TRIAL STATISTICAL ANALYSIS PLAN

c22042336-05

BI Trial No.:	1245-0168
Title:	A phase III randomised, double-blind trial to evaluate the effect of 12 weeks treatment of once daily EMPagliflozin 10 mg compared with placebo on ExeRcise ability and heart failure symptoms, In patients with chronic **H**eArt **F**aiLure with **r**educed **E**jection **F**raction (**HFrEF**) (EMPERIAL – reduced) Revised protocol version 3.0 including Protocol Amendment 2 [c17550230-03]
Investigational Product(s):	Empagliflozin
Responsible trial statistician(s):	 Phone: Fax:
Date of statistical analysis plan:	21 OCT 2019 REVISED
Version:	REVISED

1. TABLE OF CONTENTS

LIST OF TABLES

2. LIST OF ABBREVIATIONS

Term	Definition / description
6MWT	Six-minute walking test
6MWTD	Six-minute walking test distance
ACE	Angiotensin converting enzyme
AE	Adverse Event
AESI	Adverse Event of Special Interest
ALT	Alanine transaminase
ANCOVA	Analysis of covariance
ARB	Angiotensin Receptor Blocker
ARNI	Angiotensin Receptor-Neprilysin Inhibitor
ASA	acetylsalicylic acid
AST	Aspartate transaminase
ATC	Anatomical-Therapeutic-Chemical classification
AUC	Area Under the receiver operating characteristic Curve
BI	Boehringer Ingelheim
BIcMQ	Boehringer Ingelheim customised MedDRA Query
BMI	Body Mass Index
BP	Blood Pressure
bpm	beats per minute
BRPM	Blinded Report Planning Meeting
CEC	Clinical event committee
CHF	Chronic Heart Failure
CHQ	Chronic Heart Failure Questionnaire
CHQ-SAS	Chronic Heart Failure Questionnaire - Self-Administered Standardised format
CKD	Chronic Kidney Disease
CKD-EPI_{cr}	Chronic Kidney Disease Epidemiology Collaboration Equation
CML	Local Clinical Monitor
CTP	Clinical Trial Protocol
CTR	Clinical Trial Report

Term	Definition / description
CRA	Clinical Research Associate
CV	Cardiovascular
DBL	Database lock
DBP	Diastolic Blood Pressure
DKA	Diabetic ketoacidosis
DM	Diabetes Mellitus
DM&SM	Boehringer Ingelheim Data Management And Statistics Manual
ECDF	Empirical cumulative distribution functions
ECG	Electrocardiogramm
eCRF	Electronic Case Report Form
EF	Ejection Fraction
eGFR	Estimated Glomerular Filtration Rate
EMA	European Medicines Agency
EoT	End-of-Text
gCV	Geometric coefficient of variation
HF	Heart Failure
HFrEF	Heart Failure with reduced Ejection Fraction
HHF	Hospitalisation for Heart Failure
HLT	High-Level Term
IC	Informed Consent
ICH	International Conference On Harmonisation
iPD	Important Protocol Deviation
JVD	Jugular Venous Distension
KCCQ	Kansas City Cardiomyopathy Questionnaire
KCCQ-CSS	Kansas City Cardiomyopathy Questionnaire – Clinical Summary Score
KCCQ-TSS	Kansas City Cardiomyopathy Questionnaire - Total Symptom Score
LLT	Lowest Level Term
LVEF	Left Ventricular Ejection Fraction
MedDRA	Medical Dictionary For Regulatory Activities
MMRM	mixed model repeated measure analysis
MRA	Mineralocorticoid receptor antagonist

Term	Definition / description
NT-proBNP	N-Terminal Pro-Brain Natriuretic Peptide
NYHA	New York Heart Association
OC-AD	Observed Case including data After Discontinuation
OC-OT	Observed Case On Treatment
PD	Protocol Deviation
PDMAP	Project Data Management and Analysis Plan
PGI-C	Patient Global Impression of Change
PGI-S	Patient Global Impression of Severity
PRO	Patient Reported Outcome
PT	Preferred Term
pt-yrs	patient years
Q1	Lower Quartile
Q3	Upper Quartile
ROC	Receiver Operating Characteristic
RS	Randomised Set
SAE	Serious Adverse Events
SBP	Systolic Blood Pressure
SCR	Screened Set
SD	Standard Deviation
SE	Standard Error
SEM	Standard Error of Measurement
SensRS	Sensitivity analysis set
SMQ	Standardised MedDRA Query
SOC	System Organ Class
T1DM	Type 1 Diabetes Mellitus
TBILI	Total Bilirubin
TS	Treated Set
TSAP	Trial Statistical Analysis Plan
ULN	Upper Limit of Normal
UTI	Urinary Tract Infections
WCI	Worst Case Imputation

Term	Definition / description
WHO DD	World Health Organization Drug Dictionary
Y/N	Yes/No

3. INTRODUCTION

As per ICH E9 (1), the purpose of this document is to provide a more technical and detailed elaboration of the principal features of the analysis described in the protocol, and to include detailed procedures for executing the statistical analysis of the primary and secondary variables and other data.

This TSAP assumes familiarity with the Clinical Trial Protocol (CTP), including Protocol Amendments. In particular, the TSAP is based on the planned analysis specification as written in CTP Section 7 "Statistical Methods and Determination of Sample Size". Therefore, TSAP readers may consult the CTP for more background information on the study, e.g., on study objectives, study design and population, treatments, definition of measurements and variables, planning of sample size, randomisation."

SAS® Version 9.4 will be used for all analyses.

4. CHANGES IN THE PLANNED ANALYSIS OF THE STUDY

4.1 CHANGES NEWLY INTRODUCED FROM THE CLINICAL TRIAL PROTOCOL

Analyses of the following endpoints have been added:

- Systolic blood pressure
- Diastolic blood pressure
- Body weight

5. ENDPOINTS

5.1 PRIMARY ENDPOINT

The primary endpoint is defined in CTP Section 2.1.2.

5.2 SECONDARY ENDPOINTS

5.2.1 Key secondary endpoints

The key secondary endpoint is defined in CTP Section 2.1.3. The scoring algorithm for KCCQ-TSS and CHQ-SAS dyspnea can be found in Appendix 9.4.

5.2.2 Secondary endpoints

Secondary endpoints are listed in CTP Section 2.1.3. The scoring algorithm for the clinical congestion score can be found in Appendix 9.5.

5.3 FURTHER ENDPOINTS

Further endpoints are listed in CTP Section 2.2.2.

Response in KCCQ domains at week 12: Response will be defined as an improvement of at least 5 points.

Response in CHQ domains at week 12: Response will be defined as an improvement of at least 0.5 points. Scoring algorithm for all KCCQ domains and CHQ-SAS domains can be found in Appendix 9.4.

5.4 OTHER VARIABLES

5.4.1 Demographic and other baseline characteristics

Blood pressure: Baseline SBP and DBP are taken from the vital signs assessment. If no vital signs assessment is available, the value before the start of the 6MWT will be used.

Region will be defined according to Table 9.1:1

Heart rate will display values entered in the electronic case report form (eCRF) as ‘pulse’.

5.4.2 Treatment compliance

Overall compliance (%) to study medication will be calculated as a weighted average of reported compliance. Weights will be determined based on actual visit dates. Compliance will be displayed for the treated set.

5.4.3 Treatment exposure

Duration of exposure will be calculated as day of last intake of study drug – day of first intake of study drug + 1 day.

6. GENERAL ANALYSIS DEFINITIONS

6.1 TREATMENTS

There will be 3 study intervals in this trial: screening, study treatment phase (with either empagliflozin 10mg or matching placebo) and follow-up. The purpose of the definitions below is to describe the different study intervals.

Table 6.1: 1 Treatment regimens / study intervals

Label in database	**Label in tables and listings**	**Interval**	**Start date**
SCREENING	Screening	Screening	Date of informed consent
PLACEBO/ BI 10773 10 MG	Placebo/Empa 10 mg	Treatment	Date of first administration of double-blind study treatment
FOLLOW-UP	Post Pbo/Post Empa	Follow up	Date of last administration of study drug + 1 day

Details on the definition of the treatment period for different endpoints are listed in Table 6.1:2. The efficacy analyses will follow the intention-to-treat (ITT) principle. For the assignment of patients to treatment groups this means that patients will be analysed as randomised.

For safety analyses, patients will be also assigned to the treatment groups as randomised.

If a patient erroneously receives the wrong trial drug, all future medication packs will be dispensed as scheduled for the treatment group to which the patient was randomised. Therefore, the adverse events will be analysed as per randomised treatment, which is expected to reflect the prevailing treatment. In exceptional cases that a patient took the wrong treatment, all adverse events that occur while being on the wrong treatment will be listed and analyses of this data are described in section 7.8.1.6.

Table 6.1: 2 Endpoint-specific assignment to the on-treatment phase

Endpoint	Last day of assignment to on-treatment phase (days after last intake of study medication)
6MWT distance	1
PRO endpoints (key secondary endpoints)	7
Adverse events	7
Safety laboratory measurements	3
Heart rate	1
Body weight	1
Creatinine and estimated glomerular filtration rate (eGFR)	1
Blood pressure	1

6.2 IMPORTANT PROTOCOL DEVIATIONS

A protocol deviation (PD) is important, if it affects the rights or safety of the study patients or if it can potentially influence the primary outcome measures for the respective patients in a way that is neither negligible nor in accordance with the study objectives.

Important protocol deviations (iPDs) will be described in the clinical trial report (CTR). A listing of patients with medication code broken will be provided.

Table 6.2: 1 Important protocol deviations

Category / Code		Description	Requirements	Excluded from
A		**Entrance criteria not met**		
A1		**Target indication not met**		
	A1.06	No chronic HF or no NYHA class II-IV	Inclusion criterion #5 not met	None
	A1.07	Conditions on ejection fraction (EF) not met	Inclusion criterion #6 not met	None
	A1.08	Conditions on NT-proBNP not met	Inclusion criterion #7 not met	None

Table 6.2: 1 Important protocol deviations (cont.)

Category / Code		Description	Requirements	Excluded from
A2/3		**Other entrance criteria not met**		
	A2.02	Age out of range	Inclusion criterion #1 not met	None
	A2.08	Specific inclusion criterion for women of child-bearing potential not met	Inclusion criterion #2 not met	None
	A2.14	6MWT distance above 350m at screening or baseline or below 100m at baseline	Inclusion criterion #4 or exclusion criterion #10 not met	SensRS1
	A2.12	Inclusion criterion on stable concomitant medication not met	Inclusion criterion #8 not met	None
	A2.13	Use of medical devices not appropriate or device implanted recently or intent to implant	Inclusion criterion #10 or exclusion criterion #18 not met	None
	A3.44	Patient with unstable conditions	Exclusion criteria #1, #2, #6. #9, #13, #17, #20, #21 or #25 not met, inclusion criterion #9 not met	None
	A3.49	Participation in another Empagliflozin HF trial	Exclusion criterion #3 not met	None
	A3.51	Type 1 diabetes	Exclusion criterion #4 not met	None
	A3.09	Renal insufficiency or renal impairment (assessed by eGFR)	Exclusion criterion #5 not met	None
	A3.33	Systolic blood pressure at visit 1 or 2 out of range	Exclusion criterion #7 not met	None
	A3.31	Atrial fibrillation or atrial flutter with a resting heart rate >110bpm	Exclusion criterion #8 not met	None
	A3.48	Presence of condition that precludes exercise testing	Exclusion criterion #11 not met	None
	A3.47	Participation in structured Exercise training	Exclusion criterion #12 not met	None
	A3.28	Currently implanted left ventricular assist device (LVAD)	Exclusion criterion #14 not met	None
	A3.29	Patients with cardiomyopathies excluded by the protocol	Exclusion criterion #15 or #19 not met	None
	A3.30	Any severe (obstructive or regurgitant) valvular heart disease excluded by the protocol	Exclusion criterion #16 not met	None
	A3.06	Indication of liver disease	Exclusion criterion #22 not met	None
	A3.34	Haemoglobin at visit 1 below cut-off	Exclusion criterion #23 not met	None
	A3.35	History of Ketoacidosis	Exclusion criterion #24 not met	None

Table 6.2: 1 Important protocol deviations (cont.)

Category / Code		Description	Requirements	Excluded from
	A3.36	Gastrointestinal (GI) surgery or GI disorder that could interfere with study medication absorption in the investigator's opinion	Exclusion criterion #26 not met	None
	A3.37	Documented or active malignancy	Exclusion criterion #27 not met	None
	A3.39	Intake of restricted medications or any drug considered likely to interfere with the safe conduct of the trial	Exclusion criterion #28 not met	None
	A3.40	Treatment with any SGLT-2 inhibitor or SGLT-1 and 2 inhibitor	Exclusion criterion #29 not met	None
	A3.50	Treatment with i.v. iron therapy or erythropoietin (EPO) within 3 months prior to screening	Exclusion criterion #30 not met	None
	A3.11	Intake of an investigational drug in another trial within 30 days prior to intake of study medication in this trial	Exclusion criterion #31 not met	None
	A3.41	Known allergy or hypersensitivity to empagliflozin	Exclusion criterion #32 not met	None
	A3.13	Relevant alcohol or drug abuse and other conditions affecting study compliance	Exclusion criterion #33 not met	None
	A3.12	Specific exclusion criterion for pre-menopausal women not met	Exclusion criterion #34 not met	None
	A3.14	Any other clinical condition unsafe for participation that would jeopardise patient safety while participating in this clinical trial	Exclusion criterion #35 not met	None
B		**Informed consent**		
	B1	Informed consent not available/not done	Informed consent date missing	All
	B2	Informed consent too late	Informed consent date <*actual consent date*> was after Visit 1 date <*Visit 1 date*>	None

Table 6.2: 1 Important protocol deviations (cont.)

Category / Code		Description	Requirements	Excluded from
C		**Trial medication and randomisation**		
	C1.02	Incorrect trial medication taken	Wrong medication taken for more than 20% of the overall treatment duration. Can only be finally judged after data base lock (DBL) since unblinding information is required.	None
	C3.01	Non-compliance with study drug intake	Compliance <80% or >120%	None
	C4.01[1]	Medication code broken inappropriately	Medication code was broken for no valid reason. Final decision at the DBL meeting based on medical judgement.	None
D		**Concomitant medication**		
	D2.01	Prohibited medication use	Review of eCRF for prohibited medication. Final decision at the DBL meeting based on medical judgement.	None
G		**Study-specific analysis**		
	G3.48[1]	Unsuitable set up or conduct for 6MWT at baseline or week 12	Important deviations from the set-up or conduct as defined in the protocol	SensRS1
	G3.49[1]	Not compliant with contraindications of the 6MWT	6MWT not stopped when it should have been based on stopping criteria, 6MWT not rescheduled appropriately when patient presents with relative contraindications	SensRS1
	G3.50[1]	Assessment order not followed that could influence PRO at baseline or week 12	PRO measurements not before any other assessment at baseline or week 12	None
I		**Other safety-related deviations**		
	I2.01	Pregnancy		None

KEY: 6MWT= six minute walk test, HF=Heart Failure, NT-proBNP=N-Terminal Pro-Brain Natriuretic Peptide, NYHA= New York heart association, EF=ejection fraction, T1DM=Type 1 diabetes mellitus, bpm=beats per minute.

1 needs to be checked at site by Local Clinical Monitors (CMLs) / Clinical Research Associates (CRAs) as these cannot be checked programmatically

6.3 SUBJECT SETS ANALYSED

The following patient sets are defined in section 7.3 of the protocol:

- Randomised set (RS)
  This patient set includes all randomised patients, whether treated or not. Treatment will be evaluated as randomised.

- Treated set (TS)
  This patient set includes all patients who were dispensed study medication and were documented to have taken at least one dose of investigational treatment. Treatment will be evaluated as randomised.

In addition the following patient set is defined:

- Screened Set (SCR)
  Consists of all patients screened for the trial, with informed consent given and who completed at least one screening procedure at Visit 1.

For sensitivity analyses the following patients set is defined:

- Sensitivity analysis set 1 (SensRS1)
  Randomised patients excluding patients with an iPD related to 6MWT, i.e. iPD categories A2.14, G3.48 and G3.49 of Table 6.2: 1
- Sensitivity analysis set 2 (SensRS2)
  Randomised patients excluding patients for which at least one on-treatment value of urine glucose test result was included in the safety laboratory reports posted to a portal by the central laboratory, and the reports were accessed before database lock.
- Sensitivity analysis set 3 (SensRS3)
  Randomised patients excluding non-diabetic patients for which at least one on-treatment value of urine glucose test result was included in the reports posted and accessed before database lock.

Demographics and baseline characteristics as well as the efficacy analysis will be based on the RS. Demographics will be repeated on the TS if the number of patients in the analysis sets differ by more than 5%. The safety analysis will be based on the TS.

Overall number of patients participating in the study (screened, randomised, screened but not randomised, etc.) by region and country will be based on the SCR.

Table 6.3: 1 Subject sets analysed

	Subject set		
Class of endpoint	Screened Set	Randomised set	Treated set
Primary and key secondary endpoints		X	X
(other) Secondary and further endpoints		X	
Safety endpoints			X
Demographic/baseline endpoints		X	X[1]
Extent of exposure			X
Disposition	X		

1 Demographics will be repeated on the TS if the number of patients in the analysis sets differ by more than 5%.

6.4 SUBGROUPS

Subgroups are provided in Table 6.4: 1. Missing categories for subgroup variables will not be considered in the respective analysis.

Subgroups with fewer than 14 patients in total (yielding approximately 7 patients per treatment group) will generally not be included in the analyses of subgroups. Small subgroups can be combined into other subgroups. Final decision will be made at final Blinded Report Planning Meeting (BRPM).

Although several subgroup analyses are prespecified in Table 6.4: 1, the subgroup analysis by Diabetes at baseline (diabetic, non-diabetic) is the medically and academically most important subgroup analysis.

Table 6.4: 1 Categories of covariates for displays of baseline characteristics and subgroup analyses

Variable	Categorisation	Demo-graphics[1]	Subgroups for Efficacy endpoints[2]	Safety
Age (years)	<50 50 to <65 65 to <75 75 to <85 ≥ 85	X		
	< 65 ≥65	X	X	
Sex	Male Female	X	X	
Region[3]	Europe, Australia, North America	X		
	Europe or Australia North America		X	
Ethnicity	Hispanic/ Latino Not Hispanic/ Latino	X	X	
Race	American Indian or Alaska Native Asian Black or African-American Native Hawaiian or other Pacific Islander White	X		
	White Black/ African-American Asian Other including mixed race		X	
BMI (kg/m^2)	<30 ≥30	X	X	
eGFR at baseline ($mL/min/1.73m^2$)	≥90 60 to <90 45 to <60 30 to <45 20 to <30 <20	X		
	≥60 45 to <60 30 to <45 <30		X	

Table 6.4: 1 Categories of covariates for displays of baseline characteristics and subgroup analyses (cont.)

Variable	Categorisation	Demo-graphics[1]	Subgroups for Efficacy endpoints[2]	Safety
Diabetes at baseline[4]	DM, pre-DM, non-DM (excl. pre-DM)	X		
	DM, non-DM		X	X (only hypoglycemia)
History of hypertension	Yes / No	X		
Baseline BP	SBP<140 and DBP<90 vs. SBP≥ 140 or DBP≥90	X	X	
Atrial Fibrillation at baseline[5]	Yes/ No	X	X	
Baseline LVEF	<20, >=20 to <=30, >30 to <=35, >35	X	X	
History of valvular disease	Yes / No	X		
Cause of HF	Ischemic Hypertensive Valvular heart disease Diabetic Alcoholism Idiopathic Cardiomyopathy Other	X		
Time since diagnosis of HF	≤ 1 year, >1 to ≤ 5 years, > 5 to ≤ 10years, > 10 years	X		
NYHA at baseline	I/II/III/IV	X		
	II[6] vs. III/IV		X	
NT-proBNP (separated by atrial fibrillation status) at baseline[7]	atrial fibrillation at baseline and < / ≥ median; no atrial fibrillation at baseline and < / ≥ median		X	

Table 6.4: 1 Categories of covariates for displays of baseline characteristics and subgroup analyses (cont.)

Variable	Categorisation	Demo-graphics[1]	Subgroups for Efficacy endpoints[2]	Safety
Baseline KCCQ-TSS	By quartiles		X (only KCCQ-TSS and 6MWT)	
Baseline CHQ-SAS Dyspnea domain	By quartiles		X (only CHQ-SAS dyspnea and 6MWT)	
Baseline PGI-S Dyspnea	5 categories	X	X (only CHQ-SAS dyspnea)	
Baseline PGI-S HF symptoms	5 categories	X	X (only KCCQ-TSS)	
Baseline distance walked in 6 minutes	< / >= median		X (only 6MWT)	
Baseline use of ACE-inhibitor, ARB or ARNi	Yes / No	X[8]	X	
Baseline use of diuretics other than MRA	Yes / No	X[8]	X	
Baseline use of MRA	Yes / No	X[8]	X	
Baseline use of loop or high-ceiling diuretics	Yes / No	X[8]	X	
baseline haemoglobin	< / >= median		X	

1 The column demographics shows categories displayed for overall demographics. Demographics are not planned by subgroup.
2 Subgroups planned for the following endpoints: primary and key secondary endpoints.
3 Region categorisation: see Table 9.1: 1.
4 Diabetes at baseline categorisation: see Table 9.2: 1; for subgroup analysis ‘pre-DM’ and ‘non-DM (excluding pre-DM)’ will be combined into the category ‘non-DM’
5 atrial fibrillation at baseline means atrial fibrillation in last ECG before treatment intake
6 patients with NYHA class I are counted in subgroup NYHA class II
7 Median NT-proBNP will be calculated for randomised patients with/without atrial fibrillation at baseline separately and subgroups will be defined based on the respective median for patients with/without atrial fibrillation at baseline accordingly
8 part of the presentation of baseline concomitant therapy as outlined in Section 7.2

6.5 POOLING OF CENTRES

Not applicable because centre/country is not included in the statistical model.

6.6 HANDLING OF MISSING DATA AND OUTLIERS

The imputation strategy for all confirmatory analyses, i.e. primary analysis of primary and key secondary endpoints, will be based on the following principle:

Patients with an available week 12 measurement of the endpoint, regardless of whether on or off-treatment, will be ranked based on change from baseline at week 12. If a patient performs the visit but the 6MWT is not done this will not be regarded as missing data. This case will be evaluated as a measured distance of 0m.

All imputed data due to missing data will be lower than the measured values and will be ranked in the following orders: Patients who do not have a clinical event (death, serious adverse event or any AE leading to discontinuation) and have a value at week 6 are ranked by their values. Patients who do not have a clinical event and have no valid post-baseline value are ranked below. Patients who have a clinical event are given a lower rank and death is ranked below any other clinical event by time to death from randomisation.

A treatment-emergent SAE with outcome “recovered with sequelae” leading to an amputation is considered to be an ongoing clinical event.

Imputation will not be applied for any parametric analysis, descriptive statistics and derivation of responder thresholds for the key secondary endpoints.

6.6.1 Primary endpoint

Handling of missing data is described in section 7.5 of the clinical trial protocol for the primary analysis and subgroup analysis of the primary endpoint change from baseline in 6MWTD.

6.6.2 Key Secondary endpoint

Missing individual items within a questionnaire will be handled according to the rules in Appendix 9.4.

Patients with an available week 12 measurement, regardless of whether on or off treatment, will be ranked based on change from baseline at week 12 regardless of clinical events.

For any missing week 12 assessment of the KCCQ-TSS, the below imputation will be applied for the primary analysis and subgroup analysis in accordance to the strategy for the primary endpoint:

Since the lowest possible value for change from baseline in KCCQ-TSS is 0 minus the highest possible baseline value (100), all imputed values for change from baseline are below -100 as described in Table 6.6.2: 1 below.

Table 6.6.2: 1 Imputation Rules for missing KCCQ-TSS Assessments at Week 12

	Category	Ranking	Case description	(Imputed) Change from baseline in KCCQ-TSS for described case
1	Missing data on KCCQ-TSS at week 12 and no clinical event: Patients with missing KCCQ-TSS at week 12 and no event as in category 2 or 3	If post-baseline value is available, rank by last available value.	Patient has a change from baseline to last available value of x points, no clinical event	-102 +x/1000
		Rank below if no post-baseline score is available	No information of post-baseline KCCQ-TSS and no clinical event	-104
2	Missing data and clinical event (except death). A clinical event is defined as ongoing SAE at week 12 or AE leading to treatment discontinuation and patient is alive at week 12	Ranked above patients who died, but below patients with missing data and no clinical event.	Patient has ongoing SAE or AE leading to treatment discontinuation, no week 12 KCCQ-TSS assessment and patient alive at week 12.	-106
3	Patient died before the assessment of KCCQ-TSS at week 12 was conducted.	Ranked by time to death from randomisation	Patient died y days after randomisation	-108+y/1000

For any missing week 12 assessment of the CHQ-SAS, the below imputation will be applied for the primary analysis and subgroup analysis in accordance to the strategy for the primary endpoint:

Since the lowest possible value for change from baseline in CHQ-SAS dyspnea (fatigue) score is 1 minus the highest possible baseline value (7) all imputed values for change from baseline are below -6 as described below in Table 6.6.2: 2.

Table 6.6.2: 2 Imputation Rules for missing CHQ-SAS Assessments at Week 12

	Category	Ranking	Case description	(Imputed) Change from baseline in CHQ-SAS for described case
1	Missing data on CHQ-SAS at week 12 and no clinical event: Patients with missing CHQ-SAS at week 12 and no event as in category 2 or 3	If post-baseline value is available, rank by last available value.	Patient has a change from baseline to last available value of x points, no clinical event	-12 +x/1000
		Rank below if no post-baseline score is available	No information of post-baseline CHQ and no clinical event	-14
2	Missing data and clinical event (except death). A clinical event is defined as ongoing SAE at week 12 or AE leading to treatment discontinuation and patient is alive at week 12	Ranked above patients who died, but below patients with missing data and no clinical event.	Patient has ongoing SAE or AE leading to treatment discontinuation, no week 12 CHQ assessment and patient alive at week 12.	-16
3	Patient died before the assessment of CHQ at week 12 was conducted.	Ranked by time to death from randomisation	Patient died y days after randomisation	-18+y/1000

Missing change from baseline due to missing baseline questionnaire will not be imputed.

6.6.3 Secondary endpoints

Handling of missing data for the change from baseline to week 6 in exercise capacity as measured by the distance walked in 6 minutes will be analogous to the primary endpoint, see Table 6.6.3:1

Table 6.6.3: 1 Imputation rules for missing exercise capacity at week 6

	Category	Ranking	Case description	Imputed change from baseline in 6MWTD for described case (in m)*
1	Missing data on 6MWT at week 6 and no clinical event: Patients with missing 6MWT at week 6 and no event as in category 2 or 3	Ranked above patients who died or patients with a clinical event, but below patients with available measurement at week 6.	No information of post-baseline 6MWT distance and no clinical event	-354
2	Missing data and clinical event (except death). A clinical event is defined as ongoing SAE or AE leading to treatment discontinuation and patient is alive at week 6	Ranked above patients who died, but below patients with missing data and no clinical event.	Patient has ongoing SAE or AE leading to treatment discontinuation, no information of week 6 6MWT distance and patient alive at week 6.	-356
3	Patient died before the measurement of 6MWT at week 6 was conducted.	Ranked by time to death from randomisation	Patient died y days after randomisation	-358+y/1000

6.6.4 Longitudinal Analyses

There will be different methods of analysing continuous longitudinal data.

Observed case on-treatment (OC-OT):
Only the available data that were observed while patients were on study medication (defined as time from first drug intake until last permanent treatment stop date plus the endpoint specific follow-up time (as defined in Table 6.1:2)) are considered.
Measurements are assigned to planned weeks according to Table 6.7: 2
Imputed records are not included

Observed case including data after treatment discontinuation (OC-AD):

All available data are considered, including values obtained on treatment or post-treatment. Imputed records are not included.

Measurements are assigned to planned weeks according to Table 6.7: 1 and Table 6.7: 2

Worst case imputation (WCI):
All available data as described for OC-AD and additionally including worst case imputation as described in Sections 6.6.1 – 6.6.3. This imputation will only be considered in non-parametric analyses.

6.6.5 Safety Analyses

There will be no imputation of data for safety analyses.

6.6.6 Missing dates

Adverse event data
Missing or partial date information for AEs will be replaced according to general Boehringer Ingelheim (BI) rules described in the BI guidance for handling of missing and incomplete AE dates (2).

Partial onset dates from clinical event committee (CEC):
In the unlikely case that only the year is documented, the day and month will be imputed as 01 Jan unless the subsequently derived date is before randomisation; in this case the date of randomisation will be used as start date.
If year and month is present the day will be imputed as first of the month unless the subsequently derived date is before randomisation; in this case the date of randomisation will be used.

Death date
For patients with a record of death captured on the electronic case report form (eCRF) with missing or only partial death date from all available sources, the death date will be derived as the latest date of any dates as of: event onset and end dates from either the AE page, or CEC adjudicated onset dates, by using also imputed AE dates, range or of possible days based on partial death date and date of last contact.

Missing information on the date of first administration of trial drug
If the date of first drug administration is missing but the patient was randomised, the date of the first drug administration will be set to the date of randomisation. If the date of first administration is partially missing with the month and year present, the day will be set to the date of randomisation if randomisation was in the same month. If randomisation was in the month prior to the first drug administration the missing day will be imputed as the first day of the month.

Missing date of last contact date:

If the date is partially or completely missing, use the minimum of the following dates:

- the date of Informed Consent (IC) withdrawal (if applicable)
- date of death
- If the patient did not die and did not withdraw consent, the latest of:
    - visit end dates,
    - drug administration dates,
    - last drug stop date,
    - last assessment date (ECG date, date of KCCQ, date of CHQ)
    - AE onset and AE end dates
- (in case for partially missing date) Last day of the year/month given as partial date

In case of a partially missing date, if the imputed date is before the first day of the month/year given as partial date, the first day of the month/year will be used.

All other cases need to be assessed by the trial team on an individual patient basis, using the above points as guidance.

Missing information on the date of trial medication stop

If the date is partially or completely missing, use the minimum of the following dates:

- End of treatment visit date (if available)
- Date of death (if applicable)
- Trial completion (last contact date)
- Longest treatment duration (assuming 1 tablet/day)
- (in case for partially missing date) Last day of the year/month given as partial date

In case of a partially missing date, if the imputed date is before the first day of the month/year given as partial date, the first day of the month/year will be used.

All other cases need to be assessed by the trial team on an individual basis, using the points above as guidance.

Missing information on concomitant therapy dates

For incomplete date information the midpoint of the possible interval will be used. If only the year is present the day and month will be imputed as 01 July, if year and month is present the day will be imputed as 15. If the year is missing, the date will be considered missing.
If this leads to contradictions for the start or end date of a concomitant therapy (e.g. imputed end date before documented start date) a partial end date will be imputed as the end of the interval or a partial start date will be imputed as the start of the interval in the database to resolve this contradiction.

All other cases or conflicting cases resulting from these imputation rules need to be assessed by the trial team on an individual basis.

6.7 BASELINE, TIME WINDOWS AND CALCULATED VISITS

For serum creatinine and values based upon this measurement such as eGFR, baseline will be defined as the mean of all available measurements from the screening visit until start of randomised trial medication. For all other endpoints, baseline will be defined as the last available measurement before start of randomised trial medication.

Since the protocol specifies that all measurements are taken at visit 2 before any intake of trial medication, all measurements at the first day of drug intake are assumed to qualify as baseline measurements. For laboratory values, clock time is recorded. If a lab measurement is done after treatment start, this will not qualify as baseline measurements, except in case of HbA1c, which is considered to be a stable measurement and qualifies for baseline on the whole day of first treatment intake.

For randomised patients without any treatment intake: For serum creatinine and values based upon this measurement such as eGFR, baseline will be defined as the mean of all available measurements from the screening visit until and including the day of randomisation. For all other endpoints, baseline will be defined as the last available measurement before or on the day of randomisation.

Medication taken at baseline is any medication with start date continued or before date of first study medication intake (randomisation for patients not treated) and end date continued or on or after date of first medication intake (randomisation for patients not treated).

Measurements taken after the first intake of randomised trial medication will be considered on-treatment values if they have been obtained up to end of the parameter specific follow-up period as defined in Table 6.7: 1 below and will be assigned to the randomised study drug for efficacy and safety analyses.

Measurements taken after the end of the endpoint specific follow-up period after the last intake of study drug will be considered post-treatment values.

On-treatment (for OC-OT analysis) or all post-randomisation (for OC-AD analysis) efficacy and safety measurements will be assigned to visits based on time windows around the planned visit dates. These time windows are defined based on the planned number of days after the date of first administration of study drug.

The time window for the first visit after randomisation starts on the day after the first intake of study drug. For the primary analysis of the primary endpoint and the key secondary endpoints, the start of the time window of week 12 assessment will follow the protocol defined time window (refer to Table 6.7: 1 for details). For all sensitivity analyses and all other endpoints the midpoint between two on-treatment visits defines the end of a time window, with the midpoint being included in the time window of the preceding visit (Table 6.7: 2).

Table 6.7: 1 Time windows for primary analysis of primary and key secondary endpoints and Sensitivity analyses using Wilcoxon model

			Time window (days after baseline)	
Visit number	Visit label	Planned days	Start	End
2	Baseline[1]	1	NA	+1
3	Week 6	43	+2	+77[2]
4	Week 12	85	+78	NA[2]

[1] Only values taken prior or on the day of treatment start with randomised study drug can be considered baseline values. To be conservative, all data on the day of treatment start will be assumed to be before treatment start for 6MWT and PROs. Time windows will be used for assignment of measurements to scheduled visits.
[2] For patients with no value in the visit 4 window (visit 3 window respectively), determination of imputation categories will be: deaths before day 92 (day 77 respectively) will be considered for imputation category 3; value for patients with clinical event overlapping at least partially with the time window day 78 to day 92 (day 2 to day 77 respectively) will be imputed according to imputation category 2.

Table 6.7: 2 Time windows for measurements except primary analysis of primary and key secondary endpoints [OC-OT]

			Time window (days after baseline)	
Visit number	Visit label	Planned days	Start	End
Endpoints assessed at each on-site visit (e.g. creatinine / eGFR)				
2	Baseline[1]	1	NA	+1
3	Week 6	43	+2	+64
4	Week 12	85	+65	Trt stop + x* days
FU	FU	Trt stop + 7 days	Trt stop + x* + 1 day	NA
Endpoints that are not assessed on visit 3 (e.g. weight)				
2	Baseline[1]	1	NA	+1
4	Week 12	85	+2	Trt stop+ x* days
FU	FU	Trt stop + 7 days	Trt stop + x* + 1 day	NA

[1] Only values taken prior to the start of treatment with randomised study drug can be considered baseline values. Time windows will be used for assignment of measurements to scheduled visits.
* see Table 6.1: 2

Table 6.7: 3 Time windows for measurements except primary analysis of primary and key secondary endpoints [OC-AD]

			Time window (days after baseline)	
Visit number	Visit label	Planned days	Start	End
Endpoints assessed at each on-site visit (e.g. creatinine / eGFR)				
2	Baseline [1]	1	NA	+1
3	Week 6	43	+2	+64
4	Week 12	85	+65	NA
Endpoints that are not assessed on visit 3 (e.g. weight)				
2	Baseline [1]	1	NA	+1
4	Week 12	85	+2	NA

[1] Only values taken prior to the start of treatment with randomised study drug can be considered baseline values. Time windows will be used for assignment of measurements to scheduled visits.

For examinations that are not planned at every on-treatment visit, the time windows will be defined according to the same algorithm, based on the midpoint between the planned visit day of such an examination. Examples for eGFR and weight can be found in Table 6.7: 2 for OC-OT analyses and in Table 6.7: 3 for OC-AD analyses.

Only one observation per time window will be selected for analysis at an on-treatment visit – the value will be selected which is closest to the protocol planned visit day.

For the primary endpoint (6MWTD) and any measurement related to the 6MWT assessment the longest walking distance for each day will be used for the analysis for that day.

If there are two measurements which have the same time difference in days to the planned day, or if there are two measurements on the same day (endpoints not related to the 6MWT), the first value will be used.

Baseline definition for concomitant therapies

Concomitant medication taken at baseline is any medication with start date continued or before date of first study medication intake (randomisation date will be used for patients not treated) and end date continued on or after date of first study medication intake (randomisation date will be used for patients not treated).

7. PLANNED ANALYSIS

For End-Of-Text (EoT) tables, the set of summary statistics is: N / Mean / standard deviation (SD) / standard error (SE) / Min / lower quartile (Q1) / Median / upper quartile (Q3) / Max. The 1st and 99th percentiles might be substituting minimum and maximum in tables with open-ended values to safeguard against implausible extremes. Final decision will be made at the final BRPM.

Geometric means and gCV will be added to the presentation or replace the presentation of mean, standard error and standard deviation for parameters (e.g. NT-proBNP) that follow a log-normal distribution rather than a normal distribution.

Statistical parameters will be displayed to a defined number of decimal places as specified in the BI guideline "Standards for Reporting of Clinical Trials and Project Summaries" (3)

Tabulations of frequencies for categorical data will include all possible categories and will display the number of observations in a category as well as the percentage (%) relative to the respective treatment group (unless otherwise specified, all patients in the respective patient set whether they have non-missing values or not). Percentages will be rounded to one decimal place. The category missing will be displayed only if there are actually missing values.

Disposition of the patient population participating in the trial will be analysed by treatment groups and presented in the clinical trial report as a frequency distribution. The number of patients participating (screened, randomised, screened but not randomised, etc.) in the study by region, country and, for treated patients, centre, will also be analysed by treatment group and presented as a frequency distribution.

Disposition as required for reporting for the trial in EudraCT will be provided. Enrolment will be summarised by country and by age group for reporting in EudraCT.

The reasons for not randomising screened patients will be summarised descriptively.

The frequency of patients with iPDs will be presented by treatment group for the randomised set. The frequency of patients in different analysis sets will also be analysed for each treatment group.

7.1 DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS

Only descriptive statistics are planned for this section of the report.

Demographics and baseline characteristics will be analysed based on RS. Demographics will be repeated on the TS if the number of patients in the analysis sets differ by more than 5%. Standard descriptive analysis and summary tables will be presented. These summary tables will include description of variables detailed in Section 6.4. In addition, descriptive analyses of the following continuous variables measured at baseline will be presented: Age, body mass index (BMI), time since HF-diagnosis, systolic blood pressure (SBP), diastolic blood pressure

(DBP), weight, eGFR, NT-proBNP, heart rate. For patients with diabetes at baseline, HbA_{1c} at baseline will be shown descriptively.

7.2 CONCOMITANT DISEASES AND MEDICATION

Only descriptive statistics are planned for this section of the report using the randomised set. Concomitant medication use will be summarised by Anatomical-Therapeutic-Chemical classification level 3 (ATC1-ATC2-ATC3) and preferred name. Summaries will be presented for concomitant therapies taken at baseline and separately those added after baseline.

Separate summaries of use of heart failure-related drugs (e.g. ARNi, beta-blockers, ivabradine, diuretics, ACE-inhibitors, ARBs, MRAs, cardiac glycosides), anticoagulants, acetylsalicylic acid (ASA), or lipid lowering drugs at baseline and separately those for added after baseline will be presented. Definitions of these medication groups are based on World Health Organization Drug Dictionary (WHO DD) and are stored in the Project Data Management and Analysis Plan (PDMAP).

Use of devices at baseline will also be summarised.

Concomitant diseases will be summarised by system organ class and preferred term. Relevant medical history by treatment group will also be presented. Both summaries will be presented using the randomised set.

The number and percentage of patients receiving the best tolerated treatment for HF symptoms and other common concomitant diseases or symptoms (e.g. hypertension, diabetes, cardiac arrhythmia, fluid retention) according to prevailing guidelines in the opinion of the investigator will be reported.

7.3 TREATMENT COMPLIANCE

Only descriptive statistics are planned for this section of the report. The number and percentage of patients with overall study medication compliance will be reported.

7.4 PRIMARY ENDPOINT

7.4.1 Primary analysis of the primary endpoint

The primary endpoint ‘change from baseline to week 12 in exercise capacity’ as measured by the change from baseline to week 12 in distance walked in 6 minutes will be evaluated using a normal approximation of the Wilcoxon rank test using the randomised set.

For estimation of effect, non-parametric Hodges-Lehmann estimate for the median difference will be used. Corresponding 95% confidence intervals will be displayed. Any confirmatory conclusion such as for hierarchical testing will be based on the result of the Wilcoxon rank test and will not consider the Hodges-Lehmann estimate or the corresponding confidence interval, which are given to support interpretation.

If no 6MWTD is available at week 12, an imputed value as defined in Section 7.5 of the CTP will be used.

The primary analysis will be conducted on all randomised patients, regardless of whether on or off-treatment on the day of 6MWT assessment. Where more than one measurement is available for the same day (duplicate measurements), the longest distance walked will be used for analysis. In case more than one 6MWT results in the same distance walked, the first assessment will be used for analysis.

7.4.2 Sensitivity analysis of the primary endpoint

Sensitivity analysis for the primary endpoint will include;

- a mixed model repeated measure analysis (MMRM) without imputation of missing values (OC-AD).
- an MMRM restricted to on-treatment values (OC-OT).

The model will include visit-by-treatment interaction and baseline-by-visit interaction. An unstructured covariance structure will be used to model the within-patient errors. Descriptive statistics will be calculated for the value at visit and change from baseline based on OC-AD on the randomised set and based on OC-OT on the treated set. The MMRM analyses will also be shown graphically.

The primary analysis will be repeated restricted to patient in the sensitivity analysis set SensRS1, SensRS2 and SensRS3.

If the mean difference between the treatment groups in 6MWTD is above 15m at baseline, the primary analysis will be repeated, but using the ranks of the residuals from the following linear regression model

$$6MWTD_{12\text{weeks}} \sim 6MWTD_{\text{Baseline}}$$

with patients without week 12 data ranked below any patient with week 12 assessment similar to the rules for primary analysis.

7.4.3 Subgroup analysis of the primary endpoint

The primary analysis of the primary endpoint will be repeated for each subgroup as defined in Table 6.4: 1. A forest plot summarising all subgroup analyses will be shown.

7.5 SECONDARY ENDPOINTS

7.5.1 Key secondary endpoints

If the null hypothesis for the primary analysis of the primary endpoint is rejected, and the result is more favourable for empagliflozin, superiority is concluded in the primary endpoint, and the overall type I error is preserved for the test in the next step. In the next step a Hochberg procedure will be applied to account for multiplicity of testing of the key secondary endpoints:

If the largest p-value of the primary analysis of the two key secondary endpoints is below 5%, both null hypotheses of the key secondary endpoints will be rejected. Otherwise, if the smaller p-value of the primary analysis of the two key secondary endpoints is below 2.5%, only the corresponding null hypothesis can be rejected.

If the primary hypothesis of the primary endpoint is not rejected, the tests of the key secondary endpoints are conducted in an exploratory manner.

7.5.1.1 Primary analysis of the key secondary endpoints

The two key secondary endpoints 'Change from baseline to week 12 in KCCQ-TSS' and 'Change from baseline to week 12 in CHQ-SAS dyspnea score' are part of the testing strategy.

Both key secondary endpoints 'Change from baseline to week 12 in heart failure symptoms as measured by the KCCQ TSS' and 'Change from baseline to week 12 in CHQ-SAS dyspnea score' will be evaluated using a normal approximation of the Wilcoxon rank test using the randomised set. For estimation of effect the non-parametric Hodges-Lehmann estimate for the median difference will be calculated.

Change from baseline in both scores will be defined as the endpoint value at week 12 minus the last available endpoint value before start of randomised trial medication.

If no questionnaire is available at week 12, an imputed value as defined in Section 6.6 will be used.

7.5.1.2 Sensitivity analysis of the key secondary endpoints

Sensitivity analysis for the key secondary endpoint will include

- a mixed model repeated measure analysis without imputation of missing values (OC-AD).
- a mixed model repeated measure analysis restricted to on treatment values (OC-OT).

The model will include visit-by-treatment interaction and baseline-by-visit interaction. An unstructured covariance structure will be used to model the within-patient errors. Descriptive statistics will be calculated for the value at visit and change from baseline based on OC-AD on the randomised set and based on OC-OT on the treated set.

The primary analysis will be repeated restricted to patients in sensitivity analysis sets SensRS2 and SensRS3.

7.5.1.3 Subgroup analysis of the key secondary endpoints

The primary analysis of the key secondary endpoints will be repeated for each subgroup as defined in Table 6.4: 1. A forest plot summarising all subgroup analyses will be shown.

7.5.1.4 Further exploratory analyses of the key secondary endpoints

A minimal clinically meaningful within-patient change in scores (i.e. responder threshold) will be estimated from the trial population (both treatment groups pooled) for KCCQ-TSS and CHQ-SAS dyspnea as described in Appendix 9.3.

A responder analysis using a chi-square test will be evaluated in addition. The response threshold will be determined as defined in Appendix 9.3.

An empirical cumulative distribution function of descriptive change from baseline to week 12 by treatment group on OC-AD on the randomised set will be shown.

7.5.2 Other Secondary endpoints

Other secondary endpoints will not be part of the testing hierarchy. No correction for multiple testing of the other secondary endpoints will be made.

Change from baseline to week 6 in exercise capacity as measured by the distance walked in 6 minutes will be evaluated using the same model as for the primary endpoint. If no 6MWTD is available at week 6, an imputed value as defined in Section 6.6 will be used.

All other secondary endpoints will take into account all data, regardless of whether the patient was on treatment or not, and without imputation of missing values (OC-AD).

The other continuous endpoints are generally analysed in an MMRM, including visit-by-treatment interaction and baseline-by-visit interaction. An unstructured covariance structure will be used to model the within-patient errors. The MMRM analyses will also be shown graphically.
Descriptive statistics will be calculated for the value at visit and change from baseline.

Shift from baseline score to 12 week score will be tabulated descriptively for single items collected in the context of the clinical congestion score (dyspnea, orthopnea, fatigue, jugular venous distension, rales and edema).

The endpoint “Change from baseline in NT-proBNP at week 12” (after log-transformation) will be evaluated using an MMRM analysis over time with baseline log-transformed NT-proBNP by visit interaction and visit by treatment interaction as covariates without imputation of missing values. Estimates obtained from the model will then be back-transformed and reported on the original scale. The MMRM analyses will also be shown graphically.

Proportion of patients in every category of global impression of change at week 12 will be evaluated using Cochran-Mantel-Haenszel test on the difference in mean treatment scores, using modified ridit scores.

Proportions of patients categorised by the difference in scores between baseline and week 12 in patient global impression of severity score will be evaluated using a Cochran-Mantel-Haenszel test on the difference in mean treatment scores, using modified ridit scores. Difference in score will be defined as the number of categories improved/deteriorated from baseline. Shift from baseline score to 12 week score will be tabulated descriptively.

7.6 FURTHER ENDPOINTS

All further endpoints will take into account all data, regardless of whether the patient was on treatment or not, and without imputation of missing values (OC-AD).

7.6.1 Binary endpoints

For the following binary endpoints descriptive analyses will be complemented by an exploratory chi-square test to compare response between treatment groups.

- Increase in 6MWT ≥ 30 m from baseline at week 12
- Response in KCCQ domains at week 12, as defined in Section 5.3. A forest plot of response at week 12 for each domain by treatment will be shown.
- Response in CHQ domains at week 12, as defined in Section 5.3. A forest plot of response at week 12 for each domain by treatment will be shown.

Descriptive statistics are planned for all other binary endpoints:

- Occurrence of HHF and Cardiovascular (CV) death
- Occurrence of all-cause mortality
- Occurrence of CV death
- Occurrence of all-cause hospitalisation (first and recurrent)
- Occurrence of hospitalisation for HF (first and recurrent)
- Occurrence of all-cause emergency room visits (first and recurrent)
- Emergency room visit for HF (requires use of i.v. diuretics; first and recurrent)
- Occurrence of unscheduled outpatient visit (first and recurrent)
- Intensification of diuretic therapy (adding a new diuretic, increase of dose)
- Occurrence of MI (fatal or non-fatal)
- Occurrence of Stroke (fatal or non-fatal)
- Occurrence of TIA

7.6.2 Ordinal endpoints

The following endpoints will be evaluated using the same models as defined for the corresponding secondary endpoints:

- Change from baseline in Patient Global Impression of Severity of Heart Failure Symptoms at week 6.
- Change from baseline in Patient Global Impression of Severity of Dyspnea at week 6.
- Patient Global Impression of Change in Heart Failure Symptoms at week 6.

- Patient Global Impression of Change in Dyspnea at week 6.
- Change from baseline in Clinician Global Impression of Severity of Chronic Heart Failure (CHF) at Week 6 and week 12
- Clinician Global Impression of Change in CHF Severity at Week 6 and week 12

Frequencies and proportions of patients categorised descriptively by the

- shift in NYHA class from baseline to week 6 and to week 12.

7.6.3 Continuous endpoints

Further continuous endpoints are generally analysed in a mixed model with repeated measures (MMRM), including visit-by-treatment interaction and baseline-by-visit interaction. An unstructured covariance structure will be used to model the within-patient errors. The MMRM analyses will also be shown graphically.
Descriptive statistics will be calculated for the value at visit and change from baseline.

- Change from baseline in KCCQ overall summary score, CSS, and individual domains at Week 6 and week 12
- Change from baseline in CHQ individual domains at week 6 and week 12
- Change from baseline in SBP at Week 6 and week 12
- Change from baseline in DBP at Week 6 and week 12

The following endpoint is assessed as part of the main analysis or sensitivity analysis of the corresponding secondary endpoint:

- Change from baseline in KCCQ TSS at Week 6
- Change from baseline in CHQ-SAS dyspnea at Week 6
- Change from baseline in Clinical Congestion Score at week 6.
- Change from baseline in NT-proBNP at week 6.

Since bodyweight is only measured at baseline and week 12, an ANCOVA model will be presented including baseline weight as a covariate.

Descriptive statistics are planned for:

- Change from baseline to week 6 and week 12 in patient-rated assessment of dyspnea and fatigue by the Borg scale as part of the 6MWT

For the following parameters measured at the follow-up visit, additionally descriptive statistics at baseline, last value on treatment and follow-up will be presented for patients with valid measurements at all three timepoints. .

- SBP
- DBP

7.7 EXTENT OF EXPOSURE

Descriptive statistics tables based on the TS with mean, SD, SE, quartiles, median and range of the number of days a patient was on treatment will be provided. These tables will also provide the sum-total of the time (in years) that all patients were on treatment.

Frequency tables of number and percentage of patients belonging to categorical ranges of exposure weeks will be provided as well. Following are the categories of exposure-ranges (in weeks): >= 0 weeks, >=4 weeks, >= 8 weeks, >= 11 weeks.

7.8 SAFETY ANALYSIS

The safety analysis will be based on the treated set (TS).

While tables will generally display results by randomised treatment, listings will reflect whether a measurement/AE occurred on or off treatment.

7.8.1 Adverse events

Unless otherwise specified, the analyses of adverse events (AEs) will be descriptive in nature and will be based on the number of patients with AEs and not on the number of AEs.

AEs will be coded using version 22.0 of the Medical Dictionary for Regulatory Activities (MedDRA) coding dictionary.

Definitions of Boehringer Ingelheim customised MedDRA Queries (BIcMQs) for new MedDRA versions are maintained by the BI dictionary maintenance group.

For analysis multiple AE occurrence data in the data from the eCRF will be collapsed into an AE provided that all of the following applies:

- All AE attributes are identical (including Lowest Level Term (LLT), intensity, action taken, therapy required, seriousness, reason for seriousness, relationship, outcome, AE of special interest and also additional information of specific AEs or AESIs such anatomical location of UTI (upper UTI kidney versus Lower UTI bladder and below versus asymptomatic bacteriuria) or type of genital infection (fungal balanitis or vulvovaginitis versus other than fungal balinitis or vulvovaginitis)).
- The occurrences were time-overlapping or time-adjacent (time-adjacency of 2 occurrences is given if the second occurrence started on the same day or on the day after the end of the first occurrence).

For further details on summarisation of AE data, please refer to (2).

7.8.1.1 Assignment of AEs to treatment

The analysis of adverse events will be based on the concept of treatment-emergent adverse events. That means that all adverse events occurring between first drug intake till 7 days after last drug intake will be assigned to the randomised treatment. All adverse events occurring

before first drug intake will be assigned to 'pre-treatment' and all adverse events occurring after last drug intake + 7 days with be assigned to 'follow-up', except if otherwise specified.

In Section 15.3 general AE analysis tables will present only on-treatment AEs (applying the rule of 7 days for assignment as on-treatment) for the treatments groups.

Appendix 16.1.13.1 will include an analysis (overall summary table, frequency of AEs by system organ class (SOC) / preferred term (PT), frequency of serious adverse events (SAEs) by SOC/PT) where AEs and SAEs are assigned to the following phases: Screening, each treatment group, post-treatment for each treatment group.

For listings, AEs will be assigned to one of the treatment phases of Screening, Placebo, Empa 10, Post Placebo, Post Empa.

7.8.1.2 AE summaries

An overall summary of adverse events will be presented.

The frequency of patients with adverse events will be summarised by treatment, primary system organ class and preferred term. Incidence rates as defined in Section 7.8.1.7 will generally be included.

AEs will also be reported by intensity (without incidence rates). Separate tables will be provided for

- for patients with serious adverse events,
- for patients with drug related serious AEs,
- for patients with AEs leading to discontinuation
- for patients with drug-related AEs.

AEs leading to death will be listed.

The system organ classes will be sorted according to the standard sort order specified by the European Medicines Agency (EMA), preferred terms will be sorted by frequency (within system organ class).

Additionally, the following analyses will be reported in Appendix 16.1.13.1 for disclosure on EudraCT and clinicaltrials.gov:

- Frequency [N(%)] of subjects with non-serious adverse events occurring with incidence in preferred term greater than 5% by treatment,
- Adverse events per arm for disclosure on EudraCT by treatment
- Non-serious adverse events for disclosure on EudraCT by treatment
- Serious adverse events for disclosure on EudraCT by treatment

7.8.1.3 Adverse events of special interest (AESIs)

Hepatic injury

Hepatic injury AEs will be summarised based on a Standardised MedDRA Query (SMQ) based definition. From SMQ Drug related hepatic disorders (20000006) the following narrow sub-SMQs will be used:

- Narrow sub-SMQ Liver related investigations, signs and symptoms (20000008)
- Narrow sub-SMQ Cholestasis and jaundice of hepatic origin (20000009)
- Narrow sub-SMQ Hepatitis, non-infectious (20000010)
- Narrow sub-SMQ Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (20000013)

A table with frequencies of patients with these AEs by treatment, primary SOC and preferred term will be provided. Hepatic injury SAEs based on the above SMQ definition will be presented.

Patients with hepatic injury will be listed.

For presentations on adjudicated hepatic events, refer to Section 7.8.1.5.

Acute renal failure

A frequency table of patients with AEs related to acute renal failure by treatment, primary SOC and preferred term will be provided based on the narrow SMQ Acute renal failure (20000003).

SAEs and AEs leading to discontinuation based on the narrow SMQ Acute renal failure (20000003) will be presented.

In addition, frequency tables will be produced for patients with elevated creatinine ≥ 2 x baseline and > 1 x upper limit of normal (ULN).

Patients with acute renal failure will be listed.

Ketoacidosis

A frequency table of patients with AEs related to ketoacidosis will be presented by treatment, primary SOC and preferred term for investigator reported cases and for the narrow and broad BIcMQ definition of diabetic ketoacidosis (DKA).

For presentations on adjudicated events, refer to Section 7.8.1.5

Patients with DKA based on the broad BIcMQ (30000019) or investigator reported ketoacidosis will be listed.

7.8.1.4 Specific AEs

Hypoglycaemic events

The investigator will record for each AE whether it represents a hypoglycaemic event and, if so, record additional information to assess the intensity of the hypoglycaemic event. On the basis of this information the hypoglycaemic event will be categorised as follows:

- severe hypoglycaemic episode: event requiring the assistance of another person to actively administer carbohydrate, glucagon or other resuscitative actions
- documented symptomatic hypoglycaemia with a measured plasma glucose concentration < 3.0 mmol/L (< 54 mg/dL): event accompanied by typical symptoms of hypoglycaemia but no need for external assistance
- documented symptomatic hypoglycaemia with a measured plasma glucose concentration ≥ 3.0 mmol/L and ≤ 3.9 mmol/L (≥ 54 mg/dL and ≤ 70 mg/dL): event accompanied by typical symptoms of hypoglycaemia and no assistance required
- symptomatic hypoglycaemia and plasma glucose concentration > 3.9 mmol/L (70 mg/dL)
- symptomatic hypoglycaemia and plasma glucose concentration not measured
- asymptomatic hypoglycaemia: event not accompanied by typical symptoms of hypoglycaemia but with a measured plasma glucose concentration ≤ 3.9 mmol/L (≤ 70 mg/dL)

Confirmed hypoglycaemic adverse event are defined as hypoglycaemic adverse events that had a plasma glucose concentration ≤ 70 mg/dL or required assistance.

Different tables will be shown for (i) patients with investigator-defined hypoglycaemia, and (ii) patients with confirmed hypoglycaemic adverse events, i.e. hypoglycaemic adverse events that had a plasma glucose concentration ≤ 70 mg/dL or required assistance.

Subgroup analyses on confirmed events with respect to DM status at baseline (DM vs non-DM) will be performed.

In addition the number of patients with hypoglycaemia according to SMQ (20000226) will be presented.

Patients with hypoglycaemic events will be listed.

Urinary tract infections (UTI) and genital infections

The following additional specific adverse events will also be assessed and be tabulated by treatment group:

- Genital infections (BIcMQ “Infections” (30000036) narrow subsearch 1.1 “Genital tract infections predisposed to by glucosuria” (30000038) and investigator assessment)

- UTI (BIcMQ "Infections" (30000036) narrow subsearch 2.1 "UTI predisposed by glucosuria" (30000041) and investigator assessment)

Genital infections based on investigator assessment will additionally be summarised by type (fungal balanitis or fungal vulvovaginitis, genital infection other than fungal balanitis or fungal vulvovaginits), intensity (mild, moderate or severe), how the event was treated (no treatment, therapy assigned and number of antimicrobials needed to treat), whether leading to discontinuation of treatment, and the number of episodes per patient.

UTIs based on investigator assessment will additionally be summarised by intensity (mild, moderate or severe), anatomical location (upper UTI, lower UTI), occurrence of pyelonephritis or urosepsis, how the event was treated (no treatment, therapy assigned and number of antimicrobials needed to treat), whether leading to discontinuation of treatment, and the number of episodes per patient.

Complicated urinary tract infections defined as serious adverse events of BIcMQ Infection narrow, subsearch 2.1 "UTI predisposed by glucosuria" (30000041), all events of subsearch 2.1.1 'Renal infections predisposed by glucosuria' (30000042), all events of PT Urosepsis will be presented.

Complicated genital infection: defined as all serious events using the BIcMQ Infection narrow subsearch 1.1 'Genital tract infections predisposed to by glucosuria' (30000038) and all event of the subsearch 'Complicated genital tract infections predisposed to by glucosuria' (30000129) will also be presented.

Patients with UTIs or genital infections will be listed.

<u>Bone fracture events:</u>

Frequency tables of patients with bone fracture by treatment, primary system organ class (SOC) and preferred term will be provided (based on the narrow BIcMQ "Bone fractures" (30000008) and investigator reporting).

Investigator reported fractures will be reported overall and separately for each type of fracture (traumatic and non-traumatic).

Separate tables for bone fractures based on the BIcMQ, which are serious will be presented.

Patients with bone fractures will be listed.

<u>Urinary tract malignancy events:</u>

Urinary tract malignancy will be shown based on the BIcMQ 'Malignancies' (30000049) – broad sub-search 14.1 'Urinary bladder and tract malignancies' (30000057) and broad sub-search 14.2 'Renal malignancies' (30000103):

Presentation of frequency will be done by treatment, high level term and preferred term.

Patients with urinary tract malignancy will be listed.

Volume depletion

Volume depletion will be based on the BIcMQ 'Volume depletion of non-haemorrhagic cause and subsequent hypotension'– narrow subsearch 2 'Volume depletion and hypotension due to dehydration' (30000090).

A frequency table of patients with volume depletion by treatment, primary SOC and preferred term will be provided.

A separate table for serious volume depletion events will be presented.

Patients with volume depletion will be listed.

For the analysis of laboratory data, refer to Section 7.8.2.

Hypotension:

Frequency table of subjects with symptomatic hypotension as defined by the investigator on the eCRF tick box by treatment, primary system organ class and preferred term will be provided.

Symptomatic hypotension episodes will be presented by whether the intensity of diuretic medication was reduced and by whether the intensity of non-diuretic antihypertensive therapy was reduced.

A separate table for symptomatic hypotension events, which are serious, will be presented.

Additionally hypotension by treatment, primary system organ class and preferred term will be provided. Hypotension is defined as preferred terms of the BIcMQ 'Volume depletion of non-haemorrhagic cause and subsequent hypotension' – narrow sub-search 2 'Volume depletion and hypotension due to dehydration' (30000090) but excluding terms of the narrow subsearch 1 'Volume depletion due to dehydration'(30000089).

A separate table for hypotension events, which are serious, will be presented.

Patients with hypotension events will be listed.

7.8.1.5 Events qualifying for external adjudication by the adjudication committee

An independent external CEC regularly reviews events and evaluates whether pre-specified criteria for these adjudication endpoints are met. Details on composition of the CEC, responsibilities and clinical event definitions are provided in the separate CEC Charter.

Adjudication assessments will be incorporated into the database.

Details of the adjudication process are described in the CEC charter.

Hepatic adverse events:
Frequency tables summarising the relatedness and severity will be provided, including a listing showing the trigger events and adjudication results. This will include all adjudicated events up to 30 days after treatment stop.

Ketoacidosis:
Frequency tables showing adjudicated certain ketoacidosis will be provided, including a listing showing the trigger events and adjudication results. This will include all adjudicated events up to 7 days after treatment stop.

7.8.1.6 AEs while patients taking wrong medication

A listing will be provided for AEs that occurred while a patient was taking the wrong study medication.
If such a patient is identified an additional adverse event table that assigns the adverse events to the actual treatment taken will be presented. A patient who took both the assigned treatment and also at least one tablet of the wrong treatment will be counted as at risk in both treatment groups for the respective relevant time. The table will include all adverse events by SOC and PT.

7.8.1.7 Adverse event incidence rates

For AE tables showing patients with events, in addition to the frequency tabulations, time-adjusted adverse event analyses will be performed for all AEs, investigator defined drug-related AEs, AEs leading to discontinuation, serious AEs, and adverse events of special interest by SOC and PT or High-Level Term (HLT) and PT, respectively.

The time at risk in patient years for the on-treatment phase is derived as follows:

Patients with AE:
time at risk (AE) in days = date of start of AE with specified PT/SOC/HLT – study treatment start date + 1

Patients without AE:
time at risk (AE) in days = end date of time at risk – study treatment start date + 1, where end date of time at risk is the minimum of date of last study drug intake + x days and date of death, if applicable.

The standard approach will be x=7 days, but for certain AESIs in addition other approaches will be used.

Total AE-specific time at risk per treatment group is then derived as:
Time at risk (AE) [years] = Sum of time at risk [days] over all patients in a treatment group/ 365.25

For 'each row of a table' (e.g. displaying an SOC), time at risk is calculated using start of first AE summarised in this row; e.g. for patient with AE in a specified SOC, time at risk = date of start of AE with specified PT in this SOC – start of study treatment + 1.

The AE incidence rate per 100 patient years will then be calculated as follows:

Incidence rate per 100 patient years (pt-yrs) = 100 * number of patients with AE / time at risk (AE) [years].

In a similar way the time at risk and incidence rate for the post-treatment period is derived. Hereby the start date is the start date of the post-treatment phase instead of the study treatment start date.

For some specific outputs showing number of episodes, event rates will be defined as number of episodes divided by time at risk. Time at risk will be end date of time at risk – study treatment start date + 1, where end date of time at risk is the minimum of date of last study drug intake + x days and date of death, if applicable.

7.8.2 Laboratory data

Standard safety tables will not include eGFR or creatinine. Analyses for eGFR and creatinine will be done separately.

For continuous safety laboratory parameters, normalised values will be derived as well as the differences to baseline. The process of normalisation as well as standard analyses for safety laboratory data are described in the BI guidance for the Display and Analysis of Laboratory Data (BI-KMED-BDS-HTG-0042 (4)). All analyses considering multiple of times upper limit of normal (ULN) will be based on original and not normalised data.

Laboratory parameters will be shown in conventional units. Where conventional units differ from SI units, analyses will be repeated on SI units.

The analyses of laboratory data will be descriptive in nature and will be based on BI standards [see Data Management and Statistics Manual (DM&SM): Display and Analysis of Laboratory Data (BI-KMED-BDS-HTG-0042 (4))].

Laboratory measurements taken up to 3 days after the last administration of randomised study drug will be considered as on-treatment.
Default settings will be used for repeated values (using closest and then worst value).

Laboratory values will be compared to their reference ranges and frequency tables will be provided for the number of patients within and outside the reference range at baseline and the last measurement on treatment. Descriptive statistics will be provided by treatment group for baseline, last value on-treatment and for changes from baseline to last value on treatment.
For semi-quantitative or categorical laboratory values, shift tables for baseline and last value on treatment will also be provided. Frequency tables will summarise the number of patients with potentially clinically significant abnormalities.

Additionally, for urine ketones a shift table of baseline vs value at visit for each visit will be shown.

For the table on potentially clinically significant abnormalities only patients with at least one available on-treatment value will be included in the analysis of an individual laboratory parameter. Other tables are comparing baseline to on-treatment values and will only include patients with a baseline value and at least on available on-treatment value. All individual data will be presented in listings.

To support analyses of liver-related adverse drug effects, patients with Aspartate transaminase (AST) and/or Alanine transaminase (ALT) ≥3xULN with concomitant or subsequent Total Bilirubin (TBILI≥2xULN) in a 30 day period after AST/ALT elevation are of special interest. In addition, of these cases, it will be considered whether the alkaline phosphatase (AP) is less than 2 x ULN (maximum value in a 30 day period after AST/ALT elevation) or not. The start of the 30-day time span is triggered by each liver enzyme elevation above the defined thresholds. Patients who fulfil one or two of the criteria for ALT/AST or total bilirubin elevations above and have no information available for the remaining parameter(s) within the 30 day time window will not be listed under "ALT and/or AST ≥ 3xULN with TBILI ≥ 2xULN".

In addition ALT/AST will be used to investigate elevated liver enzymes:

- ALT/AST ≥ 3 x ULN
- ALT/AST ≥ 5 x ULN
- ALT/AST ≥ 10 x ULN
- ALT/AST ≥ 20 x ULN

All liver enzyme elevations within 30 days of treatment discontinuation will be shown.

Frequency tables of patients with elevated liver enzymes defined by ALT and/or AST, total bilirubin and AP combinations will be provided. Details on patients with elevated liver enzymes will be listed.

For the following parameters:

- eGFR,
- creatinine (only descriptive),

the time course of changes will be assessed. The analysis will be performed by applying MMRM models to OC-OT data on the treated set. The MMRM models that will be used are specified in Section 7.6. These analyses will be conducted on data before any normalisation. The MMRM analyses will also be shown graphically.

For the following parameters:

- haemoglobin, and
- haematocrit

the change from baseline will be assessed. The analysis will be performed by applying ANCOVA models to OC-OT data on the treated set. The ANCOVA model will include treatment and baseline as fixed effects. Theses analyses will be conducted on data before any normalisation.

To support analysis of renal function, eGFR throughout the trial will be categorised according to the following Chronic Kidney Disease (CKD) staging (Table 7.8.2: 1): All calculations for the staging of renal function will be based on the originally measured laboratory values, not on normalised values with BI standard reference ranges.

For the following parameters measured at the follow-up visit, additionally descriptive statistics at baseline, last value on treatment and follow-up will be provided for patients with valid measurements at all three timepoints.

- creatinine
- eGFR
- haematocrit
- haemoglobin

Table 7.8.2: 1 CKD staging

Stage	**eGFR (mL/min/1.73m²)**	**Description**	**Label for displays**	***Additional labels#***
1	≥90	Normal or high	≥90	≥90 (CKD 1)
2	60 to <90	Mildly decreased	60 to <90	60 to <90 (CKD 2)
3A	45 to <60	Mildly to moderately decreased	45 to <60	45 to <60 (CKD 3a)
3B	30 to <45	Moderately to severely decreased	30 to <45	30 to <45 (CKD 3b)
4	15 to <30	Severely decreased	15 to <30	15 to <30 (CKD 4)
5	<15	Kidney failure	<15	<15 (CKD 5)

A shift table from baseline to last value on treatment and to minimum value on treatment for eGFR $(CKD\text{-}EPI)_{cr}$ will be provided.

7.8.3 Vital signs

Heart rate over time will be summarised descriptively based on OC-OT on the treated set. Descriptive statistics will also be shown graphically.

Additionally descriptive statistics at baseline, last value on treatment and follow-up for patients with valid measurements at all three timepoints will be shown.

7.8.4 ECG and physical examination

Clinically relevant abnormalities found at physical examination will be considered to have already existed prior to signing of informed consent and therefore should be considered baseline conditions instead of adverse events, unless there is good reason to assume that they first appeared after signing of consent.

Outcomes of ECGs will be part of the reporting of medical history or AE reporting. Categorical findings as collected in the eCRF will also be summarised descriptively.

7.8.5 Others

Not applicable

8. REFERENCES

1	*CPMP/ICH/363/96:* "Statistical Principles for Clinical Trials", ICH Guideline Topic E9, Note For Guidance on Statistical Principles for Clinical Trials, current version.
2	*BI-KMED-BDS-HTG-0035:* "Handling of missing and incomplete AE dates", current version; IDEA for CON.
3	*BI-KMED-BDS-HTG-0045*: "Standards for Reporting of Clinical Trials and Project Summaries", current version; IDEA for CON.
4	*BI-KMED-BDS-HTG-0042*: "Handling, Display and Analysis of Laboratory Data", current version; IDEA for CON.
5	Shoemaker MJ, Curtis AB, Vangsnes E, Dickinson MG. Clinically meaningful change estimates for the six-minute walk test and daily activity in individuals with chronic heart failure. Cardiopulm Phys Ther J 24 (3), 21 - 29 (2013) [R17-1227]
6	Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol 56, 395 - 407 (2003) [R07-1182]
7	Green CP, Porter CB, Bresnahan DR, Spertus JA. Development and evaluation of the Kansas City Cardiomyopathy Questionnaire: a new health status measure for heart failure. J Am Coll Cardiol 35 (5), 1245 - 1255 (2000) [R17-2666]
8	Bennett SJ, Oldridge NB, Eckert GJ, Embree JL, Browning S, Hou N, et al. Comparison of quality of life measures in heart failure. Nurs Res 52 (4), 207 - 216 (2003) [R17-2917]
9	Ambrosy AP, Pang PS, Khan S, Konstam MA, Fonarow GD, Traver B, Maggioni AP, Cook T, Swedberg K, Burnett Jr JC, Grinfeld L, James E. Udelson JE, Faiez Zannad F, Gheorghiade M. Clinical course and predictive value of congestion during hospitalization in patients admitted for worsening signs and symptoms of heart failure with reduced ejection fraction: findings from the EVEREST trial. European Heart Journal (2013) 34, 835–843 [R19-0744]

9. ADDITIONAL SECTIONS

9.1 REGIONS AND COUNTRIES

Countries will be assigned to regions, as outlined in Table 9.1: 1.

Table 9.1: 1 Regions and countries

Region	Country
Australia	Australia
Europe	Germany Greece Italy Norway Poland Portugal Spain Sweden
North America	Canada US

9.2 DIABETES STATUS AT BASELINE

Patients will be assigned to diabetes status category, as outlined in Table 9.2:1.

Table 9.2: 1 DM at baseline

Diabetes status at baseline	**Definiton**
DM	Patients who fulfil one of the following: - investigator-reported medical history of diabetes on the medical history page OR - any pre-treatment HbA1c value >= 6.5%
Pre-DM	Patient who fulfil all of the following: - investigator reported 'no' for the medical history of diabetes on the medical history page AND - a pre-treatment HbA1c value of >=5.7% and all pre-treatment HbA1c <6.5%
Missing	Patients who fulfil one of the following: - missing assessment of history of diabetes on the medical history page AND all pre-treatment HbA1c <6.5% OR - no pre-treatment measurement of HbA1c AND investigator reported 'no' for the medical history of diabetes on the medical history page OR - no pre-treatment measurement of HbA1c and missing assessment of history of diabetes on the medical history page
Non-DM (excl pre-DM)	Patients not meeting the above criteria

9.3 DERIVATION OF RESPONDER THRESHOLD FOR KCCQ-TSS AND CHQ-SAS DYSPNEA

A minimal clinically meaningful within patient-change in scores (i.e. responder threshold) will be estimated from the trial population (both treatment groups pooled) for KCCQ-TSS and CHQ-SAS Dyspnea using the below algorithm. All observed values will be considered in this analysis without imputation of missing values (OC-AD).

Anchor

The Patient Global Impression of Severity (PGI-S) of Dyspnea and PGI-S of Heart Failure Symptoms will each be used as the primary anchor for the CHQ-SAS Dyspnea scale and KCCQ-TSS, respectively. The Patient Global Impression of Change (PGI-C) of Dyspnea and PGI-C of Heart Failure Symptoms will be considered as secondary anchor scales for the CHQ-SAS Dyspnea scale and KCCQ-TSS, respectively.

A minimal clinically meaningful improvement for each PGI-S scale will be considered to be at least a one category improvement in change from baseline in PGI-S at week 12.

A minimal clinically meaningful improvement for each PGI-C scale will be considered to be at least “a little better” in PGI-C at week 12.

As an ancillary consideration, the distance walked in 6 minutes will be evaluated as a clinical indicator of patient improvement as well. Patients who improve at least 30 meters from baseline in distance walked in 6 minutes will be considered to be clinical responders [R17-1227] (5).

Responder Threshold

To determine the threshold of the Patient Reported Outcome (PRO) measure that best identifies patients with minimal clinically meaningful difference as specified above (responders / non-responders), the following procedure will be applied:

A receiver operating characteristic (ROC) curve of change from baseline of the PRO measure will be derived from the logistic regression model with the binary anchor (at least one category improvement in PGI-S Y/N) as dependent variable.

The mean of the cut-offs closest to the balance point (i.e. where sensitivity equals specificity) will be used as the responder threshold for each respective scale’s PRO score.

For further analysis of the key secondary endpoint as defined in Section 7.5.1.4 a patient will be considered a responder if the PRO score is greater than or equal to the response threshold determined by the anchor of one category improvement in corresponding PGI-S questionnaire.

Graphical presentations

The ROC curve will be used to illustrate the ability of the PRO score to discriminate between responders and non-responders. The area under the ROC curve (AUC) will be reported in addition.

Descriptive empirical cumulative distribution functions (ECDF) will also be plotted by each anchors categories for each PRO.

Supportive Analyses

Distribution-based analyses will be performed for the CHQ-SAS Dyspnea scale and the KCCQ-TSS on data of change from baseline to week 12 to support the responder definition as defined above.

One standard error of measurement (SEM) as well as 0.2*SD and 0.5*SD of the CHQ-SAS Dyspnea scale and the KCCQ-TSS baseline measurement will be used as benchmarks [R07-1182](6).

The SEM is defined as $SD * \sqrt{(1-r)}$ with reliability coefficient r=0.88 for the KCCQ-TSS (Cronbach's alpha in [R17-2666] (7)) and r=0.91 for the CHQ-SAS Dyspnea (Cronbach's alpha in [R17-2917] (8)).
For patients who achieved a one-category improvement in PGI-S Dyspnea at week 12 from baseline, the descriptive median (with inter-quartile range) of CHQ-SAS Dyspnea will be displayed by baseline PGI-S Dyspnea category. The table will be repeated separately for one-category deterioration, two-category improvement, two category deterioration, and no change in PGI-S dyspnea.

For patients who achieved a one-category improvement in PGI-S in heart failure symptoms at week 12 from baseline, the median (with inter-quartile range) of KCCQ-TSS will be displayed by baseline PGI-S in heart failure symptoms category. The table will be repeated for one-category deterioration, two-category improvement, two category deterioration, and no change in PGI-S in heart failure symptoms.

9.4 DERIVATION OF KCCQ AND CHQ-SAS DOMAINS

The following algorithm will be used to score the KCCQ-TSS and the CHQ-SAS Dyspnea.

9.4.1 Kansas City Cardiomyopathy Questionnaire (KCCQ)

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

1. Physical Limitation

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

- Extremely limited = 1

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

If at least three of Questions 1a-f are not missing, then the Physical Limitation Score is calculated as follows:

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

2. Symptom Stability

Code the response to Question 2 as follows:

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

If Question 2 is not missing, then the Symptom Stability Score is calculated as follows:

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

3. Symptom Frequency

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

Question 3

- Every morning = 1
- 3 or more times a week but not every day = 2
- 1-2 times a week = 3
- Less than once a week = 4

- Never over the past 2 weeks = 5

Questions 5 and 7

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

Question 9

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

If at least two of Questions 3, 5, 7 and 9 are not missing, then the Symptom Frequency Score is calculated as follows:

S3 = [(Question 3) – 1]/4

S5 = [(Question 5) – 1]/6

S7 = [(Question 7) – 1]/6

S9 = [(Question 9) – 1]/4

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

4. Symptom Burden

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

- Extremely bothersome = 1

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

If at least one of Questions 4, 6 and 8 is not missing, then the Symptom Burden Score is defined as follows:

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

5. Total Symptom Score

The Total Symptom Score is defined as the mean of the following available summary scores:

Symptom Frequency Score and Symptom Burden Score

6. Self-Efficacy

Code responses to Questions 10 and 11 as follows:

Question 10

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

Question 11

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

If at least one of Questions 10 and 11 is not missing, then the Self-Efficacy Score is calculated as:

Self-Efficacy Score = 100*[(mean of Questions 10 and 11 actually answered) – 1]/4

7. Quality of Life

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

Question 12

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

Question 13

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

Question 14

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

If at least one of Questions 12, 13 and 14 is not missing, then the Quality of Life Score is calculated as:

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

8. Social Limitation

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

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

If at least two of Questions 15a-d are not missing, then the Social Limitation Score is calculated as:

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

9. Overall Summary Score

The Overall Summary Score is defined as the mean of the following available summary scores:

Physical Limitation Score, Total Symptom Score, Quality of Life Score and Social Limitation Score

10. Clinical Summary Score

The Clinical Summary Score is defined as the mean of the following available summary scores:

Physical Limitation Score and Total Symptom Score

Note: references to “means of questions actually answered” imply the following.

If there are n questions in a scale, and the subject must answer m to score the scale, but the subject answers only n-i, where n-i >= m, calculate the mean of those questions as

(sum of the responses to those n-i questions) / (n-i)

not

(sum of the responses to those n-i questions) / n

The change from baseline to week 12 in KCCQ-TSS will be calculated as the difference of the score at follow-up (week 12) and at baseline.

9.4.2 Chronic Heart Failure Questionnaire (CHQ) Self-Administered Standardised format (SAS)

The CHQ-SAS dyspnea score is defined as the mean of Questions 1-5 within the CHQ-SAS.

The CHQ-SAS fatigue score is defined as the mean of question 7, 9, 12, 14.

The CHQ-SAS emotional function score is defined as the mean of questions 6, 8, 10, 11, 13, 15, 16.

The scores for each question of each dimension are added together and divided by the number of questions. Only items that are answered will be scored. That is items that are scored as “not done” or missing will not be included in the summary (average) score. The average score of the completed items will be reported.

The change in the score of each domain will be calculated as the difference of the mean score at follow-up (week 12) and baseline.

9.5 DERIVATION OF CLINICAL CONGESTION SCORE

Following the scoring algorithm in Ambrosy et al. the clinical congestion score is defined a summary score of the 3 items jugular venous distension (JVD), orthopnea and edema [R19-0744 (9)].

Code responses to orthopnea as follows:

- None = 0
- Seldom = 1
- Frequent = 2
- Continuous = 3

Code responses to JVD as follows:

- ‘≤ 6’ = 0
- ‘6<JVD<10’ = 1

- '10≤JVD<15' = 2
- '≥15' = 3

Code responses to Edema as follows:

- absent/trace=0
- Slight=1
- Moderate=2
- Marked=3

If at least two of the three items JVD, orthopnea and edema are not missing, then the Clinical Congestion Score is calculated as:

Clinical Congestion Score = (mean of items JVD, orthopnea and edema actually answered)*3

10. HISTORY TABLE

This is a revised TSAP including the following modifications to the final TSAP

Table 10: 1 History table

Version	**Date (DD-MMM-YY)**	**Author**	**Sections changed**	**Brief description of change**
FINAL	**5-APR-19**		None	This is the final TSAP
REVISED	**21-OCT-19**		Section 6.3 / Section 7.4 / Section 7.5	Definition of Senstivity analysis set 2 and Sensitivity analysis set 3 Reason for change: For a subset of patients, on-treatment urine dipstick test results for urine glucose had been transferred from the central laboratory to the sites before DBL. A positive urine glucose test can be suggestive of treatment allocation. Therefore, an additional sensitivity analysis is performed based on the subset of patients who did not have an on-treatment urine glucose measurement transferred to and accessed by the sites before DBL. In patients with diabetes mellitus at baseline, a positive urine dipstick test result is of limited value with respect to bearing potential information on treatment allocation. Therefore, a second additional sensitivity analysis is performed, which excludes those non-diabetic patients for whom at least one on-treatment urine glucose measurement was transferred to and accessed by the sites before DBL.
				Reason for change: Clarification. Reason for change:

Table 10: 1 History table (cont.)

Version	Date (DD-MMM-YY)	Author	Sections changed	Brief description of change
			Section 6.6	Clarification of 6MWTD if a patient performs the visit at week 12 but does not perform the 6MWT. Clarification of handling of missing baseline questionnaires for WCI analyses. Clarification of handling of SAE with outcome "recovered with sequelae" leading to an amputation. Reason for change: Clarifications consistent with protocol.
			Section 6.7	Clarification of time windows for laboratory values at day of treatment start, repeated values, and time windows for sensitivity analyses using primary analysis model. Reason for Change: For laboratory values, clock time is recorded and therefore can be taken into account. All models using Wilcoxon will use the same time window
			Section 7.8	AEs will be coded using version 22.0 of the MedDRA coding dictionary. Reason for Change: Clarification
			Section 8	Update of References Reason for change: References have been updated to the current versions of BI How to guides
				Reason for change: Clarification