Statistical Analysis Plan	
Study Code	D169CC00001
Edition Number	1.0
Date	27/August/2018

An International, Double-blind, Randomised, Placebo-Controlled Phase III Study to Evaluate the Effect of Dapagliflozin on Reducing CV Death or Worsening Heart Failure in Patients with Heart Failure with Preserved Ejection Fraction (HFpEF)

DELIVER - Dapagliflozin Evaluation to improve the LIVEs of patients with pReserved ejection fraction heart failure

An International, Double-blind, Randomised, Placebo-Controlled Phase III Study to Evaluate the Effect of Dapagliflozin on Reducing CV Death or Worsening Heart Failure in Patients with Heart Failure with Preserved Ejection Fraction (HFpEF)

DELIVER - Dapagliflozin Evaluation to improve the LIVEs of patients with pReserved ejection fraction heart failure

Study Statistician [redacted]

[redacted]

Date

An International, Double-blind, Randomised, Placebo-Controlled Phase III Study to Evaluate the Effect of Dapagliflozin on Reducing CV Death or Worsening Heart Failure in Patients with Heart Failure with Preserved Ejection Fraction (HFpEF)

DELIVER - Dapagliflozin Evaluation to improve the LIVEs of patients with pReserved ejection fraction heart failure

Biometrics Team Leader [redacted]

Date

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
AE	Adverse event
ATC	Anatomical Therapeutic Chemical
BP	Blood pressure
CDF	Cumulative distribution function
CEA	Clinical event adjudication
CKD-EPI	Chronic kidney disease epidemiology collaboration equation
CMH	Cochran-Mantel-Haenszel test
CV	Cardiovascular
DAE	Adverse events leading to discontinuation of investigational product
DKA	Diabetic Ketoacidosis
DMC	Data monitoring committee
eCRF	Electronic case report form
eGFR	Estimated glomerular filtration rate
EQ- 5D-5L	EuroQol five-dimensional five-level questionnaire
FAS	Full analysis set
HF	Heart failure
HFpEF	Heart Failure with Preserved Ejection Fraction
HR	Hazard ratio
IP	Investigational Product (dapagliflozin or matching placebo)
ITT	Intention to treat
IxRS	Interactive Voice/Web Response System
KCCQ	Kansas City Cardiomyopathy Questionnaire
KM	Kaplan-Meier
LTFU	Lost to follow-up
LVEF	Left ventricular ejection fraction
MAR	Missing at random
MedDRA	Medical dictionary for regulatory activities
NYHA	New York Heart Association
PACD	Primary analysis censoring date
PGIS	Patient global impression of severity
PT	MedDRA preferred term
PTDV	Premature treatment discontinuation visit

Abbreviation or special term	Explanation
AE	Adverse event
ATC	Anatomical Therapeutic Chemical
SAE	Serious adverse event
SCV	Study Closure Visit
SOC	MedDRA system organ class
T2D	Type 2 diabetes
TSS	KCCQ total symptom score
WoC	Withdrawal of consent

AMENDMENT HISTORY

Date	Brief description of change
<< >>	<< >>
	N/A

1 STUDY DETAILS

1.1 Study objectives

1.1.1 Primary objective

Primary objective:	Endpoint/variable:
To determine whether dapagliflozin is superior to placebo, when added to standard of care, in reducing the composite of CV death and HF events (hospitalisation for HF or urgent HF visit) in patients with HF and preserved systolic function.	Time to the first occurrence of any of the components of this composite: 1. CV death 2. Hospitalisation for HF 3. Urgent HF visit (e.g., emergency department or outpatient visit)

1.1.2 Secondary objectives

Secondary objective:	**Endpoint/variable:**
To determine whether dapagliflozin is superior to placebo in reducing the total number of recurrent HF hospitalisations and CV death	Total number of (first and recurrent) hospitalisations for HF and CV death
To determine whether dapagliflozin is superior to placebo in improving Patient Reported Outcomes measured by KCCQ	Change from baseline in the total symptom score (TSS) of the KCCQ at 8 months
To determine whether dapagliflozin is superior to placebo in reducing the proportion of patients with worsened NYHA class	Proportion of patients with worsened NYHA class from baseline to 8 months
To determine whether dapagliflozin is superior to placebo in reducing all-cause mortality	Time to the occurrence of death from any cause

1.1.3 Safety objectives

Safety Objective:	**Outcome Measure :**
To evaluate the safety and tolerability of dapagliflozin compared to placebo in patients with HFpEF	• Serious adverse events (SAEs) • Adverse events leading to treatment discontinuation (DAEs) • Amputations, adverse events (AEs) leading to amputation and potential risk factor AEs for amputations affecting lower limbs

1.1.4 Exploratory objectives

Exploratory Objective:	Endpoint/Variable:
To determine whether dapagliflozin is superior to placebo in reducing all-cause hospitalisation	Time to the first occurrence of hospitalisation from any cause
To compare the effect of dapagliflozin versus placebo on health status assessed by EuroQol five-dimensional five-level questionnaire (EQ- 5D-5L) to support health economic analysis and health technology assessment	Changes in health status measured by EQ-5D-5L
To compare the effect of dapagliflozin versus placebo on health status assessed by Patient global impression of severity (PGIS) questionnaires	Changes in health status measured by PGIS
To determine whether dapagliflozin compared with placebo will have an effect on systolic BP	Change in systolic BP from baseline
To determine whether dapagliflozin compared with placebo will have an effect on body weight	Change in body weight from baseline
To determine whether dapagliflozin compared with placebo will have an effect on eGFR	Change in eGFR from baseline
To explore whether dapagliflozin compared to placebo improves KCCQ summary scores, subscores of TSS (Symptom frequency and symptom burden) and domains	Change in Clinical summary score, TSS sub-scores, Overall summary score, QoL score
To collect and store blood samples for future exploratory genetic research	Not applicable. Results will be reported separately

1.2 Definitions

1.2.1 Primary analysis censoring date

The executive committee and AstraZeneca will monitor the accrual of endpoint events and when appropriate define the primary analysis censoring date (PACD) at which time at least the pre-defined target number of 844 events for the primary composite endpoint is expected to have occurred. The study sites will be instructed to plan for study closure visits to be performed after PACD.

Analyses of efficacy endpoint events will include events with onset on or prior to PACD. Event free patients who have not been prematurely censored due to incomplete information (see Section 3.1) will be censored at PACD. HF events and deaths with onset after PACD will also be adjudicated.

1.2.2 Withdrawal of consent

Withdrawal of consent (WoC) should only occur if the patient has received appropriate information about options for modified study follow-up and does not agree to any kind of further assessment or follow-up. Information regarding vital status (dead or alive) at the end of the study will be collected from public sources, to be included in the analysis of death from any cause as a sole outcome and in patient disposition summaries.

1.2.3 Discontinuation from study drug

Discontinuation from study drug does not mean discontinuation from study follow-up or WoC. Patients who discontinue from study drug should continue study visits according to plan until study closure. If the patient does not agree to this approach, modified follow-up capturing the essential information for the objectives of the study should be arranged. Data will be included in the ITT analyses irrespective of whether the event occurred before or following discontinuation of study drug.

1.2.4 Vital Status

Known vital status at the end of the study will be defined when the patient is dead or has date last know alive on or after the PACD.

For patients who have withdrawn consent, the investigator will attempt to collect vital status from publicly available sources at study closure in compliance with local privacy laws/practices.

1.2.5 Lost to follow-up

The term lost to follow-up (LTFU) will be limited to patients with unknown vital status at the end of the study as defined in section 1.2.4. Other measures will be used to describe completeness of follow-up of the primary endpoint (section 4.1.5)

1.3 Study design

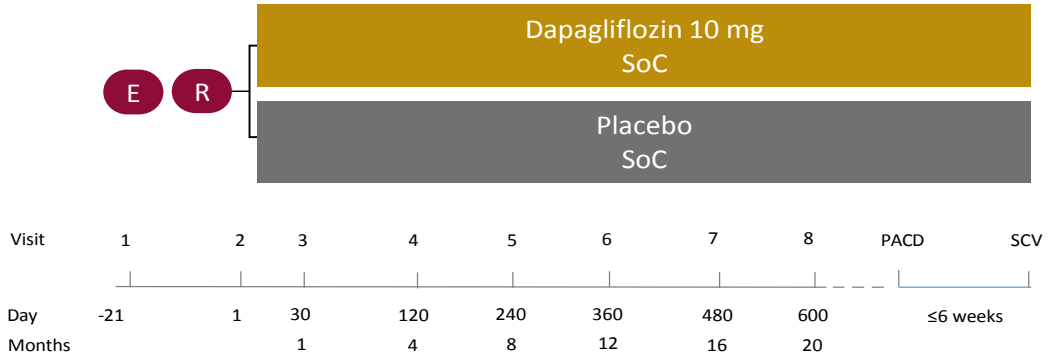
This is an international, multicentre, parallel-group, event-driven, randomised, double-blind, placebo-controlled study in patients with heart failure with preserved ejection fraction (HFpEF), evaluating the effect of dapagliflozin 10 mg versus placebo, given once daily in addition to background regional standard of care therapy, including treatments to control co-morbidities, in reducing the composite of CV death or heart failure events.

HFpEF is defined for the purposes of this study as left ventricular ejection fraction (LVEF) >40% and evidence of structural heart disease. Adult patients with HFpEF, aged ≥40 years and with NYHA class II-IV will be randomised in a 1:1 ratio to receive either dapagliflozin 10 mg or placebo once daily. A proportion of patients, here denoted as the subacute group, will be randomised during hospitalisation for heart failure or within 21 days of discharge from hospitalisation for heart failure.

It is estimated that approximately 8000 patients at approximately 400-500 sites in 20-25 countries will be enrolled to reach the target of approximately 4700 randomised patients.

In this event driven trial, study closure procedures will be initiated when the predetermined number of primary endpoints are predicted to have occurred (n=844), i.e. the PACD (section 1.2.1 and Figure 1). Patients should be scheduled for a Study Closure Visit (SCV) within 6 weeks of the PACD. The anticipated total study duration is approximately 33 months dependent on randomisation rate and event rate. The number of patients randomised, the study duration, or both, may be changed if the randomisation rate or the event rate is different than anticipated.

Figure 1 Study design



E=Enrolment; R=Randomization; SoC= Standard of Care; PACD=Primary Analysis Censoring Date; SCV=Study Closure Visit; FU=Follow Up

1.3.1 Randomisation

Patients will be randomised 1:1 to either dapagliflozin 10 mg or placebo once daily. The treatment allocation in this study will be double-blind. Randomisation will be stratified by Type 2 diabetes (T2D) status at randomisation (2 levels: with T2D; without T2D). For the purpose of stratification, T2D is defined as established diagnosis of T2D or HbA1c $\geq$ 6.5% (48 mmol/mol) at enrolment (visit 1; single measure) central laboratory test.

Randomisation will be performed in balanced blocks of fixed size. The randomisation codes will be computer generated and loaded into the IxRS (Interactive Voice/Web Response System) database.

The number of randomised patients with T2D will be monitored in order to ensure a minimum of 30% patients in each group of patients with and without T2D. Randomisation may be capped (i.e., no more patients can be randomised in a specific sub-population) if the pre-determined limit is reached.

Randomisation of patients based on geographic region will be monitored to ensure global representation. LVEF value, NYHA class, subacute/non-subacute group, and atrial fibrillation status at visit 1 may be capped in IxRS to avoid over- or under-representation of these patient subgroups.

1.4 Number of subjects

The primary objective of the study is to determine the superiority of dapagliflozin versus placebo added to standard of care in reducing the composite of CV death and heart failure events (hospitalisation for HF or urgent HF visit). Assuming a true hazard ratio (HR) of 0.80 between dapagliflozin and placebo, using a two-sided alpha of 5%, 844 primary endpoint events will provide a statistical power of 90% for the test of the primary composite endpoint. This is based on an overall 1:1 allocation between dapagliflozin and placebo.

The HR was chosen as a conservative assumption based on the observed HR 0.72 (95% confidence interval 0.50-1.04) for the composite of HF hospitalisation and CV death in patients with HF at baseline in the EMPA-REG OUTCOME trial (Fitchett et al 2016) and HR 0.61 (0.46-0.80) for patients with history of HF in the CANVAS program (Rådholm et al 2018) considering that these were post-hoc analyses in subgroups with limited documentation of baseline HF diagnosis, not characterised by ejection fraction.

The event rate assumptions are based on sub analyses of the TOPCAT and I-PRESERVE studies by geographic region, NT-proBNP levels, prior hospitalisation for HF, and T2D status (Pfeffer et al 2015, Kristensen et al 2015 Kristensen et al 2017). The sample size calculation builds on the assumption of an annual event rate of 9% in the placebo group for the majority of prevalent HFpEF patients, importantly all with NT-proBNP ≥300 pg/ml by inclusion criterion. Additionally, a subgroup of patients due to be discharged or recently discharged from a HF hospitalisation (here denoted ‘subacute’ patients) with a higher event rate is planned to be included. Assuming 20% of patients from the subacute category with an annual event rate of 24% during the first year and 9% thereafter for the remainder of the study, (corresponding to an annualised rate of approximately 17% for sub-acute patients), approximately 4700 patients are estimated to provide the required number of 844 patients with a primary event during an anticipated recruitment period of 18 months and a minimal follow-up period of 15 months (total study duration 33 months, average follow-up 24 months). The study is event driven and the number of patients or duration may change if the event rate is lower or higher than anticipated.

In addition, the expected number of patients who will be lost to follow-up is expected to be small; hence, these are not considered in the determination of the sample size.

2 ANALYSIS SETS

2.1 Definition of analysis sets

2.1.1 Full analysis set

All patients who have been randomised to study treatment will be included in the full analysis set (FAS) irrespective of their protocol adherence and continued participation in the study. Patients will be analysed according to their randomised IP assignment, irrespective of the treatment actually received. The FAS will be considered the primary analysis set for the intention to treat (ITT) analysis of primary and secondary variables and for the exploratory efficacy variables.

2.1.2 Safety analysis set

All randomised patients who received at least 1 dose of randomised treatment will be included in the safety analysis set. Patients will be analysed according to the treatment actually received. For any patients given incorrect treatment, ie randomised to one of the treatment groups, but actually given the other treatment, the treatment group will be allocated as follows: Patients who got both incorrect and correct treatment will be analysed according to their randomised treatment. Patients who got only the incorrect treatment will be analysed according to that treatment.

The Safety analysis set will be considered the primary analysis set for all safety variables.

2.2 Violations and deviations

The important protocol deviations listed below will be summarised by randomised treatment group

- Patients who were randomised but did not meet inclusion and exclusion criteria
- Patients who received the wrong study treatment at any time during the study.
- Patients who received prohibited concomitant medication, which for this study is limited to open label SGLT2 inhibitors taken in combination with IP.

As the primary analysis is ITT analysis, protocol deviation will not imply exclusion from the primary analysis.

3 PRIMARY AND SECONDARY VARIABLES

Deaths and potential HF events will be adjudicated by an independent clinical event adjudication (CEA) committee. The CEA committee members will not have access to the

treatment codes for any patient. The CEA procedures and event definitions will be described in the CEA charter according to the CDISC definitions (Hicks et al 2018).

Only HF hospitalisations and urgent HF visits confirmed by the CEA will be used in the analysis of the primary and secondary endpoints and their components.

The primary analyses of the endpoints concerning CV deaths, either as a component of a composite or on its own, will include deaths adjudicated as CV cause. Deaths adjudicated as 'cause undetermined' will be considered as non-CV deaths in these analyses.

Adjudicated events occurring from randomisation until WoC or PACD will be included in the analysis of primary and first secondary endpoint. The analysis of all-cause death as a sole outcome will in addition include any deaths (not adjudicated) after WoC, but on or before PACD.

3.1 Primary variable

The primary efficacy variable is time from randomisation to the first occurrence of any event in the composite of CV death, hospitalisation for HF or an urgent HF visit.

Patients who did not have an adjudicated primary endpoint event on or prior to PACD will be censored at the earliest of date of WoC or non-CV death when applicable, and otherwise at the date of the last clinical event assessment or the PACD, whichever occurs first. It is expected that patients alive and under study follow-up will have a clinical event assessment at their SCV after PACD. Last clinical event assessment is defined as the last date when the event assessment question for a potential heart failure event was completed on the eCRF event assessment page.

In analysis of the individual components hospitalisation for HF and urgent HF visit, to examine their contribution to the composite endpoint, date of death from any cause will be an additional point of censoring.

For analysis of time to first event, data will be expressed as two variables:

- A binary variable indicating whether the event in question occurred, or the patient was censored.
- An integer variable for the number of days from randomisation to the first occurrence of an event (start date of the event – randomisation date + 1), or for event free patients, from randomisation to censoring (censoring date – randomisation date + 1).

3.2 Secondary variables

The secondary endpoints are included in a hierarchical testing sequence following the primary endpoint as described in section 4.1.3.

3.2.1 Total number of (first and recurrent) hospitalisations for HF and CV death

The first secondary endpoint is the total number of first and recurrent hospitalisations for HF and CV death, not including urgent HF visit.

For the analysis of recurrent heart failure hospitalisation and CV death, the data will be expressed in counting process style for input to the analysis as described in Section 4.2.4.14.2.4 , as follows. The time from randomisation to end of follow-up/censoring will be split into one or more interval with variables for start of interval, end of interval and a variable indicating if an event occurred at the end of each respective interval, or if the patient was censored.

Patients who did not have the endpoint will be censored by the same rule as for the primary endpoint.

3.2.2 Change from baseline at 8 months in the KCCQ total symptom score

The efficacy variable is the change from baseline at 8 months of the Kansas City Cardiomyopathy Questionnaire (KCCQ) total symptom score (TSS).

The KCCQ is a self-administered disease specific instrument for patients with HF **(**Green et al 2000, Spertus et al 2005). The KCCQ consists of 23 items measuring HF-related symptoms, physical limitations, social limitations, self-efficacy, and health-related quality of life. The TSS incorporates the symptom burden and symptom frequency domains into a single score. Scores are transformed to a range of 0-100. Higher scores represent better outcomes.

Baseline is defined as the value at randomisation visit (visit 2). Change from baseline at each post-baseline analysis time point will be calculated as the value at the corresponding post-baseline analysis time point minus the baseline value. The KCCQ is assessed by the patient at randomisation, at the visits targeted 1, 4 and 8 months following randomisation and at premature treatment discontinuation visit (PTDV) and SCV. By the ITT principle, the analysis will include all data irrespective of whether the patient has discontinued study drug.

In order to account for patients who die prior to the 8-month assessment and to accommodate non-normal distribution of KCCQ scores, a composite rank-based endpoint will be used. The values of change from baseline to 8 months in TSS of patients who survive to 8 months will be converted to ranks (across both treatment groups combined) with lower ranks attributed to worse outcomes (i.e., lower ranks corresponding to negative or smaller values of change from baseline). Patients who die prior to the 8-month assessment will be assigned the worst rank, i.e., worse than any patient surviving to 8 months. All patients deceased prior to the 8-month assessment will be assigned the same worst rank regardless of the relative timing of their

deaths. This is done to reduce the impact of treatment differences in time to CV death on the assessment of this KCCQ secondary endpoint.

3.2.3 Proportion of patients with worsened NYHA class at 8 months

The efficacy variable is the proportion of patients with worsened NYHA class from baseline to 8 months.

The NYHA classification will be evaluated by the investigator and collected in eCRF at enrolment and randomisation visits, at 1, 4 and 8 months visits, at PTDV and SCV. Baseline is defined as the value at randomisation (visit 2). The analysis will include all data irrespective of whether the patient has discontinued study drug.

For the primary analysis the data will be dichotomised into patients with worsened NYHA class at 8 months (the NYHA class is higher than baseline), including patients who died due to any cause prior to 8 months, versus other patients with improved or unchanged class compared to baseline.

3.2.4 Death from any cause

The efficacy variable is time to from randomisation to death from any cause. All deaths on or prior to PACD, including any deaths after WoC, will be included. Patients who are alive will be censored at the earliest of date last known alive and PACD.

3.3 Safety variables

The safety and tolerability of dapagliflozin in patients with HFpEF will be evaluated from serious adverse events (SAEs), adverse events leading discontinuation of IP (DAEs), adverse events(AE) leading to amputation and AEs reflecting potential risk factors for lower limb amputations (“preceding events”).

In addition to amputation, non-serious and serious events potentially placing the patient at risk for a lower limb amputation, in this document denoted “preceding events”, should also be recorded in the eCRF as AE/SAE, whether or not an amputation has taken place. Preceding events will be defined for analysis by a predefined list of preferred terms. Additional information about amputations with underlying conditions and preceding events will be collected on dedicated eCRF pages.

SAEs will be collected from time of informed consent until and including the patent’s last visit. Non-serious AEs will be collected from randomisation until and including the patient’s last visit. Collection of non-serious AEs is limited to AE leading to amputation, preceding events, AEs leading to a potential endpoint, DAEs and AEs which are the reason for interruption of study drug.

Efficacy endpoints (deaths and potential HF events) will be adjudicated. These events will be recorded as SAEs in the database, but will not be reported as SAEs to health authorities to avoid unnecessary unblinding. However, if it is determined by the CEA committee that a potential endpoint does not meet the endpoint criteria, the event will be reported to AZ patient safety data entry site and if applicable to the health authorities.

For SAEs or DAEs reported by the Investigator as Diabetic Ketoacidosis (DKA) additional information will be recorded on specific eCRF pages in addition to the AE/SAE form.

For myocardial infarctions, unstable angina and stroke additional information will be recorded on specific eCRF pages in addition to the AE/SAE form.

3.4 Laboratory values and vital signs

Blood samples will be taken for central laboratory assessment of creatinine and calculation of eGFR at enrolment visit, at the visits targeted 1, 4 months and 12 months following randomisation, then annually and at PTDV and SCV. eGFR will be calculated (in mL/min/1.73 m^2) using the CKD-EPI formula (Levey at al 2009).

Central laboratory assessment of NT-proBNP and HbA1c will be taken at visit 1.

Systolic blood pressure (SBP), diastolic blood pressure (DBP) and pulse rate will be measured at visit 1, visit 2, at 1 and 12 months visit, then annually and at PTDV and SCV.

Weight will be measured at visit 1, at the 12 months visit, then annually and at PTDV and SCV.

3.4.1 Baseline laboratory values and vital signs

In principle baseline will be defined as the last value on or prior to date of first dose of randomised study drug, or for patients who did not receive treatment, the last value on or prior to date of randomisation. Except for cases of rescreening this will be visit 1 measurement of weight, NT-proBNP, eGFR and HbA1c, and visit 2 measurement of SBP, DBP and pulse rate.

4 ANALYSIS METHODS

4.1 General principles

No multiplicity adjustment will be made to confidence intervals as they will be interpreted descriptively and used as a measure of precision. All p-values will be unadjusted. P-values for variables not included in the confirmatory testing sequence, or following a non-significant test in the sequence will be regarded as nominal.

Primary and secondary analyses of HF events and death include adjudicated events occurring on or prior to PACD.

Stratification of analyses for T2D status will be performed using the stratification values as entered in IxRS to determine the randomisation assignment.

Incomplete dates

If only the year part of a date is available (YY), then the date will be set to YY0701. If only the year and month is available (YYMM), then the date will be set to YYMM15. Additional imputation rules will be defined as appropriate to ensure that eg, dates will not be imputed as prior to randomisation, after death or start date after end date.

Study drug compliance

The percentage of study drug compliance for the overall treatment period will be derived for each patient based on pill counts as the number of pills taken (dispensed – returned), relative to the expected number of pills taken. The expected number of pills taken is defined as 1*(date of last dose – date of first dose +1), excluding days of interruption.

Study drug compliance will be presented descriptively, including mean, median, quartiles and 5% and 95% percentiles.

4.1.1 Estimand for primary and secondary outcomes

The primary and secondary event based objectives will be evaluated under the treatment policy estimand including differences in outcomes over the entire study period until PACD to reflect the effect of the initially assigned randomised study drug, irrespective of exposure to study drug, concomitant treatment as well as subsequent treatment after discontinuation of study drug. The analysis will be performed for the full analysis set including all events that occurred on or prior to PACD, including events following premature discontinuation of study drug. The time-to-first event analysis by Cox proportional hazards regression and the analysis of recurrent events (Section 4.2.4) assume that missing data is at random.

4.1.2 Hypotheses

To control the overall type I error rate at 5% two-sided, the significance level will be adjusted for interim analysis of efficacy performed by the DMC (Section 5) using the Haybittle-Peto function implemented in the software East (Copyright © Cytel Inc) . For one planned interim analysis including 67% of the target number of primary endpoints, the significance level will be 4.980%. The following null hypothesis will be tested for the primary endpoint

H0: HR [dapagliflozin:placebo] =1

versus the alternative hypothesis

H1: HR [dapagliflozin:placebo] ≠1

The secondary endpoints included in confirmatory statistical testing using a closed testing procedure (section 4.1.3) will be based on similar two-sided alternative hypotheses for the respective treatment difference.

4.1.3 Confirmatory testing procedure

A closed testing procedure including a pre-specified hierarchical ordering of the primary and secondary endpoints will be utilised. The Type I error will be controlled at an overall two-sided 5% level for multiplicity across primary and secondary endpoints and in consideration of the planned interim analysis. With one interim analysis at 67% of events the two-sided significance level in final analysis, α, will be 4.980%. Statistical significance will be assessed in the pre-specified order of the endpoints as specified in section 1.1.1 and section 1.1.2 . If the primary endpoint is significant at level α, then the first secondary endpoint, recurrent HF hospitalisations and CV death, will be tested at level α. If the first secondary endpoint is significant, then the α will be split between KCCQ total symptom score and NYHA class. If one of them is significant at level α/2, then the other can be tested at level α. If both KCCQ and NYHA class reach statistical significance, then all-cause mortality will be tested at significance level α.

If the study is stopped in the efficacy interim analysis (section 5), testing of secondary endpoints will be performed with the same testing procedure as described in this section above with a two-sided α= 0.002.

4.1.4 Presentation of time-to-event analyses

In general, summary tables of time-to-event analyses will include the number and percent of patients with event per treatment group, event rate, hazard ratio with 95% confidence interval and p-value. The event rate will be derived as the number of patients with event divided by the total duration of follow-up across all patients in the given group.

Kaplan-Meier (KM) estimates of the cumulative proportion of patients with events will be calculated and plotted per treatment group, with the number of patients at risk indicated below the plot at specific time points. The KM plots will be presented for all time to event analyses, including the individual components of the composite endpoints.

4.1.5 Vital status and follow-up of endpoints

Potential HF endpoints and deaths will be collected and adjudicated from randomisation throughout the study until and including the patient's last visit. The investigator will attempt to collect vital status (dead or alive) at the end of the study for all patients, including vital status from publicly available sources for patients who have withdrawn consent, in compliance with local privacy laws/practices.

Known vital status at the end of the study will be defined when the patient is dead or has date last know alive on or after the PACD. In patient disposition the number of patients who are dead, alive or with unknown vital status will be reported separately for patients who did/did not withdraw consent. The term lost to follow-up (LTFU) will be limited to only patients with unknown vital status.

Follow-up of the primary endpoint will be defined in terms of completion of the event assessment question for a potential HF event as described for censoring in section 3.1. Thus, a patient that is not LTFU, ie with known vital status, may have incomplete follow-up of endpoints.

Complete follow up of the primary endpoint will be defined when the patient had a primary endpoint event, died from non-CV death or had complete event assessment on or after the PACD (ie, the patient was not censored du to incomplete follow-up of endpoints).

In addition to the number and percent of patients with complete follow-up, the proportion of total patient time with complete follow-up will be reported per treatment group.
Patient time with complete follow-up will be defined as time from randomisation until the earliest of first primary endpoint event, death, WoC, censoring where last complete event assessment is prior to PACD or PACD. The denominator, representing maximum complete follow-up, will be the time to first primary endpoint event, death or PACD.

4.2 Analysis methods

4.2.1 Demographics and baseline characteristics

Demographic and baseline characteristics, including medical history, will be summarized, using frequency distributions and summary statistics based on the FAS, for each treatment group as well as for all patients combined. No statistical test will be performed for comparison of any baseline measurement among treatment groups.

4.2.2 Concomitant and baseline medication

Baseline medication is defined as medication with at least one dose taken before date of randomisation and with no stop date before date of randomisation.

Concomitant medication is defined as medications taken post randomisation, irrespective of study drug.

The frequency of baseline and concomitant medication will be presented for the FAS per ATC class and treatment group. Summaries of prohibited medication, in this study limited to SGLT2 inhibitor taken while on IP, will be presented.

4.2.3 Analysis of the primary efficacy variable

The primary variable is the time to first event included in the primary composite endpoint. The primary analysis will be based on the ITT principle using the FAS, including events with onset on or prior to PACD, adjudicated and confirmed by the CEA committee.

In the analysis of the primary composite endpoint, treatments (dapagliflozin versus placebo) will be compared using a Cox proportional hazards model with a factor for treatment group, stratified by T2D status at randomisation. The analysis will use WoC, non-CV death, last clinical event assessment and PACD for censoring of patients without any primary event as described in Section 3.1. The Efron method for ties and p-value based on the score statistic will be used. Event rates, p-value, HR, and 95% confidence interval will be reported.

The contribution of each component of the primary composite endpoint to the overall treatment effect will be examined. In the analysis of the components, all first event of the given type will be included irrespective of any preceding non-fatal composite event of a different type. Consequently, the sum of the number of patients with individual events in the component analysis will be larger than the number of patients with a composite outcome. Methods similar to those described for the primary analysis will be used to separately analyze the time from randomisation to the first occurrence of each component of the primary composite endpoint.

Kaplan-Meier estimates of the cumulative proportion of patients with event will be calculated and plotted, for the composite endpoint and for the individual components.

4.2.3.1 Subgroup analysis of the primary endpoint

Exploratory subgroup analyses of the primary composite endpoint will be performed for the characteristics listed in Table 1. Cox proportional hazard model stratified for T2D with factors for treatment group, the subgroup variable and the interaction between treatment and subgroup will be used to examine treatment effects within relevant subgroups separately. In addition to the number and percent of patients with event, event rate estimate, HR with 95% confidence interval and p-value for each subgroup, the interaction p-value will be presented. HRs with confidence interval will be presented in a forest plot, also including the event rate and interaction p-value. The p-values for the subgroup analyses and interaction will not be adjusted for multiple comparisons as the tests are exploratory and will be interpreted descriptively.

Table 1 Characteristics and categories for sub group analysis of the primary endpoint

Characteristic	Categories
Age (years)	<= median, > median

Sex	Male, female
Race	White, Black or African, Asian, Other
Geographic region	Asia (China, Japan, Taiwan, Vietnam) Europe and Saudi Arabia (Belgium, Bulgaria, Czech Republic, France, Hungary, Netherlands, Poland, Romania, Russia, Saudi Arabia, Spain) North America (Canada, US) Latin America (Argentina, Brazil, Mexico, Peru)
NYHA class at enrolment	II, III/IV
LVEF at enrollment (%)	41-49, ≥50
NT-proBNP at enrollment (pg/ml)	<= median, > median
Randomised during hospitalisation for HF or within 21 days of discharge.	Yes, No
eGFR at enrolment (ml/min/1.73m^2)	<60, ≥60
BMI at enrolment (kg/m2)	<30, ≥30
Type 2 diabetes at enrolment*	Yes, No
Systolic blood pressure at randomisation	<= median, > median
Atrial fibrillation or flutter at enrolment ECG	Yes, No

* The subgroup analysis by T2D status will be based on eCRF medical history record and exclude T2D as a stratification factor from the model.

The subgroup analyses will be repeated for the CV death component of the primary composite endpoint.

4.2.3.2 Sensitivity analysis of the primary endpoint

Undetermined cause of death

A sensitivity analysis of the primary analysis where deaths adjudicated as 'undetermined' cause are considered as CV deaths and included as endpoint events will be performed.

Missing data and informative censoring

The time-to-event analysis using the Cox regression depends on the assumption of non-informative or ignorable censoring, corresponding to the missing-at-random assumption. The missing data in this context are patients who are prematurely censored due to WoC, LTFU or otherwise incomplete follow-up of endpoints. The amount of missing data will be described eg, in terms of the number of patients and patient time with incomplete follow-up as described in Section 4.1.5.

Patient retention and follow-up are at the forefront of study planning and conduct, and the amount of incomplete follow-up is expected to be small. To assess the impact of missing data and the robustness of the results with regard to the assumption of non-informative censoring, sensitivity analysis will be planned based on the evaluation of the missing follow-up and discussed in relation to the observed efficacy signal. This may include analysis where scenarios in terms of increased risk in censored patients are explored to identify a 'tipping point' where statistical significance would be lost.

4.2.4 Analysis of the secondary efficacy variables

4.2.4.1 Analysis of recurrent HF events and CV death

The composite outcome of recurrent HF hospitalizations andr CV death will be analysed by the semi-parametric proportional rates model (Lin et al 2000; known as the LWYY method) to test the treatment effect and to quantify the treatment difference in terms of the rate ratio with 95% confidence interval and p-value.

In addition, the two components in the composite endpoint (total HF hospitalizations and CV death) will be analysed separately to quantify the respective treatment effects and check the consistency between the composite and the components. For the analysis of total HF hospitalizations component, occurrence of CV death can be regarded as semi-competing risk (informative censoring) and may introduce a bias in the treatment effect estimate for HF hospitalizations (dilution of effect size if the drug has a positive effect on both components). To address this concern and to account for the correlation between the two components, the joint modelling (frailty model) approach (Rogers et al 2016) will be used for the component analyses. Non-parametric estimates of HF hospitalization rates over time allowing for death as terminal event will be provided as well (Ghosh and Lin 2000).

4.2.4.2 Analysis of change from baseline to 8 months in the KCCQ total symptom score

Hypothesis testing

The composite rank-based endpoint representing the patients' vital status at 8 months and the change from baseline to 8 months in TSS in surviving patients, as defined in Section 3.2.2, will be analysed using the rank ANCOVA method (Stokes et al 2012) to test the null hypothesis of no differences in the distributions of ranked outcomes between the two treatment groups. Analysis will be stratified by T2D status at randomisation, and adjusted for the baseline TSS value as follows.

First the change from baseline to 8 months in TSS and vital status at 8 months, as well as values of the baseline TSS covariate will be transformed to standardized ranks within each T2D randomization stratum, using fractional ranks and mean method for ties. Ranking for the composite endpoint will be done so that patients who died prior to the 8-month assessment are assigned the worst ranks within each stratum. This will be implemented by assigning a temporary value of -101 to subjects who died prior to 8-month assessment before deriving fractional ranks. Then, separate regression models will be fit to the ranked data for each randomization stratum using a regression model for the ranked composite variable as dependent variable, adjusting for the ranked baseline covariate. Residuals from this regression model will be captured for further testing of differences between treatment groups. The Cochran-Mantel-Haenszel (CMH) test, stratified for the T2D status at randomization, using the values of the residuals as scores will be used to compare treatment groups.

The p-value from the CMH test of treatment effect at 8 months will be the used for the confirmatory testing of the secondary endpoint in the multiple testing procedure described in section 4.1.3.

Estimation of treatment effect

Win ratio:

For a summary statistic that uses the same ranking as that used in the hypothesis test, but has a clinical interpretation, the win ratio (WR) and the corresponding 95% confidence interval (Wang and Pocock 2016) will be reported. It is noted that the WR differs from the statistic used for hypothesis testing, so that exact consistency is not expected as between these two analyses, e.g. on rare occasions, the confidence interval for WR could exclude unity while the pre-planned hypothesis test could be non-significant, or the hypothesis test could be significant with the confidence interval for WR including unity. Formal inference for the superiority of the treatment over control will be made only from the preplanned hypothesis test.

The win ratio represents the odds of having a more favourable outcome versus a less favourable outcome when assigned to the dapagliflozin 10 mg treatment group as opposed to

placebo. More specifically, each patient in the dapagliflozin group is compared with each patient in the placebo group and each pair is labelled as "winner", "loser", or "tie", depending on whether the patient on dapagliflozin has a more favourable, less favourable, or the same outcome, respectively, with respect to the composite ranked endpoint compared to the patient on placebo. Win ratio is defined as the ratio of the number of "winner" pairs to the number of "looser" pairs for the dapagliflozin arm. If the estimated win ratio is greater than 1 then the treatment effect is in favour of dapagliflozin.

The win ratio statistic adjusted for the randomization stratification factor and baseline TSS will be obtained using the methodology in (Kawaguchi et al 2011) for the stratified Mann-Whitney estimators for the comparison of two treatments with randomization based covariance adjustment. The win ratio statistic will be calculated as Mann-Whitney odds, i.e., WR=MW/((1-MW)), where MW is the adjusted Mann-Whitney estimate. The 95% confidence interval for the win ratio will be obtained as

$$\exp\{\ln(WR) \pm 1.96 * SE(\ln(WR))\}$$

where the standard error of the logarithm of WR is obtained as

$$SE(\ln(WR)) = SE(MW)/(MW * (1 - MW))$$

and the $SE(MW)$ is the standard error of the adjusted Mann-Whitney estimate. The adjusted Mann-Whitney estimates and its standard error will be obtained using the "sanon" package in R (Kawaguchi and Koch 2015).

Responder analysis:

Number and percentage of patients in each treatment group will be summarized across the following categories:

5 point improvement from baseline to 8 months in TSS vs no significant improvement:
- Change from baseline in TSS ≥ 5 points, vs
- Death prior to the 8 months assessment or change from baseline in TSS < 5 points.

5 point deterioration from baseline to 8 months in TSS vs no significant deterioration:
- Death prior to the 8 months assessment or change from baseline in TSS ≤ -5 points, vs
- Change from baseline to 8 months in TSS > -5 points.

Cumulative distribution function (CDF) plots will be presented by treatment group to summarize the distribution of change from baseline to 8 months in TSS values, where patients who die prior to the 8-month assessment will be represented with the value of -101 (a value below the worst possible change from baseline).

Handling of missing data

The number of patients with missing vital status at 8 months is expected to be negligible. If some patients are LTFU or patients who withdrew consent have unknown vital status, the main analysis will be done with these patients assigned the worst ranks (same as deaths).

In the context of analysing the composite ranked endpoint as described above, missing data may arise when patients miss the 8-month KCCQ assessment while remaining in the study during the 8-month assessment window, or when patients withdraw consent from the study prior to 8 months. If a patient is known to have died prior to the 8-month assessment, the patient is considered to have a non-missing composite outcome and will be handled as described above (assigned the worst rank). Otherwise, patients who are alive at 8 months and have missing baseline or 8-month KCCQ assessments will have their missing TSS imputed using the multiple imputation (MI) methodology as follows.

Missing TSS values at baseline or at 8 months will be imputed under the Missing at Random (MAR) assumption. The imputation will be done using a predictive mean matching multiple imputation model and a method of Fully Conditional Specification as implemented in the SAS Procedure MI (FCS statement). The predictive mean matching method ensures that the imputed values remain in the permissible range of the TSS values. The imputation model will include the treatment group, T2D randomization stratum, TSS at baseline, month 1, 4, and 8, and three auxiliary binary variables representing occurrences of any HF events in the intervals from randomization to 1 month, from 1 to 4 months, and from 4 to 8 months, respectively. Occurrences of HF events will be determined based on the investigator-reported potential HF events. Auxiliary variables related to HF events are included in the imputation model to improve the imputation accuracy, because the occurrence of HF events is associated with quality of life assessed by KCCQ.

The number of closest observations used to sample an imputed value by the predictive mean matching method will be 5 (SAS default setting).

Each imputed dataset will be analysed using the methods described in the "Hypothesis testing" and "Estimation of treatment effect" sub-sections above. The results from multiple imputed datasets will be combined using Rubin's rule as implemented in the SAS Procedure MIANALYZE.

- In the analysis of rank ANCOVA, the CMH tests statistic used for the hypothesis test has a chi-square distribution. In order to apply Rubin's combination rule, which assumes approximate normal distribution of the statistics being combined, a normalizing Wilson-Hilferty transformation will be applied to the CMH test statistics from each imputed dataset (Ratitch et al 2013). The standardized transformed statistic will be computed as follows:

$$st_{wh\_cmh}{}^{(m)} = \frac{\sqrt[3]{\frac{cmh^{(m)}}{df}} - \left(1 - \frac{2}{9 \times df}\right)}{\sqrt[2]{\frac{2}{9 \times df}}}$$

where $cmh^{(m)}$ is the CMH statistic from the m^{th} imputed dataset and df is the number of degrees of freedom associated with the statistic (in this case equal 1). The transformed statistics are approximately normally distributed with mean of 0 and variance of 1 and can be combined using Rubin's rule.

- For the estimation of the win ratio, a combined Mann-Whitney estimate (MW) and its standard error ($SE(MW)$) will first be obtained by applying Rubin's rule to the corresponding estimates from multiple imputed datasets. Then the win ratio and its 95% confidence interval will be obtained based on the combined Mann-Whitney estimate and its standard error as previously described.

- For the summaries of number and percentage of subjects in the categories of significant improvement and deterioration from baseline as well as CDF plots, as discussed in the "Estimation of treatment effect" sub-section above, the average number and percent of subjects in each category across all multiple imputed datasets will be reported.

Supportive analyses and sensitivity analyses

The number and percent of patients who die prior to the 8-month assessment will be summarized by treatment group.

Descriptive statistics of scores and change from baseline at 1,4, and 8 months and SVC will be presented for total symptom score, overall summary score, clinical summary score and domains (Physical limitation, symptom stability, symptom frequency, symptom burden, quality of life, self efficacy and social limitation).

The testing and estimation described for change from baseline at 8 months in TSS, will be repeated in an exploratory fashion for change from baseline in TSS at 1 and 4 months, and for the overall summary score and clinical summary scores at 1, 4 and 8 months.

To assess the impact on TSS change from baseline of a treatment effect on mortality, an alternative ranking my be applied where patients who die prior to the 8 months assessment will be assigned worse ranks than any patient surviving to 8 months, but among the deceased the relative ranking will be based on their last value of change from baseline in TSS while alive.

4.2.4.3 Analysis of worsened NYHA class from baseline to 8 months

The proportion of patients with worsened (higher) NYHA class at 8 months compared to baseline, including patients who died prior to 8 months in the worsened category, versus patient with improved or unchanged NYHA class, will be analyzed by logistic regression with treatment group, baseline NYHA class and T2D status randomization as factors. The odds ratio between treatment groups and its 95% confidence interval and corresponding two-sided p-value will be presented. Frequencies of NYHA class and change from baseline as well as the odds ratio for treatment effect will be presented for all post baseline visits with scheduled NYHA class evaluation. The p-value for the test of treatment effect at 8 months will be used for the confirmatory testing of the secondary endpoint in the multiple testing procedure described in section 4.1.3.

Missing NYHA assessments will be handled with the same multiple imputation methodology as described above for the analysis of KCCQ TSS in section 4.2.4.2

To assess the impact of a treatment effect of death, a sensitivity analysis will be performed where the last NYHA assessment prior to death will be carried forward.

4.2.4.4 Analysis of all-cause mortality

The 4th secondary endpoint, time to death from any cause will be analysed using Cox regression in the same manner as the primary composite endpoint, with stratification for T2D status at randomisation. The analysis will include deaths occurring on or prior to PACD. Patients who are alive will be censored at PACD, or for any patients who are LTFU, at last date known to be alive.

4.2.5 Analysis of safety variables

Analysis set

For safety analyses, all summaries will be based on the safety analysis set (Section 2.1.2).

Exposure

The total exposure to study drug will be defined as the length of period on study drug, calculated for each patient as date of last dose – date of first dose +1.

An alternative measure where days of interruption are removed will be calculated and termed actual exposure.

Total and actual exposure will be presented descriptively.

Treatment periods

The summaries for the on-treatment period will include events with an onset date on or after first dose of randomized study drug and on or before 30 days after last dose of study drug. Additional presentations will include all events with onset on or after first dose of study drug

regardless of whether patients are on or off study treatment at the time of the event (the ‘on +off ‘ treatment period.). Patients who complete the study on study drug will discontinue treatment on the SCV. Thus there will in general be no events after completion of the study drug period, and censoring of events for on-treatment analysis affects only patients who prematurely and permanently discontinue study drug.

All summaries of AEs described in Section 4.2.5.1 to 4.2.5.4 below will be presented for the on-treatment period. Additional summaries based on the on+off treatment period will be presented for SAEs, amputations and preceding events as defined in Section 3.3.

4.2.5.1 Adverse events

Summaries of AEs will primarily be based on the on-treatment period.

In addition to SAEs, the collection of AEs that are not serious is limited to DAEs, AEs leading to interruption of IP, amputations and preceding events (see section 3.3) . Thus, summaries of AEs will be limited to these categories and general summaries of all non-serious AEs are not planned.

AEs will be classified according to MedDRA by the medical coding team at AstraZeneca data management centre, using the most current version of MedDRA.

Summaries by system organ class (SOC) and preferred term (PT) will be sorted by international order for SOC and by descending order of PT in the dapagliflozin treatment group.

No statistical tests to compare crude AE frequencies between treatment groups are planned.

A summary table of the total number and percent of patients with SAE, DAE, AE leading to temporary interruption, amputations and preceding events per treatment group will be provided.

4.2.5.2 Serious adverse events

SAEs will be presented as described below both on treatment and on+off treatment.

The number and percent of patients with SAEs will be presented by SOC, PT and treatment group. The most common SAEs will also be presented by PT only.

AEs with outcome death will be presented separately by SOC and PT.

4.2.5.3 Adverse events leading to discontinuation or interruption of IP

The number and percent of patients will be presented by SOC and PT for AEs leading discontinuation of IP and AEs leading to temporary interruption (separately for the two

categories based action taken “Drug Permanently Discontinued” and “Drug Interrupted” respectively, recorded in the CRF AE module).

4.2.5.4 Amputations and preceding events

Amputations and preceding events (see section 3.3) will be presented in summary tables including the number and percent of patients with any event in the AE category, SAE, DAE and AE leading to interruption, and tabulated with frequency by SOC and PT.

In addition to presentations of the number of patients with event, the total number of events counting multiple events per subject will be presented.

In addition to the presentation of on-treatment events, on+off presentations will be provided amputations and preceding events.

4.2.5.5 Laboratory evaluation and vital signs

Summaries of creatinine and calculated eGFR will be based on creatine samples analyzed at the central laboratory.

The result and the change from baseline of creatinine, eGFR and vital signs, will be summarized by treatment group at each visit with scheduled measurement (see section 3.4) using descriptive statistics, including n, mean, SD, median and quartiles.

4.2.6 Analysis of exploratory objectives

Time to the first occurrence of hospitalisation from any cause will be analysed with the same method as the primary endpoint, based on information on the SAE eCRF form.

Change from baseline to each scheduled assessment visit (see section 3.4) for body weight, systolic blood pressure and eGFR will be analysed with a repeated measures model. All non-missing visit data will be used, including measurements after discontinuation of study drug. The model will include terms for treatment group, visit, visit*treatment group and the baseline measurement and T2D stratification factor as covariates. The model will be used to derive a least-squares estimate of the treatment difference with 95% confidence interval and corresponding two-sided p-value. Missing data will not be imputed.

For eGFR, the model above will additionally be used to derive the “total” slopes (between randomisation and eg 1 year and 2 years respectively) and the “chronic” slopes (between a post randomization time point to eg 1 year and 2 years respectively) will be estimated via linear contrasts.

The analysis of change in KCCQ clinical summary score, overall summary score, QoL score and sub-scores is described under ‘Supportive analyses and sensitivity analyses’ in section 4.2.4.2

EQ-5D-5L derived utility score will be summarised by descriptive statistics by visit and treatment group, and will be used to support modelling in a separate health economic report.

Patient global impression of severity (PGIS) will be tabulated by visit and treatment group, and will be used in anchor based analyses to support threshold for clinically important change of KCCQ total symptom score.

5 INTERIM ANALYSES

An interim analysis is planned to be performed including approximately 67% of the target number of patients with adjudicated primary endpoint events. There will in principle be one planned interim analysis for efficacy, with the possibility of the data monitoring committee (DMC) to conduct subsequent interim analysis if they deem necessary. The significance level for final analysis will be determined by the Haybittle-Peto function based on the actual number of interim analyses, using the East software (Copyright © Cytel Inc). The interim analysis will assess superiority of dapagliflozin to placebo. The interim analysis will have a nominal two-sided alpha level of 0.2%. At the interim analysis, the primary composite endpoint will be tested first at the specified alpha level. If superiority is achieved for the primary endpoint, then the superiority of dapagliflozin to placebo on CV deaths will be tested at a two-sided level of 0.2%. If CV death is significant, then an action is triggered whereby the DMC will evaluate the totality of the efficacy data and safety data, to determine if benefit is unequivocal and overwhelming such that the DMC recommends ending the study.

If the interim analysis leads to a decision to terminate the study early based on pre-defined stopping guidelines, the interim analysis database will become the basis of statistical inference for the primary endpoint and CV death. Following such a decision, the executive committee will define a PACD, on or after which study closure visits will commence. Analysis based on the final database will be conducted to support the full reporting of the study. The consistency between the interim analysis database and the subsequently locked database will be assessed.

If the study is stopped in the efficacy interim analysis, testing of secondary endpoints will be performed on the final database with the same testing procedure as described in section 4.1.3 with two-sided significance level 0.002.

A futility analysis is planned to be performed at the same time as the planned interim analysis. The study may be stopped for futility if the observed HR for the primary endpoint is > 0.946, corresponding to a predictive power of 5%. If the futility criterion of the primary endpoint is met, then DMC will evaluate the totality of data, including potential benefits on patient reported outcomes to consider recommending ending the study for futility.

6 CHANGES OF ANALYSIS FROM PROTOCOL

NA

7 REFERENCES

Fitchett et al 2016
Fitchett D, Zinman B, Wanner C, Lachin JM, Hantel S, Salsali A, et al. Heart failure outcomes with empagliflozin in patients with type 2 diabetes at high cardiovascular risk: results of the EMPA-REG OUTCOME® trial. Eur Heart J 2016;14;37(19):1526-34.

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

Green et al 2000
C. Patrick Green, MD, Charles B. Porter, MD, FACC, Dennis R. Bresnahan, MD, FACC, John A.Spertus, MD, MPH, FACC Development and Evaluation of the Kansas City Cardiomyopathy Questionnaire: A New Health Status Measure for Heart Failure. JACC, Vol.35 No 5, April 2000:1245-55

Hicks et al 2018
Hicks KA, Mahaffey KW, Mehran R, Nissen SE, et al. Cardiovascular and Stroke Endpoint Definitions for Clinical Trials. J Am Coll Cardiol 2018;71:1021–34

Kawaguchi et al 2011
Kawaguchi A, Koch G, Wang X. Stratified Multivariate Mann-Whitney Estimators for the Comparison of Two Treatments with Randomization Based Covariance Adjustment. Statistics in Biopharmaceutical Research, 2011;3(2), 217–231.

Kawaguchi and Koch 2015
Kawaguchi A, Koch, GG. sanon: An R Package for Stratified Analysis with Nonparametric Covariable Adjustment. Journal of Statistical Software, 2015;67(9).

Kristensen et al 2015
Kristensen SL, Jhund PS, Køber L, McKelvie RS, Zile MR, Anand IS, et al. Relative importance of history of heart failure hospitalisation and N-terminal Pro–B-type natriuretic peptide level as predictors of outcomes in patients With heart failure and preserved ejection fraction. JACC Heart Fail 2015 Jun;3(6):478-86.

Kristensen et al 2017
Kristensen SL, Mogensen UM, Jhund PS, Petrie MC, Preiss D1, Win S, et al. Clinical and echocardiographic characteristics and cardiovascular outcomes according to diabetes status in patients with heart failure and preserved ejection fraction: a report from the I-Preserve Trial (Irbesartan in Heart Failure With Preserved Ejection Fraction). Circulation. 201721;135(8):724-35.

Levey at al 2009
Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF 3rd, Feldman HI, Kusek JW, Eggers P, VanLente F, Greene T, Coresh J; CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration). A new equation to estimate glomerular filtration rate. Ann Intern Med. 2009 May 5; 150(9):604-12

Lin et al 2000
Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2000; 62(4):711–730.

Neal et al 2017
Neal B, Perkovic V, Mahaffey KW, de Zeeuw D, Fulcher G, Erondu N, et al, for the CANVAS Program Collaborative Group. Canagliflozin and cardiovascular and renal events in type 2 diabetes. N Engl J Med 2017; 377: 644-657.

Pfeffer et al 2015
Pfeffer MA, Claggett B, Assmann SF, Boineau R, Anand IS, Clausell N, et al. Regional Variation in Patients and Outcomes in the Treatment of Preserved Cardiac Function Heart Failure With an Aldosterone Antagonist (TOPCAT) Trial. Circulation. 2015;131:34-42.

Ratitch et al 2013
Ratitch B, Lipkovich I, O'Kelly M. Combining Analysis Results from Multiply Imputed Categorical Data. PharmaSUG 2013 - Paper SP03

Rogers et al 2016
Jennifer K Rogers, Alex Yaroshinsky, Stuart J Pocock, David Stokar and Janice Pogoda. Analysis of recurrent events with an associated informative dropout time: Application of the joint frailty model. Statist. Med. 2016, 35 2195–2205

Rådholm et al 2018
Rådholm K, Figtree, Perkovic V, Solomon SD, Mahaffey KW, et al. Canagliflozin and Heart Failure in Type 2 Diabetes Mellitus: Results From the CANVAS Program (Canagliflozin Cardiovascular Assessment Study). Circulation. 2018;137
https://doi.org/10.1161/CIRCULATIONAHA.118.034222

Spertus et al 2005
Spertus J, Peterson E, Conard MW, Heidenreich PA, Krumholz HM, Jones P, et al. Monitoring clinical changes in patients with heart failure: A comparison of methods. Am Heart J 2005;150:707-15.

Stokes et al 2012
Stokes ME, Davis CS, Koch GG.Categorical Data Analysis Using the SAS System. Carry, NC: SAS Institute Inc.

Wang and Pocock 2016
Wang D, Pocock S. A win ratio approach to comparing continuous non-normal outcomes in clinical trials. Pharmaceut. Statist. 2016, 15 238–245

Statistical Analysis Plan	
Study Code	D169CC00001
Edition Number	5.0
Date	8 December 2021

An International, Double-blind, Randomised, Placebo-Controlled Phase III Study to Evaluate the Effect of Dapagliflozin on Reducing CV Death or Worsening Heart Failure in Patients with Heart Failure with Preserved Ejection Fraction (HFpEF)

DELIVER - Dapagliflozin Evaluation to Improve the LIVEs of Patients with PReserved Ejection Fraction Heart Failure

An International, Double-blind, Randomised, Placebo-Controlled Phase III Study to Evaluate the Effect of Dapagliflozin on Reducing CV Death or Worsening Heart Failure in Patients with Heart Failure with Preserved Ejection Fraction (HFpEF)

DELIVER - Dapagliflozin Evaluation to improve the LIVEs of patients with pReserved ejection fraction heart failure

Study Statistician [redacted]

An International, Double-blind, Randomised, Placebo-Controlled Phase III Study to Evaluate the Effect of Dapagliflozin on Reducing CV Death or Worsening Heart Failure in Patients with Heart Failure with Preserved Ejection Fraction (HFpEF)

DELIVER - Dapagliflozin Evaluation to improve the LIVEs of patients with pReserved ejection fraction heart failure

Global Product Statistician [redacted]

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
AE	Adverse event
ATC	Anatomical therapeutic chemical
BP	Blood pressure
CDF	Cumulative distribution function
CEA	Clinical event adjudication
CKD-EPI	Chronic kidney disease epidemiology collaboration equation
CMH	Cochran-Mantel-Haenszel test
CMWPC	Clinically meaningful within-patient change
COVID-19	Corona Virus Disease 2019
CSP	Clinical study protocol
CV	Cardiovascular
DAE	Adverse events leading to discontinuation of investigational product
DBP	Diastolic blood pressure
DKA	Diabetic ketoacidosis
DMC	Data monitoring committee
eCDF	Empirical cumulative distribution function
eCRF	Electronic case report form
eGFR	Estimated glomerular filtration rate
FAS	Full analysis set
HbA1c	Haemoglobin A1c
HF	Heart failure
HFpEF	Heart failure with preserved ejection fraction
HR	Hazard ratio
IP	Investigational product (dapagliflozin or matching placebo)
ITT	Intention to treat
IxRS	Interactive Voice/Web Response System
KCCQ	Kansas City Cardiomyopathy Questionnaire
KM	Kaplan-Meier
LTFU	Lost to follow-up
LVEF	Left ventricular ejection fraction
MCID	Minimal clinically important difference
MedDRA	Medical dictionary for regulatory activities
MMRM	Mixed Model Repeated Measures

Abbreviation or special term	Explanation
MTP	Multiple testing procedure
NT-proBNP	N-terminal pro b-type natriuretic peptide
NYHA	New York Heart Association
PACD	Primary analysis censoring date
PGIS	Patient global impression of severity
PRAC	European Medicines Agency's Pharmacovigilance Risk Assessment Committee
PT	MedDRA preferred term
PTDV	Premature treatment discontinuation visit
QoL	Quality of life
SAE	Serious adverse event
SAS	Safety analysis set
SBP	Systolic blood pressure
SCV	Study closure visit
SD	Standard deviation
SEM	Standard error of measurement
SGLT2	Sodium-glucose co-transporter 2
SOC	MedDRA system organ class
T2D	Type 2 diabetes
TSS	Total symptom score
WHO	World Health Organization
WoC	Withdrawal of consent
WR	Win ratio

AMENDMENT HISTORY

Date	Brief description of change
Version 1 27 August 2018	Version 1.0 signed

Version 2 6 November 2020	[1.1 Study objectives] Updated primary objective with dual primary analyses: Primary analysis to be analysed in full study population and subpopulation with LVEF < 60% Updated secondary objectives: First secondary to be analysed in full study population and subpopulation with LVEF < 60%. Adding urgent HF visits to total number of HF events (first and recurrent) and CV death. Moved NYHA class from secondary objective to exploratory. Added CV death as secondary objective. Updated exploratory objectives: Added NYHA class objective from secondary objective and removed PGIS objective. Rewording of EQ-5D-5L objective and endpoint. [1.2.1 Primary analysis censoring date] Increased target number of primary endpoint events from 844 to 1117. [1.3 Study design] Updated definition of subacute patients, increasing hospitalisation from within 21 days to within 30 days. Increased number of randomised patients from 4700 to 6100 and number of enrolled patients from 8000 to 11000. Updated target number of primary endpoint events from 844 to 1117. Updated anticipated total study duration from 33 months to 39 months. [1.4 Number of subjects] Updated power, study duration, number of events and proportion of subacute. [2.1.1 Full analysis set] Updated with subpopulation information: “A subset of the full analysis set consisting of patients with baseline LVEF of < 60% (or LVEF < 60% subpopulation) will be analysed separately as part of the confirmatory statistical testing procedure.” [3.2 Secondary variables] Updated with dual primary endpoints. Updated with new definition of total number of events, including urgent HF visits. Added Figure 2 with updated multiple testing procedure with dual primary analyses. [3.2.1 Total number of (first and recurrent) hospitalisations for HF and CV death] Updated with definition of total number of events. Updated with information regarding prioritisation, which event to be counted in recurrent event analysis, if HF event and CV death occur at same day. [3.2.2 Change from baseline at 8 months in the KCCQ total symptom score] Added definition regarding ranking. [Previous 3.2.3 Proportion of patients with worsened NYHA class at 8 months] Removed entire paragraph. [3.2.3 Cardiovascular death] Added paragraph with secondary objective concerning CV death. [3.3 Safety variables]

Date	Brief description of change
	Added adjudication of potential DKA events. Added major hypoglycaemic events to list of safety variables. [4.1.1 Estimand for primary and secondary outcomes] Added estimand for KCCQ TSS. [4.1.2 Hypotheses] Added dual primary hypotheses. [4.1.3 Confirmatory testing procedure] Updated with handling of alpha for split primary analyses. Added Figure 2. [4.2.3.2 Sensitivity analysis of the primary endpoint] Updated with information that sensitivity analyses related to impact of COVID-19 will be added at next SAP update prior to interim analysis. [4.2.4.1 Analysis of recurrent HF events and CV death] Updated definition of HF events, including urgent HF visits. Added handling on priority of events occurring on the same day. [4.2.4.2 Analysis of change from baseline to 8 months in the KCCQ total symptom score] Added information on how to handle analysis under COVID-19 pandemic. Added information on ranking. Added information on handling of missing response for reasons other than death. Estimation of treatment effect updated. Added update on handling of ceiling and floor effects. Information on imputation updated. Updated information on TSS responder analyses. [4.2.4.3 NYHA] Section removed and moved to 4.2.6 Analysis of exploratory objectives. [4.2.4.3 CV death] Section on analysis of CV death added. [4.2.5.4 Amputations and preceding events] Section renamed to “Specific adverse events” and paragraphs on DKA, major hypoglycaemic events and genital infections added. [4.2.6 Analysis of exploratory objectives] Section on NYHA added (moved from previous Section 4.2.4.3). Section on PGIS removed. [5 Interim analysis] Removed futility analysis. [Reference] Added references: FDA guidance during COVID-19 2020 and Spiessen and Debois 2010 Removed references: Kawaguchi and Koch 2015 and Neal et al 2017

Date	Brief description of change
Version 3.0 9 December 2020	[4.2.4.2 Analysis of change from baseline to 8 months in the KCCQ total symptom score] Added information on responder analysis: "Additional responder analysis will be performed in the same way as for 5 points improvement and deterioration described above, using the thresholds of clinically meaningful within-patient change from baseline TSS derived from anchor-based analyses of blinded study data as described in Appendix A, with "ceiling" and "floor" values handled consistently." [Reference] Added reference: Coon and Cook 2018. [Appendix] Added Appendix A describing how to estimate clinically meaningful thresholds for KCCQ total symptom score, using PGIS.

Version 4.0 20 May 2021	Minor edits done throughout entire document. [1.2.1 Primary analysis censoring date] Updated to be consistent with CSP, that SCV should be performed within 6 weeks after PACD, which can be extended if decided by Global Study Team. Added that patients will stop taking IP at the SCV. [1.4 Number of subjects] Added information that final allocation of alpha and full testing procedure can be found in section 4.1.3. Added text that the power considerations stated in this section are examples for the dual primary analysis. [3.2.1 Total number of HF events (first and recurrent) and CV death] Removed: "Recurrent HF events (hospitalisation for HF or urgent HF visit), CV death and censoring processes all have continuous distributions so that HF events and death cannot happen at the same time." Updated for clarification: "For patients who did not have a HF event or CV death, and following last event in patients with one or more HF events, censoring will follow the same rule as for the primary endpoint." [3.2.3 Cardiovascular death] Added "or died after WoC" for specification on patients to be censored. [3.2.4 Death from any cause] Added "or with unknown vital status" for specification on patients to be censored. [3.3 Safety variables] Updated list of safety variables, adding myocardial infarction, unstable angina, stroke, major hypoglycaemic events, potential diabetic ketoacidosis and amputations. Updated for clarification: "These events will be recorded as **AEs or if they fulfil seriousness criteria as** SAEs in the database, but **SAEs** will not be reported to health authorities to avoid unnecessary unblinding." [4.1 General principles] Added for clarification: "**If the number of tablets dispensed or the number of tablets returned is missing for at least one observation, compliance is not calculated for that patient."** and "IP compliance will be presented descriptively, including mean, **SD,** median, quartiles and 5% and 95% percentiles **for safety analysis set by treatment group**." [4.1.1 Estimand for primary and secondary outcomes] Sentence removed: "The time-to-first-event analysis by Cox proportional hazards regression and the analysis of recurrent events (Section 4.2.4) assume that missing data is at random." [4.1.2 Hypotheses] Removed reference to Haybittle-Peto function as that method will not to be used. Updated alpha level for final analysis and added/removed details for clarification: "**With alpha 0.2% allocated to** one planned interim analysis~~ including 67% of the target number of primary endpoints~~, the significance level **in the final analysis** will be **4.8%**, to be split between the dual hypothesis." [4.1.3 Confirmatory testing procedure] Section updated with details on significance levels. Added table: "Table 1 Level of α_1 depending on proportion of events in LVEF < 60% subpopulation". Updated for clarity: "If the study is stopped at the efficacy interim analysis (Section 5), testing of remaining secondary endpoints will be performed in the full study

	population only, in fixed sequence at two-sided alpha of 0.2% in the order described in the right branch of Figure 2." [4.1.5 Vital status and follow-up of endpoints] Added for clarification: "The denominator, representing maximum complete follow-up, will be the time **from randomisation until the earliest of** first primary endpoint event, death or PACD." [4.2.2 Concomitant and baseline medication] Added for clarification: "The **proportion** of **patients taking** baseline and concomitant medication will be presented for the FAS per ATC class and treatment group." [4.2.3.1 Subgroup analysis of the primary endpoint] Added for clarification: "A test of interaction between randomised treatment group and the subgroup variable will be performed using Cox proportional hazard model stratified **by** T2D **status at randomisation** with factors for treatment group, the subgroup variable and the interaction between treatment and subgroup." Added: "Hazard ratio estimates, confidence intervals and p-values are not presented for subgroups with less than 15 events in total, both arms combined." Table 1 renamed to Table 2 Table 2: Updated subgroups for LVEF at enrollment to ≤ 49%, 50% to 59%, ≥ 60% [4.2.3.2 Sensitivity analysis of the primary endpoint] Added information that further sensitivity analyses will be added at a later update: "**We will monitor the blinded study data to assess the impact of COVID-19 on the study and will add supportive and** sensitivity analyses related to the impact of COVID-19 on both primary and secondary endpoints in a SAP update prior to clinical data lock. **Also, additional covariates might be added to analyses, if deemed necessary based on blinded data.**" [4.2.4.1 Analysis of total number of HF events (first and recurrent) and CV death] Added for consistency: "The composite outcome of **total number of** HF events (first and recurrent) and CV death **with onset on or prior to PACD, adjudicated and confirmed by the CEA committee,**" Sentence removed: "Recurrent HF events, CV death and censoring processes all have continuous distributions so that a HF event and death cannot happen at the same time." [4.2.4.2 Analysis of change from baseline at 8 months in the KCCQ total symptom score] Added for clarification: "In the ranking, patients who die prior to the first follow-up visit where KCCQ-KSS is assessed, at 1 month, will be defined as having a zero change from baseline while alive." Added cut-off date to define population to be used in primary KCCQ-TSS analysis: "As a consequence, the main analysis of this endpoint will be done in the population with patients who had a planned visit 5 (8 months) prior to the major COVID-19 outbreak, **defined as 11th March 2020 (the date when WHO declared COVID-19 a pandemic)** thus unaffected by the pandemic's possible impact on health-related quality of life" Removed: "The section regarding these analyses and exact date for data cut-off will be updated prior to the interim analysis." Added that formal inference will be based on Win ratio method. Section on responder analysis updated. Section on handling of missing KCCQ data updated, including numbers from anchor-

Date	Brief description of change
	based analyses. Clarifications made in section on "Handling of missing KCCQ data". [4.2.4.3 Analysis of CV death] Clarifications that CV deaths are confirmed in adjudication and how censoring is handled. [4.2.4.4 Analysis of all-cause mortality] Clarification that analysis includes deaths from any cause. [4.2.5 Analysis of safety variables] Updated that summaries of AEs will be presented both for the on-treatment period and on- and off-treatment period. [4.2.5.1 Adverse events] Updated list of safety variables, adding myocardial infarction, unstable angina, stroke, major hypoglycaemic events, potential DKA and amputations. [4.2.5.4 Specific adverse events] Added: "AEs leading to amputations" to list. Added that event rate will be presented for AEs leading to amputations and preceding events, DKA and major hypoglycaemic events, as well as definition of event rate calculation. Added for clarification: "**Events of genital area infections and necrotising fasciitis potential** of Fournier's gangrene". [4.2.5.5 Laboratory evaluation and vital signs] Removed PTDV and SCV from list of visits and added range to descriptive statistics. [4.2.6 Analysis of exploratory objectives] Added: "Only NYHA assessments made at site or through phone visits with the patients to be used in analyses." Added clarification on exploratory KCCQ analyses. [5 Interim Analyses] Removed reference to Haybittle-Peto function as that method will not to be used. [6 Changes of Analysis from Protocol] Added: "**The alpha for final analysis adjusted for interim analysis at alpha 0.2% will be set to 5% minus 0.2% = 4.8%, rather than 4.98% as determined by the Haybittle-Peto function for 67% of events (sections 9.1 and 9.5 of the protocol).**" [References] Added reference: **Burman et al 2009.** [Appendix A] Earlier Appendix A renamed A1 Methods. [Appendix A2] Added appendix including summary of results of anchor-based analysis on blinded study data. [Appendix B] Added appendix with R code for calculation of significance level.

Date	Brief description of change
Version 5.0 08 December 2021	Formatting updated throughout entire document. [3.3 Safety Variables] Minor clarifications added [4.1.3 Confirmatory Testing Procedure] Sentence added: "**For the calculation of α_1, the correlation will be based on the square root of the lower bound of a two-sided 95% confidence interval for the proportion of events in the subpopulation with LVEF<60%, using a normal approximation confidence interval for the proportion.**" Table 1 updated presenting number of events in LVEF < 60% subpopulation and full population instead of presenting proportion of events in the subpopulation. Confidence intervals added and numbers for α_1 in the different scenarios updated. Last bullet in the list clarified. [4.1.5 Vital Status and Follow-up of Endpoints] Clarified that non-CV death includes undetermined. [4.2.3.1 Subgroup Analysis of the Primary Endpoint] Updated that subgroup analysis will be done both for full population and LVEF < 60% subpopulation. [4.2.3.2 Sensitivity Analysis of the Primary Endpoint] Added description of a sensitivity analysis where patients with premature censoring have imputed time to event information and more detailed information about the planned tipping point analysis. Added sensitivity analysis where patients and events are censored at the onset date of AE associated with COVID-19 infection. [4.2.4.1 Analysis of Total Number of HF Events (First and Recurrent) and CV Death] Added sensitivity analysis where patients and events are censored at the onset date of AE associated with COVID-19 infection. [4.2.4.2 Analysis of Change from Baseline at 8 Months in the KCCQ Total Symptom Score] Added that both **planned and** performed 8 month assessments are to be included in COVID-19 supplementary analysis for KCCQ TSS. [4.2.4.3 Analysis of CV Death] Added sensitivity analysis where patients and events are censored at the onset date of AE associated with COVID-19 infection. [4.2.5.1 Adverse Events] Clarification that on-treatment period will be used for primary analysis of all safety variables, except for amputations and preceding events. Added that MedDRA 24.1 will be used. Information previously in section "4.2.5.4 Specific adverse events" added to this section. [4.2.5.4 Specific Adverse Events] Text moved to be included in Section 4.2.5.1 and section removed. [4.2.6 Analysis of Explorative Objectives] Updated that KCCQ QoL will be reported descriptively only. [Appendix B] Updated to include R and SAS code.

1 STUDY DETAILS

1.1 Study Objectives

1.1.1 Primary Objective

Primary objective	Endpoint/variable
To determine whether dapagliflozin is superior to placebo, when added to standard of care, in reducing the composite of CV death and HF events (hospitalisation for HF or urgent HF visit) in patients with HF and preserved systolic function, in • full study population • subpopulation with LVEF < 60%	Time to the first occurrence of any of the components of this composite: 1. CV death 2. Hospitalisation for HF 3. Urgent HF visit (eg, emergency department or outpatient visit)

1.1.2 Secondary Objectives

Secondary objective	Endpoint/variable
To determine whether dapagliflozin is superior to placebo in reducing the total number of recurrent HF events (hospitalisations for HF or urgent HF visit) and CV death, in • full study population • subpopulation with LVEF < 60%	Total number of HF events (first and recurrent) and CV death
To determine whether dapagliflozin is superior to placebo in improving Patient Reported Outcomes measured by KCCQ	Change from baseline in the TSS of the KCCQ at 8 months
To determine whether dapagliflozin is superior to placebo in reducing CV death	Time to the occurrence of CV death
To determine whether dapagliflozin is superior to placebo in reducing all-cause mortality	Time to the occurrence of death from any cause

1.1.3 Safety Objectives

Safety Objective	Outcome Measure
To evaluate the safety and tolerability of dapagliflozin compared to placebo in patients with HFpEF	• SAEs • DAEs • Amputations, AEs leading to amputation and potential risk factor AEs for amputations affecting lower limbs

1.1.4 Exploratory Objectives

Exploratory Objective	Endpoint/Variable
To determine whether dapagliflozin is superior to placebo in reducing all-cause hospitalisation	Time to the first occurrence of hospitalisation from any cause
To determine whether dapagliflozin is superior to placebo in reducing the proportion of patients with worsened NYHA class	Proportion of patients with worsened NYHA class from baseline to 8 months
To describe health status assessed by EQ-5D-5L to support health economic analysis and health technology assessment	Results will be reported separately in a health economic report
To determine whether dapagliflozin compared with placebo will have an effect on SBP	Change in SBP from baseline
To determine whether dapagliflozin compared with placebo will have an effect on body weight	Change in body weight from baseline
To determine whether dapagliflozin compared with placebo will have an effect on eGFR	Change in eGFR from baseline
To explore whether dapagliflozin compared to placebo improves KCCQ summary scores, sub-scores of TSS (symptom frequency and symptom burden) and domains	Change in Clinical summary score, TSS sub-scores, Overall summary score, QoL score
To collect and store blood samples for future exploratory genetic research	Not applicable. Results will be reported separately

1.2 Definitions

1.2.1 Primary Analysis Censoring Date

The executive committee and AstraZeneca will monitor the accrual of endpoint events and when appropriate define the PACD at which time at least the pre-defined target number of 1117 events for the primary composite endpoint is expected to have occurred. The study sites will be instructed to plan for SCV to be performed within 6 weeks after PACD, which can be extended if decided by the Global Study Team. Patients will stop taking IP at the SCV.

Analyses of efficacy endpoint events will include events with onset on or prior to PACD. Event free patients who have not been prematurely censored due to incomplete information (see Section 3.1) will be censored at PACD. HF events and deaths with onset after PACD will also be adjudicated.

1.2.2 Withdrawal of Consent

Withdrawal of consent should only occur if the patient has received appropriate information about options for modified study follow-up and does not agree to any kind of further assessment or follow-up. Information regarding vital status (dead or alive) at the end of the study will be collected from public sources, to be included in the analysis of death from any cause as a sole outcome and in patient disposition summaries.

1.2.3 Discontinuation of Investigational Product

Discontinuation of IP does not mean discontinuation from study follow-up or WoC. Patients who discontinue from IP should continue study visits according to plan until study closure. If the patient does not agree to this approach, modified follow-up capturing the essential information for the objectives of the study should be arranged. Data will be included in the ITT analyses irrespective of whether the event occurred before or following discontinuation of IP.

1.2.4 Vital Status

Known vital status at the end of the study will be defined when the patient is dead or has date last know alive on or after the PACD.

For patients who have withdrawn consent, the investigator will attempt to collect vital status from publicly available sources at study closure in compliance with local privacy laws/practices.

1.2.5 Lost to Follow-up

The term LTFU will be limited to patients with unknown vital status at the end of the study as defined in Section 1.2.4. Other measures will be used to describe completeness of follow-up of the primary endpoint (Section 4.1.5).

1.3 Study Design

This is an international, multicentre, parallel-group, event-driven, randomised, double-blind, placebo-controlled study in patients with HFpEF, evaluating the effect of dapagliflozin 10 mg versus placebo, given once daily in addition to background regional standard of care therapy, including treatments to control co-morbidities, in reducing the composite of CV death or HF events.

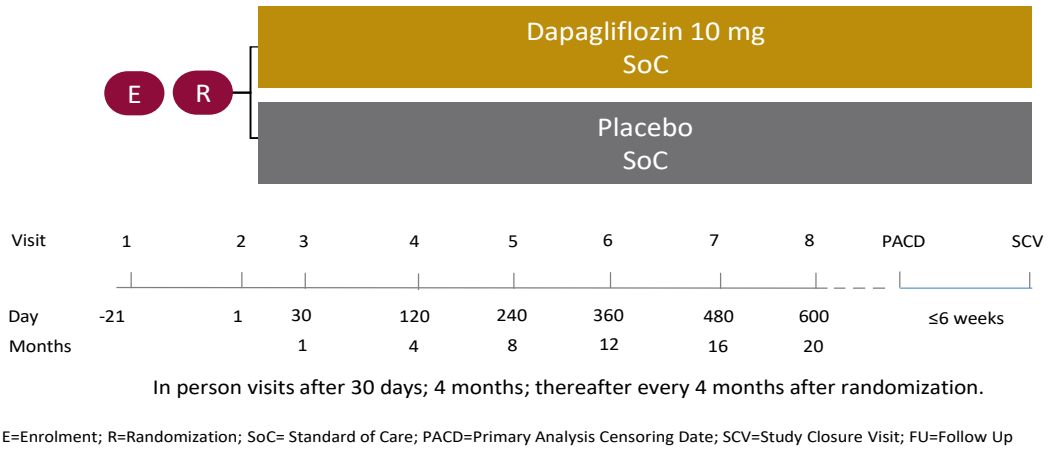
HFpEF is defined for the purposes of this study as LVEF > 40% and evidence of structural heart disease. Adult patients with HFpEF, aged ≥ 40 years and with NYHA class II to IV will be randomised in a 1:1 ratio to receive either dapagliflozin 10 mg or placebo once daily. A proportion of patients, here denoted as the subacute group, will be randomised during hospitalisation for HF or within 30 days of discharge from hospitalisation for HF.

Originally, 4700 patients were planned to be randomised with a study duration of approximately 33 months, when 844 primary events had occurred. Based on the ongoing blinded monitoring of event accrual (including the percentage of patients from the subacute category), the sample size was increased from original 4700 to approximately 6100 patients.

It was estimated that approximately 11000 patients at approximately 400 to 500 sites in 20 to 25 countries will be enrolled to reach the target of approximately 6100 randomised patients.

In this event driven trial, study closure procedures will be initiated when the predetermined number of primary endpoints are predicted to have occurred (n = 1117), ie, the PACD (Section 1.2.1 and Figure Figure 1 Study Design). Patients should be scheduled for a SCV within 6 weeks of the PACD, which can be extended if decided by Global Study Team. The maximum treatment duration is expected to be approximately 39 months, dependent on randomisation rate and event rate. The number of patients randomised, the study duration, or both, may be changed if the randomisation rate or the event rate is different than anticipated.

Figure 1 Study Design



E=Enrolment; R=Randomization; SoC= Standard of Care; PACD=Primary Analysis Censoring Date; SCV=Study Closure Visit; FU=Follow Up

1.3.1 Randomisation

Patients will be randomised 1:1 to either dapagliflozin 10 mg or placebo once daily. The treatment allocation in this study will be double-blind. Randomisation will be stratified by T2D status at randomisation (2 levels: with T2D; without T2D). For the purpose of stratification, T2D is defined as established diagnosis of T2D or HbA1c ≥ 6.5% (48 mmol/mol) at enrolment (Visit 1; single measure) central laboratory test.

Randomisation will be performed in balanced blocks of fixed size. The randomisation codes will be computer generated and loaded into the IxRS database.

The number of randomised patients with T2D will be monitored in order to ensure a minimum of 30% patients in each group of patients with and without T2D. Randomisation may be capped (ie, no more patients can be randomised in a specific sub-population) if the pre-determined limit is reached.

Randomisation of patients based on geographic region will be monitored to ensure global representation. LVEF value, NYHA class, subacute/non-subacute group, and atrial fibrillation status at Visit 1 may be capped in IxRS to avoid over- or under-representation of these patient subgroups.

1.4 Number of Subjects

The primary objective of the study is to determine the superiority of dapagliflozin versus placebo added to standard of care in reducing the composite of CV death and HF events (hospitalisation for HF or urgent HF visit). Two hypotheses will be tested simultaneously (ie, dual primary analyses) for this primary objective: (1) in the full population and (2) in an LVEF < 60% subpopulation, with alpha allocated to each test.

Originally, assuming a true HR of 0.80 between dapagliflozin and placebo, using a two-sided alpha of 5%, 844 primary endpoint events were targeted in order to provide a statistical power of 90% for the test of the primary composite endpoint.

To allow testing for the dual primary analyses, alpha will be allocated to each test to ensure strong control of the overall type I error rate. The target number of patients with a primary endpoint has been increased to 1117 in order to provide adequate statistical power for each test. The power to reject the dual primary hypotheses depends on how alpha is allocated between the two hypotheses and the proportion of primary events in the LVEF < 60% subpopulation. It is anticipated that at least 70% of the primary endpoint events (ie, approximately 780 events) will be available for the LVEF < 60% subpopulation. The final allocation of alpha and full testing procedure is specified in Section 4.1.3 and the alpha levels used in the following text are just examples used to illustrate the power considerations for the dual primary analysis. For illustration, testing the effect on the primary endpoint in the LVEF < 60% subpopulation, a true HR of 0.80 and approximately 1117 primary endpoint events in the full population (at least 780 events in the subpopulation) would then provide at least:

- 80% power for a two-sided nominal alpha of 2.4%
- 85% power for a two-sided nominal alpha of 3.7%

For testing the effect on the primary endpoint in the full study population, a true HR of 0.80 and approximately 1117 primary endpoint events would also provide:

- 90% power for a two-sided nominal alpha of 1.5%

- 93% power for a two-sided nominal alpha of 2.4%.

This is based on an overall 1:1 allocation between dapagliflozin and placebo.

The HR 0.80 was originally chosen as a conservative assumption based on the observed HR 0.72 (95% confidence interval 0.50-1.04) for the composite of HF hospitalisation and CV death in patients with HF at baseline in the EMPA-REG OUTCOME trial (Fitchett et al 2016) and HR 0.61 (0.46-0.80) for patients with history of HF in the CANVAS program (Rådholm et al 2018) considering that these were post-hoc analyses in subgroups with limited documentation of baseline HF diagnosis, not characterised by ejection fraction.

The event rate assumptions are based on sub analyses of the TOPCAT and I-PRESERVE studies by geographic region, NT-proBNP levels, prior hospitalisation for HF, and T2D status (Pfeffer et al 2015, Kristensen et al 2015, Kristensen et al 2017). The sample size calculation builds on the assumption of an annual event rate of 9% in the placebo group for the majority of prevalent HFpEF patients, importantly all with NT-proBNP ≥ 300 pg/ml by inclusion criterion. Additionally, a subgroup of patients due to be discharged or recently discharged from a HF hospitalisation (here denoted 'subacute' patients) with a higher event rate is planned to be included. Assuming 20% of patients from the subacute category with an annual event rate of 24% during the first year and 9% thereafter for the remainder of the study, the original sample size of 4700 was estimated to provide 844 events during a recruitment period of 18 months and a minimum follow-up of 15 months.

Based on the ongoing blinded monitoring of event accrual (including the percentage of patients from the subacute category), the sample size was increased from original 4700 to approximately 6100 randomised patients. Accordingly, the recruitment period was anticipated to increase from the original 18 months to 26 months. Recruitment might be marginally prolonged in a few countries to meet local targets. The study is event driven and the number of patients or duration may further change.

With the same event rate assumptions as above, assuming 11% of patients from the subacute category, approximately 6100 patients were estimated to provide the required number of 1117 patients with a primary event in the full study population, during an anticipated recruitment period of 26 months and a minimum follow-up period of 13.5 months (total study duration 39 months).

In addition, the expected number of patients who will be LTFU is expected to be small; hence, these are not considered in the determination of the sample size.

2 ANALYSIS SETS

2.1 Definition of Analysis Sets

2.1.1 Full Analysis Set

All patients who have been randomised to IP will be included in the FAS irrespective of their protocol adherence and continued participation in the study. Patients will be analysed according to their randomised IP assignment, irrespective of the treatment actually received. The FAS will be considered the primary analysis set for the ITT analysis of primary and secondary variables and for the exploratory efficacy variables. A subset of the FAS consisting of patients with baseline LVEF of < 60% (or LVEF < 60% subpopulation) will be analysed separately as part of the confirmatory statistical testing procedure (see CSP Section 4.2 for justification of testing LVEF < 60% subpopulation).

2.1.2 Safety Analysis Set

All randomised patients who received at least 1 dose of randomised treatment will be included in the SAS. Patients will be analysed according to the treatment actually received. For any patients given incorrect treatment, ie, randomised to one of the treatment groups, but actually given the other treatment, the treatment group will be allocated as follows: Patients who got both incorrect and correct treatment will be analysed according to their randomised treatment. Patients who got only the incorrect treatment will be analysed according to that treatment.

The SAS will be considered the primary analysis set for all safety variables.

2.2 Violations and Deviations

The important protocol deviations listed below will be summarised by randomised treatment group

- Patients who were randomised but did not meet inclusion criteria, or met exclusion criteria
- Patients who received the wrong IP at any time during the study.
- Patients who received prohibited concomitant medication, which for this study is limited to open label SGLT2 inhibitors taken in combination with IP.

As the primary analysis is ITT analysis, protocol deviation will not imply exclusion from the primary analysis.

3 PRIMARY AND SECONDARY VARIABLES

Deaths and potential HF events will be adjudicated by an independent CEA committee. The CEA committee members will not have access to the treatment codes for any patient. The CEA procedures and event definitions will be described in the CEA charter according to the CDISC definitions (Hicks et al 2018).

Only HF hospitalisations and urgent HF visits confirmed by the CEA will be used in the analysis of the primary and secondary endpoints and their components.

The primary analyses of the endpoints concerning CV deaths, either as a component of a composite or on its own, will include deaths adjudicated as CV cause. Deaths adjudicated as "cause undetermined" will be considered as non-CV deaths in these analyses.

Adjudicated events occurring from randomisation until WoC or PACD will be included in the analysis of primary and secondary endpoints. The analysis of all-cause death as a sole outcome will in addition include any deaths (not adjudicated) after WoC, but on or before PACD.

3.1 Primary Variable

The primary efficacy variable is time from randomisation to the first occurrence of any event in the composite of CV death, hospitalisation for HF or an urgent HF visit.

Patients who did not have an adjudicated primary endpoint event on or prior to PACD will be censored at the earliest of date of WoC or non-CV death when applicable, and otherwise at the date of the last clinical event assessment or the PACD, whichever occurs first. It is expected that patients alive and under study follow-up will have a clinical event assessment at their SCV after PACD. Last clinical event assessment is defined as the last date when the event assessment question for a potential HF event was completed on the eCRF event assessment page.

In analysis of the individual components hospitalisation for HF and urgent HF visit, to examine their contribution to the composite endpoint, date of death from any cause will be an additional point of censoring.

For analysis of time to first event, data will be expressed as two variables:

- A binary variable indicating whether the event in question occurred, or the patient was censored.
- An integer variable for the number of days from randomisation to the first occurrence of an event (start date of the event – randomisation date + 1), or for event free patients, from randomisation to censoring (censoring date – randomisation date + 1).

3.2 Secondary Variables

The secondary endpoints are included in hierarchical testing sequences following the dual primary analysis as described in Section 4.1.3 and depicted in Figure 2.

3.2.1 Total Number of Heart Failure Events (First and Recurrent) and Cardiovascular Death

The efficacy variable is the total number of first and recurrent HF events (hospitalisations for HF or urgent HF visits) and CV death.

For the analysis of first and recurrent HF events and CV death, the data will be expressed in counting process style for input to the analysis as described in Section 4.2.4.1, as follows. The time from randomisation to end of follow-up/censoring will be split into one or more interval with variables for start of interval, end of interval and a variable indicating if an event occurred at the end of each respective interval, or if the patient was censored. If a HF event and CV death occurred at the same day, then only the CV death will be counted.

For patients who did not have a HF event or CV death, and following last event in patients with one or more HF events, censoring will follow the same rule as for the primary endpoint.

3.2.2 Change from Baseline at 8 Months in the KCCQ Total Symptom Score

The efficacy variable is the change from baseline at 8 months of the KCCQ-TSS.

The KCCQ is a self-administered disease specific instrument for patients with HF (Green et al 2000, Spertus et al 2005). The KCCQ consists of 23 items measuring HF-related symptoms, physical limitations, social limitations, self-efficacy, and health-related quality of life. The TSS incorporates the symptom burden and symptom frequency domains into a single score. Scores are transformed to a range of 0 to 100. Higher scores represent better outcomes.

Baseline is defined as the value at randomisation visit (Visit 2). Change from baseline at each post-baseline analysis time point will be calculated as the value at the corresponding post-baseline analysis time point minus the baseline value. The KCCQ is assessed by the patient at randomisation, at the visits targeted 1, 4 and 8 months following randomisation and at PTDV and SCV. By the ITT principle, the analysis will include all data irrespective of whether the patient has discontinued IP.

In order to account for patients who die prior to the 8-month assessment and to accommodate non-normal distribution of KCCQ scores, a composite rank-based endpoint will be used. The values of change from baseline at 8 months in TSS of patients who survive to 8 months will be converted to ranks (across both treatment groups combined) with lower ranks attributed to worse outcomes (ie, lower ranks corresponding to negative or smaller values of change from

baseline). Patients who die prior to the 8-month assessment will be assigned the worst rank, ie, worse than any patient surviving to 8 months, but among the deceased the relative ranking will be based on their last value of change from baseline in TSS while alive.

3.2.3 Cardiovascular Death

The efficacy variable is time from randomisation to CV death, confirmed in adjudication. All CV deaths on or prior to PACD will be included. Patients who are alive or died after WoC will be censored at the earliest of date of WoC, last known alive and PACD. Patients who die of any other cause are censored at their date of death.

3.2.4 Death from Any Cause

The efficacy variable is time from randomisation to death from any cause. All deaths on or prior to PACD, including any deaths after WoC, will be included. Patients who are alive or with unknown vital status will be censored at the earliest of date last known alive and PACD.

3.3 Safety Variables

The safety and tolerability of dapagliflozin in patients with HFpEF will be evaluated from SAEs, DAEs, amputations, AEs leading to amputation and AEs reflecting potential risk factors for lower limb amputations ("preceding events").

In addition to amputation, non-serious and serious events potentially placing the patient at risk for a lower limb amputation, in this document denoted "preceding events", should also be recorded in the eCRF as AE/SAE, whether or not an amputation has taken place. Preceding events will be defined for analysis by a predefined list of PRAC PTs. Additional information about amputations with underlying conditions and preceding events will be collected on dedicated eCRF pages.

SAEs will be collected from time of informed consent until and including the patient's last visit. Non-serious AEs will be collected from randomisation until and including the patient's last visit. Collection of non-serious AEs includes cardiac ischaemic events (myocardial infarction and unstable angina), stroke, major hypoglycaemic events, potential DKA, amputations, AE leading to amputation, and preceding events, AEs leading to a potential endpoint, DAEs and AEs which are the reason for interruption of IP.

Efficacy endpoints (deaths and potential HF events) will be adjudicated. These events will be recorded as AEs or, if they fulfil seriousness criteria, as SAEs in the database, but SAEs will not be reported to health authorities to avoid unnecessary unblinding. However, if it is determined by the CEA committee that a potential endpoint does not meet the endpoint criteria, the event will be reported to AstraZeneca patient safety data entry site and if applicable to the health authorities.

For SAEs or DAEs reported by the Investigator as potential DKA, additional information will be recorded on specific eCRF pages in addition to the AE/SAE form. All potential DKA events will be adjudicated by an independent committee and adjudicated outcomes will be considered the main analysis for DKA events.

For myocardial infarctions, unstable angina, stroke, major hypoglycaemic events and amputations, additional information will be recorded on specific eCRF pages in addition to the AE/SAE form.

3.4 Laboratory Values and Vital Signs

Blood samples will be taken for central laboratory assessment of creatinine and calculation of eGFR at enrolment visit, at the visits targeted 1, 4, and 12 months following randomisation, then annually and at PTDV and SCV. eGFR will be calculated (in mL/min/1.73 m^2) using the CKD-EPI formula (Levey at al 2009).

Central laboratory assessment of NT-proBNP and HbA1c will be taken at Visit 1.

Systolic blood pressure, DBP, and pulse rate will be measured at Visit 1, Visit 2, at 1 and 12 months visit, then annually and at PTDV and SCV.

Weight will be measured at Visit 1, at the 12 months visit, then annually and at PTDV and SCV.

3.4.1 Baseline Laboratory Values and Vital Signs

In principle, baseline will be defined as the last value on or prior to date of first dose of randomised IP, or for patients who did not receive treatment, the last value on or prior to date of randomisation. Except for cases of rescreening this will be Visit 1 measurement of weight, NT-proBNP, eGFR and HbA1c, and Visit 2 measurement of SBP, DBP, and pulse rate.

4 ANALYSIS METHODS

4.1 General Principles

No multiplicity adjustment will be made to confidence intervals as they will be interpreted descriptively and used as a measure of precision. All p-values will be unadjusted. P-values for variables not included in the confirmatory testing sequence, or following a non-significant test in the sequence, will be regarded as nominal.

Primary and secondary analyses of HF events and death include adjudicated events occurring on or prior to PACD.

Stratification of analyses for T2D status will be performed using the stratification values as entered in IxRS to determine the randomisation assignment.

Incomplete dates

If only the year part of a date is available (YY), then the date will be set to YY0701. If only the year and month is available (YYMM), then the date will be set to YYMM15. Additional imputation rules will be defined as appropriate to ensure that eg, dates will not be imputed as prior to randomisation, after death or start date after end date.

IP compliance

The percentage of IP compliance for the overall treatment period will be derived for each patient based on pill counts as the number of pills taken (dispensed – returned), relative to the expected number of pills taken. The expected number of pills taken is defined as 1 × (date of last dose – date of first dose + 1), excluding days of interruption. If the number of tablets dispensed or the number of tablets returned is missing for at least 1 observation, compliance is not calculated for that patient.

IP compliance will be presented descriptively, including mean, SD, median, quartiles and 5% and 95% percentiles for SAS by treatment group.

4.1.1 Estimand for Primary and Secondary Outcomes

The primary and secondary event-based objectives will be evaluated under the treatment policy estimand including differences in outcomes over the entire study period until PACD to reflect the effect of the initially assigned randomised IP, irrespective of exposure to IP, concomitant treatment as well as subsequent treatment after discontinuation of IP. The analysis will be performed for the FAS including all events that occurred on or prior to PACD, including events following premature discontinuation of IP.

The estimand for the change from baseline in KCCQ-TSS at 8 months will employ a combination of a treatment policy strategy and a composite strategy. For the intercurrent event of death (due to any cause) prior to the KCCQ assessment at 8 months, a composite strategy will be used, where death will be considered unfavorable and represented by a lowest (worst) rank of a combined outcome variable as described in Section 3.2.2. For all other types of intercurrent events, including but not limited to a premature discontinuation of randomised treatment, a treatment policy strategy will be used.

4.1.2 Hypotheses

The primary endpoint will be tested twice, simultaneously: (1) in the full study population, and (2) in the LVEF < 60% subpopulation.

To control the overall type I error rate at 5% two-sided, the significance level will be adjusted for interim analysis of efficacy performed by the DMC (Section 5). With alpha 0.2% allocated to one planned interim analysis, the significance level in the final analysis will be 4.8%, to be

split between the dual hypotheses. The following null hypothesis will be tested for both the dual analyses of the primary endpoint

H0: HR [dapagliflozin:placebo] = 1

versus the alternative hypothesis

H1: HR [dapagliflozin:placebo] ≠ 1

The secondary endpoints included in confirmatory statistical testing using a closed testing procedure (Section 4.1.3) will be based on similar two-sided alternative hypotheses for the respective treatment difference.

4.1.3 Confirmatory Testing Procedure

A closed testing procedure including a pre-specified hierarchical ordering of the primary and secondary endpoints will be utilised, with recycling of alpha following the framework of Burman et al 2009. The Type I error will be controlled at an overall two-sided 5% level across primary and secondary endpoints and in consideration of the planned interim analysis. Two-sided nominal p-values will be reported for each hypothesis. Statistical significance for a given hypothesis will be declared if the point estimate is in favour of the dapagliflozin arm, in addition to the two-sided p-value meeting the corresponding p-value threshold.

At the final analysis, statistical significance will be assessed in two branches in the pre-specified order of the endpoints and populations as specified in Figure 2. The total significance level, alpha, will be split for the two primary analyses of the primary endpoint, allocating α_1 to test the subpopulation and α_2 to test the full population.

For derivation of the two-sided nominal p-value thresholds α_1 and α_2, in the first step of the MTP, a two-sided alpha of 0.2% will be allocated to the interim analysis and 4.8% to the final analysis. The significance level α_2 (for the primary analysis in the full population at the final analysis) will be fixed at 2.4% two-sided. The inherent correlation structure between the full population and the LVEF < 60% subpopulation, where the corresponding test statistics for the primary endpoint are bivariate normal with correlation equal to the proportion of events in the LVEF < 60% subpopulation, will be taken into account when calculating α_1 (Spiessen and Debois 2010). For the calculation of α_1, the correlation will be based on the square root of the lower bound of a two-sided 95% confidence interval for the proportion of events in the subpopulation with LVEF < 60%, using a normal approximation confidence interval for the proportion. The threshold α_1 will be such that for α_2 = 2.4% two-sided; the two-sided probability of rejecting at least one true null hypothesis at the final analysis will be 4.8%. It then follows that if the primary endpoint in the full population at interim analysis is assessed versus a two-sided p-value of 0.2%, the two-sided probability of rejecting at least one true primary null hypothesis at any analysis can be no larger than 5%. Table 2 shows how the two-

sided nominal p-value threshold α_1 depends on the proportion of events in the LVEF < 60% subpopulation at the final analysis. R and SAS code for calculating α_1 is provided in Appendix B.

Table 2 Level of α_1 Depending on Proportion of Events in LVEF < 60% Subpopulation

Patients with event (LVEF < 60% / overall)	Proportion (95% CI)	Correlation = sqrt of lower confidence limit	Two-sided alpha (%) for primary endpoint		
			Interim analysis Full population	Final analysis (α_2) Full population	Final analysis (α_1) Subpopulation LVEF < 60%
780/1117	0.698 (0.671, 0.725)	0.819	0.2	2.4	3.647
790/1117	0.707 (0.681, 0.734)	0.825	0.2	2.4	3.674
800/1117	0.716 (0.690, 0.743)	0.831	0.2	2.4	3.701
810/1117	0.725 (0.699, 0.751)	0.836	0.2	2.4	3.730
820/1117	0.734 (0.708, 0.760)	0.842	0.2	2.4	3.758
830/1117	0.743 (0.717, 0.769)	0.847	0.2	2.4	3.788

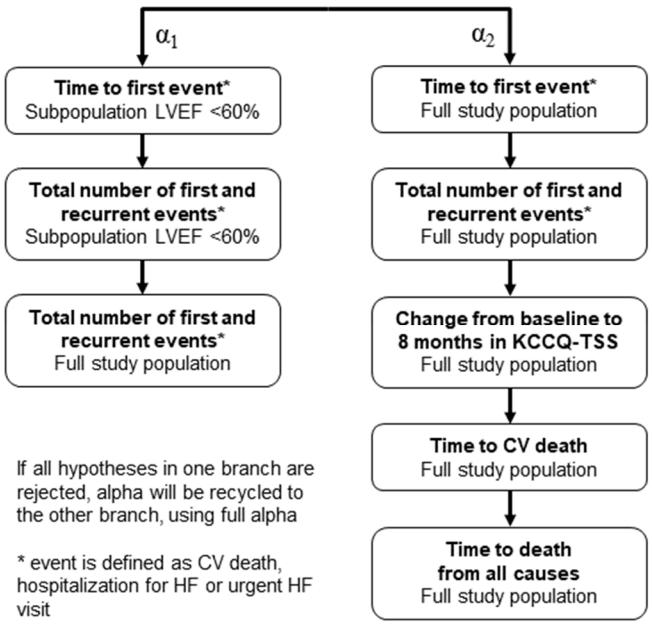
CI, Confidence interval; LVEF, left ventricular ejection fraction; sqrt square root.

- If both the primary null hypotheses can be rejected, the following hypotheses in each branch will be tested at 2.4%, in the order described in Figure 2.
- The following will apply if only one of the tests of the primary endpoint can be rejected at respective levels 2.4% (in the full population) and α_1 (in the LVEF < 60% subpopulation): the remaining hypotheses in the branch where the primary hypothesis was rejected will be tested in fixed sequence at the following two-sided significance levels
  - 4.8% – 2.4% = 2.4% in the left branch only (in case the primary endpoint in the subpopulation was significant at level α_1 but not in the full population at level 2.4%)
  - 4.8% – α_1 in the right branch only (in case the primary endpoint in the full population was significant at level 2.4% but not in the subpopulation at level α_1)
- If all hypotheses in one branch are rejected, alpha will be recycled to the other branch, where remaining unrejected hypotheses can be tested at full alpha adjusted for interim analysis (ie, 4.8%) in the order described in Figure 2.
- If the first secondary hypothesis (recurrent HF events and CV death) in full study population is rejected in one of the branches, it does not have to be re-tested in the other branch. If the primary hypothesis is rejected in both branches and the first secondary

hypothesis (recurrent events) is rejected in the LVEF < 60% subpopulation, then the first secondary hypothesis in full population can be tested at full alpha adjusted for interim analysis (4.8%).

If the study is stopped at the efficacy interim analysis (Section 5), testing of remaining secondary endpoints will be performed in the full study population only, in fixed sequence at two-sided alpha of 0.2% in the order described in the right branch of Figure 2.

Figure 2 Testing Procedure



CV, cardiovascular; HF, heart failure; KCCQ, Kansas city cardiomyopathy questionnaire; LVEF, left ventricular ejection fraction; TSS, total symptom score

4.1.4 Presentation of Time-to-Event Analyses

In general, summary tables of time-to-event analyses will include the number and percent of patients with event per treatment group, event rate, HR with 95% confidence interval and p-value. The event rate will be derived as the number of patients with event divided by the total duration of follow-up across all patients in the given group.

Kaplan-Meier estimates of the cumulative proportion of patients with events will be calculated and plotted per treatment group, with the number of patients at risk indicated below the plot at specific time points. The KM plots will be presented for all time to event analyses, including the individual components of the composite endpoints.

4.1.5 Vital Status and Follow-up of Endpoints

Potential HF endpoints and deaths will be collected and adjudicated from randomisation throughout the study until and including the patient's last visit. The investigator will attempt to collect vital status (dead or alive) at the end of the study for all patients, including vital status from publicly available sources for patients who have withdrawn consent, in compliance with local privacy laws/practices.

Known vital status at the end of the study will be defined when the patient is dead or has date last know alive on or after the PACD. In patient disposition the number of patients who are dead, alive or with unknown vital status will be reported separately for patients who did/did not withdraw consent. The term LTFU will be limited to only patients with unknown vital status.

Follow-up of the primary endpoint will be defined in terms of completion of the event assessment question for a potential HF event as described for censoring in Section 3.1. Thus, a patient that is not LTFU, ie, with known vital status, may have incomplete follow-up of endpoints.

Complete follow up of the primary endpoint will be defined when the patient had a primary endpoint event, died from non-CV death (including undetermined death) or had complete event assessment on or after the PACD (ie, the patient was not censored due to incomplete follow-up of endpoints).

In addition to the number and percent of patients with complete follow-up, the proportion of total patient time with complete follow-up will be reported per treatment group.

Patient time with complete follow-up will be defined as time from randomisation until the earliest of first primary endpoint event, death, WoC, censoring where last complete event assessment is prior to PACD or PACD. The denominator, representing maximum complete follow-up, will be the time from randomisation until the earliest of first primary endpoint event, death or PACD.

4.2 Analysis Methods

4.2.1 Demographics and Baseline Characteristics

Demographic and baseline characteristics, including medical history, will be summarized, using frequency distributions and summary statistics based on the FAS, for each treatment group as well as for all patients combined. No statistical test will be performed for comparison of any baseline measurement among treatment groups.

4.2.2 Concomitant and Baseline Medication

Baseline medication is defined as medication with at least one dose taken before date of randomisation and with no stop date before date of randomisation.

Concomitant medication is defined as medications taken post randomisation, irrespective of IP.

The proportion of patients taking baseline and concomitant medication will be presented for the FAS per ATC class and treatment group. Summaries of prohibited medication, in this study limited to open label SGLT2 inhibitor taken while on IP, will be presented.

4.2.3 Analysis of the Primary Efficacy Variables

Dual primary analyses will be performed simultaneously for the primary composite endpoint, (1) in the full population based on the FAS as well as (2) in the LVEF < 60% subpopulation. The same procedure described below will be used for both of these analyses.

The primary variable is the time to first event included in the primary composite endpoint. The primary analysis will be based on the ITT principle using the FAS, including events with onset on or prior to PACD, adjudicated and confirmed by the CEA committee.

In the analysis of the primary composite endpoint, treatments (dapagliflozin versus placebo) will be compared using a Cox proportional hazards model with a factor for treatment group, stratified by T2D status at randomisation. The analysis will use WoC, non-CV death, last clinical event assessment and PACD for censoring of patients without any primary event as described in Section 3.1. The Efron method for ties and p-value based on the Wald statistic will be used. Event rates, p-value, HR, and 95% confidence interval will be reported.

The contribution of each component of the primary composite endpoint to the overall treatment effect will be examined. In the analysis of the components, all first event of the given type will be included irrespective of any preceding non-fatal composite event of a different type. Consequently, the sum of the number of patients with individual events in the component analysis will be larger than the number of patients with a composite outcome. Methods similar to those described for the primary analysis will be used to separately analyse the time from randomisation to the first occurrence of each component of the primary composite endpoint.

Kaplan-Meier estimates of the cumulative proportion of patients with event will be calculated and plotted, for the composite endpoint and for the individual components.

4.2.3.1 Subgroup Analysis of the Primary Endpoint

Exploratory subgroup analyses of the primary composite endpoint will be performed for the characteristics listed in Table 3 for both full population and LVEF < 60% subpopulation. A

test of interaction between randomised treatment group and the subgroup variable will be performed using Cox proportional hazard model stratified by T2D status at randomisation with factors for treatment group, the subgroup variable and the interaction between treatment and subgroup. In addition to the number and percent of patients with event, event rate estimate, HR with 95% confidence interval and p-value for each subgroup, the interaction p-value will be presented. Hazard ratio estimates, confidence intervals and p-values are not presented for subgroups with less than 15 events in total, both arms combined. HRs with confidence interval will be presented in a forest plot, including number of patients with event and interaction p-value. The p-values for the subgroup analyses and interaction will not be adjusted for multiple comparisons as the tests are exploratory and will be interpreted descriptively.

Table 3 Characteristics and Categories for Subgroup Analysis of the Primary Endpoint

Characteristic	Categories
Age at enrolment (years)	≤ median, > median
Sex	Male, Female
Race	White, Black or African American, Asian, Other
Geographic region	Asia (China, Japan, Taiwan, Vietnam) Europe and Saudi Arabia (Belgium, Bulgaria, Czech Republic, France, Hungary, Netherlands, Poland, Romania, Russia, Saudi Arabia, Spain) North America (Canada, US) Latin America (Argentina, Brazil, Mexico, Peru)
NYHA class at enrolment	II, III/IV
LVEF at enrollment (%)	≤ 49, 50 to 59, ≥ 60
NT-proBNP at enrollment (pg/mL)	≤ median, > median
Randomised during hospitalisation for HF or within 30 days of discharge.	Yes, No
eGFR at enrolment (mL/min/1.73m^2)	< 60, ≥ 60
BMI at enrolment (kg/m2)	< 30, ≥ 30
T2D at enrolment [a]	Yes, No
SBP at randomisation	≤ median, > median
Atrial fibrillation or flutter at enrolment ECG	Yes, No

[a] The subgroup analysis by T2D status will be based on eCRF medical history record and exclude T2D as a stratification factor from the model

BMI, body mass index; ECG, electrocardiogram; eGFR, estimated glomerular filtration rate; HF, heart failure; LVEF, left ventricular ejection fraction; NT-proBNP, N-terminal pro b-type natriuretic peptide; NYHA, New York Heart Association; SBP, systolic blood pressure; T2D, type 2 diabetes

The subgroup analyses will be repeated for CV death and the HF event (hospitalisation for HF and urgent HF visit) component of the primary composite endpoint.

4.2.3.2 Sensitivity Analysis of the Primary Endpoint

Undetermined cause of death

A sensitivity analysis of the primary analysis where deaths adjudicated as 'undetermined' cause are considered as CV deaths and included as endpoint events will be performed.

Missing data and informative censoring

The time-to-event analysis using the Cox regression depends on the assumption of non-informative or ignorable censoring, corresponding to the missing-at-random assumption. The missing data in this context are patients who are prematurely censored due to WoC, LTFU or otherwise incomplete follow-up of endpoints. The amount of missing data will be described eg, in terms of the number of patients and patient time with incomplete follow-up as described in Section 4.1.5.

Patient retention and follow-up are at the forefront of study planning and conduct, and the amount of incomplete follow-up is expected to be small.

To assess the effect of incomplete follow up of the primary endpoint, a sensitivity analysis may be performed where time to event information is imputed for patients with premature censoring (censored before PACD due to WoC or incomplete primary event assessment). Event rates will be estimated separately in the two T2DM strata by an exponential distribution with constant hazard rate over time. Using the hazard ratio from the primary analysis, the event rates will be calculated for the dapagliflozin group, separately for the T2DM strata (by multiplying the corresponding placebo group rates by the hazard ratio estimated in the primary analysis). Using the estimated event rates, new event times will be simulated for patients with premature censoring from the exponential distribution. If the simulated time is in the interval from the censoring date to PACD (or death date, whichever came first), a new event will be imputed at the resulting event time. Otherwise, if the simulated time is outside the interval from the original censoring to PACD or death, the patient will be considered censored at PACD or death. The primary analysis will thereafter be conducted again, supplemented by the simulated time-to-event information. The process is to be repeated 1000 times and the resulting hazard ratios and standard errors will be combined using the Rubin's rule.

A tipping point analysis may be conducted to assess the robustness of the statistical significance of the primary analysis. While keeping the placebo event rates constant at the estimated values, the event rates in the dapagliflozin group will gradually be increased by increasing the hazard ratio from the primary analysis until the test of the primary endpoint no longer is statistically significant.

COVID-19

Subjects affected by COVID-19 infection will be defined by pre-specified preferred terms for adverse events associated with COVID-19 infection. A COVID-19 sensitivity analysis of the primary endpoint (and components) will be performed where the main analysis of the primary endpoint will be done, where patients and events are censored at the onset date of AE associated with COVID-19 infection. In this setting, onset of COVID-19 can be assumed to be unrelated to randomised treatment and as such should not introduce informative censoring while accounting for impact of COVID-19 infection in the main analysis.

4.2.4 Analysis of the Secondary Efficacy Variables

4.2.4.1 Analysis of Total Number of HF Events (First and Recurrent) and CV Death

The composite outcome of total number of HF events (first and recurrent) and CV death with onset on or prior to PACD, adjudicated and confirmed by the CEA committee, will be analysed by the semi-parametric proportional rates model (Lin et al 2000; known as the LWYY method) to test the treatment effect and to quantify the treatment difference in terms of the rate ratio with 95% confidence interval and p-value. If a HF event and CV death occurred at the same day, then only CV death will be counted.

In addition, the two components in the composite endpoint (total number of HF events and CV death) will be analysed separately to quantify the respective treatment effects and check the consistency between the composite and the components. For the analysis of total number of HF events component, occurrence of CV death can be regarded as semi-competing risk (informative censoring) and may introduce a bias in the treatment effect estimate for HF events (dilution of effect size if the drug has a positive effect on both components). To address this concern and to account for the correlation between the two components, the joint modelling (frailty model) approach (Rogers et al 2016) will be used for the component analyses. Non-parametric estimates of HF event rates over time allowing for death as terminal event will be provided as well (Ghosh and Lin 2000).

COVID-19

A COVID-19 sensitivity analysis of the first secondary endpoint (and components) will be performed where the main analysis LWYY will be applied and where patients and events are censored at the onset date of AE associated with COVID-19 infection. In this setting, onset of COVID-19 can be assumed to be unrelated to randomised treatment and as such should not introduce informative censoring while accounting for impact of COVID-19 infection in the main analysis.

4.2.4.2 Analysis of Change from Baseline at 8 Months in the KCCQ Total Symptom Score

Hypothesis testing

The composite rank-based endpoint representing the patients' vital status at 8 months and the change from baseline at 8 months in KCCQ-TSS in surviving patients, as defined in Section 3.2.2, will be analysed using the rank ANCOVA method (Stokes et al 2012) to test the null hypothesis of no difference in the distributions of ranked outcomes between the two treatment groups. Analysis will be stratified by T2D status at randomisation, and adjusted for the baseline KCCQ-TSS value as follows.

First the change from baseline at 8 months in KCCQ-TSS and vital status at 8 months, as well as values of the baseline KCCQ-TSS covariate will be transformed to standardised ranks within each T2D randomisation stratum, using fractional ranks and mean method for ties. Ranking for the composite endpoint will be done so that patients who died prior to the 8-month assessment are assigned the worst ranks within each stratum. Among the deceased, the relative ranking will be based on their last value of change from baseline in KCCQ-TSS while alive before deriving fractional ranks. In the ranking, patients who die prior to the first follow-up visit where KCCQ-TSS is assessed, at 1 month, will be defined as having a zero change from baseline while alive. Then, separate regression models will be fit to the ranked data for each randomisation stratum using a regression model for the ranked composite variable as dependent variable, adjusting for the ranked baseline covariate. Residuals from this regression model will be captured for testing of differences between treatment groups. The CMH test, stratified by T2D status at randomisation, using the values of the residuals as scores will be used to compare treatment groups.

KCCQ data missing for reasons other than death will be imputed as described in Section "Handling of missing KCCQ data".

The p-value from the CMH test of treatment effect at 8 months will be the used for the confirmatory testing of the secondary endpoint in the MTP described in Section 4.1.3.

COVID-19

Due to COVID-19 pandemic, on-site assessments could not be performed in a substantial number of sites, where some were done remotely and some cancelled. Furthermore, it could be assumed that lock-downs and other measures could impact PRO assessments. As a consequence, the main analysis of this endpoint includes the population with patients who had a planned or performed 8 month assessment (Visit 5) prior to the major COVID-19 outbreak, defined as 11th March 2020 (the date when WHO declared COVID-19 a pandemic) thus unaffected by the pandemic's possible impact on health-related quality of life (FDA 2020). The KCCQ-TSS in the presence of COVID-19 pandemic will be described.

Estimation of treatment effect

Win ratio

For a summary statistic that uses the same ranking as that used in the hypothesis test, but has a clinical interpretation, the WR and the corresponding 95% confidence interval (Wang and Pocock 2016) will be reported. It is noted that the WR differs from the statistic used for hypothesis testing, so that exact consistency is not expected between these two analyses, eg on rare occasions, the 95% confidence interval for WR could exclude unity while the p-value for the pre-planned hypothesis test could be > 0.05, or the hypothesis test could be < 0.05 with the confidence interval for WR including unity. Formal inference for the superiority of the treatment over control will be made only from the pre-planned hypothesis test based on the WR.

The win ratio represents the odds of having a more favourable outcome versus a less favourable outcome when assigned to the dapagliflozin 10 mg treatment group as opposed to placebo. More specifically, each patient in the dapagliflozin group is compared with each patient in the placebo group and each pair is labelled as "winner", "loser", or "tie", depending on whether the patient on dapagliflozin has a more favourable, less favourable, or the same outcome, respectively, with respect to the composite ranked endpoint compared to the patient on placebo. Win ratio is defined as the ratio of the number of "winner" pairs to the number of "loser" pairs for the dapagliflozin arm. If the estimated win ratio is greater than 1 then the treatment effect is in favour of dapagliflozin.

The win ratio statistic adjusted for the randomisation stratification factor and baseline KCCQ-TSS will be obtained using the methodology in (Koch et al 1998, Kawaguchi et al 2011) for the stratified Mann-Whitney estimators for the comparison of two treatments with randomisation based covariance adjustment. The win ratio statistic will be calculated as Mann-Whitney odds, ie, WR = MW / (1 – MW), where MW is the adjusted Mann-Whitney estimate. This transformation is monotonous in the domain of the Mann-Whitney estimate. The 95% confidence interval for the win ratio will be obtained by transforming the bounds of the confidence interval (Koch et al 1998) for the Mann-Whitney estimate, using the same transformation as for the win ratio.

Responder analysis

Number and percentage of patients in each treatment group will be summarised across the following categories, where change from baseline is defined as KCCQ-TSS at 8 months minus KCCQ-TSS at baseline:

Thirteen point improvement from baseline at 8 months in KCCQ-TSS, identified as a clinically meaningful improvement in anchor-based analyses (see Appendix A), vs no clinically meaningful improvement:

- Change from baseline in KCCQ-TSS ≥ 13 points, vs
- Death prior to the 8 months assessment or change from baseline in KCCQ-TSS < 13 points.

Five point deterioration from baseline at 8 months in KCCQ-TSS, identified as a clinically meaningful deterioration in anchor-based analyses (see Appendix A), vs no clinically meaningful deterioration:

- Death prior to the 8 months assessment or a negative change from baseline in KCCQ-TSS ≥ 5 points, vs
- Change from baseline at 8 months in KCCQ-TSS that is positive or, if negative, is smaller than 5 points.

Patients who had a baseline value of KCCQ-TSS ≥ 100 – 13 = 87 points (ie, too close to the "ceiling" to have a clinically meaningful improvement based on the instrument), will be defined as having achieved "responder status" for improvement only if the following conditions are both met: KCCQ-TSS remains ≥ 87 points at 8 months and KCCQ-TSS ≥ baseline at 8 months (ie, they had no deterioration from their baseline score). Similarly, for clinically meaningful deterioration, patients who had a baseline value of KCCQ-TSS ≤ 5 points (ie, too close to the "floor" to have a clinically meaningful deterioration based on the instrument), will be defined as having achieved "responder status" for deterioration only if KCCQ-TSS remains ≤ 5 points at 8 months and KCCQ-TSS ≤ baseline at 8 months (ie, they had no improvement from their baseline score).

The proportion of patients in the different KCCQ-TSS responder categories will be compared between treatment groups using a logistic regression model including treatment group, stratification variable (T2D at randomisation) and baseline KCCQ-TSS value. The observed number and proportion of KCCQ-TSS responders, odds ratio between treatment groups, its corresponding 2-sided 95% confidence interval and p-value estimated from each imputed dataset will be combined using Rubin's rule, and the combined results will be presented.

Additional responder analysis will be performed in the same way as described above, for 17 points improvement ("large improvement") and 14 points deterioration ("large deterioration"). These thresholds of clinically meaningful change from baseline KCCQ-TSS were derived from anchor-based analyses of blinded study data as described in Appendix A. In these analyses, "ceiling" and "floor" values are handled in an analogous way as for the analysis of 13 points improvement and 5 points deterioration.

Empirical cumulative distribution function plots will be presented by treatment group to summarize the distribution of change from baseline at 8 months in KCCQ-TSS values, where patients who die prior to the 8-month assessment will be represented with the value of -101 (a value below the worst possible change from baseline).

Handling of missing KCCQ data

The number of patients with missing vital status at 8 months is expected to be negligible. If some patients are LTFU or withdrew consent and have unknown vital status, the main analysis will be done with these patients assigned the worst ranks (same as deaths, described below).

In the context of analysing the composite ranked endpoint as described above, missing data may arise when patients miss the 8-month KCCQ assessment while remaining in the study during the 8-month assessment window (+/- 14 days will be used), or when patients withdraw consent from the study prior to 8 months. If a patient is known to have died prior to the 8-month assessment, the patient is considered to have a non-missing composite outcome and will be handled as described above (assigned the worst rank). Otherwise, patients who are alive at 8 months and have missing baseline or 8-month KCCQ assessments will have their missing KCCQ-TSS imputed using the multiple imputation methodology as follows.

Missing KCCQ-TSS values at baseline or at 8 months will be imputed under the Missing at Random assumption. The imputation will be done using a predictive mean matching multiple imputation model and a method of Fully Conditional Specification as implemented in the SAS Procedure MI (FCS statement). The predictive mean matching method ensures that the imputed values remain in the permissible range of the KCCQ-TSS values. Imputation will be done sequentially, ie, imputing each time point in their chronological order and the imputations at a given time point will be informed by preceding imputed time points. The imputation model will include the treatment group, T2D randomisation stratum, prior KCCQ-TSS (at baseline, month 1 and month 4), and three categorical variables representing the number of HF events (categorised as 0, 1 or ≥ 2) in the intervals from randomisation to 1 month, from 1 to 4 months, and from 4 to 8 months, respectively, depending on the time point being imputed. Occurrences of HF events will be determined based on the investigator-reported potential HF events. Auxiliary variables related to HF events are included in the imputation model to improve the imputation accuracy, because the occurrence of HF events is expected to be associated with HF symptoms as assessed by KCCQ-TSS.

The number of closest observations used to sample an imputed value by the predictive mean matching method will be 5 (SAS default setting).

Each imputed dataset will be analysed using the methods described in the "Hypothesis testing" and "Estimation of treatment effect" sub-sections above. The results from multiple imputed datasets will be combined using Rubin's rule as implemented in the SAS Procedure MIANALYZE.

- In the analysis of rank ANCOVA, the CMH tests statistic used for the hypothesis test has a chi-square distribution. In order to apply Rubin's combination rule, which assumes approximate normal distribution of the statistics being combined, a normalising Wilson-

Hilferty transformation will be applied to the CMH test statistics from each imputed dataset (Ratitch et al 2013). The standardized transformed statistic will be computed as follows:

$$st_{wh\_cmh}{}^{(m)} = \frac{\sqrt[3]{\frac{cmh^{(m)}}{df}} - \left(1 - \frac{2}{9 \times df}\right)}{\sqrt[2]{\frac{2}{9 \times df}}}$$

where $cmh^{(m)}$ is the CMH statistic from the m[th] imputed dataset and df is the number of degrees of freedom associated with the statistic (in this case equal 1). The transformed statistics are approximately normally distributed with mean of 0 and variance of 1 and can be combined using Rubin's rule.

- For the estimation of the win ratio, a combined Mann-Whitney estimate and its standard error will first be obtained by applying Rubin's rule to the corresponding estimates from multiple imputed datasets. Then the win ratio and its 95% confidence interval will be obtained based on the combined Mann-Whitney estimate and its standard error as previously described.
- For the summaries of number and percentage of subjects in the categories of significant improvement and deterioration from baseline, the number and percent of subjects with actual observed improvement and observed deterioration/death respectively will be reported. The estimation of odds ratio and confidence intervals for the KCCQ-TSS responder analyses will use the imputation datasets created for the main analysis. Therefore, deaths will be defined as non-responders, and responder status will be determined based on the imputed KCCQ-TSS values for the patients who have missing KCCQ-TSS due to reasons other than death.

Supportive analyses and sensitivity analyses for KCCQ

The number and percent of patients who die prior to the 8-month assessment will be summarized by treatment group.

Descriptive statistics of scores and change from baseline at 1, 4 and 8 months will be presented for TSS, overall summary score, clinical summary score and domains (physical limitation, symptom stability, symptom frequency, symptom burden, quality of life, self-efficacy and social limitation).

The testing and estimation described for change from baseline at 8 months in KCCQ-TSS, will be repeated in an exploratory fashion for change from baseline in KCCQ-TSS at 1 and 4 months, and for the overall summary score and clinical summary scores at 1, 4 and 8 months.

4.2.4.3 Analysis of CV death

Time to CV death will be analysed using Cox regression in the same manner as the primary composite endpoint, with stratification for T2D status at randomisation. The analysis will include CV deaths, confirmed in adjudication, occurring on or prior to PACD. Patients who did not die from CV death, will be censored at the earliest of death due to other cause, WoC, PACD, or for any patients who are LTFU, at last date known to be alive.

COVID-19

As part of the COVID-19 related sensitivity analysis of the primary endpoint, the component CV death will be reported (Section 4.2.3.2).

4.2.4.4 Analysis of All-Cause Mortality

Time to death from any cause will be analysed using Cox regression in the same manner as the primary composite endpoint, with stratification for T2D status at randomisation. The analysis will include deaths from any cause occurring on or prior to PACD. Patients who are alive will be censored at PACD, or for any patients who are LTFU, at last date known to be alive.

4.2.5 Analysis of Safety Variables

Analysis set

For safety analyses, all summaries will be based on the SAS (Section 2.1.2).

Exposure

The total exposure to IP will be defined as the length of period on IP, calculated for each patient as date of last dose – date of first dose +1.

An alternative measure where days of interruption are removed will be calculated and termed actual exposure.

Total and actual exposure will be presented descriptively.

Treatment periods

The summaries for the on-treatment period will include events with an onset date on or after first dose of randomised IP and on or before 30 days after last dose of IP. Additional presentations will include all events with onset on or after first dose of IP regardless of whether patients are on or off IP at the time of the event (the "on- and off-" treatment period.). Patients who complete the study on IP will discontinue treatment on the SCV. Thus, there will in general be no events after completion of the IP period, and censoring of events for on-treatment analysis affects only patients who prematurely and permanently discontinue IP.

All summaries of AEs described in Section 4.2.5.1 to 4.2.5.4 below will be presented for the on-treatment period and on- and off- treatment period.

4.2.5.1 Adverse Events

The on-treatment period was used for primary analysis of all safety variables, except for amputations and preceding events, for which the on- and off-treatment period was considered the primary approach.

In addition to SAEs, the collection of AEs that are not serious includes myocardial infarction, unstable angina, stroke, major hypoglycaemic events, potential DKAs, amputations, AEs leading to amputation, and preceding events, AEs leading to a potential endpoint, DAEs, and AEs which are the reason for interruption of IP (see Section 3.3). Thus, summaries of AEs will be limited to these categories and general summaries of all non-serious AEs are not planned.

AEs will be classified according to MedDRA by the medical coding team at AstraZeneca data management center, using MedDRA 24.1.

Summaries by SOC and PT will be sorted by international order for SOC and by descending order of PT in the dapagliflozin treatment group.

No statistical tests to compare crude AE frequencies between treatment groups are planned. A summary table of the total number and percent of patients with AE with outcome death, AEs of definite or probable DKA, any major hypoglycemic event, SAE, DAE, AE leading to temporary interruption of IP, AEs possibly related to IP, amputations and preceding events per treatment group will be provided.

Amputations, AEs leading to amputations, and preceding events (see Section 3.3) will be presented in summary tables including the number and percent of patients with any event in the AE category, SAE, DAE and AE leading to interruption, and tabulated with frequency by PT.

In addition to presentations of the number of patients with event, the total number of events counting multiple events per subject will be presented.

All potential events of DKA will be submitted to an independent DKA Adjudication Committee. The adjudicated outcome, definite or probable, will be considered the main analysis for DKA.

For major hypoglycaemic events a summary table including the total number of subjects with events, the number and percent of patients with event in the AE intensity category, SAE, DAE, AE leading to interruption, possible relation to IP will be presented. The presentation of on-treatment events, on- and off-treatment presentations will be provided for all major hypoglycaemic events.

For AEs leading to amputations and preceding events, DKA and major hypoglycaemic events, event rate per 100 subject years will also be presented, calculated as 100 times the number of patients with event divided by the total duration of treatment (including 30 days after last dose) in the given group for the on-treatment presentation, and total duration of follow-up in the given group for on and off treatment.

Events of genital area infections and necrotising fasciitis to be medically assessed in a blinded fashion prior to clinical data lock as potential events of Fournier's gangrene will be presented in a summary table including the number and percent of patients with any event in the SAE or DAE category, and tabulated with frequency by PT.

4.2.5.2 Serious Adverse Events

SAEs will be presented as described below both on treatment and on and off treatment.

The number and percent of patients with SAEs will be presented by SOC, PT and treatment group. The most common SAEs will also be presented by PT only.

AEs with outcome death will be presented separately by SOC and PT.

4.2.5.3 Adverse Events Leading to Discontinuation or Interruption of Investigational Product

The number and percent of patients with event will be presented by SOC and PT for AEs leading to discontinuation of IP and AEs leading to temporary interruption (separately for the two categories based on action taken "Drug Permanently Discontinued" and "Drug Interrupted" respectively, recorded in the CRF AE module).

4.2.5.4 Laboratory Evaluation and Vital Signs

Summaries of creatinine and calculated eGFR will be based on creatine samples analysed at the central laboratory.

The result and the change from baseline of creatinine, eGFR and vital signs, will be summarized by treatment group at each visit (excluding PTDV and SCV) with scheduled measurement (see Section 3.4) using descriptive statistics, including n, mean, SD, range, median, and quartiles.

4.2.6 Analysis of Exploratory Objectives

Time to the first occurrence of hospitalisation from any cause will be analysed with the same method as the primary endpoint, based on information on the SAE eCRF form.

The proportion of patients with worsened (higher) NYHA class at 8 months compared to baseline, including patients who died prior to 8 months in the worsened category, versus patient with improved or unchanged NYHA class, will be analyzed by logistic regression with

treatment group, baseline NYHA class and T2D status at randomisation as factors, presented as an odds ratio with corresponding 95% confidence interval. Only NYHA assessments made at site or through phone visits with the patient to be used in analyses.

Change from baseline to each scheduled assessment visit (see Section 3.4) for body weight, SBP and eGFR will be analysed with a MMRM. All non-missing visit data will be used, including measurements after discontinuation of IP. The model will include terms for treatment group, visit, visit by treatment group interaction, the baseline measurement and T2D stratification status at randomisation as covariates. The model will be used to derive a least-squares estimate of the treatment difference with 95% confidence interval and corresponding two-sided p-value. Missing data will not be imputed.

For eGFR, the MMRM model above will additionally be used to derive the "total" slopes (between randomisation and eg, 1 year and 2 years respectively) and the "chronic" slopes (between a post randomisation time point to eg, 1 year and 2 years respectively) will be estimated via linear contrasts.

The analysis of change from baseline in KCCQ clinical summary score, overall summary score and KCCQ-TSS sub-scores (symptom burden and symptom frequency) will follow the analysis of KCCQ-TSS in Section 4.2.4.2. QoL score will be summarised using descriptive statistics.

EQ-5D-5L derived utility score will be summarised by descriptive statistics, and used for health economic modelling and reported in a separate health economic report.

5 INTERIM ANALYSES

An interim analysis is planned to be performed including approximately 67% of the target number of patients with adjudicated primary endpoint events (approximately 748 events). There will in principle be one planned interim analysis for efficacy, with the possibility of the DMC to conduct subsequent interim analysis if they deem necessary. The significance level for final analysis will be based on the actual number of interim analyses. The interim analysis will assess superiority of dapagliflozin to placebo. The interim analysis will have a nominal two-sided alpha level of 0.2%. At the interim analysis, the primary composite endpoint will be tested in the full study population at the specified alpha level. If superiority is achieved for the primary endpoint, then the superiority of dapagliflozin to placebo on CV deaths will be tested in the full study population at a two-sided level of 0.2%. If CV death is significant, then an action is triggered whereby the DMC will evaluate the totality of the efficacy data and safety data, to determine if benefit is unequivocal and overwhelming such that the DMC recommends ending the study.

If the interim analysis leads to a decision to terminate the study early based on pre-defined stopping guidelines, the executive committee will define a PACD, on or after which SCVs will commence. The study report will be based on all events occurring on prior to the PACD.

If the study is stopped at the efficacy interim analysis, testing of remaining secondary endpoints will be performed on the final database in the full population only, in fixed sequence described in the right branch of Figure 2 (Section 4.1.3) at two-sided significance level 0.2%.

6 CHANGES OF ANALYSIS FROM PROTOCOL

The alpha for final analysis adjusted for interim analysis at alpha 0.2% will be set to 5% minus 0.2% = 4.8%, rather than 4.98% as determined by the Haybittle-Peto function for 67% of events (Sections 9.1 and 9.5 of the CSP).

7 REFERENCES

Burman et al 2009
Burman CF, Sonesson C, Guilbaud O. KF. A recycling framework for the construction of Bonferroni-based multiple tests. Stat Med. 2009;28:739–61.

Coon and Cook 2018
Coon CD, Cook KF. Moving from significance to real-world meaning: methods for interpreting change in clinical outcome assessment scores. Qual Life Res 2018 Jan;27(1):33–40.

FDA 2020
Statistical considerations for clinical trials during the COVID-19 public health emergency. Guidance for industry. June 2020. US Department of Health and Human Services. Food and Drug Administration.

Fitchett et al 2016
Fitchett D, Zinman B, Wanner C, Lachin JM, Hantel S, Salsali A, et al. Heart failure outcomes with empagliflozin in patients with type 2 diabetes at high cardiovascular risk: results of the EMPA-REG OUTCOME® trial. Eur Heart J 2016;14;37(19):1526-34.

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

Green et al 2000
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. JACC, 2000 Apr;35(5):1245–55.

Hicks et al 2018
Hicks KA, Mahaffey KW, Mehran R, Nissen SE, et al. Cardiovascular and Stroke Endpoint Definitions for Clinical Trials. J Am Coll Cardiol 2018;71:1021–34.

Kawaguchi et al 2011
Kawaguchi A, Koch G, Wang X. Stratified Multivariate Mann-Whitney Estimators for the Comparison of Two Treatments with Randomization Based Covariance Adjustment. Statistics in Biopharmaceutical Research, 2011;3(2):217–231.

Koch et al 1998
Koch GG, Tangen CM, Jung JW, Amara IA. Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them. Statist. Med 1998;17(15-16):1863–1892.

Kristensen et al 2015
Kristensen SL, Jhund PS, Køber L, McKelvie RS, Zile MR, Anand IS, et al. Relative importance of history of heart failure hospitalisation and N-terminal Pro–B-type natriuretic peptide level as predictors of outcomes in patients With heart failure and preserved ejection fraction. JACC Heart Fail 2015 Jun;3(6):478–86.

Kristensen et al 2017
Kristensen SL, Mogensen UM, Jhund PS, Petrie MC, Preiss D, Win S, et al. Clinical and echocardiographic characteristics and cardiovascular outcomes according to diabetes status in patients with heart failure and preserved ejection fraction: a report from the I-Preserve Trial (Irbesartan in Heart Failure With Preserved Ejection Fraction). Circulation. 2017;135(8):724–35.

Levey at al 2009
Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF 3rd, Feldman HI, Kusek JW, Eggers P, VanLente F, Greene T, Coresh J; CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration). A new equation to estimate glomerular filtration rate. Ann Intern Med. 2009 May 5;150(9):604–12.

Lin et al 2000
Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2000;62(4):711–730.

Pfeffer et al 2015
Pfeffer MA, Claggett B, Assmann SF, Boineau R, Anand IS, Clausell N, et al. Regional Variation in Patients and Outcomes in the Treatment of Preserved Cardiac Function Heart Failure With an Aldosterone Antagonist (TOPCAT) Trial. Circulation. 2015;131:34–42.

Ratitch et al 2013
Ratitch B, Lipkovich I, O'Kelly M. Combining Analysis Results from Multiply Imputed Categorical Data. PharmaSUG 2013 – Paper SP03

Rogers et al 2016
Jennifer K Rogers, Alex Yaroshinsky, Stuart J Pocock, David Stokar and Janice Pogoda. Analysis of recurrent events with an associated informative dropout time: Application of the joint frailty model. Statist. Med. 2016;35:2195–2205.

Rådholm et al 2018
Rådholm K, Figtree, Perkovic V, Solomon SD, Mahaffey KW, et al. Canagliflozin and Heart Failure in Type 2 Diabetes Mellitus: Results From the CANVAS Program (Canagliflozin Cardiovascular Assessment Study). Circulation. 2018;138:458–468. https://doi.org/10.1161/CIRCULATIONAHA.118.034222

Spertus et al 2005
Spertus J, Peterson E, Conard MW, Heidenreich PA, Krumholz HM, Jones P, et al. Monitoring clinical changes in patients with heart failure: A comparison of methods. Am Heart J 2005;150:707–15.

Spiessen and Debois 2010
Spiessen B, Debois M. Adjusted significance levels for subgroup analyses in clinical trials. Cont Clin Trials 2010; 31:647–656.

Stokes et al 2012
Stokes ME, Davis CS, Koch GG. Categorical Data Analysis Using the SAS System. Cary, NC: SAS Institute Inc.

Wang and Pocock 2016
Wang D, Pocock S. A win ratio approach to comparing continuous non-normal outcomes in clinical trials. Pharmaceut. Statist. 2016;15:238–245.

Appendix A Estimation of Clinically Meaningful Thresholds for KCCQ Total Symptom Score

A 1 Methods

Thresholds for CMWPC will be estimated according to predefined algorithms using an anchor-based approach, supplemented with graphical visualisations of the distribution across anchor categories. Clinically meaningful thresholds will be estimated for change from baseline KCCQ-TSS at 8 months.

This appendix describes the methods which were applied to blinded study data prior to database lock and unblinding of the study, with results and derived thresholds presented in this SAP prior to the interim analysis. The threshold analyses were performed on the FAS population used in the main analysis for KCCQ (the population with patients who had a planned Visit 5, ie, at 8 months, prior to the major COVID-19 outbreak; see 4.2.4.2), on blinded study data across both treatment arms only including patients with complete data at baseline and 8 months.

Anchor-based approaches

Anchor-based approaches estimate a threshold by 'anchoring' the results on a separate variable, often a patient-reported outcome. The anchor-based analysis will employ the PGIS in HF symptoms. Meaningful change will be evaluated using observed scores according to a predefined algorithm. The responses to PGIS at baseline and 8 months will be used in the analysis.

Categorisation of anchors

The change from baseline PGIS at 8 months will be categorized and categories will be collapsed in different ways, to provide a clearer distinction between patients who have and have not experienced a meaningful change according to this anchor.

The ordinal responses to PGIS at baseline and 8 months will be assigned the following numeric values:

- 1 ('no symptoms')
- 2 ('very mild')
- 3 ('mild')
- 4 ('moderate')
- 5 ('severe')
- 6 ('very severe')

Change from baseline PGIS at 8 months will be categorized as small, moderate or large improvement/deterioration or stable as defined in Table A1.

Table A1 Categories of Change from Baseline PGIS in Heart Failure Symptoms at 8 Months

		PGIS at 8 months					
		No symptoms	**Very mild**	**Mild**	**Moderate**	**Severe**	**Very Severe**
PGIS at baseline		**1**	**2**	**3**	**4**	**5**	**6**
No symptoms	1	0 Stable	+1 SD	+2 MD	+3 LD	+4 LD	+5 LD
Very mild	2	-1 SI	0 Stable	+1 SD	+2 MD	+3 LD	+4 LD
Mild	3	-2 MI	-1 SI	0 Stable	+1 SD	+2 MD	+3 LD
Moderate	4	-3 LI	-2 MI	-1 SI	0 Stable	+1 SD	+2 MD
Severe	5	-4 LI	-3 LI	-2 MI	-1 SI	0 Stable	+1 SD
Very severe	6	-5 LI	-4 LI	-3 LI	-2 MI	-1 SI	0 Stable

LD, large deterioration; LI, large improvement; MD, moderate deterioration; MI, moderate improvement; SD, small deterioration; SI, small improvement

The categories in Table A1 will be further collapsed as

- 'moderate or large deterioration' in the categorisation with 5 categories (version A)
- 'small or moderate deterioration' in the categorisation with 5 categories (version B)
- 'small or moderate improvement' in the categorisation with 5 categories (version B)
- 'moderate or large improvement' in the categorisation with 5 categories (version A)

The change from baseline KCCQ-TSS at 8 months, will be used repeatedly in the anchor-based analyses. To explore the adequateness of each anchor categorisation, the Spearman correlation coefficient between change from baseline KCCQ-TSS and change from baseline PGIS at 8 months will be assessed.

The larger the correlation coefficient between an anchor and the endpoint, the greater the confidence in the classifications. An anchor is considered adequate if it has a correlation coefficient of 0.3 or greater (Coon and Cook 2018).

Descriptive statistics (mean, SD, median, quartiles, minimum and maximum) and an eCDF is presented for each categorisation in Section A 2. The eCDF curves display a continuous plot

of the change from baseline on the horizontal axis, and the cumulative proportion of patients experiencing changes from baseline up to that level, on the vertical axis. If the eCDF curves show very poor distinction between categories, they may be complemented with curves illustrating the probability density function for that categorisation.

Establishing the clinically meaningful threshold

The various estimates from the different streams of evidence (tables and plots of the distribution) will be examined for convergence in an effort to triangulate onto a single threshold value which represents CMWPC (for improvement and deterioration, respectively) and the KCCQ-TSS responder analysis will be performed for this threshold. However, if the values are too disparate, a range of clinically relevant thresholds may be identified. CMWPC thresholds identified will be indicated in the eCDF for change from baseline KCCQ-TSS by treatment, in the unblinded results, and responder analysis will be performed for the thresholds.

A 2 Summary of Results of Anchor-Based Analysis on Blinded Study Data

The anchor-based analysis of change from baseline KCCQ-TSS at 8 months in different categories of change from baseline PGIS at 8 months, is presented in Table A2. As this analysis is done on blinded study data and only includes patients with observed values for both KCCQ-TSS and PGIS at 8 months (patients who died and all other patients with missing data are excluded), the “mean” is selected as a representation of the average of a group. This anchor-based analysis indicates that small or moderate improvement corresponds to a mean increase in KCCQ-TSS of 13 points. A large improvement in PGIS corresponds to a mean increase in KCCQ-TSS of about 17 points. A large deterioration in PGIS corresponds to a mean decrease in KCCQ-TSS of about 14 points, whereas a moderate deterioration in PGIS corresponds to a mean decrease in KCCQ-TSS of 5 points. It is important to note that the group of patients who were categorized as being “stable” in terms of their HF symptoms at 8 months had a mean increase in KCCQ-TSS of almost 5 points.

In the responder analysis of the third secondary efficacy endpoint, change from baseline measured at 8 months in the TSS of the KCCQ (Section 4.2.4.2), an increase of 13 points or more in KCCQ-TSS will be considered a clinically meaningful improvement and a decrease of 5 points or more will be considered a clinically meaningful deterioration. The anchor-based analysis and the distribution curves indicate that a “small” improvement cannot be distinguished from a “moderate” improvement, while they are both clearly separated from the “stable” category. Likewise, the anchor-based analysis and distribution curves indicate that a “small” deterioration cannot be distinguished from the “stable” category. The Spearman’s correlation coefficient between change from baseline at 8 months in KCCQ-TSS and PGIS was around 0.29-0.30, where a correlation of 0.3 or greater between an anchor and the anchored scale is considered adequate (Coon and Cook 2018).

Table A2 Distribution of Change from Baseline KCCQ-TSS at 8 months by Change from Baseline PGIS 8 Months

	N	(%)	Mean	SD	Min	Q1	Median	Q3	Max	Correlation [a]
PGIS at 8 Months: 7										0.29
Large Improvement	120	(6)	17.4	22.51	-54.2	0.5	15.1	32.8	70.8	
Moderate Improvement	275	(13)	12.9	20.13	-76.0	0.0	12.5	25.0	72.9	
Small Improvement	453	(21)	13.0	19.63	-47.9	0.0	11.5	24.0	85.4	
Stable	811	(38)	4.5	19.01	-64.6	-5.2	2.1	16.7	66.7	
Small Deterioration	277	(13)	1.7	17.32	-37.5	-8.3	0.0	11.5	55.2	
Moderate Deterioration	111	(5)	-4.7	20.43	-59.4	-16.7	-4.2	6.3	58.3	
Large Deterioration	64	(3)	-13.7	27.85	-91.7	-30.2	-7.8	4.2	29.2	
PGIS at 8 Months: 5 Categories (collapsing "moderate" and "large")										0.29
Moderate or Large	395	(19)	14.3	20.96	-76.0	0.0	12.5	27.1	72.9	
Small Improvement	453	(21)	13.0	19.63	-47.9	0.0	11.5	24.0	85.4	
Stable	811	(38)	4.5	19.01	-64.6	-5.2	2.1	16.7	66.7	
Small Deterioration	277	(13)	1.7	17.32	-37.5	-8.3	0.0	11.5	55.2	
Moderate or Large	175	(8)	-8.0	23.74	-91.7	-20.8	-4.2	5.2	58.3	
PGIS at 8 Months: 5 Categories (collapsing "small" and "moderate")										0.30
Large Improvement	120	(6)	17.4	22.51	-54.2	0.5	15.1	32.8	70.8	
Small or Moderate	728	(34)	13.0	19.81	-76.0	0.0	11.5	25.0	85.4	
Stable	811	(38)	4.5	19.01	-64.6	-5.2	2.1	16.7	66.7	
Small or Moderate	388	(18)	-0.1	18.46	-59.4	-10.4	0.0	10.4	58.3	
Large Deterioration	64	(3)	-13.7	27.85	-91.7	-30.2	-7.8	4.2	29.2	

[a] Absolute value of the Spearman correlation coefficient for change from baseline KCCQ-TSS at 8 months and change from baseline PGIS at 8 months with each categorisation.

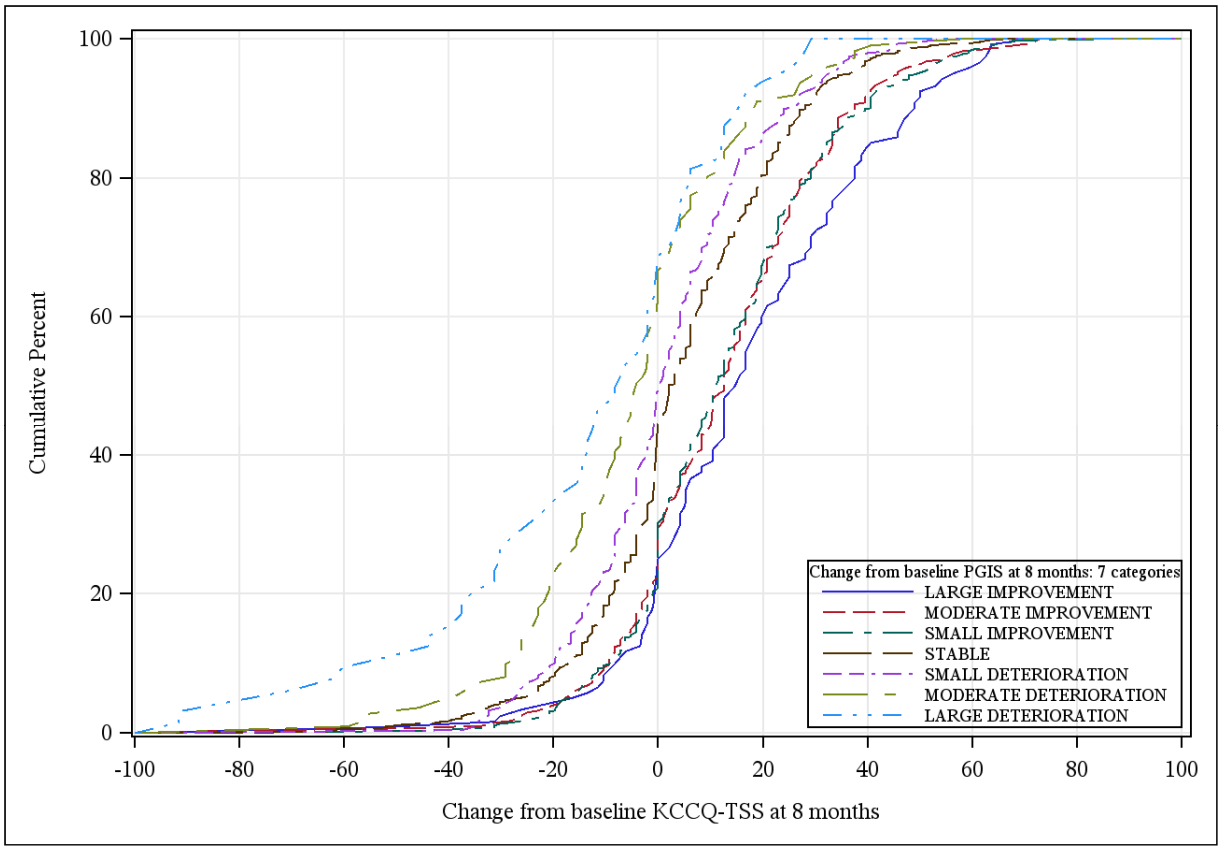
Categories of change from baseline PGIS at 8 months as defined in Table A1.

KCCQ, Kansas City Cardiomyopathy Questionnaire; PGIS, patient global impression of severity; TSS, total symptom score

The eCDF curves in Figure A1 demonstrate a clear separation between all categories of improvement and the "stable" category, in the interval between 5 and 40 points increase in KCCQ-TSS at 8 months, where separation is expected for these curves. However, the separation is less distinct between the categories of "small" and "moderate" improvement. For deterioration, the "large" and "moderate" deterioration categories are clearly separated from the "small" and the "stable" category, in the interval between 5 and 40 points decrease in KCCQ-TSS at 8 months, where separation is expected for these curves. The combined "moderate or large" categories of deterioration and improvement in Figure A2 are separated

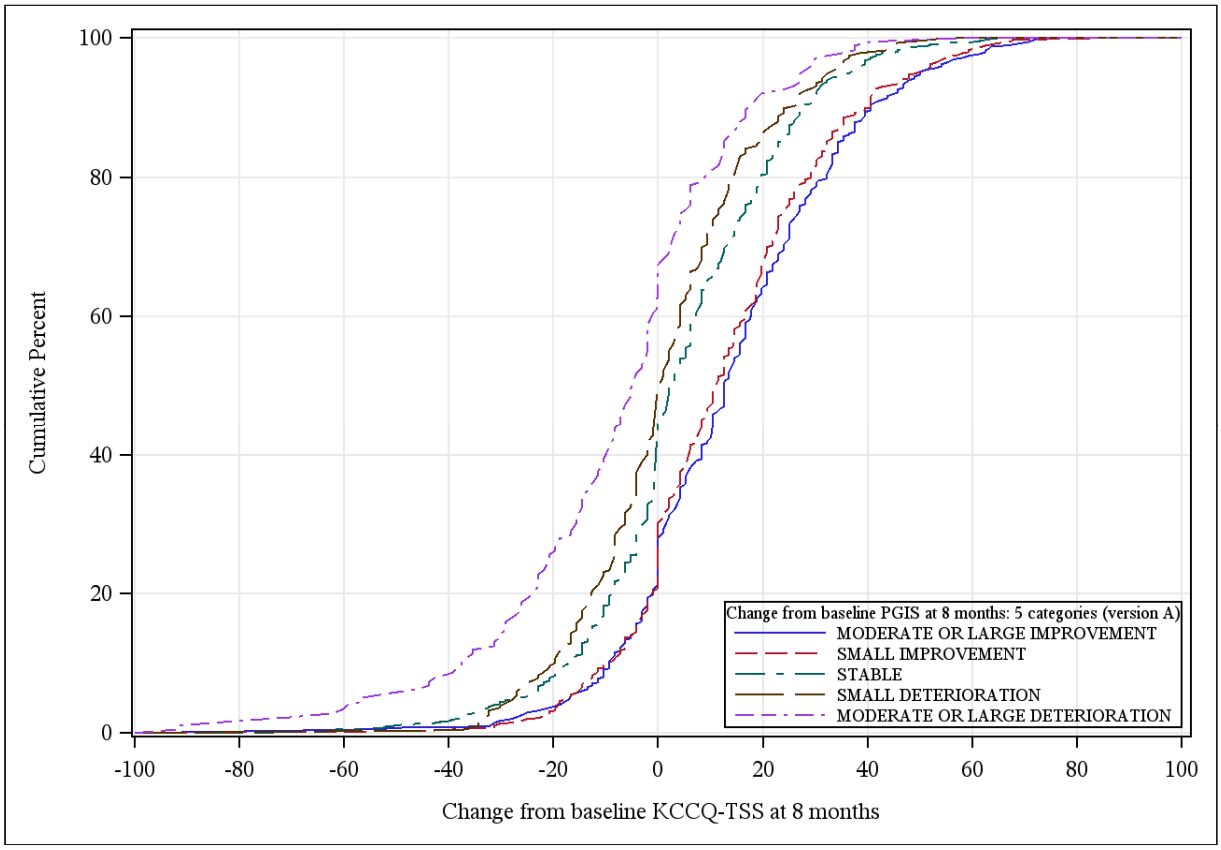
from the "stable" category. This is also observed for combined "small or moderate" categories of deterioration and improvement in Figure A3.

Figure A1 Empirical Cumulative Distribution Function for Change from Baseline KCCQ-TSS at 8 Months versus Change from Baseline PGIS at 8 Months with 7 Categories



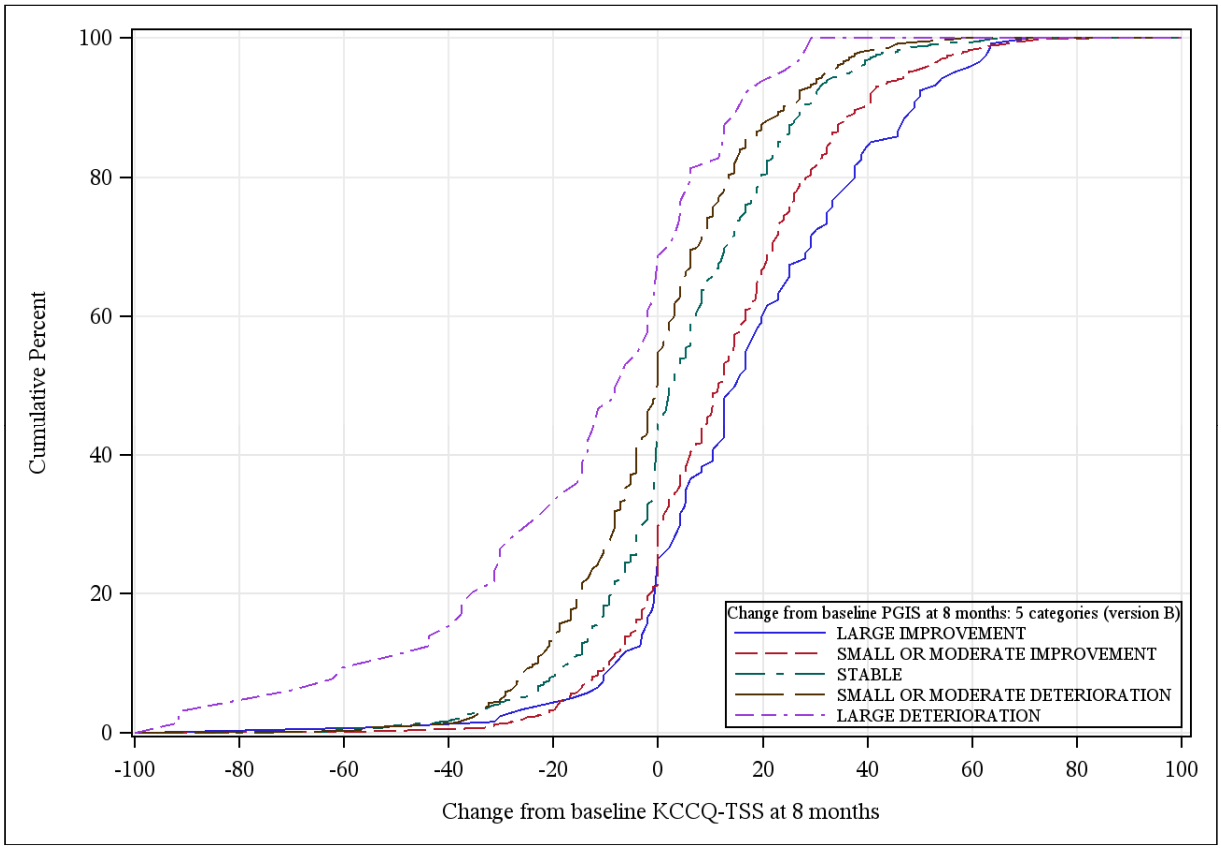
KCCQ, Kansas City Cardiomyopathy Questionnaire; PGIS, patient global impression of severity; TSS, total symptom score

Figure A2 **Empirical Cumulative Distribution Function for Change from Baseline KCCQ-TSS at 8 Months Versus Change from Baseline PGIS at 8 Months with 5 Categories (Collapsing "Moderate" and "Large")**



KCCQ, Kansas City Cardiomyopathy Questionnaire; PGIS, patient global impression of severity; TSS, total symptom score

Figure A3 Empirical Cumulative Distribution Function for Change from Baseline KCCQ-TSS at 8 months Versus Change from Baseline PGIS at 8 Months with 5 Categories (Collapsing "Small" and "Moderate")



KCCQ, Kansas City Cardiomyopathy Questionnaire; PGIS, patient global impression of severity; TSS, total symptom score

A 3 Summary of Results of Distribution-Based Analysis on Blinded Baseline Study Data

Distribution-based methods (0.5 SD and 1 SEM) were used to explore the MCID in the KCCQ-TSS, in patients with HFpEF. The MCID is a value to which between-group differences in average change from baseline are compared, to assess clinical relevance of the difference between treatment groups.

The SEM was calculated as $SEM = \sigma_x * \sqrt{1 - r_{xx}}$, where σ_x is the SD at baseline and r_{xx} is the reliability (internal consistency) of the scale at baseline. The internal consistency of the KCCQ-TSS was assessed by Cronbach's alpha. The distribution-based analyses indicated that 0.5 SD (based on Cohen's "medium" effect size) of the baseline KCCQ-TSS score was equal to 11.0 and that 1 SEM was equal to 8.6 (Table A3). Based on these distribution-based analyses, a rounded mid-point between these values of 10 points is expected to represent a

MCID for KCCQ-TSS in patients with HFpEF. The MCID will not be used to inform responder analyses, as the MCID is not based on within-patient change and is therefore not appropriate for assessing an individual's "response".

Table A3 Distribution-Based Cut-offs for a Minimal Clinically Important Difference in the KCCQ-TSS

	N_1	One-half SD	N_2	1 SEM
Baseline KCCQ-TSS	4730	11.0	4562	8.6

N_1 The SD is based on the number of patients with an observed baseline KCCQ-TSS.

N_2 The SEM is based on the number of patients with an observed scorable responses to each of the items in the KCCQ-TSS.

KCCQ, Kansas City Cardiomyopathy Questionnaire; N, number of patients in treatment group; SD, standard deviation, SEM, standard error of measurement; TSS, total symptom score

Academic Statistical Analysis Plan	
Study Code	D169CC00001
Edition Number	1.3
Date	April 15, 2022

DELIVER: Academic Statistical Analysis Plan

Dapagliflozin Evaluation to Improve the LIVEs of Patients with PReserved Ejection Fraction Heart Failure

An International, Double-blind, Randomised, Placebo-Controlled Phase III Study to Evaluate the Effect of Dapagliflozin on Reducing CV Death or Worsening Heart Failure in Patients with Heart Failure with Preserved Ejection Fraction (HFpEF)

DELIVER: Academic Statistical Analysis Plan

Dapagliflozin Evaluation to Improve the LIVEs of Patients with PReserved Ejection Fraction Heart Failure

An International, Double-blind, Randomised, Placebo-Controlled Phase III Study to Evaluate the Effect of Dapagliflozin on Reducing CV Death or Worsening Heart Failure in Patients with Heart Failure with Preserved Ejection Fraction (HFpEF)

DELIVER CO-CHAIRS, STEERING COMMITTEE

[redacted]

Apr 15, 2022
Date

[redacted]

Apr 15, 2022
Date

Table of Contents

1. INTRODUCTION

DELIVER is an international, multicentre, parallel group, event-driven, randomized, double-blind trial in patients with chronic heart failure and left ventricular ejection fraction (LVEF) >40%, comparing the effect of dapagliflozin 10 mg once daily, vs. placebo, in addition to standard of care. Patients with or without diabetes, with signs and symptoms of heart failure, a LVEF >40%, elevation in natriuretic peptides and evidence of structural heart disease are eligible. The primary endpoint is time-to-first cardiovascular death or worsening heart failure event (heart failure hospitalization or urgent heart failure visit), and will be assessed in dual primary analyses – the full population and in those with LVEF <60%. The study is event-driven and will target 1117 primary events. A total of 6,263 patients have been randomized.

The DELIVER executive committee has developed this academic statistical analysis plan (ASAP) that describes pre-specified analyses that were not described in the DELIVER regulatory SAP (rSAP). General principles outlined in the regulatory SAP will be followed unless specified otherwise here. This document is meant to supplement and complement the regulatory SAP and delineate all analyses that were pre-specified prior to database lock. When relevant, analyses will be conducted based on the pooled DAPA-HF and DELIVER dataset to examine the effects of dapagliflozin in a broad range of patients with HF.

2. CLINICAL ENDPOINTS OF INTEREST

In addition to the efficacy and safety variables listed in the rSAP, the effect of dapagliflozin on the following endpoints will be explored. These events that are imbalanced between arms may be analyzed as time-to-event to better understand the time course. All endpoints will be assessed in the full cohort and in the LVEF < 60% subgroup. These include:

- Days alive and out of the hospital
- Quality of life-adjusted days alive and out of the hospital
- Investigator reported vs. CEC-adjudicated endpoints
- Time to onset of benefit of dapagliflozin
- New diuretic initiation, discontinuation, and dose changes
- New onset atrial fibrillation
- In the T2D subgroup, new glucose lowering therapy initiation and changes in insulin dose (in those on insulin at baseline)
- In the non-T2D subgroup, new diagnosis of diabetes
- Signs and symptoms of HF
- Patient Global Impression of Severity
- Target risk factor control (for blood pressure, smoking, antiplatelet/anticoagulant therapy)
- Cardiac ischemic events including myocardial infarction, unstable angina, unplanned coronary revascularization, and stroke
- Hyperkalemia as a reported adverse event and initiation of new potassium-lowering therapy
- Acute kidney injury as a reported adverse event and initiation of dialysis
- Anemia and requirement for blood transfusion as reported adverse events
- Gout as an adverse event and initiation of new uric acid-lowering therapy
- KCCQ Overall Summary Score at 1, 4 and 8 months

- KCCQ Clinical Summary Score at 1, 4 and 8 months
- KCCQ Physical Limitations Score at 1, 4 and 8 months
- KCCQ Social Limitations Score at 1, 4 and 8 months
- Proportion of patients with clinically meaningful deterioration (5 point or greater worsening), and small (≥5 point), moderate (≥10 point) and large (≥20 point) improvement in KCCQ-TSS, CSS, OSS, PL, QoL and Social Limitations Scores.

COVID-19 Related Endpoints

In addition, the following COVID-19 related endpoints will be evaluated:

- Occurrence of COVID-19 infection (documented as AE or SAE)
- Occurrence of COVID-19 related hospitalizations (overall and among patients with COVID-19 infection)
- Occurrence of COVID-19 related hospitalizations requiring ICU admission (overall and among patients with Covid-19 infection)
- Occurrence of COVID-19 related deaths (overall and among patients with Covid-19 infection)
- Acute kidney injury and initiation of dialysis reported as an adverse event during hospitalization for COVID-19
- Requirement for mechanical ventilation reported as an adverse event during COVID-19 hospitalization
- Requirement for vasopressor support reported as an adverse event during COVID-19 hospitalization
- Sudden cardiac death/cardiac arrest requiring resuscitation during COVID-19 hospitalization
- Worsening heart failure reported during or following COVID-19 hospitalization
- Use of systemic corticosteroids for COVID-19
- Diabetic ketoacidosis reported as an adverse event during or following COVID-19 hospitalization
- Among patients with documented COVID-19 infection, total events of COVID-10 related hospitalizations and COVID-19 related deaths

3. LABORATORY-BASED ENDPOINTS OF INTEREST

In addition, the following laboratory-based endpoints will be assessed:

- eGFR-based
    - Composite of confirmed sustained decline in eGFR, ESRD, and/or renal death. Sustained decline in eGFR will be defined as ≥40%, ≥50%, ≥57% decline from baseline
    - Acute, chronic, and total eGFR slope analysis, including with blanking period to account for acute, expected eGFR changes
    - Focused examination of the "eGFR dip", the acute changes in eGFR in the days-to-weeks after randomization
    - Recalculation of eGFR based on variable calculators (including the 2009 CKD-EPI Equation and 2021 CKD-EPI Equation)

4. BREAKDOWN OF ENDPOINTS

- Mode of death including focused examination of sudden death (as a composite with ventricular arrhythmias reported as adverse events)
- Reasons for hospitalization (total all-cause hospitalization, non-CV hospitalization, HF-related hospitalization, and other CV hospitalizations)
- 30-day readmission (all-cause and HF-related)
- Breakdown of worsening HF events (including urgent visits / Emergency Department stays / oral loop diuretic escalation)

Unknown deaths will not be included as a component of CV deaths in the primary analysis as outlined in the rSAP. In a prespecified exploratory analysis, we will apply a probabilistic model (predetermined prior to database lock) to better distinguish unknown deaths as either CV or non-CV in etiology. This probabilistic model will be built based on known clinical factors that differentially predict adjudicated known cases of CV vs. non-CV deaths.

5. SUBGROUPS

In addition to the subgroups listed in the rSAP, the following subgroups of interest will be explored to examine event rates and for consistency of efficacy and safety of dapagliflozin. All subgroups will be identified based on randomization or pre-randomization data unless otherwise specified. For each subgroup, we will assess the treatment effect and interaction with treatment for the primary endpoint and each of the secondary endpoints, including components of the primary endpoint, measures of quality of life (KCCQ), NYHA class, and the renal composite endpoint. In addition, all subgroups will be assessed in the LVEF < 60% subgroup.

- Improved/recovered LVEF (those who had LVEF ≤40% at any time prior to randomization)
- LVEF subgroups in the rSAP are specified according the following cutpoints (≤ 49%, 50 to 59%, ≥60%). Additional LVEF subgroups to limit digit preference will be considered and treatment effects will be examined across LVEF as a continuous function. In addition, the two-way interaction between sex and LVEF will be examined.
- Age subgroups in the rSAP are specified according to the following cutpoints (median age). Specific evaluation of older age categories will be considered and treatment effects will be examined across age as a continuous function
- BMI subgroups in the rSAP are specified according to the following cutpoints (30kg/m^2). BMI categories will additionally be evaluated according to the full WHO classification and treatment effects will be examined across BMI as a continuous function
- Other anthropometric indices e.g., waist-to-height ratio using quantiles and recognized cutpoints
- eGFR subgroups in the rSAP are specified according to the following cutpoints (60mL/min/1.73m^2). eGFR categories will additionally be evaluated according the full KDIGO classification and treatment effects will be examined across eGFR as a continuous function
- Focused examination of Stage IV CKD (if eGFR was less than 30mL/min/1.73m^2 at randomization or at any post-randomization measurement)
- Further breakdown of glycemic categories into no diabetes, prediabetes, and T2D and examination of treatment effects across HbA1c as a continuous measure
- Time from prior HF hospitalization

- Time from index HF diagnosis
- Background HF therapies including focused examination of patients on various combinations of therapies (including the Heart Failure Collaboratory score) and on/off MRA and on/off ARNI at randomization
- In T2D subgroup, background anti-hyperglycemic therapies including focused examination of patients on various combinations of therapies
- Patients with COPD
- Patients with OSA
- Patients with history of coronary artery disease / prior MI
- Patients with metabolic syndrome (using standard definitions)
- Subgroups based on baseline use and dosing of diuretics
- Patients with multimorbidity and frailty
- Patients with baseline risk as determined by the MAGGIC and other risk scores
- Subgroups based on baseline evidence of congestion and congestion scores
- Regional subgroups based on socioeconomic differences based on the GINI coefficient
- Subgroups based on KCCQ-TSS and other KCCQ domains at baseline.

5. ALTERNATIVE ANALYTIC APPROACHES

Unless otherwise specified, these alternative approaches will be considered for the primary endpoint and each of the secondary endpoints, including components of the primary endpoint, measures of quality of life (KCCQ), NYHA class, and the renal composite endpoint.

- Win ratio using different clinically relevant hierarchies e.g., death, heart failure hospitalization, urgent heart failure visit requiring IV therapy, outpatient therapy for worsening HF, quality of life, and kidney endpoints
- Multi-state modeling of changes in transitional states (ranging from alive and well to death)
- Estimation of time to first statistically significant benefit
- Forecasting lifetime benefit of dapagliflozin if treatment effects were assumed to be maintained long-term
- Absolute risk reductions and NNT calculation overall and across key subgroups
- Cost effectiveness based on US perspective, European perspective, and Other Regions of the World perspective
- Assessment of DELIVER trial and label eligibility in the GWTG-HF registry and other "real-world" datasets
- "Real world" application of the DELIVER trial findings to the GWTG-HF registry and other datasets to estimate projected benefit if dapagliflozin was implemented in usual care

6. COVID-19 META-ANALYSES

- Together with the subset of patients in DELIVER with COVID-19, a meta-analysis will be performed using available phase 3/4 published trials of sodium–glucose cotransporter 2 inhibitor therapies in COVID-19 (*including but not limited to DARE-19*)
- Analyses evaluating outcomes after post-randomization COVID-19 diagnosis will be performed (for instance, increase in mortality or HF event risk after COVID-19 diagnosis)

Systemic Search	To ensure trials beyond DARE-19 and DELIVER were not missed, a systemic search via PubMed and EMBASE will be conducted of • Randomized, placebo-controlled trials of SGLT2 inhibitors in COVID-19 • Published between March 1st, 2020 to August 1, 2022
Rationale	DARE-19 randomized non-critically ill patients with one or more cardiometabolic risk factors (including T2D, HTN, ASCVD, HF or CKD) hospitalized with COVID-19 to dapagliflozin versus placebo, with one of the primary outcomes being respiratory/ cardiovascular/ kidney organ failure or death from any cause. DELIVER randomized patients with HF and LVEF above 40% to dapagliflozin or placebo, and due to the time course of the trial had many patients experiencing COVID-19 related hospitalizations and deaths. Neither trial was adequately powered to assess the effects of dapagliflozin on all-cause mortality, and specific end-organ complications. This pre-specified meta-analysis will allow for greater power to evaluate the effects of dapagliflozin on a range of COVID-19 related clinical endpoints.
Overall Aim	Using study-level published data from DARE-19 and participant-level data from DELIVER, we aim to estimate the effect of SGLT2 inhibitors on all-cause mortality and specific end-organ complications overall, and in clinically-relevant subgroups
Primary Endpoint	COVID-19 related death (this includes COVID-19 related deaths in DELIVER and all deaths in DARE-19)
Secondary Endpoints	• Acute kidney injury and initiation of dialysis during or following hospitalization for COVID-19 • Requirement for mechanical ventilation during COVID-19 hospitalization • Requirement for vasopressor support during COVID-19 hospitalization • Sudden cardiac death/resuscitated cardiac arrest requiring resuscitation during COVID-19 hospitalization • Worsening heart failure during or following COVID-19 hospitalization • Composite of COVID-19 related death and organ failure (acute kidney injury, initiation of dialysis, mechanical ventilation, vasopressor support, cardiac death/ resuscitated cardiac arrest, worsening heart failure) • Composite of COVID-19 related death, acute kidney injury and initiation of dialysis. • Diabetic ketoacidosis during or following COVID-19 hospitalization
Subgroups	• With or without diabetes • With or without ASCVD • With or without CKD (eGFR < 60) • With or without HTN • Age, sex, race, BMI, geographic region
Statistical Analysis	• Intention-to-treat analyses from both trials will be considered and include all randomized participants • All effect sizes will be extracted as point estimates (95% CI). • Statistical heterogeneity will be assessed between trials
Risk of Bias	Study quality will be evaluated using the Cochrane Risk of Bias Tool
Reporting	This planned meta-analysis will be conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis statement
Registration	This meta-analysis will be registered on PROSPERO

7. META-ANALYSIS OF SGLT2 INHIBITOR HFPEF TRIALS AND OTHER SGLT2 INHIBITOR TRIALS

A meta-analysis will be performed using available phase 3/4 published trials of other sodium–glucose cotransporter 2 inhibitor therapies in HFpEF, including but not limited to EMPEROR-Preserved.

Systemic Search	To ensure trials beyond EMPEROR-Preserved and DELIVER were not missed, a systemic search via PubMed and EMBASE will be conducted of • Randomized, placebo-controlled CV and kidney outcomes trials of SGLT2 inhibitors • Published between January 1, 2015 to July 1, 2022 • Only studies including >1,000 patients with HF and LVEF >40%
Rationale	Both EMPEROR-Preserved and DELIVER were similarly designed in evaluating patients with HF, an LVEF above 40%, and elevated natriuretic peptides. Neither trial was powered for mortality or kidney disease outcomes. This pre-specified meta-analysis of the 2 largest trials of HFmrEF and HFpEF will allow for greater power to evaluate a broad range of clinical endpoints and within subgroups of interest than either trial could provide alone.
Overall Aim	Using study-level published data from EMPEROR-Preserved and participant-level data from DELIVER, we aim to estimate the effect of SGLT2 inhibitors on cardiovascular events, kidney events, and mortality outcomes overall, and in clinically-relevant subgroups
Primary Endpoint	Time from randomization to the occurrence of the composite of death adjudicated as CV cause or unplanned HF hospitalization
Secondary Endpoints	• Time from randomization to the occurrence of the composite of death adjudicated as CV cause or a worsening HF event (including either unplanned hospitalization or urgent HF visit requiring IV therapy) • Total number of worsening HF events and cardiovascular death • Time from randomization to the occurrence of deaths adjudicated as CV cause • Time from randomization to death from any cause • Time from randomization to renal composite outcome (50% or higher sustained decline in eGFR, end stage kidney disease, or renal death) • Proportion of patients with clinically meaningful deterioration (5 point or greater worsening), and small (≥5 point), moderate (≥10 point), and large (≥15 point) improvement in KCCQ-TSS, CSS, OSS
Subgroups	• LVEF (<50%, ≥50 to <60%, ≥60%) • With or without diabetes • Use of no use of ACEi/ARB/ARNI at baseline • Use and no use of MRA at baseline • Age (≥70 and <70 years), sex (male, female), race (White, Black, Asian, Other), BMI (<30 and ≥30 kg/m^2), eGFR (≥60 and <60mL/min/1.73m^2), systolic blood pressure, history of AF/AFL, hospitalization for HF within 12 months, NYHA class (II and III/IV)
Statistical Analysis	• Fixed effects model • Only intention-to-treat analyses from both trials will be considered and include all randomized participants • All effect sizes will be extracted as point estimates (95% CI). For the time-to-first event endpoints, Cox proportional hazards models will be used for hazard ratio (HR) and 95% CI. Recurrent event analyses will be based on the Lin-Wei-Yang-Ying model and summarized as

	rate ratio (RR) and 95% CI. Responder analyses for KCCQ changes will be based on logistic regression analyses summarized as odds ratios with 95% CIs. • The continuous association between LVEF and treatment effects on the primary endpoint will be assessed with restricted cubic spline analyses. Data from these published splines in the EMPEROR program will be digitized using a validated, semiautomatic tool (DigitizeIt software https://www.digitizeit.xyz/). • Statistical heterogeneity will be assessed between trials
Risk of Bias	Study quality will be evaluated using the Cochrane Risk of Bias Tool
Reporting	This planned meta-analysis will be conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis statement
Registration	This meta-analysis will be registered on PROSPERO

Meta-analyses will also be performed using available phase 3/4 published trials of other sodium–glucose cotransporter 2 inhibitor therapies in different disease states to provide a comprehensive assessment of the value of SGLT2 inhibitors across the disease spectrum.

Appendix B Programming Code for Calculating Significance Level for LVEF <60% Subpopulation

Let Z_1 and Z_2 denote the standardized test statistic for testing the hypothesis of treatment effect in the LVEF < 60% subgroup and the full population respectively. Z_1 and Z_2 are bivariate normal with correlation equal to the proportion of events in the LVEF < 60% subpopulation (Spiessen and Debois 2010). To control the familywise error rate below α, for a pre-specified significance level α_2 for the full population, we need to define α_1 for the subgroup such that, under the null hypothesis,

(1) $P(Z_1 > z_{\alpha_1} \text{ OR } Z_2 > z_{\alpha_2}) = \alpha$

where z_{α_1} and z_{α_2} are the corresponding critical values from the standard normal distribution.

Equation (1) can be rewritten as

$P(Z_2 > z_{\alpha_2}) + P(Z_1 > z_{\alpha_1}, Z_2 \leq z_{\alpha_2})$

$= \alpha_2 + P(Z_1 > z_{\alpha_1}, Z_2 \leq z_{\alpha_2})$

$= \alpha$

Thus we need to find α_1 such that

$P(Z_1 > z_{\alpha_1}, Z_2 \leq z_{\alpha_2}) = \alpha - \alpha_2$

As noted by Spiessen and Debois 2010, this corresponds to error spending for group sequential methods where Z_2 is the test statistic at interim analysis and Z_1 is the test statistic at the final analysis. Accordingly, standard software for group sequential designs can be used to calculate the significance level α_1 as shown below using the R package gsDesign or the SAS procedure SEQDESIGN

For the proportion of events in the LVEF <60% subgroup we use the lower bound of a 95% confidence interval for the estimated proportion calculated using normal approximation as

$$p - z_{2.5}\sqrt{p(1-p)/e}$$

where $z_{2.5}$ is the upper 2.5% percentile of the standard normal distribution and $p = e_{60}/e$ for e_{60} events in the subgroup and e events in total.

In an example of 810 (72.5%) in the subgroup out of a total of 1117 event, the lower confidence limit for the proportion is 0.699, which will be used in the example R and SAS code below.

R gsDesign package

Example with lower proportion 0.699 (lower confidence limit) events in the subgroup:

```
gsd <-
   gsDesign(k=2, timing=c(0.699,1),
      # 69.9% events at interim, corresponding to proportion in subgroup
   test.type=1, sfu=sfLinear,
   sfupar=c(0.699,1,0.5,1),
      # proportion 0.5 of total alpha spent at interim, corresponding to
      # alpha2 = 0.5*0.048=0.024 two-sided set for the full population
    alpha=0.024)
      #total one-sided alpha 0.048/2
                        #
(alpha1 <- 2*100*(1-pnorm(gsd$upper$bound[2])))
       # alpha1 now holds the two-sided significance level for final
       # analysis, corresponding to alpha1 for the subgroup
```

SAS proc SEQDESIGN

Example with lower proportion 0.699 (lower confidence limit) events in the subgroup:

```
proc seqdesign bscale=pvalue;
     design nstages=2 info=cum(69.9 100)
     /* 69.9% events at interim, corresponding to proportion in subgroup */
     method=peto(pvalue=0.012)
     /* alpha 0.012 one-sided spent at interim, corresponding to
        0.024 two-sided set for the full population */
         stop=reject
         alt=lower alpha=0.024;
     /*   0.048/2 total one-sided alpha  */
     ods output boundary = bound;
  run;

  data alpha1;
    set bound;
    if _stage_=2;
    alpha1=100*2*bound_la;
  run;

   /* alpha1 now holds the two-sided significance level for final analysis,
      corresponding to alpha1 for the subgroup */
```

SIGNATURE PAGE

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature

Document Name: d169cc00001-statistical-analysis-plan-v5		
Document Title:	Statistical Analysis Plan Version 5	
Document ID:	Doc ID-003872995	
Version Label:	5.0 CURRENT LATEST APPROVED	
Server Date **(dd-MMM-yyyy HH:mm 'UTC'Z)**	**Signed by**	**Meaning of Signature**
15-Dec-2021 13:51 UTC	[redacted]	Content Approval
15-Dec-2021 12:21 UTC	[redacted]	Content Approval

Notes: (1) Document details as stored in ANGEL, an AstraZeneca document management system.