Statistical Analysis Plan

Protocol Title: Dapagliflozin Effect on Symptoms and Biomarkers in Patients with Heart Failure (DEFINE-HF)

Protocol Number: D1690C00032

ClinicalTrials.gov registration: NCT02653482

Version Date: June 26, 2019

Abbreviation or special term	Explanation
AE	Adverse Event
AFib	Atrial fibrillation
AZ	AstraZeneca
BMI	Body mass index
BNP	B-type natriuretic peptide
CRF	Case Report Form (electronic/paper)
DBP	Diastolic Blood Pressure
DM	Diabetes mellitus
EPS	Enrolled Patients Set
GFR	Glomerular filtration rate
HbA1c	Hemoglobin A1c
HDL	High-Density Lipoprotein
HF	Heart failure
KCCQ	Kansas City Cardiomyopathy Questionnaire
LVEF	Left Ventricular Ejection Fraction
ITTS	Intention to treat set
IP	Investigational Product
MCAR	Missing completely at random
MDRD	Modification of Diet in Renal Disease
NSVT	Nonsustained ventricular tachycardia
NTproBNP	N-terminal (NT)-pro hormone BNP
NYHA	New York Heart Association
PPS	Per protocol set
PAC	Premature atrial contractions
PTDV	Premature treatment discontinuation visit
PVC	Premature ventricular contractions
SAE	Severe Adverse Event
SAFS	Safety Analysis Set
SAS	Statistical Analysis Software
SBP	Systolic Blood Pressure
SD	Standard deviation
SE	Standard error

Abbreviation or special term	Explanation
SGLT-2	Sodium-glucose cotransporter 2
T2DM	Type 2 Diabetes Mellitus
VF	Ventricular fibrillation
VT	Ventricular tachycardia

Contents

1 INTRODUCTION

1.1 Study Objectives

Primary Study Objectives	• To estimate the effect of dapagliflozin vs. placebo on 6- and 12-week NT pro-BNP. • To compare the proportion of patients that achieve a meaningful change from baseline in quality of life (≥ 5-point increase in average of 6- and 12-week KCCQ overall summary scores) or NT pro-BNP (≥ 20% decrease in average of 6- and 12-week NT pro-BNP) between dapagliflozin and placebo.
Secondary Study Objectives	• To compare the proportion of patients with a ≥ 5-point increase in the average of 6- and 12-week Kansas City Cardiomyopathy Questionnaire (KCCQ) scores, from baseline, between dapagliflozin and placebo. • To compare the proportion of patients with a ≥ 20% decrease in the average of 6- and 12-week NT pro-BNP, from baseline, between dapagliflozin and placebo. • To compare the proportion of patients with both a ≥ 5-point increase in the average of 6- and 12-week KCCQ overall summary scores and a ≥ 20% decrease in the average of 6- and 12-week NT pro-BNP, from baseline, between dapagliflozin and placebo. • To estimate the effect of dapagliflozin vs. placebo on 6- and 12-week NT pro-BNP. • To estimate the effect of dapagliflozin vs. placebo on 6- and 12-week 6-minute walk test. • To estimate the effect of dapagliflozin vs. placebo on 6- and 12-week BNP. • To estimate the effect of dapagliflozin vs. placebo on 6- and 12-week HbA1c (evaluated separately in patients with and without type 2 diabetes). • To estimate the effect of dapagliflozin vs. placebo on 6- and 12-week weight. • To estimate the effect of dapagliflozin vs. placebo on 6- and 12-week systolic blood pressure.
Exploratory objectives	• To estimate the effect of dapagliflozin vs. placebo on daily loop diuretic dose (furosemide equivalent). • To estimate the effect of dapagliflozin vs. placebo on hospitalizations for heart failure. • To estimate the effect of dapagliflozin vs. placebo on urgent outpatient heart failure visits. • To estimate the effect of dapagliflozin vs. placebo on hospitalizations for heart failure or urgent outpatient heart failure visits. • To compare the change in NYHA Class from baseline at 6 and 12 weeks between dapagliflozin and placebo.

1.2 Study Design

Design Configuration and Subject Population	• A 12-week randomized, double-blind, placebo-controlled trial to evaluate the effects of once-daily dapagliflozin 10 mg on heart failure disease-specific biomarkers (BNP and NT pro-BNP), symptoms, health status, and quality of life in patients with type 2 diabetes and chronic heart failure with reduced systolic function. Substudies will also be conducted for exploratory biomarker analyses and effects on arrhythmia burden.
Treatment Groups	• Dapagliflozin 10 mg administered orally vs. placebo
Inclusion criteria	• Age > 18 and < 120 at the screening visit • Established diagnosis of heart failure (for at least 16 weeks prior to the screening visit) with reduced systolic function (LVEF≤40% due to either ischemic or non-ischemic etiology) documented by an imaging modality (echocardiography, nuclear imaging, LV angiography, magnetic resonance imaging) within the past 24 months. Any local measurement of LVEF by any modality within the eligibility range made within the past 24 months is acceptable provided there has been no subsequent LVEF measurement above 40%. • No change in diuretic management for 1 week prior to screening visit or between the screening and randomization visit • NYHA class II or III heart failure symptoms at the screening and randomization visit • BNP ≥100 pg/mL and/or NT pro-BNP ≥ 400 pg/mL at the screening visit (For patients with permanent atrial fibrillation inclusion thresholds will be BNP ≥ 125 pg/mL or NTproBNP ≥ 600 pg/mL) • Ability to provide informed consent prior to initiating screening visit procedures
Exclusion criteria	• Decompensated heart failure (hospitalization for heart failure within the 30 days prior to screening or NYHA class IV heart failure symptoms at screening) • History of type 1 diabetes • Estimated glomerular filtration rate (eGFR) < 30 at the screening visit by modified MDRD equation GFR in mL/min per 1.73 m^2 = 175 x $SerumCr^{-1.154}$ x $age^{-0.203}$ x 1.212 (if patient is African American) x 0.742 (if female) • Admission for an acute coronary syndrome (ST-elevation MI, non-ST-elevation MI, or unstable angina), percutaneous coronary intervention, or cardiac surgery within 30 days prior to the screening visit.

	• Admission for cardiac resynchronization therapy (CRT) within 90 days prior to the screening visit • Planned cardiovascular revascularization (percutaneous intervention or surgical) or major cardiac surgery (coronary artery bypass grafting, valve replacement, ventricular assist device, cardiac transplantation, or any other surgery requiring thoracotomy) or CRT within the 90 days after the screening visit. • Participation in any interventional clinical trial (with an investigational drug or device) that is not an observational registry within the 8 weeks prior to the screening visit. • History of hypersensitivity to dapagliflozin • For women of child-bearing potential: Current or planned pregnancy or currently lactating. • Women who are surgically sterile or those who are postmenopausal for at least 1 year are not considered to be of child-bearing potential. Women of child-bearing potential, who are sexually active, must agree to use a medically-accepted method of birth control for the duration of the study. Acceptable birth control methods include: (1) surgical sterilization (such as a hysterectomy or bilateral tubal ligation), (2) progesterone hormonal contraceptives (birth control pills or implants), (3) barrier methods (such as a condom or diaphragm) used with a spermicide, or (4) an intrauterine device (IUD). Women of child-bearing potential will have a urine pregnancy test at every clinic visit and it must be negative to continue study participation. • Life expectancy <1 year at the screening visit • Patients who are volume depleted based upon physical examination at the time of the screening or randomization visit • BNP <100 pg/mL and NT pro-BNP<400 pg/mL at the screening visit (For patients with permanent atrial fibrillation exclusion thresholds will be BNP<125 pg/mL and NTproBNP<600 pg/mL) • Patients currently being treated with any SGLT-2 inhibitor (dapagliflozin, canagliflozin, empagliflozin) or having received treatment with any SGLT-2 inhibitor within the 12 weeks prior to the screening visit. • Average supine systolic BP <90 mmHg at the screening or randomization visit • Past or current history of bladder cancer • Active Hematuria • Donation of blood or bone marrow 12 weeks prior to the screening visit and no planned donations during the study period • Heart failure due to restrictive cardiomyopathy, active myocarditis, constrictive pericarditis, severe stenotic valve disease, and HOCM (hypertrophic obstructive cardiomyopathy).

1.3 Sample Size and Power

Planned Sample Size	• Approximately 110 patients will be enrolled in each arm.
Power Statement	• The first dual primary endpoint of this study is the average of 6- and 12- week NT pro-BNP. A generalized linear mixed model with patient as a random effect will be used to estimate the average effect over 6 and 12 weeks controlling for baseline NT pro-BNP and other pre-specified covariates. A gamma distribution with log link will be used to account for the skewed nature of NT pro-BNP. Several sensitivity analyses will be performed, one of which includes site as a random effect to account for clustering of patients within centers. For the first dual primary endpoint a sample size of 110 for each group will achieve 80% power with α=0.05 to detect a difference in NT pro-BNP between the two groups of at least 302 pg/mL from baseline to 12 weeks. The assumptions for this calculation were derived from the BATTLESCARRED trial where the estimated standard deviation for NT pro-BNP was 1250 pg/mL. We expect the standard deviation in the DEFINE-HF Trial to be somewhat lower (961 pg/ml) given lower NT pro-BNP threshold. Of note, 302pg/mL reduction in NT pro-BNP is equivalent to 31.5% of the standard deviation in NT pro-BNP (based on the above assumption). There is an anticipated non-completion rate of approximately 13% so the final sample size per group will be 125. • The second dual primary endpoint, the proportion of patients with a ≥ 5-point KCCQ overall summary score increase or a ≥ 20% decrease in NT pro-BNP, averaging over 6 and 12 weeks, will be analyzed using generalized linear mixed models controlling for baseline NT pro-BNP and KCCQ overall summary score as well as several additional covariates as specified below. A binomial distribution with logit link will be used. Several sensitivity analyses will be performed, one of which includes site as a random effect to account for clustering of patients within centers. Power was determined based on comparisons of proportions between two independent groups, where the anticipated control group percent change is 30%. A sample size of 110 for each group will achieve 80% power with α=0.05 to detect an absolute difference in proportions between the two groups of 18% from baseline to 12 weeks for the second dual primary endpoint.

1.4 Substudies

Substudies	• **Arrhythmia Substudy**: Patient participation will be optional. At the screening visit, those likely to be enrolled will be given the option to participate in the arrhythmia substudy. If they elect to enroll they will wear a Holter monitor for 14 days to establish a baseline. The 14-day Holter will be repeated at the 6-week visit to compare change in arrhythmia burden. Outcomes of VT, NSVT episodes, VF, A fib, and PVCs will be compared between dapagliflozin and placebo. • **SensiVest Substudy**: Patients at selected sites who elected to participate will have blinded lung fluid volume measurements at rest during the randomization (week 0), 6, 12 and 13 week visits. Lung fluid volume will be compared between the treatment group and placebo group over the 12 weeks of treatment. • **Biomarker Substudy**: A biomarker substudy will be conducted. Specimens will be collected for exploratory biomarker testing from patients at the Randomization, Week 6 and Week 12 visits to assess the impact of dapagliflozin versus placebo on known markers of myocardial fibrosis/necrosis, inflammation, and oxidative stress.

2 GENERAL CONSIDERATIONS FOR DATA ANALYSES

2.1 Analysis Data Sets

Efficacy and safety analyses will be performed on data from the following analysis sets:

- Enrolled patients set (EPS) will consist of all patients who signed the informed consent. This data set will be used to summarize the patient disposition data.

- Modified intention to treat set (modified ITTS) is defined as all patients who have been randomized to study treatment, have received at least one dose of study medication, and have sufficient evaluable data for endpoint ascertainment during follow up. For clarity, patients with no evaluable data for a particular outcome during follow up will be excluded from the analyses of these respective endpoints. In patients who undergo initiation of intravenous home inotropic support and/or receive left ventricular assist device or heart transplant during the study or who are started on a prohibited medication (open label SGLT-2is), for efficacy analyses, the last available data point(s) prior to these events will be used. Subjects will be analyzed according to the randomization group. The modified ITTS data set will be used for the primary, secondary, and selected exploratory (NYHA class, loop diuretic dose) efficacy endpoints.

- On-treatment set (OTS) includes all subjects in the modified ITTS. Primary and secondary endpoint measurements will be excluded for the follow-up time point(s) when subjects were temporarily or permanently off the study drug at the time the corresponding measurements were obtained. The OTS data set will be used for a sensitivity analyses for the primary, secondary, and selected exploratory (NYHA class, loop diuretic dose) efficacy endpoints.

- The per protocol set (PPS) includes all subjects in the modified ITTS who did not have any major protocol deviations. Major protocol deviations, which are detailed in section 2.2, will be determined prior to unblinding of treatment groups. Subjects will be summarized according to actual treatment received regardless of the allocated treatment. This will be used for a sensitivity analysis for the primary efficacy endpoints only.

- The safety analysis set (SAFS) includes all patients who received at least 1 dose of study medication. Throughout the safety results sections, erroneously treated patients (eg, those randomized to dapagliflozin but actually given placebo) will be accounted for in the actual treatment group. If a patient received study drug from the wrong kit for only a part of the treatment duration and then switched to another, the associated actual treatment group for that patient will be the treatment group the patient had the longest exposure to. The main safety analyses will be restricted to adverse events that occurred between randomization and the 12-week visit. Adverse events that occurred between 12 weeks and 13 weeks (after discontinuation of study treatment) will be collected, and presented in a separate, supplemental analysis. The safety analysis set will be used to summarize safety data and patient demography and their baseline characteristics, and to analyze selected exploratory endpoints (heart failure hospitalizations, urgent heart failure visits, and a composite of heart failure hospitalizations or urgent heart failure visits). If there is a difference between the SAFS and modified ITTS, baseline and demography data will also be presented for the IITS.

2.2 Protocol Deviations and Major Eligibility Violations

Important Protocol Deviations (IPDs) are defined as those important deviations from the protocol that are likely to have an impact on the efficacy and/or safety of study treatments.

Protocol deviations will be reviewed in a blinded fashion by the study team prior to database lock. All decisions to exclude patients and/or data from the modified ITTS or PPS will be made prior to the unblinding of the study and agreed by the study team.

Error! Reference source not found.9.2 specifies the criteria for IPDs Patients having IPDs will be summarized and listed by treatment group and overall. The list of protocol deviations (including IPDs) will be compiled separately and finalized prior to database lock.

2.3 Strata, Covariates, and pre-specified subgroup analyses

All efficacy and safety endpoints will be analyzed in the entire cohort, and then within the subgroups of patients with and without diabetes. Analyses for the primary and secondary endpoints will be adjusted for the corresponding baseline measurements as well as history of diabetes (DM), eGFR, and age. A sensitivity analysis will be conducted for each endpoint adjusting for the following additional baseline covariates: AFib type (No AFib, persistent/permanent AFib, paroxysmal AFib) and LVEF (as continuous variable). Restricted cubic splines will be used for continuous variables to accommodate non-linear effects.

Additional sensitivity analysis will repeat the primary analysis but will also include site as a random effect to account for potential clustering by enrolling center. A different sensitivity analysis will use the same model as the primary analysis, but adjust only for the corresponding baseline measurements, without additional covariates.

Additionally, the following pre-specified subgroup analyses will be performed for the primary endpoints (stratified by the baseline variables below):

- Baseline NTproBNP (< median, ≥ median)
- Baseline LVEF (≤30%, > 30%)
- Atrial fibrillation type (No AFib, permanent/persistent AFib, paroxysmal AFib)
- Baseline KCCQ overall summary score (<70, ≥70)
- Baseline eGFR (<60, ≥60 mL/min/1.73 m^2)
- Type of cardiomyopathy (ischemic, non-ischemic)
- Age (<65, ≥65)
- Sex (male, female)
- Race (white, black, other)
- Baseline RAASi type (ARNI, ACE/ ARB, neither)
- Loop diuretic dose (Furosemide equivalent mean daily dose: ≤ 40 mg, >40 mg)

Subgroup analyses will be carried out by augmenting the primary analysis model with terms for subgroup and a subgroup-by-treatment interaction. Adjusted point estimates and 95% confidence intervals will be calculated for the effect of dapagliflozin compared with placebo within each subgroup. An interaction p-value will be provided.

Due to the large number of study sites and the expected low number of patients per site, site effects will not be explored, although study site will be included as a random effect in efficacy sensitivity analyses to account for within-site correlations, as specified above.

2.4 Endpoints Definitions

Primary endpoints:

- Average of 6- and 12-week NT pro-BNP values.
- Proportion of patients that achieve a meaningful change from baseline in quality of life (≥ 5-point increase in average of 6- and 12-week KCCQ overall summary scores) or NT pro-BNP (≥ 20% decrease in average of 6- and 12-week NT pro-BNP).

Secondary endpoints:

- Proportion of patients with a ≥ 5-point increase in average of 6- and 12-week KCCQ overall summary scores
- Proportion of patients with a ≥ 20% decrease in average of 6- and 12-week NT pro-BNP
- Proportion of patients with a ≥ 5-point increase in average of 6- and 12-week KCCQ overall summary scores and a ≥ 20% decrease in average of 6- and 12-week NT pro-BNP.
- NT pro-BNP at 6 weeks and at 12 weeks separately.
- Average of 6- and 12-week KCCQ overall summary scores.
- Average of 6- and 12-week 6-minute walk tests. Data points for 6-minute walk test assessment that had associated notations of marked non-HF related limitations in exercise capacity (e.g., orthopedic injury, arthritis, etc.) will be excluded.
- Average of 6- and 12-week BNP levels.
- Average of 6- and 12-week HbA1c levels.
- Average of 6- and 12-week weight(lbs).
- Average of 6- and 12-week supine systolic blood pressure measurements. Supine systolic blood pressure will be taken three times at each time point. The average will be used.

Exploratory Outcome Variables

- Daily loop diuretic dose (Furosemide equivalent). The following equations will be used to convert doses to Furosemide equivalent:
  40 mg Furosemide = 20 mg Torsemide = 2 mg Bumetanide = 50 mg Ethacrynic Acid
  When patient is not on a loop diuretic, dose = 0 mg.
  For subjects with dosing schedule reported as “as needed (prn)”, we will assume a twice a week dosing regimen for loop diuretic at the recorded dose.
  In addition to comparing daily loop diuretic dose between treatment arms, we will also compare the proportion of patients who had their loop diuretic dose reduced or discontinued during the treatment period.
- Proportion of patients experiencing hospitalizations for heart failure. Only events positively adjudicated by the clinical event committee will be included. Events that occurred between screening and

randomization, and between 12 weeks and 13 weeks will be excluded from this analysis. Events that occurred between the week 12 and week 13 study visits will be reported in a separate, supplemental analysis.

- Proportion of patients experiencing urgent outpatient heart failure visits. Only events positively adjudicated by the clinical event committee will be included. Events that occurred between screening and randomization, and between 12 weeks and 13 weeks will be excluded from this analysis. Events that occurred between the week 12 and week 13 study visits will be reported in a separate, supplemental analysis.
- Proportion of patients experiencing hospitalizations for heart failure or urgent outpatient heart failure visits. Events that occurred between screening and randomization, and between 12 weeks and 13 weeks will be excluded from this analysis. Events that occurred between the week 12 and week 13 study visits will be reported in a separate, supplemental analysis.
- Change in NYHA class between baseline and follow-up at 12 weeks is defined as a 3-level categorical variable: 1) Decrease, 2) No change, 3) Increase. For patients missing NYHA class at 12 weeks, value at 6 weeks will be used.

Substudy Endpoints

- **Arrhythmia Substudy:**
  - Number of sustained VT episodes per day during the 14 days when patients wear the Holter monitor starting at 6 weeks.
  - Number of non-sustained VT episodes per day during the 14 days when patients wear the Holter monitor starting at 6 weeks.
  - Number of VF episodes per day during the 14 days when patients wear the Holter monitor starting at 6 weeks
  - Percentage of time in AFib during the 14 days when patients wear the Holter monitor starting at 6 weeks, among patients without known permanent or persistent AFib at baseline
  - Proportion of PVC heart beats during the 14 days when patients wear the Holter monitor starting at 6 weeks
  - Proportion of PAC heart beats during the 14 days when patients wear the Holter monitor starting at 6 weeks, among patients without known permanent or persistent AFib at baseline

- **Sensivest Substudy:**
  - Average of 6- and 12-week lung fluid volumes
  - Change in lung fluid volumes between weeks 12 and 13

- **Biomarker Substudy:**
  - Average of 6- and 12-week oxidative stress: Uric Acid
  - Average of 6- and 12-week fibrosis: galectin-3 (Gal3) and ST-2
  - Average of 6- and 12-week myocyte necrosis: High sensitivity cardiac troponin T(hs-cTnT)
  - Average of 6- and 12-week inflammation: high-sensitivity C reactive protein(hsCRP) and interleukin-6 (Il-6)

- Average of 6- and 12-week advanced glycation end-products: soluble Receptor for Advanced Glycation End Products(sRAGE) and N(6)-Carboxymethyllysine (CML).

2.5 Multiple Testing

Both primary hypotheses will be tested at the 2-sided 5% significance level. In the case of non-significant findings for the first dual primary endpoint, the second dual primary endpoint will be tested at the 2-sided 5% significance level but will be regarded as an exploratory endpoint, and a nominal P value will be produced. No adjustments for multiplicity will be made for secondary and exploratory endpoints.

2.6 Missing Data

Primary endpoint assessments (NT pro-BNP and KCCQ overall summary score) at baseline, 6-week and 12-week are expected to have no or negligible missing data because of the study management procedures. In the rare case where the baseline NT pro-BNP is not available, the screening value will be used. For the first primary endpoint, analysis will be conducted among patients with a baseline assessment and with at least one follow-up assessment using repeated measure models adjusting for baseline NT pro-BNP and KCCQ overall summary score as well as history of DM, eGFR, and age. This approach uses maximum likelihood methods, which accommodates missing at random data conditional the covariates and on other observed follow-up assessments. For the second primary endpoint, proportion of patients with a meaningful change in average of 6- and 12-week KCCQ overall summary score or average of 6- and 12-week NT pro-BNP, the time point at which the outcome is assessable will be used in the situation where the outcome is assessable at only one of the two timepoints. Additionally, patients who have missing 6-week or 12-week assessments will be compared to those with complete data to evaluate for any selection bias. If imbalance is noted, sensitivity analysis will be conducted in which patients may be weighted by the probability of missing data to minimize the selection bias.

There are four possible sources for missing data:

1. Administrative error
2. Premature study terminations
3. Partial completion of KCCQ
4. Deaths

The first source is believed to be negligible due to the study management procedures and is also believed to be missing completely at random (MCAR). For patients who drop out of the study prematurely, a premature treatment discontinuation visit (PTDV) will be scheduled for them, whenever possible, to get the last assessment. The PTDV assessment will be carried forward to the next follow-up time point. For partially completed KCCQs, the scoring algorithm accommodates a limited number of skipped responses. Missing data due to deaths is expected to be very low (fewer than 10 patients) because of the short follow-up period and will be ignored in the primary analysis. A sensitivity analysis for KCCQ overall summary score will be conducted by setting KCCQ summary score to 0 for patients who experience death.

2.7 Data Handling Conventions and Transformations

Diabetes duration will be calculated as (Consent date – DM diagnosis date) / 365.25. It will be rounded to the whole year.

Body mass index (BMI) will be calculated as weight (kg) / [height (m)]^2. It will be displayed to 1 decimal place in listings, but will not be rounded or truncated prior to summarization

Estimated glomerular filtration rate (eGFR) will be calculated using the MDRD-4 equation: GFR in mL/min per 1.73 m^2 = 175 x $SerumCr^{-1.154}$ x $age^{-0.203}$ x 1.212 (if patient is African American) x 0.742 (if female). Patients with missing data in Creatinine Serum, age, race, or gender will not have eGFR calculated.

If urine microalbumin value is reported as being too high for the laboratory to calculate a UACR, 5000 mg value for UACR will be used.

3 Analyses of Baseline Characteristics

Patient demographics, baseline clinical characteristics, medical histories, and baseline labs will be described overall and by treatment group. Continuous measures will be summarized by mean $\pm$ standard deviation and compared using Student's T-tests. Categorical variables will be summarized by frequency and percent and compared using χ^2or Fisher's exact tests, as appropriate.

3.1 Demographics and Baseline Characteristics

- Age
- Sex
- Race/Ethnicity

3.2 Medical History

3.2.1 Diabetes History

- Diabetes duration
- History of diabetic peripheral neuropathy
- History of diabetic autonomic neuropathy
- History of diabetic retinopathy
- History of diabetic ketoacidosis (DKA) event
- History of hyperosmolar hyperglycemic syndrome (HHS) event
- History of severe hypoglycemic event(s)
- History of amputation

3.2.2 Other Medical History

- History of heart failure
- NYHA Class
- Most Recent LVEF Assessment
- History of PAD
- History of hypertension
- History of coronary artery disease

- History of dyslipidemia
- History of angina
- History of atrial fibrillation
- History of atrial flutter
- History of MI
- History of PCI
- History of CABG
- History of ventricular tachycardia
- ICD implanted
- Permanent pacemaker implanted
- History of valve disease

3.3 Physical examination

- Body Mass Index
- Sitting Pulse and Blood Pressure
- Supine Pulse and Blood Pressure
- Standing Pulse and Blood Pressure

3.4 Lab results

- HbA1c
- BNP
- NT pro-BNP
- Glucose (mg/dL)
- BUN - urea nitrogen (mg/dL)
- Creatinine (mg/dL)
- Estimated Glomerular Filtration Rate (eGFR)
- Sodium (mmol/L)
- Potassium (mmol/L)
- Chloride (mmol/L)
- Carbon Dioxide (mmol/L)
- Calcium (mg/dL)
- Phosphate - as phosphorus (mg/dL)
- Albumin (g/dL)
- Random Urine Creatinine (mg/dL)
- Random Urine Microalbumin < 0.2 mg/dL
- Random Urine Microalbumin (mg/dL)
- Enter Urine Albumin result, if measurable.
- Random Urine Albumin/Creatinine Ratio (mcg/mg creat)

4 SUBJECT DISPOSITION

4.1 Subject Enrollment

The number and percent of subjects randomized for each site will be summarized overall and by treatment group. The denominator for the percent calculation will be number of subjects screened.

4.2 Disposition of Subjects

A summary of subject disposition will be provided overall and by treatment group, as appropriate. This summary will present the number of subjects screened, randomized, included in the safety analysis set, and the number and percent of subjects meeting the following criteria:

- Completed the study drug treatment,
- Did not complete the study drug treatment (with summary of reasons for not completing the study treatment)
- Completed the study (includes the study treatment period and any post treatment follow up period), and
- Did not complete the study, (with summary of reasons for not completing the study).

The denominator for the percent of subjects in each category will be the number of subjects in the safety analysis set.

No inferential statistics will be generated. A data listing of reasons for premature study treatment/study discontinuation will be provided.

4.3 Extent of Exposure

4.3.1 Duration of Exposure to Study Drug

Duration of exposure to study drug will be defined as (last dose date – first dose date of double-blind treatment phase + 1), regardless of temporary interruptions in study drug administration, and will be expressed in weeks (shown to one decimal place, e.g., 4.5 weeks).

Duration of exposure to study drug will be summarized using descriptive statistics (sample size, mean, standard deviation, median, first quartile (Q1), third quartile (Q3), minimum, and maximum).

Summaries will be provided by treatment group for the safety analysis set.

4.3.2 Adherence with Study Drug

- The actual number of tablets ingested will be calculated using the Study Drug Accountability CRF:

Total number of tablets dispensed - total number of tablets returned.

- The expected number of tablets ingested will be calculated based on the Study Drug Dispensed CRF and Study Drug Accountability CRF:

  Total number of days supposed on study drug ([date of last dose of study drug - start date + 1])

- Adherence rate

  100(total number of tablets ingested) / (expected number of tablets ingested).

Adherence will be capped at 100% for summarization. Patients with unreturned bottles will be excluded from adherence calculations.

Descriptive statistics for adherence (sample size, mean, standard deviation, median, Q1, Q3, minimum, and maximum) will be provided by treatment group and overall for the safety analysis set.

5 EFFICACY ANALYSES

5.1 Analysis of the Primary Efficacy Endpoint

Analysis of the primary efficacy endpoints will be performed on the modified ITT data set, first on the entire cohort, then within the subgroups of patients with and without diabetes, and then within other subgroups as specified in section 2.3. Several sensitivity analyses will be performed as outlined in section 2.3 for both dual primary efficacy endpoints; in addition, supportive analyses will be repeated using the same models on the on-treatment set (OTS) and per-protocol set (PPS).

5.1.1 The first dual primary endpoint: NT pro-BNP

Mean, standard deviation, median, and interquartile range (IQR) will be reported for NT pro-BNP and for its change from baseline at 6- and 12-week follow-up, for the entire cohort and by treatment group.

A generalized linear mixed model will be used to estimate the effect of treatment on the average of 6- and 12-week NT pro-BNP values, adjusting for log baseline NT pro-BNP, history of DM, eGFR, and age. Random effects will be included for patient. A gamma distribution and log link function will be used to account for the skewed nature of NT pro-BNP. The model is as follows:

$$\mathrm{E}(y_{ijk}) = \mu_{ijk}, \qquad \mathrm{Var}(y_{ijk}) = \mu_{ijk}^2 \phi,$$

$$\log(\mu_{ijk}) = \beta_0 + \beta_1 \mathrm{Trt}_{ij} + \beta_2 t_k + \beta_3 \mathrm{Trt}_{ij} * t_k + \boldsymbol{\beta x} + \gamma_{ij},$$

$$\gamma_{ij} \sim \mathrm{N}(0, \sigma^2),$$

where, μ_{ijk} denotes the expected NT pro-BNP level for patient j from site i at time k, Trt_{ij} is a 0/1 variable denoting treatment group (dapagliflozin vs. placebo), t_k indicates the follow-up assessment time ($t_6 = 0, t_{12} = 1$), and $\mathbf{x}$ is the design matrix for log baseline NT pro-BNP, age, DM, and baseline eGFR. γ_{ij} are patient random effects with variance σ^2. ϕ is a scale factor. With this parameterization, the quantity $\exp(\beta_1 + \beta_3/2)$ represents the expected relative effect of dapagliflozin vs. placebo on the geometric mean of 6- and 12-week NT pro-BNP. The quantities $\exp(\beta_1)$ and $\exp(\beta_1 + \beta_3)$ represent the corresponding relative effects at 6 and 12 weeks, respectively.

A sensitivity analysis will be performed using the same model additionally adjusting for AFib type (No AFib, persistent/permanent AFib, paroxysmal AFib) and baseline LVEF.

Additional sensitivity analysis will repeat the primary analysis but will also include site as a random effect to account for potential clustering by enrolling center. A different sensitivity analysis will use the same model as the primary analysis, but adjust only for the corresponding baseline measurements, without additional covariates.

Finally, another sensitivity analysis will repeat the primary analysis on the on-treatment set (OTS).

5.1.2 The second dual primary endpoint: meaningful change in NT pro-BNP or KCCQ

Unadjusted proportions of patients achieving meaningful change in NT pro-BNP or KCCQ (a ≥ 5-point increase in average of 6- and 12-week KCCQ overall summary scores or a ≥ 20% decrease in average of 6- and 12-week NT pro-BNP), will be reported for the treatment group and placebo group. A mixed-effect logistic regression model will be used to assess the effect of treatment on the proportion of patients achieving meaningful change on 6 and 12 weeks average KCCQ and NT pro-BNP. Models will be adjusted for log baseline NT pro-BNP, baseline KCCQ overall summary score, history of DM, eGFR, and age. The model is specified as follows:

$$\mathrm{E}(\mathrm{y}_{ij}) = \mathrm{p}_{ij}, \qquad \mathrm{Var}(\mathrm{y}_{ij}) = \mathrm{p}_{ij}(1 - \mathrm{p}_{ij})$$

$$\mathrm{logit}(\mathrm{p}_{ij}) = \beta_0 + \beta_1 \mathrm{Trt}_{ij} + \boldsymbol{\beta}\mathbf{x}$$

where p_{ij} denotes the expected probability for patient j from site i achieving meaningful change in NT pro-BNP or KCCQ, $\mathrm{Trt}_{ij}$is a 0/1 variable denoting treatment group (dapagliflozin vs. placebo), $\mathbf{x}$ is the design matrix for log baseline NT pro-BNP, baseline KCCQ summary score, history of DM, eGFR, and age. With this parameterization, the quantity $\exp(\beta_1)$ represents the odds ratio of achieving meaningful change in the treatment group vs. the placebo group.

A sensitivity analysis will be performed using the same model additionally adjusting for AFib type (No AFib, persistent/permanent AFib, paroxysmal AFib) and baseline LVEF.

Additional sensitivity analysis will repeat the primary analysis but will also include site as a random effect to account for potential clustering by enrolling center. A different sensitivity analysis will use the same model as the primary analysis, but adjust only for the corresponding baseline measurements, without additional covariates.

Additional sensitivity analyses will repeat the primary analysis on the on-treatment set (OTS); and repeat the primary analysis excluding patients with baseline KCCQ overall summary score of over 90 (as these patients will have a limited opportunity for improvement in KCCQ).

Additional supportive analyses will be performed evaluating the empirical cumulative distribution function for change in NT pro-BNP, and for change in KCCQ overall summary score at 6 weeks and at 12 weeks (responder curves).

5.2 Secondary outcome variables

The following secondary outcomes will be analyzed on the intention to treat (modifed ITT) data set, first on entire patient cohort and then within subgroups of patient with or without diabetes.

1. Proportion of patients with a ≥ 5-point increase in average of 6- and 12-week KCCQ overall summary scores
2. Proportion of patients with a ≥ 20% decrease in average of 6- and 12-week NT pro-BNP
3. Proportion of patients with a ≥ 5-point increase in average of 6- and 12-week KCCQ overall summary scores and a ≥ 20% decrease in average of 6- and 12-week NT pro-BNP.
4. NT pro-BNP at 6 weeks and at 12 weeks separately.
5. Average of 6- and 12-week KCCQ overall summary scores.
6. Average of 6- and 12-week 6-minute walk tests. Data points for 6-minute walk test assessment that had associated notations of marked non-HF related limitations in exercise capacity (e.g., orthopedic injury, arthritis, etc.) will be excluded.
7. Average of 6- and 12-week BNP levels.
8. Average of 6- and 12-week HbA1c levels.
9. Average of 6- and 12-week weight(lbs).
10. Average of 6- and 12-week supine systolic blood pressure measurements. Supine systolic blood pressure will be taken three times at each time point. The average will be used.

Outcomes 1 – 3 are binary outcomes and will be analyzed in a manner analogous to that of the second dual primary outcome. Sensitivity analyses for outcomes 1 and 3 will be performed using the same models among patients with KCCQ overall summary score $\leq$ 90. Separate models will also provide effect estimates for the 6 and 12 week time points.

Outcomes 6 – 10 are continuous outcomes and will be analyzed in a manner analogous to that of the first primary outcome, although appropriate distributions and link functions will be chosen for the given outcomes. Models will also provide separate effect estimates for the 6 and 12 week time points..

Outcome 5, KCCQ overall summary score, being semi-continuous measures and ranging between 0 and 100, will first be scaled to between 0 and 1 using formula: $Y^* = \frac{[Y(N-1)+0.5]}{100N}$, where N is total number of patients with KCCQ, and then will be analyzed in a manner analogous to that of the first primary outcome. Beta distribution and log link function will be used [2]. Analyses for KCCQ overall summary score will be performed for each of the time point (6 weeks and 12 weeks) separately, as well as for the average of 6- and 12-week measurements.

In addition, the following sensitivity analyses will be performed for all secondary endpoints:

1. A sensitivity analysis will repeat the primary analysis additionally adjusting for AFib type (No AFib, persistent/permanent AFib, paroxysmal AFib) and baseline LVEF. 2. A sensitivity analysis will repeat the primary analysis but also include site as a random effect to account for clustering by center

3. A sensitivity analysis will repeat the primary analysis, adjusting only for the corresponding baseline measurements, but not other covariates

4. A sensitivity analysis will repeat the primary analysis on the on treatment set (OTS).

5.3 Exploratory outcome variables

The following exploratory outcomes will be summarized descriptively for the treatment group and placebo group with outcomes 1-3 on the safety analysis dataset (SAF) and outcomes 4 and 5 on the intention to treat (IIT) dataset. This will be completed first on the entire cohort and then within subgroups of patients with or without diabetes. Continuous measures will be summarized by mean $\pm$ standard deviation and compared using Student's T-tests. Categorical variables will be summarized by frequency and percent and compared using χ^2or Fisher's exact tests, as appropriate

1. Proportion of patients with hospitalizations for heart failure over 12 weeks.
2. Proportion of patients with urgent outpatient heart failure visits over 12 weeks.
3. Proportion of patients with hospitalizations for heart failure or urgent outpatient heart failure visits over 12 weeks.
4. Daily loop diuretic dose (furosemide equivalent).
5. Proportion of patients with improved/worsened NYHA Class from baseline at 6- and 12- weeks.

In addition, a Cox proportional hazard model will be used to assess the effect of treatment vs. placebo on the time to the first occurrence for outcomes 1-3, adjusting for history of DM, eGFR, and age. A sensitivity analysis will be conducted additionally adjusting for AFib type (No AFib, persistent/permanent AFib, paroxysmal AFib) and baseline LVEF. A 2nd sensitivity analysis will be performed using stratified Cox proportional model conditional on site to account for clustering. A third sensitivity analysis will be performed using the same approach as the main analysis, but without adjustment for covariates

Daily loop diuretic dose will be analyzed in a manner analogous to that of the first primary endpoint, although appropriate distributions and link functions will be chosen for the given outcomes. Additionally, among patients who were on loop diuretic at randomization, we will also compare the proportion of patients who had loop diuretic dose reduced or discontinued (using the average of 6- and 12-week doses, and also separately at 6 and 12 week timepoints).

Change in NYHA will be analyzed in a manner analogous to that of the first dual primary endpoint, although appropriate distributions and link functions will be chosen for the given outcome.

5.4 Supplemental pre-specified analyses.

For all outcomes that include assessment of KCCQ overall summary score, additional sensitivity analyses will be performed using KCCQ clinical summary score, in a manner analogous to that of the KCCQ overall summary score.

In addition, additional subgroup analyses (in addition to stratification by DM status, using strata specified in section 2.3) may be repeated for all secondary efficacy outcomes.

5.5 Substudy Endpoints.

5.5.1 Arrhythmia Substudy:

The arrhythmia burden outcomes (sustained VT episodes, non-sustained VT episodes, VF episodes, AFib burden, PAC, and PVC burden) at baseline (the period of 14 days with Holter monitor on) and 6 weeks (the period of 14 days with Holter monitor on) will be summarized for the treatment group and placebo group on the Intention to teat (IIT) dataset, first on entire cohort and then within subgroups of patients with or without diabetes.

Negative binomial mixed linear models with log link will be used to assess the effect of dapagliflozin on sustained VT episodes, non-sustained VT episodes, and VF episodes at 6 weeks. The logarithm of the number of days the patient wore a Holter monitor will be used as an offset variable. Each arrhythmia burden outcome will be modeled separately and the corresponding baseline values (defined as number of episodes per day during the 14 days of baseline establish period) will be adjusted for.. Addition fixed effects will be included for age, DM, and eGFR.

AFib burden, PAC and PVC will be analyzed in a manner analogous to that of the first primary endpoint, although the appropriate distribution and link function will be chosen for the given outcome. For analyses of AFib burden and PAC burden, patients with known permanent/persistent AFib at baseline will be excluded.

A sensitivity analysis additionally adjusting for AFib type (No AFib, persistent/permanent AFib, paroxysmal AFib) and LVEF will be performed for each arrhythmia burden outcome (sustained VT episodes, non-sustained VT episodes, VF episodes, AFib burden, PAC, and PVC burden).

Furthermore, a sensitivity analysis on patients with permanent/ persistent AFib at baseline will be performed for sustained VT episodes, non-sustained VT episodes, VF episodes, and PVC burdens.

Finally, the primary analysis will be repeated with site included as a random effect to account for clustering of enrolling sites.

Additional analyses maybe performed as appropriate following evaluation of the Holter monitor data.

5.5.2 Sensivest Substudy (at selected sites):

Mean, standard deviation, median, and interquartile range (IQR) will be reported for the lung fluid volume at follow-ups and its changes from baseline to 6 weeks, baseline to 12 weeks, and 12 weeks to 13 weeks, for the entire cohort and by treatment group.

A beta-distribution mixed model will be used to estimate the effect of treatment on lung fluid volume which is measured as percentage. Fixed effects will be included for treatment group, baseline lung fluid volume, follow-up assessment time, age, DM, and baseline eGFR. Random effects will be included for patient. The model is as follows:

$$\mathrm{E}(y_{ijk}) = \mu_{ijk}, \qquad \mathrm{Var}(y_{ijk}) = \mu_{ijk}(1-\mu_{ijk})/(1+\phi),$$

$$\log(\mu_{ijt}) = \beta_0 + \beta_1 \mathrm{Trt}_{ij} + \beta_{2.1} t_{(1)} + \beta_{2.2} t_{(2)} + \beta_{3.1} \mathrm{Trt}_{ij} t_{(1)} + \beta_{3.2} \mathrm{Trt}_{ij} t_{(2)} + \mathbf{\beta x} + \gamma_{ij},$$

$$\gamma_{ij} \sim \mathrm{N}(0, \sigma^2),$$

where μ_{ijt} denotes the expected lung fluid volume for patient j from site i at time t, Trt_{ij} is a 0/1 variable denoting treatment group (dapagliflozin vs. placebo), $t_{(1)}$and $t_{(2)}$ indicate the follow-up assessment time (6weeks: $t_{(1)} = 0, t_{(2)} = 0$; 12weeks: $t_{(1)} = 1, t_{(2)} = 0$; 13weeks: $t_{(1)} = 0, t_{(2)} = 1$), $\mathbf{x}$ is the design matrix for age, DM, baseline lung fluid volume, and baseline eGFR. γ_{ij} are patient random effects with variance σ^2. ϕ is a scale factor. With this parameterization, the quantity $\exp(\beta_1)$ represents the expected dapagliglozin effect at 6 weeks. Similarly,$\exp(\beta_1 + \beta_{3.1})$ is the expected dapagliflozin effect at 12 weeks. And $\exp(\beta_{3.2} - \beta_{3.1})$ can be used to test, on cessation of dapagliflozin or placebo, whether lung fluid concentration will increase in patients treated with dapagliflozin to a greater extent than in those treated with placebo.

The association between the change in lung fluid volume from baseline to 6 and 12 weeks and clinical outcomes, including heart failure specific biomarkers (BNP, NT pro-BNP), as well as patients' symptoms and functional status (KCCQ overall summary score and 6-minute walk test), will be examined using generalized mixed linear models with clinical outcomes as dependent variables (separate models) and the change of lung fluid volume at 6 and 12 weeks as independent variables. Other potential confounders are age, eGFR, AFib type (No AFib, persistent/permanent AFib, paroxysmal AFib), DM, and LVEF.

The association between the change of lung fluid volume from baseline at 6 and 12 weeks and novel heart failure biomarkers, including Uric Acid, myeloperoxidase level (MPO), Fibrosis (galectin-3 and soluble ST-2), Myocyte necrosis (hs-cTnT), and Inflammation (hsCRP and interleukin-6), will be assessed similarly.

Sensitivity analyses will be conducted additionally adjusting for AFib type (No AFib, persistent/permanent AFib, paroxysmal AFib) and baseline LVEF . Additional analyses may be performed as appropriate following evaluation of the Sensivest data.

5.5.3 Biomarker Substudy:

- Oxidative stress: Uric Acid
- Fibrosis: galectin-3 (Gal3) and ST-2
- Myocyte necrosis: High sensitivity cardiac troponin T (hs-cTnT)

- Inflammation: high-sensitivity C reactive protein (hsCRP) and interleukin-6 (Il-6)
- Advanced glycation end-products: soluble Receptor for Advanced Glycation End Products(sRAGE) and N(6)-Carboxymethyllysine (CML).

Mean, standard deviation, median, and interquartile range (IQR) will be reported for the biomarkers listed above and for their changes from baseline at 6- and 12-week follow-up, for the entire cohort and by treatment group.

The effect of dapagliflozin on the biomarkers at 6- and 12-week will be analyzed in a manner analogous to that of the first primary outcome, although appropriate distributions and link functions will be chosen for the given outcomes.

The relationship between biomarkers and clinical outcomes, including heart failure specific biomarkers (BNP, NT pro-BNP), as well as patients' symptoms and functional status (KCCQ overall summary score and 6-minute walk test), will be examined using generalized mixed linear models with clinical outcomes as dependent variables (separate models). Independent variables will include the changes from baseline at 6- and 12-week in the aforementioned biomarkers and other potential covariates (age, eGFR, AFib type, DM, and LVEF). The biomarkers will be included in separate models one at a time. Center and patient will be included as random factors. The appropriate distributions and link functions will be chosen for the given outcomes

Additional analyses maybe performed as appropriate following evaluation of the biomarker data.

6 SAFETY ANALYSES

Safety analyses will be performed on the safety analysis set (SAF). Total number of adverse events as well as number and proportion of patients developing adverse event(s) will be compared by treatment group. For patient level analyses multiple events will be counted once only per subject in each summary. The following safety variables will be included:

1. All cause death
2. Cardiovascular death
3. Non-fatal myocardial infarction (MI)
4. Stroke
5. Acute kidney injury (defined as doubling of serum creatinine based on the modified RIFLE criteria)
6. Adverse events (AEs).

    - Adverse events of special interest
        - Diabetic Ketoacidosis
        - Volume Depletion Event (defined as hypotension, syncope, orthostatic hypotension or dehydration)
        - Severe Hypoglycemic Event
        - Lower Limb Amputations
    - Drug Adverse Event
    - Serious Adverse event
        - Resulted in death
        - In-patient hospitalization or prolonging of existing in-patient hospitalization

- Persistent or significant disability
- Life-threatening
- Congenital anomaly/birth defect
- Important medical event

Safety analyses will be restricted to adverse events that occurred between randomization and 12 weeks. Adverse events occurred between 12 weeks and 13 weeks will be presented separately in a supplemental analysis.

7 REFERENCES

1. Green CP1, Porter CB, Bresnahan DR, Spertus JA., Development and Evaluation of the Kansas City Cardiomyopathy Questionnaire: A New Health Status Measure for Heart Failure, J Am Coll Cardiol. 2000 Apr;35(5):1245-55
2. Hunger,M et all, Longtitudinal beta regression models for analyzing health-related quality of life scores over time Medical Research Methodology, 2012; 12:144

8 SOFTWARE

All analyses will be performed using SAS 9.4 or higher.

9 APPENDICES

9.1 Scoring and Interpreting the KCCQ

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

A. Physical Limitation

The Physical Limitation score corresponds to questions 1a through 1f. Responses to questions 1a through 1f should be coded numerically as follows:

1 = Extremely Limited
2 = Quite a bit Limited
3 = Moderately Limited
4 = Slightly Limited
5 = Not at all Limited
6 = Limited for other reasons or did not do the activity

If the responses to questions 1a through 1f are not values 1, 2, 3, 4 or 5 then the response is set to missing. Note that a response of 6 (Limited for other reasons or did not do the activity) is treated as a missing value. If at least three responses to questions 1a-1f are not missing, then the physical limitation score is computed by calculating the mean response and standardizing the result as follows:

Physical Limitation = 100*(Mean Response – 1)/4

B. Symptom Stability

The Symptom Stability score corresponds to question 2. Responses to question 2 should be coded numerically as follows:

1 = Much Worse
2 = Slightly Worse
3 = Not Changed
4 = Slightly Better
5 = Much Better
6 = I've had no symptoms over the last 2 weeks

If the response is 6 (no symptoms over last 2 weeks) then set the response to 3 (not changed). If question 2 is not missing then the symptom stability score is computed by standardizing the result as follows:

Symptom Stability = 100*(Response – 1)/4

C. Symptom Frequency

The Symptom Frequency score corresponds to questions 3, 5, 7 and 9. The responses should be coded sequentially (1, 2, 3…) in order of increasing health status as follows:

Question 3

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

Questions 5 and 7

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

Question 9

1 = Every night

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

If two or more responses are missing then symptom frequency cannot be computed and will be missing. Otherwise, the symptom frequency is computed by calculating the mean of the standardized responses and multiplying by 100 as follows:

$$\text{Symptom Frequency} = 100*\text{Mean}((Q3 - 1)/4, (Q5 - 1)/6, (Q7 - 1)/6, (Q9 - 1)/4)$$

D. Symptom Burden

The Symptom Burden score corresponds to questions 4, 6 and 8. The responses should be coded numerically as follows:

1 = Extremely Bothersome
2 = Quite a bit Bothersome
3 = Moderately Bothersome
4 = Slightly Bothersome
5 = Not at all Bothersome
6 = I've had no swelling (fatigue, shortness of breath)

If a response is 6 (none) then set the response to 5 (not at all). If at least one response is present then symptom burden score is computed by calculating the mean response and standardizing the result as follows:

$$\text{Symptom Burden} = 100*(\text{Mean Response} - 1)/4$$

E. Total Symptom Score

The total symptom score is calculated as the mean of the symptom frequency score and symptom burden score.

F. Self-Efficacy

The Self-Efficacy score corresponds to questions 10 and 11. Responses to questions 10 and 11 should be coded sequentially (1, 2, 3, 4, 5) in order of increasing health status, with 1 denoting the response associated with the lowest health status. If at least one question response is present then the self-efficacy score may be computed by standardizing the mean response as follows:

$$\text{Self-Efficacy} = 100*(\text{Mean Response} - 1)/4$$

G. Quality of Life

The Quality of Life score corresponds to questions 12, 13 and 14. Responses to questions 12, 13 and 14 should be coded sequentially (1, 2, 3, 4, 5) in order of increasing health status, with 1 denoting the response associated with the lowest health status. If at least one question response is present then the quality of life score may be computed by standardizing the mean response as follows:

$$\text{Quality of Life} = 100*(\text{Mean Response} - 1)/4$$

H. Social Limitation

The Social Limitation score corresponds to questions 15a through 15d. These responses should be coded numerically as follows:

1 = Severely Limited
2 = Limited Quite a bit
3 = Moderately Limited
4 = Slightly Limited
5 = Did Not Limit at All
6 = Does not apply or did not do for other reasons

If the responses to questions 15a through 15d are not values 1, 2, 3, 4 or 5 then the response is set to missing. Note that a response of 6 is treated as a missing value. If at least two question responses are present then the social limitation score may be computed by standardizing the mean response as follows:

$$\text{Social Limitation} = 100*(\text{Mean Response} - 1)/4$$

I. Clinical Summary Score

The clinical summary score is calculated as the mean of the physical limitation score and total symptom score.

J. Overall Summary Score

The overall summary score is calculated as the mean of the physical limitation score, total symptom score, quality of life score and social limitation score.

9.2 Protocol deviations

Number	Important Protocol Deviations Criteria	Major Impact on Analysis (MIA)/Other	Exclusion level
1. Did not fulfil eligibility criteria – inclusion criteria:			
1.1	Ability to provide informed consent prior to initiating screening visit procedures	MIA	Complete exclusion from PPS and MODIFIED ITTS
1.2	Age > 18 and < 120 at the screening visit	Other	No exclusion
1.3	Established diagnosis of heart failure (for at least 16 weeks prior to the screening visit) with reduced systolic function (LVEF≤40% due to either ischemic or non-ischemic etiology) documented by an imaging modality within the past 24 months. Any local measurement of LVEF by any modality within the eligibility range made within the past 24 months is acceptable provided there has been no subsequent LVEF measurement above 40%.	Other	To be decided on a case by case basis
1.4	No change in diuretic management for 1 week prior to screening visit or between the screening and randomization visit	Other	To be decided on a case by case basis
1.5	NYHA class II or III heart failure symptoms at the screening and randomization visit	Other	To be decided on a case by case basis

Number	Important Protocol Deviations Criteria	Major Impact on Analysis (MIA)/Other	Exclusion level
1.6	BNP ≥100 pg/mL and/or NT pro-BNP ≥ 400 pg/mL at the screening visit (For patients with permanent atrial fibrillation inclusion thresholds will be BNP ≥ 125 pg/mL or NTproBNP ≥ 600 pg/mL)	Other	To be decided on a case by case basis
2. Did not fulfil eligibility criteria – exclusion criteria:			
2.0	Decompensated heart failure (hospitalization for heart failure within the 30 days prior to screening or NYHA class IV heart failure symptoms at screening)	Other	To be decided on a case by case basis
2.1	History of type 1 diabetes	Other	To be decided on a case by case basis
2.2	Estimated glomerular filtration rate (eGFR) < 30 at the screening visit by modified MDRD equation GFR in mL/min per 1.73 m^2 = 175 x $SerumCr^{-1.154}$ x $age^{-0.203}$ x 1.212 (if patient is African American) x 0.742 (if female)	Other	To be decided on a case by case basis
2.3	Admission for an acute coronary syndrome (ST-elevation MI, non-ST-elevation MI, or unstable angina), percutaneous coronary intervention, or cardiac surgery within 30 days prior to the screening visit.	Other	To be decided on a case by case basis

Number	Important Protocol Deviations Criteria	Major Impact on Analysis (MIA)/Other	Exclusion level
2.4	Admission for cardiac resynchronization therapy (CRT) within 90 days prior to the screening visit	Other	To be decided on a case by case basis
2.5	Planned cardiovascular revascularization (percutaneous intervention or surgical) or major cardiac surgery (coronary artery bypass grafting, valve replacement, ventricular assist device, cardiac transplantation, or any other surgery requiring thoracotomy) or CRT within the 90 days after the screening visit.	Other	To be decided on a case by case basis
2.6	Participation in any interventional clinical trial (with an investigational drug or device) that is not an observational registry within the 8 weeks prior to the screening visit.	Other	To be decided on a case by case basis
2.7	History of hypersensitivity to dapagliflozin	Other	To be decided on a case by case basis
2.8	For women of child-bearing potential: Current or planned pregnancy or currently lactating.	Other	To be decided on a case by case basis

Number	Important Protocol Deviations Criteria	Major Impact on Analysis (MIA)/Other	Exclusion level
2.9	Women who are surgically sterile or those who are postmenopausal for at least 1 year are not considered to be of child-bearing potential. Women of child-bearing potential, who are sexually active, must agree to use a medically-accepted method of birth control for the duration of the study. Women of child-bearing potential will have a urine pregnancy test at every clinic visit and it must be negative to continue study participation.	Other	To be decided on a case by case basis
2.10	Life expectancy <1 year at the screening visit	Other	To be decided on a case by case basis
2.11	Patients who are volume depleted based upon physical examination at the time of the screening or randomization visit	Other	To be decided on a case by case basis
2.12	BNP <100 pg/mL and NT pro-BNP<400 pg/mL at the screening visit (For patients with permanent atrial fibrillation exclusion thresholds will be BNP<125 pg/mL and NTproBNP<600 pg/mL)	Other	To be decided on a case by case basis
2.13	Patients currently being treated with any SGLT-2 inhibitor (dapagliflozin, canagliflozin, empagliflozin) or having received treatment with any SGLT-2 inhibitor within the 12 weeks prior to the screening visit.	Other	To be decided on a case by case basis

Number	Important Protocol Deviations Criteria	Major Impact on Analysis (MIA)/Other	Exclusion level
2.14	Average supine systolic BP <90 mmHg at the screening or randomization visit	Other	To be decided on a case by case basis
2.15	Past or current history of bladder cancer	Other	To be decided on a case by case basis
2.16	Active Hematuria	Other	To be decided on a case by case basis
2.17	Donation of blood or bone marrow 12 weeks prior to the screening visit and no planned donations during the study period	Other	To be decided on a case by case basis
2.18	Heart failure due to restrictive cardiomyopathy, active myocarditis, constrictive pericarditis, severe stenotic valve disease, and HOCM (hypertrophic obstructive cardiomyopathy).	Other	To be decided on a case by case basis

3. Patient developed discontinuation of investigational product criteria but dosing continued			
3.0	Subject experienced an Adverse Event which in the opinion of the Investigator and/or AstraZeneca, contraindicated further dosing	Other	To be reviewed on the case by case basis
3.1	Pregnancy confirmed by a positive pregnancy test or other examinations	Other	Complete exclusion from PPS
3.2	Donation of blood or bone marrow during the study period	Other	To be decided on a case by case basis
3.3	Admission for an acute coronary syndrome (ST-elevation MI, non-ST-elevation MI, or unstable angina), percutaneous coronary intervention, or cardiac surgery during the study period.	Other	To be decided on a case by case basis
4. Patient received prohibited concomitant medication but study treatment not discontinued:			
4.4.1	SGLT2	Other	Only the efficacy endpoints before the start of SGLT2 will be included in modified IITS, OTS, and PPS
5. Patient received incorrect investigational treatment/dose			
5.0	Patient took incorrect treatment or damaged Investigational Drug Kit	MIA	To be decided on a case by case basis
5.1	Major compliance issues <80% or >120% compliance with IP dosing in double blind treatment period	MIA	To be decided on a case by case basis
5.3	Patient randomized but never took IP	MIA	To be decided on a case by case basis

5.4	Overdose	MIA	To be decided on the case by case basis
6. Protocol-required procedure not adhered to:			
6.0	Study data collection has not been stopped when subject decided to withdraw their consent from the study completely	Other	No exclusion from PPS Data collected after consent withdrawal to be excluded from data base
6.1	Patients with absence of NT-proBNP data at randomization visit and having values at Screening visit.	Other	No exclusion from PPS Values from screening visit will be used to impute the values at randomization visit.
6.2	Patients with absence of baseline NT-proBNP data at both screening and randomization visits	Other	Excluded from the NT-proBNP analyses
6.3	Patients with no NT-proBNP values after randomization	Other	Excluded from the NT-proBNP analyses
6.4	Patients with absence of baseline KCCQ summary score at randomization visit	Other	Excluded from the KCCQ analyses
6.5	Patients with no KCCQ overall summary score after randomization	Other	Excluded from KCCQ analyses
6.6	Patients with absence of baseline 6-minute walk assessment at randomization visit	Other	Excluded from the 6-minute walk analyses
6.7	Patients with no 6-minute walk assessment after randomization	Other	Excluded from 6-minute walk analyses
6.8	Patients with missing values in the covariates: age, gender, Hx AFib, AFib type, Hx DM, eGFR at randomization, lvef at randomization	Other	No exclusion from PPS Excluded from the first and second primary analysis