evaluating the impact of the informative censoring of CV death on the estimate of the relative rate reduction through performing additional analyses including:
   a. Checking of existence of any non-negligible imbalance in CV mortality between treatment groups through analysis for CV mortality alone. This analysis will be performed using Cox's proportional hazards model for time-to-CV death. If it deems necessary, pooling appropriate patient (sub-) population from PARADIGM-HF (CLCZ696B2314) and this study will be performed using the same Cox model to further assess any adverse signal on CV mortality between treatments;
2) Providing a series of sensitivity analyses to investigate the robustness and consistency of the primary efficacy results:
   b. Performing the primary analysis (the LWYY model) and the component analyses (the joint frailty model) on the PPS;
   c. Analysis using the WLW method (Wei et al 1989, Li and Lagakos 1997) on time to first (the conventional time-to-first event analysis) and time to 2nd, 3rd, 4th, 5th, and 6th composite events; an average effect (stratified by event number) will be provided for each of those analyses. This analysis will partially evaluate the robustness of the overall estimate of effect size against the impact from an (expected) small subgroup of patients with higher number of hospitalizations for HF. If the patient's first event is a CV death, then his 2nd, 3rd,…, Kth event will be counted as CV death as well. This analysis avoids informative censoring by counting CV death repeatedly as event. It is acknowledged that some of the analyses, time-to-1st event, for example, may not be powered;
   d. Analysis using a negative binomial regression model (McCullagh and Nelder 1989) on total number of primary composite endpoints with an offset of logarithm of time from randomization to censoring or death. The common rate ratio of composite events will be provided;
   e. Analysis using a multi-state model (Castaneda and Gerritse 2010) to assess the hazard ratios of transition intensities between different states defined by hospitalization status and CV death;

f. Performing the analysis for recurrent heart failure hospitalizations using the same LWYY model used for the primary endpoint with the conditional rate ratio. A weighted-pooled effect size and its 95% CI, using the results from this analysis and the Cox regression analysis for CV death outlined in 1), will also be provided. The estimated optimal weights will be obtained through the estimated covariance matrix of the two estimates of the component effect sizes.

3) Evaluating treatment effect on time to first primary composite event (cardiovascular death or hospitalizations for heart failure). The time to first primary composite event will be analyzed using Cox proportional hazards model with treatment as fixed factor and stratified by region. Additionally, a 1og-rank test will be performed.
4) Some additional supportive analyses may be performed.

Analysis using the multi-state models

The multi-state model utilized in this study is described in Figure 9-1. The transition intensities are defined as follows.

$h_{ij}^{EH}(t) = Y_{ij}^{H}(t)\rho_{0j}^{EH}(t)\exp(\alpha x_{ij})$ for being hospitalized for heart failure (H) at time t from the state of event free (E);

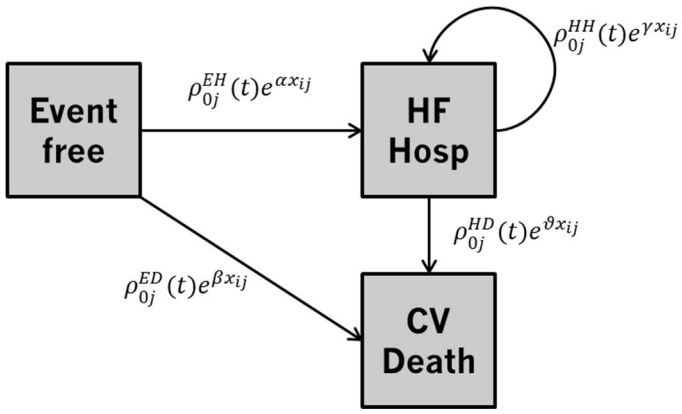
$h_{ij}^{ED}(t) = Y_{ij}^{D}(t)\rho_{0j}^{ED}(t)\exp(\beta x_{ij})$ for experiencing CV death (D) at time t from the state of event free (E);

$h_{ij}^{HH}(t) = Y_{ij}^{H}(t)\rho_{0j}^{HH}(t)\exp(\gamma x_{ij})$ for being re-hospitalized for heart failure (H) at time t after being previously hospitalized for heart failure (H);

$h_{ij}^{HD}(t) = Y_{ij}^{D}(t)\rho_{0j}^{HD}(t)\exp(\vartheta x_{ij})$ for experiencing CV death (D) at time t after being previously hospitalized for heart failure (H);

where i and j stand for subject i in rejoin j, $Y_{ij}(t)$s are the at-risk (to D or to H) indicators, and x_{ij} is the treatment group indicator defined above taking value 0 or 1.

Figure 9-1 Scheme of the multi-state model



In this model, we have assumed that the transition intensities for recurrent hospitalization for HF are proportional; and that all transition intensities for recurrent hospitalization for HF to CV death are proportional. We have four transition intensities to be modeled.

The data structure for analysis has a form below:

patid	time_start	time_stop	status	from	to	transition	time	treatment	region
1	0.00	8.70	0	E	H1	E -> H	8.70	1	1
1	0.00	8.70	0	E	D	E -> D	8.70	1	1
2	0.00	6.30	0	E	H1	E -> H	6.30	0	1
2	0.00	6.30	1	E	D	E -> D	6.30	0	1
3	0.00	2.25	1	E	H1	E -> H	2.25	1	2
3	0.00	2.25	0	E	D	E -> D	2.25	1	2
3	2.25	6.75	1	H1	H2	H -> H	4.50	1	2
3	2.25	6.75	0	H1	D	H -> D	4.50	1	2
3	6.75	11.36	0	H2	H3	H -> H	4.61	1	2
3	6.75	11.36	0	H2	D	H -> D	4.61	1	2
4	0.00	2.25	1	E	H1	E -> H	2.25	0	2
4	0.00	2.25	0	E	D	E -> D	2.25	0	2
4	2.25	6.75	1	H1	H2	H -> H	4.50	0	2
4	2.25	6.75	0	H1	D	H -> D	4.50	0	2
4	6.75	11.36	0	H2	H3	H -> H	4.61	0	2
4	6.75	11.36	1	H2	D	H -> D	4.61	0	2

where `status` is `0` for censored and `1` for event.

The below R code can be used for the analysis.

```
> library(survival)
> fit1 <- coxph(Surv(start, stop, status) ~ factor(transition_trt) +
cluster(patid) + strata(transition_region) , data=recurent_data)
```

or

```
> fit1 <- coxph(Surv(start, stop, status) ~ factor(trt) + cluster(patid) +
strata(region)   ,   data=recurent_data,   subset=(transition=='E2H'))
```

where `transition_region` is the derived from combination of `transition` and `region`, `transition_trt` is the derived from combination of `transition` and `trt`.

On-treatment LWYY analysis

On-treatment analysis is referring to the supportive analysis for the primary endpoint based on the FAS and with events censored after premature discontinuation of study drug. Specifically, the patient-specific follow-up period will be truncated at the date of last study drug intake + 29 days. Accordingly, the date of last study drug intake + 29 days will be used as censoring if smaller than the censoring date defined in Section 9.2.

Various death components used in composite endpoint

The LWYY analysis will be repeated replacing the CV death component in the primary endpoint with the following (less disease-specific) components as supportive analyses:

- Death due to CV or unknown cause;

- All-cause death

Time-dependent systolic blood pressure (SBP) covariate adjusted analyses

The analyses will be performed as supportive analyses to explore whether the treatment effect is driven by SBP changes:

- LWYY analysis for the primary composite endpoint of CV death and total heart failure hospitalizations;
- Cox proportional hazards model mentioned in (3) for the first primary composite endpoint.

9.3.1.3 Subgroup analysis

A full set of subgroups will be performed for subgroup analysis for the primary endpoint and its components, as outlined at Section 3.

Subgroup analyses will be performed for the FAS. To explore beneficial effects in subgroups or homogeneity of beneficial effects among subgroups, the estimated rate ratio, two-sided 95% confidence interval, and within subgroup p-value and p-value for the test for the treatment-by-subgroup interaction will be provided for each of the subgroups based on the proportional rates model in which treatment, subgroup, and treatment-by-subgroup are included as fixed-effect factors. Positive findings from these subgroup analyses have to be interpreted with caution since there is a non-negligible chance of false positives. Rate ratios from each subgroup will be graphically displayed using a forest plot.

9.3.2 Secondary analysis

9.3.2.1 Multiplicity adjustment among secondary comparisons

The four secondary hypotheses to be tested are:

- Comparison of change from baseline in KCCQ clinical summary score at Month 8 (denoted as H_{210});
- Comparison of change from baseline to Month 8 NYHA class (denoted as H_{220}, the parameter is the category change from baseline to Month 8 in NYHA class);
- Time from randomization to first occurrence of composite renal endpoint events (denoted as H_{230}, the parameter is the hazard ratio forthe composite renal endpoint events);
- All-cause mortality (denoted as H_{240}, the parameter is the hazard ratio for all-cause death).

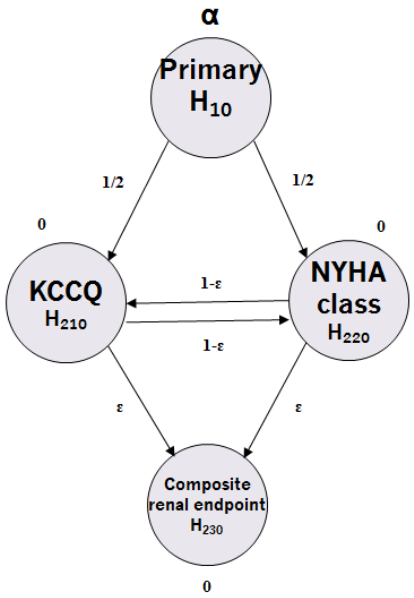
The secondary null hypotheses will be tested and statistical inferences will be made only if the primary null hypothesis is rejected. The four secondary efficacy hypotheses will be tested for superiority of LCZ696 to valsartan for the FAS.

The sequentially rejective multiple test procedures (Bretz et al 2009) will be used for testing the hypotheses of the primary endpoint and the first three secondary endpoints as indicated in Figure 9-2, to control the alpha level. The all-cause mortality (H_{240}) will be tested at a full level of alpha, after the rejection of the primary hypothesis, considering as a hard-endpoint.

In Figure 9-2, the fractions marked in the graph denote the weights or proportions of alpha to be added to the alphas in the next nodes according to where the arrows point once the current node of hypothesis is rejected.

- The initial allocation of (local) significance levels are α, 0, 0, and 0 for the comparisons of the primary endpoint and the first three secondary endpoints, respectively.
- The general algorithm for the testing procedure goes as follows: Test the hypotheses H_{10}, H_{210}, H_{220}, and H_{230} each at its local significance level defined above. If a hypothesis can be rejected, reallocate its level to one of the other hypotheses according to a pre-specified rule represented by the weighted graph in Figure 9-2. Update the reallocation weights in the reduced graph (the rejected hypothesis is removed from the graph) and repeat the testing step for the remaining no-rejected hypotheses with the updated local significance levels. This possibly leads to further rejected null hypotheses with associated reallocation of the local significance levels. The procedure is repeated until no further hypothesis can be rejected. The reallocation of the local alpha levels is fully determined by the initial graph given in Figure 9-2 and the update algorithm for this sequentially rejective multiple test procedure.
    - In the notation of Bretz et al 2009, a weight of ε for an edge indicates an infinitesimally small weight. If a hypothesis (vertex) with such an outgoing edge is rejected and the vertex removed, no significance level is passed on along such an edge as long as there are other outgoing edges with positive weights. If after removal of another vertex only infinitesimal outgoing edges remain, then the algorithm of Bretz et al 2009 turns them into edges with positive weights that sum to 1. In this specific procedure, this implies that no significance level is passed from the secondary null hypotheses for KCCQ (H_{210}) or NYHA class (H_{220}) to the secondary null hypothesis for composite renal endpoint (H_{230}) until both KCCQ and NYHA class null hypotheses have been rejected. To ensure the sum of weights on all outgoing edges of a vertex is 1 when such edges are present, some edges are given a weight of $1 - \varepsilon$.
-

Figure 9-2 Illustration of weights for alpha relocation in the sequentially rejective multiple test procedure for the secondary hypotheses



More specifically, the multiple testing procedure will be carried out in the following steps:

1). First the null hypothesis for the primary endpoint H_{10} is tested at full alpha (e.g. at one-sided level of 0.024). If not rejected, stop; if rejected, go to 2)

2). The null hypotheses for KCCQ (H_{210}) and NYHA (H_{220}) are tested at ½ alpha simultaneously.
   a. If none of them are rejected, stop;
   b. if both of them are rejected, go to 3);
   c. if only one of them is rejected (e.g. H_{210} for KCCQ), the other one (e.g. H_{220} for NYHA) can be re-tested at full alpha. If this is still not rejected, stop; if this is rejected, go to 3);

3). The null hypothesis H_{230} for composite renal endpoint is tested at full alpha level.

Note: The power analyses for each hypothesis in Figure 9-2 are presented in the sections for the sample size discussions.

For publications the secondary variables may be assessed using a full level of alpha without consideration of multiplicity adjustment.

9.3.2.2 Analysis of change in KCCQ clinical summary score at Month 8

The KCCQ instrument includes multiple domains. Only the domains that address HF symptoms and physical limitations will be analyzed for the secondary objective. The clinical summary score of KCCQ is computed as the mean of the following two domain scores:

- Physical limitation score
- Total HF symptom score

If a patient dies, a worst score (score of 0) will be imputed for the clinical summary score at all subsequent scheduled visits after the date of death where the clinical summary score would have been assessed.

Changes from baseline in KCCQ clinical summary score at Month 8 will be analyzed based on a repeated measures ANCOVA model in which treatment, region, visit, and treatment-by-visit interaction will be included as fixed-effect factors and baseline value as a covariate, with a common unstructured covariance matrix among visits for each treatment group. Treatment comparisons and effect size estimates at Month 8 will be provided. The analysis will be performed based on all available data up to Month 8 in the FAS and based on likelihood method with an assumption of missing at random (MAR) for missing data.

To investigate the sensitivity of the analysis results based on the MAR assumption, we will perform an analysis using a pattern mixture model approach which assumes, besides the imputation with 0 for the KCCQ scores after death, that missing KCCQ scores after HF hospitalizations were missing not at random (MNAR). This sensitivity analysis would first create multiple (100) imputations of missing KCCQ values under a MAR assumption, resulting in multiple (100) complete data sets. The imputation model assumes a multivariate normal model for the 3 KCCQ measurements at baseline, Month 4 and Month 8. This model will be estimated separately within each treatment by region stratum. For each completed data set, the imputed values at a given visit of patients who had a HF hospitalization since the preceding visit, a penalty factor (<1) would be multiplied to the imputed KCCQ scores to reflect that the MAR approach may have overestimated the missing KCCQ values. To the resulting 100 datasets, the original analysis model described above will be fitted, yielding 100 sets of parameter estimates and associated covariance matrices. These are then combined by using Rubin's rules to derive overall estimates, confidence intervals that adequately reflect missing data uncertainty as well as associated p-values. The analysis will be repeated for varying penalty factors in a range between 0.6 and 1.

In addition, a responder analysis for KCCQ clinical summary score change from baseline at Month 8 (defined separately for patients with at least 5 point deterioration and for patients with at least 5 point improvement) will be performed using a generalized mixed model with treatment, region, visit, and visit-by-treatment as fixed factors and baseline score as covariate, with a common compound symmetry covariance matrix among visits for each treatment group. Treatment comparisons and effect size estimates at Month 8 will be provided. The analysis will be performed based on all available data up to Month 8 in the FAS. The goal of the responder analysis is to assess the clinical relevance of the difference between the two groups in the mean change from baseline. As an additional sensitivity analysis, the above specified analyses will also be performed for the missing data due to death without any imputation.

9.3.2.3 Analysis of NYHA class at Month 8

Change from baseline to month 8 NYHA class will be analyzed using a repeated measures proportional cumulative odds model. The response variable is the category change from baseline to any given time points (4 weeks and 16 weeks) up to Month 8 (32 weeks) (expressed as improved, unchanged, worsened). The NYHA class change after patients who have died will be categorized into 'worsened'. The model will include patient as a random effect and the randomized treatment phase baseline NYHA class, region, treatment, visit and treatment-by-

visit interaction as fixed effect factors. This model assumes that the treatment effect sizes across measurement categories are the same. The analysis will be performed based on all available data up to 8 months in the FAS and based on likelihood method with an assumption of missing at random (MAR) for missing data not due to death. The estimated between treatment effect (reductions in odds) with the associated two-sided 95% confidence interval at Month 8 will be provided. As supportive analysis, the same model will be performed on data with no imputation for missing data in NYHA class change due to death.

9.3.2.4 Analysis of time-to-first occurrence of composite renal endpoint event

Time-to-first occurrence of composite renal endpoint events will be analyzed using Cox's proportional hazards model with a fixed treatment group factor and stratified by region. The estimated hazards ratio and the corresponding two-sided 95% confidence interval will be provided.

The Kaplan-Meier curves by treatment group will be presented. Additionally, the frequency and percentage of patients with composite renal endpoint events will be provided by treatment group.

Additional to the above major analysis for the composite renal endpoint, which is based on the criterion that dialysis for ≥ 30 days without known recovery of renal function, a supportive analysis with the same model defined above will be performed on a further restrictive endpoint which requires a verified continuation of dialysis for > 90 days.

9.3.2.5 Analysis of time to all-cause mortality

As indicated above, all-cause mortality endpoint will be tested at a full level of alpha, after the rejection of the primary hypothesis, considering as a hard-endpoint.

It will be analyzed using Cox's proportional hazards model with treatment factor and stratified by region. The estimated hazards ratio and the corresponding two-sided 95% confidence interval will be provided. The p-value for the score test will be utilized for inference.

The Kaplan-Meier curves by treatment group will be presented. Additionally, the frequency and percentage of all-cause mortality will be provided by treatment group.

9.3.2.6 Subgroup analysis for the secondary endpoints

Subgroup analysis for each of the secondary variables will also be performed similarly as described in Section 9.3.1.3 for the primary variable.

9.3.3 Exploratory analysis

In general, exploratory variables (Section 9.2) will be analyzed in the FAS unless specified otherwise. Statistical tests will be performed at the two-sided significance level of 0.05. To better satisfy the normality assumption, the log-transformation will be applied to data on each biomarker prior to statistical analysis. There will be no multiplicity adjustment for analysis of exploratory variables.

9.3.3.1 Analysis for time-to-event variables

There are two types of time-to-event exploratory variables:

- Time-to-first-event variables
    - Time to new onset of DM;
    - Time to new onset of AF event.
- Time-to-recurrent-event variable
    - composite endpoint of CV death, total HF hospitalizations, total strokes, and total MIs

Time-to-first event variables will be analyzed using a Cox proportional hazards model with treatment factor and stratified by region. The estimated hazards ratio and the corresponding two-sided 95% confidence interval will be provided. The Kaplan-Meier estimates may also be provided by treatment group, in tables and graphs.

New onset of AF refers to the adjudication-confirmed AF experienced after randomization. For this endpoint, only patients who have no investigator-reported medical history of AF prior to screening visit and no adjudication-confirmed AF between screening and randomization visits will be included in the analysis population.

For time-to-recurrent-event variable, the composite endpoint of CV death, total HF hospitalizations, total strokes, and total MIs will be analyzed using the same proportional rates model (LWYY). In the component analysis, the four components will be grouped into two components: CV death and the composite of total HF hospitalizations, total strokes, and total MIs. The estimated rate ratios and the corresponding two-sided 95% confidence intervals will be provided for the composite endpoint and its two grouped components.

9.3.3.2 Analysis of continuous variables

Continuous variables, which can be analyzed using the linear models with or without transformation on the responses, are:

- Rate of change (slope) in eGFR from baseline to endpoint;
- Changes in health-related quality of life (assessed by clinical summary score, total score and individual scores of the sub-domains from the KCCQ and scores of the EQ-5D for health status (EQ-5D index derived from EQ-5D descriptive system and EQ-5D VAS) from baseline to pre-defined time-points;
- Changes (in log-scale when skewed) from baseline to pre-defined time-points in pre-selected biomarkers (e.g., cardiac, vascular, renal, collagen, metabolism, and inflammatory biomarkers);
- Summarize pharmacokinetics (PK) parameters at steady-state using population modelling;
- Days alive out of the hospital during the first 12 months (48 weeks), and during the entire double-blind period;
- Change from baseline in MMSE summary score at year 2 (week 96).

Analysis for rate of change (slope) in eGFR

For the rate change in eGFR, the eGFR slope will be estimated from a repeated measures ANCOVA model including treatment, region, time (when the eGFR is assessed in months), and treatment-by-time as fixed effects with random intercept and slope (time) and a common unstructured covariance. The least-squared means of slopes for within and between treatment groups, and the corresponding two-sided 95% confidence intervals will be provided.

Analysis of changes in health-related quality of life change

Changes from baseline in KCCQ mean scores (for all 10 scores), similarly for changes from baseline in EQ-5D VAS and EQ-5D index (if available), will be analyzed based on a repeated measures ANCOVA model in which treatment, region, visit, and treatment-by-visit interaction will be included as fixed-effect factors and baseline value as a covariate, with a common unstructured covariance matrix among visits for each treatment group. Treatment comparisons and effect size estimates, including two-sided 95% confidence intervals, at different visits will be provided. The analysis will be performed based on all available data up to 3 years in the FAS and based on likelihood method with an assumption of missing at random (MAR) for missing data.

In addition, a responder analysis for KCCQ clinical summary score change from baseline at various visits (defined separately for patients with at least 5 point deterioration and for patients with at least 5 point improvement) will be performed, refer to the section for the analysis for categorical data.

In the analyses of the KCCQ scores, including the responder analysis, both the worst-case imputation (score of 0) and no imputation for missing data due to death (including subsequent visits) analyses will be performed.

For the patients with KCCQ data collected at screening and run-in visits, the above specified analyses, including the responder analysis, will be repeated using the run-in epoch baseline.

Changes from baseline and actual values will also be presented by visit and treatment group for KCCQ scores (both from run-in epoch baseline and from randomized epoch baseline) and EQ-5D (VAS) (from randomized epoch baseline) assessments using descriptive statistics of N, mean and SD, and will be graphically displayed using line plots.

Note that since the EQ-5D descriptive system contains ordinal variables (the dimensions of mobility, self-care, usual activities, pain/discomfort and anxiety/depression, with each having 3 levels (no problems, some problems, severe problems)), besides the continuous scores of index and VAS, frequencies and percentages of patients in each category will be provided by treatment group and visit.

Analysis in biomarkers

For pre-selected biomarkers, the log-transformed ratio to baseline will be calculated for each subject at every post-dose time point i.e. log (post-dose value/baseline). The change from baseline to a pre-defined time-point (Week 16 and Week 48) in logarithmic scale will be analyzed using a similar repeated measures ANCOVA model as described above. Estimates of treatment difference sizes and their two-sided 95% confidence intervals will be provided. These

estimates and confidence intervals will be back-transformed for presentation. In addition, the following summary statistics will be presented over time by treatment for each variable:

- The summary statistics: n, mean, SD, median, minimum, maximum, Q1 and Q3 will be presented for the baseline values, absolute post-dose values and change from baseline values.
- The geometric mean will be presented for the baseline values, absolute post-dose values and for the ratio to baseline values. The geometric mean of the ratio to baseline will be presented in terms of % change from baseline and will be calculated as follows: (exp(mean of the log-transformed ratio to baseline values) -1)*100
- The Coefficient of Variation (CV) will be calculated for the baseline values, the absolute post-dose values and the ratio to baseline values.

Analysis of pharmacokinetics (PK) parameters at steady state

The population-based PK parameters will be summarized at selected time-points by dose-level. The analysis set will be those in full analysis set (FAS) who participated in Pharmacokinetics sub-study.

- The summary statistics: n, mean, SD, median, minimum, maximum, Q1 and Q3 will be presented for the baseline values and absolute post-dose values.
- The geometric mean will be presented for the baseline values and absolute post-dose values.
- The Coefficient of Variation (CV) will be calculated for the baseline values and the absolute post-dose values.

Analysis of days alive out of hospital

For days alive out of hospital during the first 12 months, it will be treated as a continuous variable. The mean difference between treatment groups will be compared using an ANCOVA model with factors of region and treatment group. Estimate of treatment difference size and its two-sided 95% confidence interval will be provided.

As further exploratory analysis, the total number of DAOOH will be analyzed based on an ANCOVA with treatment group, region and potential follow-up time (dependent on time of randomization) as covariates. For patients who are lost to follow-up, the potential follow up time will be reduced accordingly. The percentage of DAOOH (relative to potential follow-up time) will be analyzed based on an ANOVA with treatment group and region as covariates.

Analysis of changes in MMSE summary score at 2 years

The change from baseline in the summary score of MMSE will be analyzed using a repeated measures ANCOVA model in which treatment, region, baseline hypertension status, baseline diabetes status, visit (Week 48 and Week 96) and treatment-by-visit interaction are included as fixed-effect factors and baseline age, MMSE baseline value and visit by MMSE baseline interaction as covariates, with a treatment-specific unstructured covariance matrix among visits for each treatment group. The adjusted mean changes at week 96 within each treatment, the difference in mean changes at week 96 between two treatments, its 95% confidence interval obtained from the above model will be presented. The analysis will be performed based on all

available data up to 96 weeks for patients with both baseline and at least one post-baseline MMSE assessment, and based on likelihood method with an assumption of missing at random (MAR) for missing data. In case that if there is any a problem with model convergence, the factors of baseline hypertension and diabetes status may be dropped from the modeling.

With a total of approximately 2500 patients expected to be eligible for the above analysis, there is more than 90% probability that the lower bound of the 95% confidence interval of the between-treatment difference of mean MMSE summary scores at Week 96 is $\geq$ -0.5 (exclude a between-treatment difference of 0.5 or more reduction), assuming that there is no true between-treatment difference and the common standard deviation is 3. A between-treatment difference of 0.5 or less is not considered clinically meaningful.

To investigate the sensitivity of the analysis results based on the MAR assumption, analyses based on multiple imputations, such as those using a pattern mixture model approach, may be performed, which assumes that missing MMSE scores after death, HF hospitalizations, or stroke were missing not at random (MNAR).

Similar analyses as above will also be performed in the subset of patients who have ApoE4 genotype assessment done. For these analyses, additional factors such as ApoE4 status (yes/no) and the interactions of treatment by ApoE4 status, and ApoE4 by visit may be added in the modeling. The ApoE4 gene has been identified as a genetic marker that predicts predisposition to cognitive decline and Alzheimer's disease (Protocol Section 6.6.9.1). This genetic marker is assessed at Visit 103 or at any visit thereafter when consent is obtained and is considered as the baseline characteristic of participating patient because it is unlikely for such genetic marker to change over a short period.

In addition, changes from baseline in the summary score of MMSE will also be analyzed using a repeated measures ANCOVA model in which treatment, geographic region, baseline hypertension status, baseline diabetes status, visit (Weeks 48, 96, 144, 192, 240), and treatment-by-visit interaction are included as fixed-effect factors, and baseline age, baseline MMSE value and visit by MMSE baseline interaction as covariates, with a common unstructured covariance matrix among visits for each treatment group. Treatment comparisons and effect size estimates and their 95% confidence intervals at different visits (only for visits at Week 48, 96, 144) and averaged over-all-visits will be provided. The analysis will be performed based on likelihood method with an assumption of missing at random for missing data.

Similar repeated measure ANCOVA model will be fitted in the subgroup of patients who have ApoE4 genotype assessment done.

A descriptive summary of MMSE absolute score and change from baseline over time will be tabulated and plotted. Changes in MMSE individual test domain scores from baseline over time will be tabulated as well.

All analyses in MMSE score will be performed in the FAS for patients who have participated in the cognitive function assessment sub-study.

9.3.3.3 Analyses of categorical and ordinal variables

Binary variables include:

- Indicator of 30 day HF rehospitalization (after a prior in-study HF hospitalization);
- Responder for at least 5-point change in deterioration and responder for at least 5-point improvement for KCCQ scores. .

The 30-day HF rehospitalization will be analyzed using logistic regression with treatment and region as fixed-effect factors. For 30 day hospital readmission, the analysis set will be restricted to patients who had hospitalization for HF at least once during double blind period.

The responder analysis for KCCQ score changes from baseline (for both run-in epoch baseline and randomized epoch baseline) at various visits (defined separately for patients with at least 5 point deterioration and for patients with at least 5 point improvement) will be performed using a generalized mixed model with treatment, region, visit, and visit-by-treatment as fixed factors and baseline score as covariate, with a common compound symmetry covariance matrix among visits for each treatment group. Treatment effect size estimates and 95% CIs at various visits will be provided. The analysis will be performed based on all available data up to Year 3 in the FAS.

Ordinal variables include:

- Assessment in the clinical composite assessment (improved, unchanged, and worsened) at post-randomization visits;
- Change in NYHA class from randomization;
- Changes in HF signs and symptoms from randomization;
- Patient global assessment at 8 months.

The first three variables will be analyzed, at Month 8, 1 year, 2 years, and 3 years, using a repeated measures proportional cumulative odds model. The model will include patient as a random effect and the randomized treatment phase baseline category (only for NYHA class and HF signs and symptoms), region, treatment, visit (all available post-randomization visits) and treatment-by-visit interaction as fixed effect factors. This model assumes that the treatment effect sizes across measurement categories are the same. The visit-wise effect size estimates and their 95% confidence intervals will also be provided. The analysis will be based on all available data in the FAS and likelihood method with an assumption of missing at random for missing data. The patient global assessment at 8 months will be analyzed similarly but using a non-repeated measures proportional cumulative odds model. The SAS procedure of GLIMMIX (SAS/STAT 9.3) can be used for the analysis.

The clinical composite assessment at a selected visit is defined as:

- Improved, if i) NYHA class decreases at least one level and Global Assessment is not worse at the selected visit and there is no major AE up to the selected visit; or ii) Global

assessment is improved and NYHA class does not increase at the selected visit and there is no major AE up to the selected visit.

- Worsened, if i) NYHA class increases at the selected visit; or ii) Global Assessment is worse at the selected visit; or iii) experiences a major AE up to the selected visit.
- Unchanged, if neither "Improved" nor "Worsened".

For Global Assessment, we converted the seven-category classification (recorded in the eCRF) to three-category classification in defining the clinical composite assessment:

- Markedly improved and Moderately improved to Improved;
- Moderately worsened and Markedly worsened to Worsened;
- Slightly improved, Unchanged, and Slightly worsened to Unchanged.

Major AEs are adjudication-confirmed CV death and heart failure hospitalizations.

9.3.3.4 Analysis of count data

Count data include:

- Number of HF events;
- Number of worsening HF events or CV deaths;
- Number of hospital admissions;
- Number of days/stays in ICU;
- Number of stays in ICU;
- Number of re-hospitalizations;
- Number of ER visits for HF;
- Number of 30 day hospital readmissions (after a prior in-study HF hospitalization).

Count variables will be analyzed using a negative binomial model (McCullagh and Nelder 1989) with the count data as the dependent variable and treatment group and region as fixed-effect factors and log(follow-up duration) as the off-set. The model estimated event rates (intensities/risk rate) and their 95% confidence intervals will be provided by treatment group. The treatment comparison will be performed through the estimated ratio of risk rates. The estimated reduction in event rate (ratio LCZ696/valsartan) and its 95% confidence interval will also be provided. Note that, in case when the follow-up durations depend on the treatments (potentially informative censorings), the estimated rate reduction should be interpreted with caution. However, when LCZ696 has a reduction in mortality, the rate reduction estimated from the negative binomial model will become conservative. As mentioned before, for 30 day hospital readmission, the analysis set will be restricted to patients who had hospitalization for HF at least once during double blind period. This implies that the baseline treatment group balance obtained from the randomization may be broken.

Descriptive statistics (n, mean, standard deviation, median, max, and min) will be provided for these variables by treatment group. For each variable, the annualized count for each patient is calculated as the total counts during the study divided by the total duration of the double blind

treatment phase, multiplied by 365.25. The annualized counts will be summarized according to treatment groups.

9.3.3.5 Subgroup analysis

Subgroup of ACEI intolerant patients

The primary and secondary variables as well as the symptom based variables, such as the clinical composite assessment, and so on, will also be analyzed using methods proposed above for the sub-group of ACEI intolerant patients.

Subgroup analysis for exploratory analysis

Subgroup analysis for the exploratory objectives will be performed for change in MMSE total score from baseline and the EQ-5D assessments for Region (North America, Western Europe, Central Europe, and the United Kingdom), age group, diabetes, AF, NT proBNP (<median or > median), and use of aldo antagonist at baseline.

9.3.4 General handling of missing data in the efficacy analyses

9.3.4.1 Time to event variables

For time-to-event variables, the censoring rules are presented in Section 9.2.

9.3.4.2 KCCQ clinical summary score

The clinical summary score is a mean of the physical limitation and total symptom scores. The total symptom score is the mean of the symptom frequency and symptom burden scores. Each scale score (the physical limitation, symptom frequency or symptom burden) is calculated as the mean of its item scores and transformed to a 0–100 scale, with higher score indicating higher level of functioning. A score of 100 represents perfect health whereas a score of 0 represents dead. For patients who die, a worst score (score of 0) will be imputed for the clinical summary score at all subsequent scheduled visits after the date of death where the clinical summary score would have been assessed.

9.3.4.3 NYHA classification change over time

For categorical analysis in NYHA class change from baseline at month 8 or over time, missing NYHA classification due to death will be imputed as NYHA class worsened.

9.3.4.4 Other variables

For the variables other than those mentioned above, double blind post-baseline LOCF will be utilized for imputing missing data unless specific rules are mentioned in the analysis sections. If it deems necessary, multiple imputations method may be used in addition to the available data analysis for each variable when the missing at random assumption is considered reasonable.

10 Safety evaluation

Safety will always be evaluated based on the VRS and LRS for run-in phase and the SAF for double-blind phase. Note that, in general, SAF is a subset of LRS and LRS is a subset of VRS.

The safety evaluation will include:

- Specific identified and potential risk and safety factors evaluation
- Adverse events
- Laboratory evaluations
- Vital signs
- Electrocardiogram

The following subgroups will be evaluated for safety, except for the specific identified and potential risk analysis:

- Age group <65, ≥65 years
- Age group <75, ≥75 years
- Gender (male, female)
- Race (Caucasian, Black, Asian, Other)
- Region

10.1 Allocation of AEs

For AE reporting, the allocation of an AE in the different treatment phases is shown in Table 10-1.

If not otherwise specified (see below specific safety evaluation), all AEs with start date before final visit and SAEs within 30 days after final visit will be included in the AE analyses.

Table 10-1 Allocation of AEs

Screening (V1)	V101 or V102 – V103 (val run-in)	V103 – V199 (LCZ run-in)	V201(Rand) – CMP	CMP – CMP + 30d (SAE only)	Phase AE to be reported in
X					Reported by the site from Informed Consent
	X				Report AE in valsartan run-in
		X			Report AE in LCZ696 run-in
			X		Report AE in DB
	X	X2			Report as two separate AEs: One with onset date X for valsartan run-in and one with onset date X2 for LCZ696 run-in
	(X)	X	X2		Report as two separate AEs:

Screening (V1)	V101 or V102 – V103 (val run-in)	V103 – V199 (LCZ run-in)	V201(Rand) – CMP	CMP – CMP + 30d (SAE only)	Phase AE to be reported in
					One with onset date X (X) for LCZ696 (valsartan) run-in and one with onset date X2 for DB
			X, X2		Report as one AE: One with onset date X for DB
	(X)	X	X		Report AE with onset date X (X) in LCZ696 (valsartan) run-in
				X (SAE)	Report SAE in the 30 days follow-up
	(X)	(X)	X	X2 (SAE)	Report as two separate AEs: One AE with onset date X (X) for DB (valsartan or LCZ696 run-in) and one SAE with onset date X2 for the 30 days follow-up

X and (X) stands for onset and possible onset date of AE, respectively.

X2 stands for same AE but with an increased severity.

10.2 Analysis for the identified and potential risks in the risk management plan

The identified and potential risks for LCZ696 are listed in Table 10-2. Besides the specifically specified analyses for each risk, which are listed in the table, the following standard analysis will be applied to all risks.

- Incidence (absolute and relative frequency) rates in terms of patient regardless of study drug relationship by treatment group
- Exposure adjusted incidence rates per 100 patient years regardless of study drug relationship, by treatment
- Listing of subject numbers per risk.

The search paths for the related preferred terms (PTs), high level group term (HLGT), high level term (HLT), in Standard Medical Queries (SMQs), or NMQ in MedDRA for the risks in Table 10-2 are store in the LCZ696 Case Retrieval Strategy (2019).

Table 10-2 Identified and potential risks

Risks	Analysis
Hyperkalemia	1. Frequency of hyperkalemia leading to treatment interruption and discontinuation regardless of study drug relationship, by treatment 2. Categorical analysis with frequency and percentage by treatment for patients with serum potassium ≥5.5 mEq/L, >6.0 mEq/L and >6.5 mEq/L by treatment.
Hypotension	1. Blood pressure summary (mean, standard deviation, median, and range) including measurement at each visit and change from baseline will be provided by treatment 2. Summary with frequency and percentage by treatment for patients with: a. systolic blood pressure (SBP) <100 mmHg; b. 30 mmHg drop in SBP;

Risks	Analysis
	c. simultaneous SBP <100 mmHg and 30 mmHg drop in systolic BP.
Angioedema	Frequency and percentage by treatment (Standard statistical analysis)
Renal impairment	1. Summary for serum creatinine and eGFR by treatment 2. Summary for frequency and percentage by treatment for patients with: a. eGFR decline by >25%, >40%, >50%, and >30 mL/min/1.73 m^2; b. Serum creatinine increase by >50%; c. Serum creatinine level >0.5 mg/dL, >2.0 mg/dL, >2.5 mg/dL, >3.0 mg/dL 3. Subgroup analyses on the relationship between eGFR < 60 and >=60 mL/min/1.73 m^2 at baseline and incidence rate/severity of AEs/laboratory parameters indicating newly occurred/worsening of renal impairment.
Hypersensitivity	Standard analysis
Anaphylaxis/anaphylactoid reactions	Standard analysis
Hepatotoxicity	1. Summary with frequency and percentage by treatment for patients with: a. ALT or AST >3x ULN; b. ALT or AST >5x ULN; c. ALT or AST >8x ULN; d. ALT or AST >10x ULN; e. ALT or AST >20x ULN; f. ALT or AST >3x ULN and TB >1.5x ULN; g. ALT or AST >3x ULN and TB >2x ULN; h. ALT or AST >5x ULN and TB >2x ULN; i. ALT or AST >8x ULN and TB >2x ULN; j. ALT or AST >10x ULN and TB >2x ULN; k. ALT or AST >20x ULN and TB >2x ULN; l. ALP > 2 x ULN m. ALP > 3 x ULN n. ALP > 5 x ULN o. TB > 1.5 x ULN p. TB > 2 x ULN q. TB > 3 x ULN r. ALP > 3x ULN and total bilirubin > 2x ULN s. ALP > 5x ULN and total bilirubin > 2x ULN t. ALT or AST > 3x ULN and total bilirubin > 2x ULN and ALP =< 2x ULN u. (ALT or AST > 3x ULN and total bilirubin > 2x ULN and ALP =< 2x ULN) or reported Hy's Law case v. ALT or AST > 3x ULN and (nausea or vomiting or fatigue or general malaise or abdominal pain or (rash and eosinophilia)) w. TB >3x ULN and AST or ALT are ≤3 x ULN and alkaline phosphatase ≤1.5x ULN; x. AP >3x ULN and AST, ALT, TB are within normal range.

Risks	Analysis
	2. A cross-tabulation of baseline and worst post-baseline values by below, within and above normal range categories will be provided. Shift tables will be provided for the parameters AST, ALT, TB and AP.
Embryo-fetal toxicity/Lethality	Standard analysis
Neonatal/infantile toxicity through exposure from breast milk	Standard analysis
Statin drug-drug interaction	Standard analysis for patients who take statin during randomized treatment epoch (at baseline and post-baseline)
Malignancy	Standard analysis
Cognitive impairment	Standard analysis

10.3 Adverse events (AEs)

Any AE occurred during the study period will be included in AE summary tables as described in Section 10.1, i.e., AEs occurred in run-ins, double-blind period, and SAEs up to one-month after the study completion.

The incidence of treatment-emergent adverse events (new or worsened) will be summarized by primary system organ class, preferred term, severity, and relationship to study drug according to the Medical Dictionary for Regulatory Activities (MedDRA).

The MedDRA version used for reporting the study will be described in a footnote.

Within each reporting phase (run-ins or double-blind, see Table 10-1), the following rules are applicable. If a subject reported more than one adverse event with the same preferred term, the adverse event with the greatest severity will be presented. If a subject reported more than one adverse event within the same primary system organ class, the subject will be counted only once with the greatest severity at the system organ class level, where applicable. Statistical analyses, which will only be performed for the double-blind or later phase, will include all post-baseline AEs up to and including the analysis cut-off (AEs up to the final visit and SAEs up to 30 days after the final visit for each patient) irrespective of how long after the last day of study treatment they occurred.

The number and percentage of subjects reporting any adverse event during each reporting phase will be summarized by primary system organ class, preferred term and treatment. The most common adverse events reported (≥ 1 % in any group for each preferred term in the SOC-PT table) will be presented in descending frequency according to its incidence in the LCZ696 group starting from the most common event.

Separate summaries, for each reporting phase (run-ins and double-blind), will be provided for study medication related adverse events, death, serious adverse event, other significant adverse events leading to study discontinuation and adverse events leading to dose adjustment.

Specific interested AEs will be summarized separately in addition to the above analysis. These specific interested AEs are: angioedema, hyperkalemia, hypotension, renal dysfunction, liver.

Due to the different durations in the two sub-run-in phases, besides providing percentages, the annualized exposure duration adjusted event rates will also be provided.

In addition to the above standard analyses, for the double-blind phase, analysis for time-to-first selected AEs by treatment group will be performed using Kaplan-Meier estimate. The annualized exposure duration adjusted event rates will also be provided.

10.4 Laboratory data

10.4.1 General laboratory data

Laboratory data will be summarized for each reporting phase as described before: Run-ins based on the VRS and LRS and double blind period based on the SAF.

The summary of laboratory evaluations will be presented for three groups of laboratory tests: hematology, serum chemistry and urinalysis, which include hemoglobin, hematocrit, RBC count, WBC count, platelet count, blood urea nitrogen (BUN), creatinine, estimated glomerular filtration rate (eGFR), total bilirubin, AST (SGOT), ALT (SGPT), alkaline phosphatase, sodium, potassium, chloride, calcium, phosphorus, total protein, albumin, and uric acid.

The eGFR is calculated according to the formula:

eGFR (mL/min/1.73 m2) = 175 x (Serum creatinine/88.4)-1.154 x (Age)-0.203

x (0.742 if female) x (1.212 if black),

where serum creatinine is in μmol/L (SI unit) and age is non-rounded at the time of the laboratory sample in years.

Descriptive summary statistics (mean, median, standard deviation, min and max) for the change from baseline to each study visit will be presented. These descriptive summaries will be presented by laboratory test and treatment group. Change from baseline will only be summarized for subjects with both baseline and post baseline values and will be calculated as:

- change from baseline = post baseline value – baseline value

Note that baselines (for run-in and double blind phase) are defined in Section 4.

In addition, shift tables will be provided for all parameters to compare a subject's baseline laboratory evaluation relative to the visit's observed value. For the shift tables, the normal laboratory ranges will be used to evaluate whether a particular laboratory test value was normal, low, or high for each visit value relative to whether or not the baseline value was normal, low, or high. These summaries will be presented by laboratory test and treatment group.

The number and percentage of subjects with clinically notable laboratory results after baseline will be presented. Clinically notable laboratory results, for those parameters where ranges are available, are given in Table 10-3 below. For the calculation the denominator are based on the evaluable post-baseline subjects who did not have the notable abnormality at baseline from the central laboratory.

Table 10-3 Clinical notable criteria for selected laboratory tests

Hematology

RBC count	>50% increase, >20% decrease
Hemoglobin	>50% increase, >20% decrease
Hematocrit	>50% increase, >20% decrease
WBC count	>50% increase, >50% decrease
Platelet count	>75% increase, >50% decrease
Blood Chemistry	
ALT (SGOT)	>150% increase
AST (SGPT)	>150% increase
BUN	>50% increase
	>14.28 mmol/L
Creatinine	>50% increase
	>136.8 µmol/L
Total bilirubin	>100% increase
CPK	>300% increase
Alkaline phosphatase	>100% increase
Sodium	>5% decrease
Potassium	>20% increase, >20% decrease
	≥6.0 mmol/L
	>5.0 mmol/L
	<3.5 mmol/L
Chloride	>10% increase, >10% decrease
Calcium	>10% increase, >10% decrease
Uric acid	>50% increase

In the above table increase and decrease are defined as compared to the baseline value. Here baseline is defined as the run-in baseline for the run-in phases and the double-blind baseline for the double-blind phase.

10.4.2 Specific laboratory data

Liver function test (LFT) data will be analyzed according to what specified in hepatotoxicity in Table 10-2 in Section 10.2.

Laboratory data related to renal impairment (serum creatinine, eGFR) will be analyzed according to what specified in renal impairment in Table 10-3 in Section 10.2.

10.5 Vital signs

Sitting systolic blood pressure, sitting diastolic blood pressure, sitting pulse rate and body weight will be summarized by treatment group and visit with standard summary statistics (mean, median, standard deviation, min, max), including changes from baseline. Baselines (run-in and double blind period) are defined in Section 4. Graphical mean plots with 95% CIs for these vital signs will also be provided.

The descriptive summaries will be presented by vital sign and treatment group and visit. Change from baseline will only be summarized for subjects with both baseline and post-baseline values and will be calculated as:

- change from baseline = post-baseline value – baseline value

The number and percentage of subjects with clinically notable vital signs changes from baseline will be presented. Clinically notable vital sign results are provided in Table 10-4 below.

Table 10-4 Clinically notable changes in vital signs

Vital Sign (unit)	Clinically notable criteria
Weight (kg)	decrease > 7% from Baseline increase > 7% from Baseline
Sitting systolic blood pressure (mmHg)	<90 and decrease from baseline of >20 >180 and increase from baseline of >20
Sitting diastolic blood pressure (mmHg)	<50 and decrease from baseline of >15 >105 and increase from baseline of >15
Pulse (bpm)	<50 and decrease from baseline of > 15 >120 and increase from baseline of >15

10.6 Electrocardiogram (ECG)

The following quantitative variables will be summarized: heart rate,QRS duration, and QTc. Summary statistics (n, mean, SD, min .Q1, median, Q3, max) for the change from baseline (Visit 201) in ECG intervals by visit and treatment will be provided. QTc categorical summary will also be provided (for both Fridericia and Bazett corrctions) (for example, >500 ms, > 500 ms (with RBBB), > 500 ms (without RBBB)) and categorical summary for maximum change from baseline for QTc (> 60 ms, > 60 ms (with RBBB), > 60 ms (without RBBB)).

In addition, shift tables comparing baseline ECG results (normal, abnormal, not available, total) with the maximum on study result (normal, clinical significant, not available, total) will be provided for each variable. Clinically significant is defined if at least one abnormality: “atrial fibrillation, atrial flutter, LBB block, RBB block, pathological Q-waves, left ventricular hypertrophy, paced rhythm or other” is answered “yes”.

11 Resource utilization

Data relating to resource utilization will be used to describe medical resources used by the study participants. Mainly descriptive statistics of resources utilization data will be provided by treatment group.

12 Interim analyses

One interim analysis for efficacy is planned. The cut-off time for this interim analysis will be when about two-thirds of the target number of primary events have been reported and adjudication-confirmed, i.e., when approximately 1231 of composite of adjudication-confirmed hospitalizations for HFs and CV deaths are available. In the interim analysis, the analysis dataset will comprise all patients who were randomized before the cutoff date. Bonferroni multiplicity adjustment will be adopted for the statistical comparisons between treatments. An alpha of 0.001 (one-sided alpha) will be spent for the comparison of primary endpoint at the interim analysis and the rest of alpha (one-sided 0.024 for the current specified boundary) will be

utilized at the final analysis. In the interim analysis, the study may be stopped for superior efficacy only when the primary endpoint and CV death both are significant at level of 0.001 (one-sided).

If the study is stopped early for claiming superior efficacy at the interim analysis, the secondary endpoints will be tested using the same hierarchical testing procedure as described in Section 9.3.2.1 before with an overall alpha level used for the primary endpoint at the interim analysis (one-sided alpha of 0.001 for the current specified boundary).

Interim safety assessments are planned to be performed every six months. No further alpha adjustment will be made for these interim safety assessments.

Interim analyses will be performed by an independent statistician who will not be involved in the trial conduct. The results will be reviewed by the independent DMC.

Investigators, Novartis employees, and others who are involved in the conduct of the trial will remain blinded to the treatment codes and interim analysis results until all monitoring decisions have been made and the database has been locked for the final analysis.

13 Power and sample size considerations

13.1 Power for the primary endpoint

Sample size is calculated through simulations for the proportional rates model. The control group rate of total hospitalizations for HF and rate of mortality are estimated in two steps in order to get a reasonable estimate for the selected patient population.

First we estimate the rates from the candesartan group of the CHARM-Preserved study (Yusuf et al 2003) for patients with EF >45%, using a joint model that joins the Poisson regression model and the exponential regression model together through a shared gamma frailty (Cowling et al 2006). An independent uniform censoring on the events are assumed during the follow-up. With the given frailty between HF hospitalization and CV death, the estimated baseline intensity in the Poisson model= 0.00032 HF hospitalizations/day per patient and the estimated hazard rate in the exponential model = 0.000136 CV deaths/day per patient. The estimated gamma shape parameter (the frailty parameter that is also assumed to be 1/scale) is 0.193. These specified parameter values produce approximately an annualized rate of 0.083 for time to first primary event and an annualized rate of 0.036 for CV mortality.

Second, we adjust these estimated rates up by 8% (and then increasing the corresponding Poisson intensity rate and hazard rate) based on the results from the recent completed HF studies and publications, which are believed to reflect a higher risk patient population than that in the CHARM-Preserved study. This leads an annualized rate of 0.09 for time to first primary event and an annualized rate of 0.04 for CV mortality for a high risk patient population.

The target reduction in RR for the primary endpoint is chosen to be about 22%, which approximately corresponds to a reduction of 30% for HF hospitalization and a reduction of 10% for CV death, given the gamma frailty defined above between these two components. With the CHARM-Preserved data structure for EF $\geq$45%, the specified rates will produce an approximately 15% reduction for time-to-first event analysis, see Table 13-1.

The patient enrollment is assumed to be uniform lasting 2 years and 5 months and the minimum follow-up is to be 2 years and 2 months.

In Table 9-1, sample sizes and powers are estimated through simulations for the framework defined above. Three thousand trial replicates are generated for each scenario with different sample sizes in the simulations. Both the power and type I error rate are estimated for the selected proportional rates model (LWYY). For our study, with a one-sided alpha level of 0.025, a total of 4600 patients will provide more than 90% of power for the LWYY method. This will require approximately 1847 primary events (see the row in table below with sample size = 4600 and HR = 0.9 and RR = 0.7). The type I error rates were preserved well in the cases we examined.

Table 13-1 Sample size and power estimations through simulations

Sample size	Simulation specs		Expected no. of composite events	Estimated HR or RR		Power
	HR for death	RR for total hosp for HF		Time to 1st composite event	LWYY	LWYY
4600	0.9	0.7	1847	0.859	0.778	0.95
		0.75	1882	0.881	0.814	0.85
	1.0	0.7	1846	0.876	0.784	0.94
		0.75	1885	0.900	0.822	0.82

3000 simulation runs were performed for each scenario.

13.2 Sample size re-estimation

A blinded sample size re-estimation will be considered around the time of the efficacy interim analysis. The pooled intensity/hazard rates for hospitalization for HF and CV death will be estimated using the same joint shared-gamma-frailty model (with one treatment group) used for the simulations for the sample size calculations in the above sections. The differences between the model parameter estimates used for the sample size calculations and the new ones obtained in the interim analysis will then be evaluated. Sample size simulations based on the new estimates will be performed using the same assumptions used in the initial sample size calculations. The new sample size and/or the duration of follow up will be determined based on the new simulations in order to preserve the target power and to achieve the required number of events in an acceptable timeframe.

13.3 Power for secondary endpoints

13.3.1 Power for the change in KCCQ clinical summary score at 8 months

For the change from randomized treatment epoch baseline to Month 8 in clinical summary score assessed by KCCQ, the planned sample size of 4600 will provide 83% power to detect a treatment difference of 2 points at the one-sided significance level of 0.025, assuming a standard deviation of 22 (observed in PARADIGM study). This estimation has taken into consideration that approximately 10% of patients may be excluded from analysis due to no KCCQ assessment done, based on PARADIGM experience.

13.3.2 Power for NYHA class

From the PARAMOUNT (CLCZ696B2214) study, which is the phase II pilot study with a similar HFpEF population, after 9 months of treatment with valsartan, the incidence rates in each category distribute as follows in Table 13-2.

Table 13-2 PARAMOUNT NYHA incidence rates

	LCZ696				Valsartan			
	Baseline	Class I	Class II	Class III/IV	Baseline	Class I	Class II	Class III/IV
Class I	1 (0.8)	0 (0.0)	1 (100.0)	0 (0.0)	1 (0.8)	0 (0.0)	1 (100)	0 (0.0)
Class II	100 (78.7)	**16 (16.0)**	80 (80.0)	4 (4.0)	102 (81.6)	**9 (8.8)**	89 (87.3)	4 (3.9)
Class III	26 (20.5)	0 (0.0)	**13 (50.0)**	13 (50.0)	22 (17.6)	0 (0.0)	**8 (36.4)**	14 (63.6)

From this table, there are approximately 7% of patients (16/100=16% vs. 9/102=8.8%) within class II and 14% of patients (13/26=50% vs. 8/22=36.4%) in class III improved from baseline to 9 months endpoint in LCZ696 compared with valsartan.

Our power evaluation is based on this data and takes the incidence rates in valsartan group as the rates in the comparator group. A one-sided alpha of 0.025 is used to control the type I error. Table 13-3 shows powers for a sample size of 4600 patients.

Table 13-3 Powers for NYHA class comparisons

	Improved from Baseline to endpoint in Class III			
Improved from baseline to endpoint in Class II	10%	12%	14%	16%
5%	98%	99%	99%	>99%
7%	>99%	>99%	>99%	>99%
9%	>99%	>99%	>99%	>99%

The results in this table are based on simulations with 100000 replicates.

From this table, there are approximately 7% of patients (16/100=16% vs. 9/102=8.8%) within class II and 14% of patients (13/26=50% vs. 8/22=36.4%) in class III improved from baseline to 9 months endpoint in LCZ696 compared with valsartan.

13.3.3 Power for composite renal endpoint

With an expected hazard reduction of 30% (LCZ696 over valsartan) in composite renal endpoint, an annual first occurrence of renal events rate of 0.63% (or 1%), an average follow-up of 3.375 years, and a one-sided alpha of 0.025, forty-six hundred patients will provide a power of 36% (or 52%). When the hazard reduction is about 35%, forty-six hundred patients will provide a power of 47% (or 66%) when the annual event rate is 0.63% (or 1%). Note that the power provided here is conditional power, when the null hypotheses for the primary endpoint, the secondary endpoints of KCCQ change and NYHA class change at month 8 are all rejected.

The power estimations are mainly based on data observed from recently completed PARADIGM-HF trial. In that trial, the observed annual event rate for composite renal endpoint is 0.63% and the hazard reduction of LCZ696 over enalapril is 37%.

13.3.4 Power for all-cause mortality

With a reduction of 20% in mortality, an annual all-cause mortality rate of 5% (25% increase from CV mortality), an average follow-up of 3.375 years, and a one-sided alpha of 0.025, forty-six hundred patients will provide a power of 81%. When the reduction is about 15%, forty-six hundred patients will provide a power of 56%.

14 References

1. Andersen PK and Gill RD (1982) Cox's regression model for counting processes: a large sample study. Ann. Statist.; 10 (4): 1100-1120.
2. Bretz F, Maurer W, Brannath W, Posch M (2009) A graphical approach to sequentially rejective multiple test procedures. Statist Med; 28:586-604.
3. Castaneda J, Gerritse B (2010) Appraisal of several methods to model time to multiple events per subject: modelling time to hospitalizations and death. Revista Colombiana de Estadistica; 33:34-61.
4. Cook RJ, Lawless JF (1997) Marginal analysis of recurrent events and a terminating event. Stat in Medicine; 16:911-924.
5. Cowling BJ, Hutton JL, Shaw JEH (2006) Joint modeling of event counts and survival times. Appl. Statist; 55(Part 1):31-39.
6. Fleiss J, Statistical methods for rates and proportions. 2nd ed. 1981: John Wiley and Sons, Inc, New York.
7. Ghosh D, Lin DY (2000) Nonparametric analysis of recurrent events and death. Biometrics; 56:554–562.
8. Li QH, Lagakos SW (1997) Use of Wei-Lin-Weissfeld Method for the Analysis of a Recurring and Terminating Event. Statistics in Medicine 16:925-940.
9. Liu L, Huang X (2008) The use of Gaussian quadrature for estimation in frailty proportional hazards models. Statistics in Medicine; **27**:2665–2683.
10. Lin DY, Wei LJ, Ying Z, Yang I (2000) Semiparametric regression for the mean and rate functions of recurrent events. J R Stat Soc; B, 62:711–730.
11. Lu L, Liu X (2008) Analysis of Correlated Recurrent and Terminal Events Data in SAS ®. NESUG.
12. McCullagh P, Nelder JA (1989) Generalized Linear Models, 2nd ed. London: Chapman and Hall.

13. Wei LJ, Lin DY, Weissfeld L (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distribution. J American Statistical Association; 84:1065–1073.

14. Yusuf S, Pfeffer MA, Swedberg K, et al for the CHARM Investigators and Committees (2003) Effects of candesartan in patients with chronic heart failure and preserved left-ventricular ejection fraction: the CHARM-Preserved trial. Lancet; 362:777-781.

15. Zeng D, and Lin DY (2009). Semiparametric Transformation Models with Random Effects for Joint Analysis of Recurrent and Terminal Events. Biometrics; 65:746-752.