Janssen Research & Development, LLC

Statistical Analysis Plan

Canagliflozin: Impact on Health Status, Quality of Life, and Functional Status in Heart Failure

Protocol 28431754HFA3002; Phase 3B

JNJ-28431754 (canagliflozin) (INVOKANA)

Status:	FINAL
Date:	19 July 2021
Prepared by:	Janssen Research & Development, LLC
Document No.:	EDMS-RIM-294497, 2.0

Compliance: The study described in this report was performed according to the principles of Good Clinical Practice (GCP).

TABLE OF CONTENTS

LIST OF IN-TEXT TABLES

LIST OF IN-TEXT FIGURES

AMENDMENT HISTORY

1st Amendment (December 30, 2020)

Final changes (July 1, 2021)

ABBREVIATIONS

AE	adverse event
API	application program interface
ANOVA	analysis of variance
CDF	cumulative distribution function
CI	confidence interval
CSR	clinical study report
ED	emergency department
EF	ejection fraction
EHR	electronic health record
EOS	end of study
EOT	end of treatment
ES	effect size
FAS	Full analysis set
FDA	Food and Drug Administration
HF	heart failure
HFpEF	heart failure preserved ejection fraction
HFrEF	heart failure reduced ejection fraction
HRU	healthcare resource utilization
ICD	International Statistical Classification of Diseases and Related Health Problems
ICH	International Conference on Harmonisation
IHN	integrated health networks
ITT	intent to treat
IxR	interactive web response system
KCCQ	Kansas City Cardiomyopathy Questionnaire
KCCQ-TSS	Kansas City Cardiomyopathy Questionnaire – Total Symptom Score
LOCF	last observation carried forward
LS	least squares
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	mixed model of repeated measurement
MI	multiple imputation
MID	minimum detectable
NYHA	New York Heart Association
PGIC	Patient Global Impression of Change
PGI-S	Patient Global Impression of Severity
PI	principal investigator
PP	per protocol
SAE	serious adverse event
SAP	statistical analysis plan
SAS	Statistical Analysis System
SD	standard deviation
SE	standard error
SGLT2	sodium glucose cotransporter-2 inhibitor
SUSAR	suspected unexpected serious adverse reaction
T2DM	type 2 diabetes mellitus
TEAE	treatment-emergent adverse event
UGT	UDP-Glucuronosyltransferase
US	United States

Amendment Page – Summary of Changes (December 30, 2020)

Section #	Original Text/Figure	Amended Text/Figure	Rationale
1.2	Figure 1	Figure 1	Changed entry criterion for KCCQ <80 incorporated
1.4	Added text	After approximately 10 months, the planned enrollment in this study is not expected to reach the target of 1900 participants by the end of January 30, 2021 (the scheduled end of study enrollment window). Instead, the final anticipated enrollment may be in the range of about 400-480 participants. Consequently, a reduction in total sample size of this magnitude will have impact on statistical power at different effect sizes, as shown by the figure below. A large effect size was reported in the recently concluded DEFINE-HF study[17]. Additional KCCQ domains were also reported in this study with robust improvements over baseline[17,18].	Statement about lower than expected sample size enrollment and its impact on statistical power
2.3.2.	Added full analysis set (FAS) definition	This includes all randomized participants who have received at least one dose of study intervention/medication and have at least one post-baseline KCCQ measurement. Analyses of the primary, secondary, and exploratory endpoints will be based on the full analysis set (FAS).	Primary, secondary, and exploratory endpoints will be based on FAS (earlier version stated as based on all randomized set)
5.2.2.	Population: all randomized set	Population: FAS - all randomized participants with HF (who either have HFrEF or HFpEF, regardless of T2DM status) with at least one dose of study intervention and at least 1 post-baseline KCCQ measurement	Changed analysis population to FAS
5.2.3.	Added to the analysis method	Mortality in this short-duration trial is expected to be low and to be comparable between groups and will be ignored in the primary analyses of mean differences in changes in KCCQ Total Symptom scores, but can be included in the responder analyses by assigning them to the lowest response (greatest deterioration) category. Should there be a >5% mortality rate and if it is unequally distributed between treatment arms, then joint modeling of survival and health status will be performed. A win-ratio approach[19,20] may also be considered by analyzing the KCCQ-TSS score as a composite, rank-based outcome, incorporating participant vital status at 12	Added to the analysis method regarding missing data due to deaths

		weeks along with a change in score from baseline to 12 weeks in surviving participants. This essentially is the rank analysis of covariance method, with a corresponding win ratio used to estimate the magnitude of treatment effect.	
5.3.2.	Population: all randomized set	Population: FAS - all randomized participants with HF (who either have HFrEF or HFpEF, regardless of T2DM status) with at least one dose of study intervention and at least 1 post-baseline KCCQ measurement	Changed analysis population to FAS
5.3.3.	Added to analysis method	The first key secondary endpoint will be analyzed following the same approach as employed for the primary effectiveness endpoint (i.e. using an MMRM model).	Added to analysis method regarding daily step counts from two-sample t-test to MMRM analysis
References	Added references	Nassif, M. et al. (2019). Dapagliflozin Effects on Biomarkers, Symptoms, and Functional Status in Patients With Heart Failure With Reduced Ejection Fraction - The DEFINE-HF Trial. *Circulation*, Vol. 140, No. 18, 1463-1476 Bhatt, D. et al. (2020). Sotagliflozin in Patients with Diabetes and Recent Worsening Heart Failure. *N Engl J Med*, 2020; Nov 16 Gasparyan, S. et al. (2019). Adjusted Win Ratio with Stratification: Calculation Methods and Interpretation, *Stat.ME* Wang, D. and Pocock, S. (2016). A win-ratio approach to comparing continuous non-normal outcomes in clinical trials, *Pharmaceutical Statistics*, online 11 March 2016	Added more references pertaining to sample size section and analysis methods such as win-ratio

1. INTRODUCTION

This Statistical Analysis Plan (SAP) describes the planned analysis and reporting for the 28431754HFA3002 **(Canagliflozin: Impact on Health Status, Quality of Life, and Functional Status in Heart Failure – CHIEF-HF)** study. This Phase 3b study is being conducted in the United States (US) in participants with symptomatic HF stratified by HF (reduced ejection fraction [HFrEF] or preserved ejection fraction [HFpEF]) and compares the effectiveness of canagliflozin 100 mg versus placebo for improvement in HF symptoms measured by the Kansas City Cardiomyopathy Questionnaire - Total Symptom Score (KCCQ-TSS) (see Appendix 1: The Kansas City Cardiomyopathy Questionnaire (KCCQ)) after 12 weeks. This document contains details of the relevant definitions, data handling conventions and methods for analyses. The document will be finalized prior to database lock.

This SAP is based on the clinical study protocol (finalized first on 7 November, and subsequently amended on 7 February 2020 (Protocol **28431754HFA3002 Amendment 1**, dated **07 February 2020 - EDMS-ERI-154820237, version 3.0)** by Janssen Scientific Affairs, LLC) that provides details on the conduct of this study and operational aspects of clinical assessments and timing for completing a participant in this study.

Based on the last protocol amendment, changes were made to this SAP on December 30, 2020 and are listed on the preceding amendment page in this document. A final version of this SAP with editorial changes is posted on 1 July 2021.

1.1. Objectives

Primary Objective

The primary objective is to determine the superiority of the effectiveness of canagliflozin 100 mg daily versus placebo in participants with symptomatic HF in improving the overall KCCQ-TSS.

Secondary Objectives

The first key secondary objective is:

- to determine the superiority of the effectiveness of canagliflozin 100 mg daily versus placebo in improving the total daily step count.

The second key secondary objective is:

- to determine the superiority of the effectiveness of canagliflozin 100 mg daily versus placebo in improving the KCCQ individual domain scores (physical limitation, quality of life, clinical summary, and overall).

Exploratory Objectives

The exploratory objectives are to assess the effectiveness of canagliflozin 100 mg daily versus placebo in participants with symptomatic HF by:

- responses on the Patient Global Impression of Change (PGIC) and the Patient Global Impression of Severity (PGI-S)
- healthcare resource utilization (HRU) and health economics data
- associations (correlations and/or categorical association measures) between digital markers (step count and floors climbed), and clinical events such as need for outpatient intravenous therapies, emergency department (ED) visits, and hospitalizations
- daily floors climbed
- time to first hospitalization/readmission for HF
- time to first hospitalization/readmission for other events
- time to death
- the participant's satisfaction with their experience with the virtual design at the end of the study (see Appendix 2: Participant Satisfaction Survey).

1.2. Study Design

OVERALL DESIGN

This is a randomized, double-blind, placebo-controlled, decentralized, virtual, superiority study conducted in the US in participants with symptomatic HF (stratified by heart failure with HFrEF or HFpEF). This study examines the improvement in the KCCQ-TSS after 12 weeks of treatment with canagliflozin 100 mg or placebo. The study will continue for an additional 6 months without further study intervention to collect participant follow-up data from the smartphone, Fitbit, and claims in a real-world setting. At the end of the 12-week, double-blind treatment period, participants will be informed about their actual treatment allocation, whether they had been randomized to canagliflozin or placebo.

The study will enroll participants from large integrated health networks (IHNs) and physician practices in the US with the participants conducting many of their study-related activities through an app on a smartphone and will provide daily activity data recorded by using a Fitbit device. There will be no in-person clinic visits required by the study (see Appendix 3: Schedule of Activities).

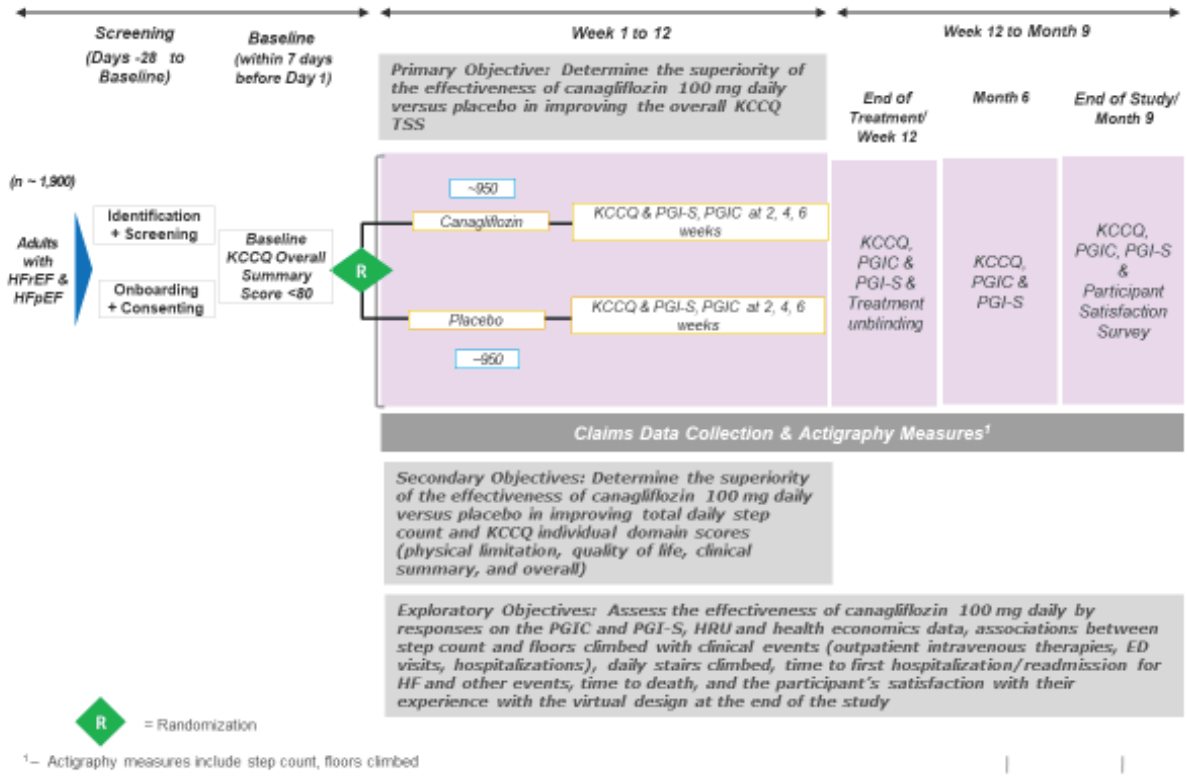
The study consists of a screening period, a 12-week double-blind treatment period, and a 6-month follow-up (no intervention) observational period for each participant. Approximately 1,900 participants are planned for randomization into the study in a 1:1 ratio to canagliflozin 100 mg daily or placebo groups (950 in each study intervention group). After Week 12 (i.e., the end of the double-blind treatment period), the study will continue for an additional 6 months when study intervention will no longer be provided. Randomization will be performed using the interactive web response system (IxR). Participants will be stratified based on HF ejection fraction (EF) type – HFrEF or HFpEF.

Data collection from the smartphone, Fitbit device, and claims will continue until Month 9 for all endpoints. Specifically, study assessments include the following: KCCQ, PGIC, PGI-S, and an

optional participant satisfaction survey at end of study (via smartphone), step count and floors climbed (via the Fitbit device), medical and pharmacy claims data (including up to 36 months history for eligibility confirmation), HRU and health economics (including hospitalization/readmission due to HF or other reason, ED visits, the duration of stay at each hospitalization/readmission), review of claims for prohibited medications and new diagnoses that may require discontinuation, and adverse events. In addition, information related to outpatient IV diuretics, inotropes and vasodilators and in-patient IV therapies as well will be considered relevant. Note that claims data that involves up to 36 months history to confirm eligibility into this study will not be part of the study database for analysis.

Participants will be of any gender, 18 years of age or older, and will have clinically stable, symptomatic HF with or without type 2 diabetes mellitus (T2DM) with a KCCQ overall summary score <80 prior to randomization, which represents New York Heart Association (NYHA) II and III participants.

Figure 1: Schematic Overview of the Study



1.3. Statistical Hypotheses for Trial Objectives

The primary effectiveness endpoint is the change in the KCCQ-TSS from baseline to Week 12. The null hypothesis to be tested here is that there is no difference between canagliflozin 100 mg and placebo in the treatment of symptomatic HF based on the primary effectiveness endpoint. For

the first key secondary effectiveness endpoint of change in daily step count from baseline to Week 12, the null hypothesis to be tested is that canagliflozin 100 mg is not different from placebo with respect to the mean change in daily step count.

The two hypotheses described above will be tested using a fixed-sequence (hierarchical) testing procedure to control the familywise type I error rate at 5% as follows:

1. (primary effectiveness endpoint): change from baseline to Week 12 in KCCQ-TSS for canagliflozin 100 mg versus the placebo
2. (first key secondary endpoint): change from baseline to Week 12 in daily step count for canagliflozin 10 mg versus placebo.

If the result of the 2-sided p-value from the first hypothesis (i.e., primary effectiveness endpoint) test is ≤0.05, then the second hypothesis test result will be interpreted inferentially at the 5% significance level. This sequential testing approach will control the familywise type I error rate at 5%.

Hypotheses related to the second key secondary effectiveness endpoint (i.e., KCCQ individual domain scores defined by changes in physical limitation, quality of life, clinical summary, and overall summary from baseline to Week 12) as well as other exploratory endpoints will be tested at the nominal significance level (5%) without adjustment for multiplicity (or controlling type I error rate).

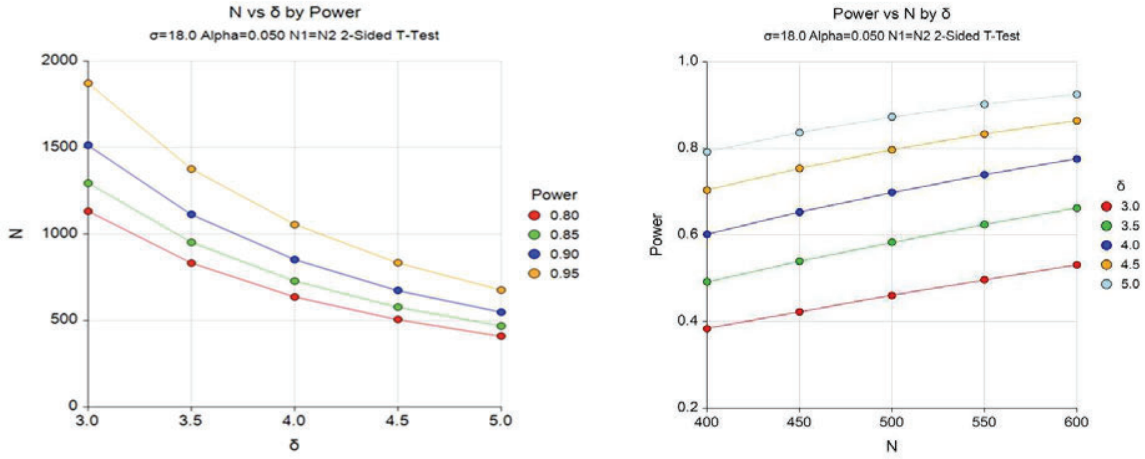
1.4. Sample Size Justification

The primary effectiveness endpoint is the change from baseline in the KCCQ-TSS at the end of the 12-week double-blind treatment period. Although a 5-point improvement in the KCCQ-TSS is generally viewed as clinically meaningful,[2,12] recent studies[7,8] have shown that improvements in the range of 3 to 5 points on the KCCQ-TSS are equally meaningful in the overall HF-related health status measured by the KCCQ over a 12-week time period. In considering the clinical significance of scores, it is important to separate mean changes for groups of participants, as compared with intra-individual changes. On the TSS, there are 7 items (4 for symptom frequency and 3 for severity/burden).[3,11] Shifting a response on KCCQ questions 5 and 7 shifts the TSS score by 2.083 points on a 100-point scale. A shift in response for questions 3 and 9 shifts the score by 3.125 points and for items 4, 6, and 8, shifts the score by 4.166 points, assuming no missing responses. Thus, an intra-individual change of at least 2 categories (3 if the only change is in items 5 and 7) would exceed the 5-point threshold that has the been shown to be clinically significant. When analyzing groups of participants, the mean difference across the entire population includes a distribution of participants with different magnitudes of individual change and a 3- to 5-point difference between groups reflects a significantly larger proportion of participants in one group, as compared with the other, who have clinically important improvements in their TSS.

Thus, based on the totality of evidence evaluating the clinical significance of the KCCQ,[5,6,9,14] a between group difference of 3 points is considered clinically meaningful and the current study is powered to detect a mean change from baseline of 3 points in the KCCQ-TSS between canagliflozin 100 mg and placebo groups at Week 12. Assuming a significance level of 5% and a

standard deviation (SD) of 18 points for mean change in KCCQ-TSS, a total of 1,900 participants will be randomized in this study providing approximately 95% power. This sample size is adequate to account for a potential 5% to 10% dropout rate - that is, with a dropout as high as 10% [n=1,710], the statistical power remains >90%.

After approximately 10 months, the planned enrollment in this study is not expected to reach the target of 1900 participants by the scheduled end of study enrollment window in the first quarter of 2021. Instead, the final anticipated enrollment may only be in the range of about 400-480 participants. Consequently, a reduction in total sample size of this magnitude will have impact on the statistical power for the assumed effect size as well as for various other plausible effect sizes, shown by the figures below. At 50% statistical power, an effect size of 3.5 points in the KCCQ change score (assuming standard deviation is 18) can be detected with a minimum sample size of 400. For power ≥80%, the required sample size has to be at least 445 if the expected mean difference in KCCQ-TSS change score over 12 weeks is as high as 4.8, assuming the pooled standard deviation is 18. An effect size of this magnitude for the KCCQ-TSS change score was reported in the recently concluded DEFINE-HF study[17]. Robust improvements over baseline were also reported for additional KCCQ domains in this study as well as in the SOLOIST study[17,18].



1.5. Randomization and Blinding

Central randomization will be implemented in this study. Participants will be randomly assigned to 1 of 2 study intervention groups based on a computer-generated randomization schedule prepared before the study by or under the supervision of the sponsor before initiation of the study.

Randomization will be balanced by using randomly permuted blocks and will be stratified by EF (HFrEF or HFpEF) at study entry as recorded in the participant's electronic health record (EHR).

The interactive web response system (IxR) will generate a randomization code, participant randomization number, and kit number once consent is completed and the participant is confirmed eligible by the virtual principal investigator. The kit number will dictate the study intervention assignment and the matching study intervention bottle(s) to be shipped to the participant directly from the drug distribution vendor.

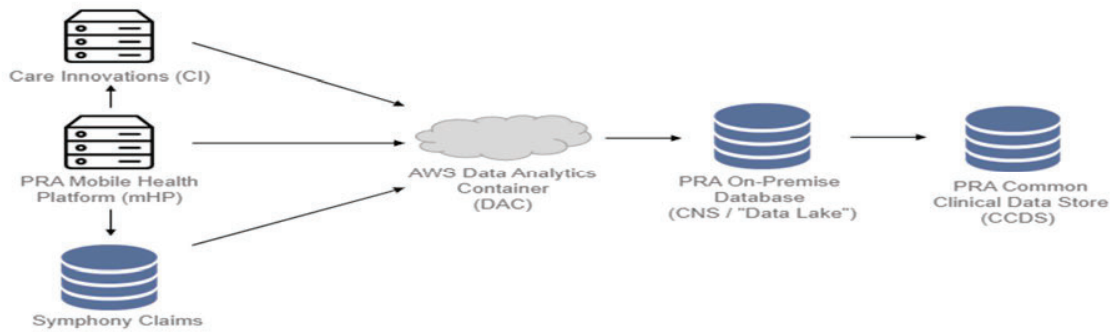
The IxR will provide the kit number of the study intervention bottle(s) to be dispensed for each randomly assigned participant of the double-blind treatment period. Based on these randomization codes, study intervention will be packaged and labeled in a manner that maintains the double-blinded nature of the study.

The drug depot will not be provided with randomization codes. The codes will be maintained within the IxR, which has the functionality to allow the virtual principal investigator or designee to break the blind for an individual participant.

Under normal circumstances, the blind should not be broken until participants have completed the 12-week, double-blind treatment period. The virtual principal investigator or designee may in an emergency determine the identity of the study intervention by contacting the IxR. In the event the blind is broken, the sponsor must be informed as soon as possible. The date and reason for the unblinding must be documented in the study database. All randomization codes will be released after completion of the study. Participants who have had their intervention assignment unblinded should continue to perform their scheduled evaluations.

1.6. Study Data Flow

Data flow in this study is facilitated by an automated integrated process that extracts clinical data from 3 sources: PRA mobile health platform [PRA mHP], Care Innovations, and Symphony. These data are loaded into PRA's Common Clinical Data Store (CCDS) where the data is accessible by JReview for data validation and analysis, as well as for extraction into RAW and SDTM datasets. This automated Extract-Transform-Load (ETL) process will utilize a data container on Amazon Web Services (AWS), referred to as the "Data Analytics Container" or "DAC", which is managed by PRA. A more detailed data flow diagram is provided in Appendix 5 of this SAP.



The PRA mHP platform serves as the main hub for source data capture on the CHIEF-HF study. It captures eConsent, information entered by Health Network Systems (HNS) from a participant's electronic medical records (EMRs), participant status overseen by Care Innovations (CI), randomization through integration with Endpoint IRT, drug shipment through integration with Marken, primary and secondary endpoint data from ePROs (KCCQ, PGI-S, PGIC), dosing compliance, as well as activity measured by the Fitbit watch. PRA mHP data will go into the DAC via a direct database connection. Medical and pharmacy claims data is contained within Symphony

Health Solutions and exists as completely de-identified data. This includes data used in the study to review and analyze adverse events, healthcare resource utilization (HRU) and health economics (including hospitalization/readmission due to Heart Failure [HF] or other reason, Emergency Department (ED) visits, the duration of stay at each hospitalization/readmission), as well as review of prohibited medications and new diagnoses that may require discontinuation.

Symphony claims data will be pulled into the DAC for participants randomized into the study via PRA mHP eConsent and will be mapped to SDTM as detailed in the SDTM specifications relevant to the study protocol and SAP. The collection timeframe for claims associated with AEs of special interest will begin on the Informed Consent Date through the End of Treatment + 30 days. The collection timeframe for Pharmacy claims of prohibited concomitant medications (SGLT2i) will begin on the ICF date through End of Treatment. The collection timeframe for claims associated with Health Resource Utilization will begin on the ICF Date through the 9-month observation period.

2. GENERAL ANALYSIS DEFINITIONS

As this is a virtual study, there are no planned in-person clinic visits for measuring key endpoints on the study participants. Data collection via smartphone, Fitbit device, and claims will continue throughout the 9-month duration of the study. While the focus of the first 12-week, double-blind treatment duration is to establish the primary and key secondary objectives related to primary and secondary effectiveness endpoints, the latter 6 months (i.e., Month 3 until Month 9) of the study (when no study intervention is provided) are aimed to explore multiple objectives by collecting additional follow-up data on PROs, daily activity, and HRUs and generate hypotheses of interest for future studies.

Unless stated otherwise, following the intent-to-treat (ITT) principle, analyses of the primary and secondary endpoints will be based on the full analysis set (FAS) and include data collected in the 12-week, double-blind treatment period. Where meaningful, these analyses may also be repeated using the per protocol analysis set. Analysis based on full analysis set (FAS) will also be extended to the full study data period (up to the end of the 9-month long study). In a similar way, exploratory endpoints will be analyzed according to the randomized intervention groups by using data from both the 12-week, double-blind treatment period and the 9-month study duration, respectively.

2.1. Time Windows

In this virtual study, the randomization date will refer to the day of randomization of an eligible participant. Once randomized, study intervention will be shipped and delivered to the randomized participant directly, and therefore, the first study intervention may be started by a participant a few days after the actual randomization date. This will be considered the date of starting study intervention, or, a reference day (Study Day 1). Scheduled study intervention will end after 12 weeks. After Week 12, participants will be followed up to Month 9 without receiving any further study intervention. Total study days will be numbered relative to the Study Day 1 for each participant up to the end of the 9-month long study. All measurements will be arranged in a chronological order relative to Study Day 1. Participants will be instructed and notified by the study app on their smartphone to confirm taking the study intervention on schedule and completing

all necessary patient reported outcome (PRO) measurements. Study app is programmed to allow PRO measurements to be recorded only within a pre-specified window of days around each scheduled time point.

The primary, secondary and exploratory variables (ie, KCCQ, and its individual domains, daily step count, daily floors climbed, PGIC and PGI-S, participant satisfaction survey, etc.) will be measured at specific time points as identified in the protocol, using either a smartphone-based mobile app or, the Fitbit device, as is relevant, registered by the study participant. Measurements taken closest to the planned (target) time points will be considered for the endpoint analyses and will be appropriately assigned to a time window to allow reasonable variations around the target day. Endpoint for analyses and summaries is the last recorded postbaseline data.

Daily activity measurements via the Fitbit device will be collected continuously from the time study started (ie, when both the smartphone and the Fitbit are linked by a participant). Fitbit daily activity data will be aggregated into averages over 2-week intervals starting with a baseline period (ie, Baseline = average of Study Day 1 to Study Day 14) and similarly over the entire post-baseline study duration for each participant. This is to allow capturing a robust length of physical activity for each participant using the Fitbit device and minimizing variability associated with measurements in shorter time spans. The end-of-treatment (EOT) average value of the step count will be the average of the last 14 days prior to the last dose of study intervention, regardless of when the last dose is taken.

Claims data will be extracted as access to participant-level information becomes available. This will include clinical outcomes (including hospitalizations for HF and/or other reasons) coded using International Statistical Classification of Diseases and Related Health Problems (ICD)-9 and ICD-10 as appropriate.

Time windows will not be applied to summarize the safety data collected through claims. Self-reported safety data will be summarized through Week 12 (Day 87).

Table 1 presents the time windows and the target days for each time of measurement defined in the protocol.

Table 1: Measurement Windows

Parameter	Time Interval (label on output)	Time Interval (Day)*	Target Time Point (Day)
KCCQ (and its individual domains); PGIC and PGI-S; participant satisfaction survey	Screening/Baseline	-21 to <1	1
	Week 2	11 to 17	14
	Week 4	25 to 31	28
	Week 6	39 to 45	42
	Week 12	81 to 87	84
	Month 6	161 to 175	168
	Month 9	245 to 259	252
Daily Step Count / Daily Floors Climbed **	Baseline to Week 2	1 to 14	
	Week 3 & Week 4	15 to 28	
	Week 5 & Week 6	29 to 42	
	Week 7 & Week 8	43 to 56	

Parameter	Time Interval (label on output)	Time Interval (Day)*	Target Time Point (Day)
	Week 9 & Week 10	57 to 70	
	Week 11 & Week 12	71 to 84	
	…	…	
	Week 35 & Week 36	238 to 252	

*Relative to Study Day 1
** Measurements over a 2-week interval with ≥ 100 daily steps to allow capture of physical activity

2.2. Pooling Algorithm for Analysis Centers

This is a virtual study with multiple IHNs competitively enrolling HF patients to meet the overall study enrollment target. Therefore, from the analytical perspective, no separate sites exist for pooling to be applicable. It will be considered a single center study.

2.3. Analysis Sets

An overview of the different analysis sets and their usage is presented in Appendix 4: Usage of the Analysis Sets.

2.3.1. All Randomized Analysis Set

The all randomized analysis set includes all participants who are randomly assigned to a study treatment. In order to adhere to the ITT principle, analyses based on all randomized analysis set will use the following rules:

- participants will be evaluated according to the study intervention they have been assigned to (which may be different from the study intervention they have received)
- all available data will be included (e.g., including assessment after study intervention discontinuation).

2.3.2. Full Analysis Set (FAS)

This includes all randomized participants who have received at least one dose of study intervention/medication and have at least one post-baseline KCCQ measurement.

- Randomized participants in this analysis set will be considered regardless of when they received/started their first dose of study intervention/medication.
- all available data will be included (e.g., including assessment after study intervention discontinuation).

Analyses of the primary, secondary, and exploratory endpoints in this study will be based on the full analysis set (FAS).

2.3.3. Per Protocol Analysis Set

The per protocol analysis set (PP) will consist of all participants who received study intervention and who complied with the protocol to allow assessment of the treatment effects. Specifically, this includes randomized participants who comply with the study intervention, have at least the key (ePRO) measurement of KCCQ completed, and do not experience major protocol deviations that may have an impact on estimating the treatment effect. The effectiveness analysis for the primary and secondary endpoints will be repeated for the PP analysis set.

2.3.4. Safety Analysis Set

The safety analysis set includes all randomized participants who received at least 1 dose of study intervention. Summary of available safety data from all sources (claims and patient reported) will be based on this analysis set.

2.4. Definition of Subgroups

Analysis for the primary effectiveness endpoint will be performed for the subgroups identified below.

Subgroup	**Definition**
Region**	US - East, West, North and South
Age Group	• 18-25 • 26-50 • 51-64 • >65
Gender	• Male • Female
Race	• White • Non-White
Diabetes type	• T2DM • No-T2DM
HF – Ejection Fraction type	• pEF (preserved ejection fraction) • rEF (reduced ejection fraction)

**Information on region will be based on details available through IHNs.

Additional subgroups may be identified (including those based on combinations of T2DM and HF ejection fraction types) before finalizing the database.

2.5. Study Day and Relative Day

Study Day 1 refers to the start of the first study intervention administration. All assessments recorded via the smartphone/mobile app and/or Fitbit device will be assigned a day relative to this date.

Relative day for a specific time-point of measurement is defined as:

- measurement date - (date of Study Day 1) +1, if measurement date is ≥ date of Study Day 1
- measurement date - date of Study Day 1, if measurement date < date of Study Day 1.

2.6. Baseline, End of Treatment, and End of Study

Baseline is defined as the last observation made prior to the start of the first study intervention administration. End of Treatment (EOT) observation will refer to the last observation made prior to and including the day of the last dose of study intervention, regardless of when the last dose is taken. For the PRO measurements, EOT observation will include the last observed value prior to the last dose of study intervention, and/or, within a measurement window allowed by the app that may overlap the last dose date (i.e., may extend the last dose date by a few days). For the daily

activity data, it will represent average from the last 14 days prior to the last dose taken, regardless of when the dose is taken.

The EOS value is defined as the observation that is taken on the date of very last available postbaseline data in the study period. Unscheduled measurement results are included in this definition and will be considered as the last timepoint value if the unscheduled measurement result is the last postbaseline result available within the analysis period.

2.7. Imputation Rules for Missing AE Date/Time of Onset/Resolution

Claims data are generally based on service dates and not on start and end dates for adverse events. Consequently, no imputation rule for missing AE dates can be applied.

3. INTERIM ANALYSIS AND DATA MONITORING COMMITTEE REVIEW

No interim analysis is planned in this study. However, data from the 12-week double-blind treatment period will be summarized/analyzed and reported out first as soon as the last randomized participant has completed the treatment period. Data from the full 9-month study period will be reported after the follow-up is completed on all study participants.

4. SUBJECT INFORMATION

The number of participants in each analysis set will be summarized and listed by intervention group, and overall. In addition, the distribution of participants by region within the US and recruiting networks will be presented, unless otherwise noted.

4.1. Demographics and Baseline Characteristics

Table 2 presents a list of the demographic variables that will be summarized by each intervention group, and both combined for the full analysis set (FAS). Demographics will also be summarized by various IHNs (or regions) using the FAS as well as the all randomized analysis set.

Table 2: Demographic Variables

Continuous Variables:	**Summary Type**
Age (years)	Descriptive statistics (N, mean, standard deviation [SD], median and range [minimum and maximum], and IQ range).
Categorical Variables	
Age (18-25 years, 26-50 years, 51-64 years, and ≥65 years)	Frequency distribution with the number and percentage of participants in each category.
Sex (male, female)	
Race[a] (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian, or, other Pacific Islander, White, Other)	
Diabetes status at baseline (T2DM or no T2DM)	
HF ejection fraction type (HFrEF or HFpEF)	

[a] If multiple race categories are indicated the Race is recorded as 'Multiple'.

4.2. Disposition Information

Total number of screened participants and number of screen failures will be provided. The total number of participants randomized into the study will be provided. Randomized participants who complete the 12-week, double-blind treatment period and who do not discontinue study intervention prematurely are considered study intervention completers.

The number of participants in the following disposition categories will be provided for the 12-week, double-blind treatment period by intervention group and overall:

- participants randomized
- Participants treated with study agent.
- participants receiving study intervention
- participants completing the double-blind treatment period
- participants who discontinued study intervention
- participants who terminated study prematurely
- reasons for discontinuation of study intervention and termination of study
- participants who were unblinded during the double-blind treatment period.

The distribution of the time to study termination for the 12-week, double-blind treatment period will be displayed with Kaplan-Meier curves. Participants who terminate study participation prematurely at any time will be considered an 'Event' and their date of study termination will be used in the time to study termination calculation. In this analysis, participants who complete the 12-week, double-blind treatment period will be considered 'censored' and the date of double-blind treatment period completion will serve as the time of censoring.

4.3. Treatment Compliance

Participants will be instructed and notified by the study app weekly to confirm the number of study intervention capsules that they took that week. Study intervention compliance data will be monitored bi-weekly for ensuring compliance is >80%.

Compliance with the study intervention is defined in this study in terms of the total number of doses received during the total treatment period of 12 weeks and not in terms of maintained drug concentration while participating in the study. Participants who take study intervention as instructed >80% of the total study intervention duration will thus be considered compliant.

For the double-blind study intervention period, compliance (%) = 100 x number of days taking study intervention (pills)/total study intervention duration (including days off study intervention).

The number and percentage of participants who have >80% study intervention compliance through Week 12 (end of double-blind treatment period) will be provided by randomized group.

4.4. Extent of Exposure

The number and percentage of participants who receive study intervention will be summarized by randomized group.

Descriptive statistics for study intervention duration (N, mean, SD, median, and range [minimum, maximum]) will be presented by randomized group based on the safety analysis set. Person-years of exposure to study intervention are calculated [as days of intervention/365.25] and will be presented by randomized group.

Duration of study intervention will be summarized in the following duration categories: [<1 week, 1-<2 weeks, 2-<3 weeks, …, 11-≤12 weeks] by randomized group and presented graphically in a histogram.

Cumulative study intervention exposure [≥1 week, ≥2 weeks, …, ≥12 weeks] will also be summarized.

Study intervention duration is defined as (date of last dose of study intervention – date of first dose of study intervention) +1.

4.5. Protocol Deviations

In general, a list of major protocol deviations will be produced for the final clinical study report (CSR) and these major deviations from protocol may have the potential to impact participants' rights, safety, or well-being, or the integrity and/or result of the clinical study. Participants with major protocol deviations will be identified prior to database lock and summarized

4.6. Prior and Concomitant Medications

In this pragmatic virtual study, other than history on prior medications up to 36 months, no additional information including specific start/end dates of using prior medications will be available. However, dates of prescription-fills for specific concomitant/prohibited medications collected through the claims/pharmacy database for the duration of the study may be available. The concomitant medication use during study may be categorized into broad classes of **glucose-lowering agents** (such as insulin, metformin, sulfonylurea, DPP-4 inhibitor, GLP-1 receptor agonist), **renal and cardiovascular (CV) protective agents** (such as RAAS inhibitor [including sac/val], statin, antithrombotics, beta-blockers, diuretics, IV in-patient diuretics, vasodilators, inotropes, anti-infectives) and **UGT inducers** (such as carbamazepine, phenytoin, phenobarbital, tiprnavir, rifampicin, testosterone, propionate, nelfinavir, ritonavir, efavirenz, lamotrigine) respectively.

5. EFFICACY

5.1. Analysis Specifications

Descriptive statistics for continuous variables include count (n), mean, SD, standard error (SE), median, minimum, and maximum. Descriptive statistics for categorical variables include participant counts and percentages.

5.1.1. Level of Significance

A 5% level of significance will be used for all analyses. All statistical tests and 95% confidence intervals presented in the analyses of endpoints will be 2-sided.

5.1.2. Data Handling Rules

This study will obtain mainly 3 types of data for statistical analysis: the smartphone/app-based PRO data, the activity data collected from the Fitbit device and the claims data on adverse events (including clinical outcomes, prescription fills, and related HRU). In addition, participant's self-reported AE data collected via a Call Center will be directly sent over to Janssen Safety separately and is described below as well. All data will be received and processed by PRA (3rd party vendor, see Appendix 5: CHIEF-HF Study Data Flow Diagram) into analysis files based on specifications provided by the sponsor/study SAP unless specified otherwise.

Smartphone Data - PRO Measures

PRO data (ie, KCCQ, PGIC and PGI-S, participant satisfaction survey) will be collected via the smartphone apps available with participants. Based on the administration of the KCCQ at screening/baseline through an app, an overall summary score will be produced to decide eligibility of individual participant into the study. The individual domains and summary scores of KCCQ will be calculated using the scoring algorithm provided in the KCCQ manual at specified time points during the study. Missingness in these PRO measures may result from participants not completing the questionnaires on the smartphone app at the scheduled times during the study period for various reasons including non-compliance with the protocol, discontinuation due to AE, loss to follow up, etc.

Missing and incomplete data in KCCQ PRO measure may be of 2 types: missing at the form-level and at the item-level. Form-level missing data refers to a participant missing an entire PRO assessment for a scheduled time point. If not otherwise specified, analyses will be conducted on all available data and form-level missing data will not be imputed. For the KCCQ, derivation of individual domain scores in the presence of missing item-level responses will follow algorithms specified in the KCCQ manual. PRA will calculate baseline KCCQ overall summary score based on the KCCQ app and include in the analysis database. For calculation of all other KCCQ domain scores (including the KCCQ-TSS) at baseline and at all post-randomization time points, scoring instructions available per the KCCQ manual with Janssen will be followed (see Appendix 8). For the PGIC and PGI-S measures, a single imputation (ie, last observed scores carried forward) for missing post-baseline scores will be used as a conservative rule. Vertical file formats will be prepared for all PRO data, reading one record per patient for each measurement time (ie, baseline, Weeks 2, 4, 6, and 12, and Months 6 and 9, respectively).

Fitbit Data - Clinical (Activity) Measures

Data from the Fitbit device pertaining to the daily total step counts and daily total floors climbed will be aggregated over 2-week intervals for the entire study duration (see Table 1). Due to the general sedentary nature of HF patients, Fitbit recorded <100 steps taken in a day will not be

considered in the calculation of averages for daily step counts . If missing >7 days (50%) in a 2-week interval, the corresponding 2-week interval average will be set to missing.

In case of premature discontinuation of the study intervention, the average of the last non-missing 14-day window prior to the last dose of study intervention will always be used as an EOT value. If the EOT value corresponding to the period (EOT-14, EOT-1) is missing (not valid), the EOT period will be rolled back by one day, ie, (EOT-15, EOT-2), (EOT-16, EOT-3), until a non-missing (valid) value is available. This corresponds to a last observation carried forward (LOCF) up to Week 12/EOT approach. If no post-baseline daily activity data have been collected for at least 7 days over a 14-day time window, no EOT value will be computed, and the post-baseline value will be set to missing.

Baseline for daily step count and daily floors climbed variables will be defined by averaging the daily values over a 14-day interval starting with the first dose date. These data will be organized into a vertical file format to read one record per participant for averages over each 2-week interval during the entire study (including the EOT/EOS values, as applicable).

Claims Data – AEs, HRUs, Clinical Outcomes

Claims data pertaining to AEs of interest (as defined in the protocol), number of prescription fills, HRUs, and clinical outcomes reported with appropriate ICD codes applied at the individual participant level will be organized with one record for each incident (ie, hospitalization, etc.) date. Adverse events of interest include all severe hypoglycemic events, acute kidney injury, fractures, diabetic ketoacidosis, and lower limb amputation.

In addition, the following adverse events will be examined: hypotension; Fournier's gangrene; genital mycotic infections; urinary tract infections (including urosepsis and pyelonephritis); increased urination; and hypersensitivity reactions including angioedema and anaphylaxis.

Self-reported AE data – Data from participants self-reporting adverse events (AEs) to a Call Center (set up by Care Innovations) will be manually submitted directly to the Sponsor (i.e. Janssen Safety [GMS]) for appropriate evaluation and analysis. Janssen Safety (GMS) will extract participant-reported AEs and transfer data to PRA for upload into the CCDS. Self-reported events will be captured from the date of the first dose of study intervention through the end of treatment (Month 3) plus 30 days. A listing of these self-reported AE may be produced based on available data.

5.2. Primary Efficacy Endpoint

5.2.1. Definition

The primary effectiveness endpoint is the change in the KCCQ-TSS from baseline to Week 12.

5.2.2. Estimand

The primary estimand, the main clinical quantity of interest to be estimated in the study, is defined by the following 5 attributes (ICH E9 [R1]).[4]

- Population: FAS: all randomized participants with HF (who either have HFrEF or HFpEF, regardless of T2DM status) with at least one dose of study intervention and at least 1 post-baseline KCCQ measurement
- Treatment: canagliflozin 100 mg or placebo
- Variable: change from baseline to Week 12 post first dose in the KCCQ-TSS
- Intercurrent events: treatment discontinuations (either due to AE or other reasons) or treatment switching; use treatment policy strategy to address intercurrent events (i.e., use all available measurements collected during Week 12 , regardless of the occurrence of the intercurrent event); events of "loss to follow up" or "study withdrawal" or "death" are not considered intercurrent events but rather missing data.
- Population-level summary: difference in mean change from baseline to Week 12 in the KCCQ-TSS between randomized groups.

In other words, the main quantity to estimate for the primary effectiveness endpoint is the mean difference between the 2 treatment groups (canagliflozin 100 mg and placebo) of the change from baseline to Week 12 in the KCCQ-TSS.

5.2.3. Analysis Methods

The primary effectiveness endpoint analysis will be based on the full analysis set (FAS).

Considering the repeated measurements of the KCCQ-TSS in the double-blind treatment period (ie, at baseline, Weeks 2, 4, 6, and 12 respectively), the primary effectiveness endpoint (change in the KCCQ-TSS from baseline to Week 12) will be analyzed by a MMRM method simultaneously adjusting for treatment, stratification factor (HFrEF or HFpEF), time, time by treatment, and baseline KCCQ-TSS value as covariates. Time is repeated within participants and an unstructured covariance structure will be assumed across treatment groups to model the within-participant errors. Fixed effects tests will be computed using the Kenward-Roger method to determine the degrees of freedom of the denominator. In case of convergence issues, alternative variance-covariance structures will be evaluated in the following order, with the first structure in the following list that converges being used in the analysis: heterogeneous Toeplitz, standard Toeplitz, and AR (1) with separate participant random effect. The primary comparison will be based on the difference in the LS mean changes from baseline to Week 12 between treatments. The treatment difference in the LS means, corresponding 95% confidence intervals (CI) and associated p-value will be calculated. The LS means ±SE of change from baseline values estimated by MMRM will be plotted by each timepoint for each intervention group.

```
PROC MIXED DATA=analysis_dataset;
CLASS treatment time ptid;
MODEL change = treatment time treatment*time baseline strat_var/DDFM=KR;
REPEATED time / SUBJECT=ptid TYPE=UN;
LSMEANS treatment;
LSMEANS treatment*time/pdiff cl;
RUN;
```

An MMRM model assumes the KCCQ-TSS data missing at random (MAR). A sensitivity analysis may be performed involving multiple imputation (MI) modeling approach for missing KCCQ-TSS values when the MAR assumption is considered not valid. Implementing the MI process requires 3 steps: 1) each missing value is imputed multiple times, based on statistical modeling, resulting in repetitions of the original dataset, along with the newly imputed values (implemented in the Statistical Analysis System [SAS] using PROC MI); 2) each dataset is analyzed using an MMRM model (implemented in SAS using PROC MIXED). The LS means and SEs from each analysis will be output to a SAS dataset; 3) results from the analyses performed on multiple datasets in step #2 are combined to provide valid statistical inference (implemented in SAS using PROC MIANALYZE).[1]

A complete SAS code for MMRM model and implementation of the MI approach described above will be added to the final SAP appendix or to the final study report. Mortality in this short-duration trial is expected to be low and to be comparable between groups and will be ignored in the primary analyses of mean differences in changes in KCCQ total symptom scores, but can be included in the responder analyses (described below) by assigning them to the lowest response (greatest deterioration) category. Should there be a >5% mortality rate in this study, and if it is unequally distributed between treatment arms, then joint modeling of survival and health status (e, g.,KCCQ) will be performed. A win-ratio approach[19,20] may also be considered by analyzing the KCCQ-TSS score as a composite, rank-based outcome, incorporating participant vital status at 12 weeks along with a change in score from baseline to 12 weeks in surviving participants. This essentially is the rank analysis of covariance method, with a corresponding win ratio used to estimate the magnitude of treatment effect.

An exploratory responder analysis will be performed when comparing proportions of participants with a 5- (or 10-) point improvement in the KCCQ-TSS from baseline to Week 12 between the groups. This will be based on a 2-sample difference in proportions test along with a 95% confidence interval for the difference in 2 proportions.

The mean differences between the groups will be augmented with both responder analyses and empirical cumulative distribution and probability density function curves. Specifically, the empirical cumulative distribution function (CDF) of the change from baseline in KCCQ-TSS will be generated. CDF shows the proportion of the population scoring less than or equal to each possible change score. CDFs are useful as they graphically characterize the treatment effect or differences between groups. It also enables a ready assessment of different magnitudes of clinical change (e.g., 5, 10, and ≥20-point changes, corresponding with small to moderate, moderate to large, and large to very large clinical changes). CDFs and histograms will be shown for changes from baseline to Week 12 in the KCCQ TSS. For all CDF plots, the x-axis is the score change from baseline to final time point (i.e., Week 12) for the KCCQ-TSS. The y-axis is the cumulative proportion of the participants in each of the corresponding categories that reach the score change on the x-axis.

According to the Food and Drug Administration's (FDA's) guidance on PRO measures, a responder is defined by the empirically determined score change in a measure, experienced by an individual patient over a predetermined time period that has been demonstrated in the target

population to have a significant treatment benefit. This is also often referred to as responsiveness or ability of the PRO measure to capture the true underlying change in the patient health status over time. Assessing the responsiveness of the PRO measure using an anchor-based approach involves comparing the changes in the PRO scores to other clinically meaningful markers or anchors that represent a valid measure of clinical change.

For the anchor-based analysis of the KCCQ-TSS endpoint, participants will be characterized based on the PGIC scores at Week 12. The mean change in KCCQ-TSS endpoint observed in the smallest improvement category of the PGIC will be examined as a key anchor-based indicator of a responder. The definition for a small improvement in PGIC is given as those who improve by one category (i.e., from "a little improved" to "somewhat improved"). Specifically, the retrospective PGIC responses of "3=a little improved" will be selected to represent time points where participants register meaningful changes in their HF symptoms. Each time a participant responds with a PGIC score of "3=a little improved", a KCCQ-TSS change score from the preceding visit is computed. For the double-blind treatment period, the PGIC will be collected at 4 time points (i.e., Weeks 2, 4, 6, and 12 respectively) which can yield between zero and 4 KCCQ-TSS change scores by each participant. For each participant with at least one PGIC response of "3=a little improved", a mean of the individual participant's KCCQ-TSS change scores is computed. A summary of KCCQ-TSS change scores across PGIC categories[16] will then be presented.

The KCCQ-TSS change scores may also be standardized into effect size (ES) statistics by dividing them by their SD at baseline. The following cutoff values will be used to interpret effect size: small ES = 0.2, moderate ES =0.5, and large ES = 0.8. Together with the CDF, results from the anchor-based methods will be examined and a single responder definition threshold will be determined, if the results from the anchor-based assessments in this study differ substantially from prior studies examining the clinical meaningfulness of changes in the KCCQ.

In addition, a distribution-based change (i.e., minimum detectable [MID] change) may be explored for which the KCCQ-TSS can be derived using the baseline SD. Generally, a value of half the baseline SD has been frequently found to correspond to the MID obtained via the anchor-based method. A value of 0.2*SD may also be examined.

Subgroup analyses for the change from baseline to Week 12 for the primary effectiveness endpoint will be conducted for subgroups (listed in Section 2.4) based on all randomized analysis set. These analyses will assess consistency of the overall treatment effect across subgroup variables.

The main MMRM analyses will be repeated with each subgroup tested separately. Intervention group differences of LS means as well as corresponding 95% CIs within each subgroup level will be presented.

The intervention-by-subgroup interaction p-value will be estimated using a separate MMRM model including intervention, baseline KCCQ value, stratification factor (HFrEF/HFpEF), subgroup, and intervention-by-subgroup interaction term(s). Interactions with p-value <0.1 will be investigated further to determine the nature of interaction (quantitative or qualitative) and the

association with other subgroups. For qualitative interactions, Gail and Simon (1985) approach[15] may be considered.

Results will be presented in a summary table and in a forest plot. The forest plot will display the "overall" treatment effect, based on the main MMRM analysis, as a reference line. LS means of the intervention group differences, the corresponding 95% CIs, and the numbers of participants in the subgroup levels will be presented in the forest plot.

5.3. Secondary Endpoints

5.3.1. Definition

The first key secondary effectiveness endpoint is change in the daily step count from baseline to Week 12 averaged over two-week intervals. Even though evidence of a meaningful change in daily step count in chronic HF patients via Fitbit device is not readily available, there are several studies that have conducted six-minute walk tests (6MWT) among HF patients in general and examined a minimum clinically important difference in daily step counts under different conditions. A change of 30 meters[10] (approximately 1000 steps) may be considered in the present study to be of clinical interest to benchmark the expected mean change in daily step count in this study over the double-blind treatment period. This expected mean change in daily step count will also be examined in relation to variables such as PRO measures (as anchors) and clinical outcomes to better characterize treatment effect on daily activity.

As HF patients are generally not consistent in their daily walk patterns, there will be variations in levels of their daily activity (whether it is a weekday or a weekend) as well as variability in the total wear time of the Fitbit device. As a rule, we will consider Fitbit recorded ≥100 steps[13] taken in a day in the calculations of averages. This will ensure capturing any meaningful level of physical activity while avoiding even smaller variations in daily steps taken by a participant wearing a Fitbit device. For each participant, baseline daily step count will be calculated by averaging daily step counts from Study Day 1 through Study Day 14. All subsequent post-baseline daily step count data will also be averaged in 2-week intervals. The change from baseline to Week 12 in daily step count will be based on subtracting the average of daily step count for Week 11 and Week 12 from the baseline average for each participant. Additionally, daily step count data will be averaged in 3-, 5- and 7-day (1-week) intervals at baseline and throughout the treatment period respectively and corresponding changes from baseline will be calculated.

The variable daily floors climbed will also be aggregated in a similar fashion.

5.3.2. Estimand

The first key secondary estimand to be estimated in the study, is defined by the following -

- Population: FAS (i.e., all randomized participants with HF who either have HFrEF or HFpEF, regardless of T2DM status with at least one dose of study intervention and at least 1 post-baseline KCCQ measurement)
- Treatment: canagliflozin 100 mg or placebo

- Variable: change from baseline to Week 12 in the daily step count averaged over 2-week intervals
- Intercurrent event: treatment discontinuations (either due to AE or other reasons) or treatment switching or temporary disruption in daily activity; use treatment policy strategy to address intercurrent events (i.e., use all available measurements collected during study, regardless of the occurrence of the intercurrent event); events of "loss to follow up" or "study withdrawal" or "death" are not considered intercurrent events but rather missing data.
- Population-level summary: difference in mean change from baseline to Week 12 in the daily step count averaged over 2-week intervals (i.e., difference is based on subtracting the average of daily step count for Week 11 and Week 12 from the baseline 14-day average daily step count) between two randomized groups.

5.3.3. Analysis Methods

Daily step counts and daily floors climbed data at baseline will be presented as 14-day averages of daily counts starting with the first dose date and ending with Day 14 during the double-blind treatment period. This method of averaging will continue until end of Month 9 (Day 252) or, until the time of end of study (EOS) observation is available.

The above two key secondary endpoints will be analyzed in two ways –

1) following the same approach as employed for the primary effectiveness endpoint (i.e. using an MMRM model adjusting for baseline covariates such as treatment, stratification factor and baseline average step count).

2) by calculating mean change from baseline to Week 12 in daily step counts using the corresponding 14-day averages at baseline and Week 12 respectively. That is, the mean difference in the 14-day averages in daily step counts or daily floors climbed is calculated for each randomized intervention group and difference in mean changes between the 2 randomized intervention groups will be compared using a 2-sided t-test.

As stated in Section 1.3, the first key secondary endpoint of daily step count will be tested once the primary endpoint is shown to be statistically significant at 5% significance level.

Other secondary endpoints refer to changes in the KCCQ individual domain scores (physical limitation, quality of life), and the KCCQ clinical summary and overall summary scores from baseline to Week 12. The individual KCCQ domain scores as well as the summary scores will be produced at each time point (based on the KCCQ algorithm) with changes from baseline calculated at each specified post-baseline time point. For these endpoints, an MMRM model (like that specified for the analysis of the primary endpoint) will be employed to summarize changes over time from baseline to Week 12. Hypotheses related to these domains will be independently tested at 5% without any control for overall type I error rate.

5.4. Exploratory Variables

5.4.1. Definition

The exploratory endpoints are the following:

- mean PGIC scores over time
- changes in PGI-S scores from baseline to Week 12
- change in HRU and health economics data
- number of outpatient intravenous therapies, ED visits, and hospitalizations
- time to first hospitalization/readmission for heart failure
- time to first hospitalization/readmission for other events
- time to death

5.4.2. Analysis Methods

The exploratory endpoints listed below will be presented using descriptive statistics at each time point measurements are taken. Where specified, 95% confidence intervals (CIs) and p-values will be presented but no multiplicity adjustment will be made. Analyses of mean changes from baseline to Week 12 and month 9 (Day 252) respectively will be conducted, as described below.

- Appendix 6: Patient Global Impression of Change (PGIC) is a validated generic tool for assessment of overall change in the severity of illness following treatment. Participants rate how they feel now compared with how they felt before receiving study intervention on a 7-point scale where 1 is “Very much improved” and 7 is “Very much worse”. A mean ± SD PGIC score for each time point will be calculated by randomized intervention group. Distribution of PGIC score at Week 12 will also be presented. Based on the PGIC symmetric scale, a dichotomous scale of “Yes” or “No” will be derived. A favorable change is score of 1-3 = ’Yes’, which means there is significant improvement with the study intervention. If the response is 4-7 = ’No’, it is considered no significant change. Change in the KCCQ-TSS from baseline to Week 12 by PGIC response at Week 12 will be summarized by treatment group, and, also graphically displayed (i.e., x-axis will display categories of PGIC response at Week 12 and y-axis will display mean change scores from baseline to Week 12 in KCCQ-TSS by treatment group).
- Distribution of participants in PGI-S (see Appendix 7: Patient Global Impression of Severity (PGI-S)) response categories will be summarized by randomized intervention group at baseline and by Week 12. Mean change in PGI-S score between baseline and Week 12 will be summarized by randomized group. Change in the KCCQ-TSS from baseline to Week 12 by PGI-S response at Week 12 will be summarized.
- For daily activity measured by daily step counts, total active days (defined as Fitbit recorded ≥100 steps) and percent of active days to total days spent in the study analysis period will be summarized for each intervention group using mean and SD. Level of daily activity will also be described using categories of total step counts per day – 1) <3000 steps per day (low activity/sedentary), 2) 3000-5999 (moderate activity), and 3) ≥ 6000 steps per day (high activity). Associations between daily step count (categorized into activity levels) and baseline characteristics (eg, age, sex, race, T2DM status, EF type, etc.) will be examined using a 1-way ANOVA for continuous data values and the Fisher exact test for categorical data.
- Correlation between total step counts per day and clinical events (eg, # of outpatient intravenous therapies, # ED visits, # hospitalizations, etc.) will be described for the double-

blind treatment period and total study duration separately. Relationships between change in step counts/floors climbed and incidence of reported clinical outcomes and PRO measures will be examined at the end of the double-blind treatment period and end of study separately.

- Daily floors climbed will be summarized by each intervention group for the double-blind treatment period and over the entire study duration separately.
- Descriptive statistics (ie, N, mean, SD, minimum and maximum) will be presented for total HRU and health economics data for each intervention group separately for the double-blind treatment period and the observational (no intervention) period. HRUs and health economics data typically include information regarding participants using health services for managing any clinical events/illnesses, hospitalizations/ED visits, surgeries, medication issues/prescription changes, and costs incurred.
- Descriptive statistics will be presented for participant's satisfaction with his/her experience with the virtual design at the end of the study.
- Using data from claims, the number of outpatient intravenous therapies, ED visits, and hospitalizations by each intervention group will be summarized. These summaries will include descriptive statistics.
- Using data from the claims and data on deaths, time to first hospitalization/readmission for heart failure, time to first hospitalization/readmission for other events, and time to death will each be analyzed using a Cox's regression model both for the double-blind treatment period and the entire study duration and include terms for study intervention and stratification factor (HFrEF/HFpEF). Hazard ratios, corresponding 95% CIs, and associated p-values will be presented. Distribution of the times to first hospitalization/readmission for HF, time to first hospitalization/readmission for other events, and time to death will be displayed with Kaplan-Meier curves. Participants who do not experience any such events of interest during the study analysis period will be considered 'censored' in these analyses.

No adjustment for multiplicity will be performed for the exploratory analyses listed above.

6. SAFETY

Summaries of safety data will be based on the safety analysis set and will use claims data.

6.1. Adverse Events

All AE data summaries will be based on the safety analysis set except where stated otherwise. A treatment emergent AE is defined as an event that occurs after the first dose of study intervention (Study Day 1) and/or worsens in severity during the double-blind treatment period.

Adverse events self-reported to the Call Center will be assessed by Janssen Global Medical Safety (GMS) for start and end dates, causality and meeting SUSAR reporting requirements. If the specific dates pertaining to the self-reported AEs are unknown, there will be limited follow-up for resolution. For any events still ongoing at the end of the study, there will be no follow up (as the Call Center will be closed at that point) possible. Adverse events will be coded according to MedDRA and will be reported by preferred term for each randomized intervention group. Janssen Data Management will request a listing of all AEs/SAEs reported to the Call Center from the Janssen GMS. This listing will be shared with PRA (external vendor) to convert the information

into the SDTM format and for inclusion in the final SDTM data package for summary by various subgroups (e.g., T2DM and non-T2DM).

Adverse events collected through the claims database will be analyzed separately for the 12-week treatment period + 30 days after last dose and the 9-month entire study period respectively. It is assumed that claims database will have dates of service provided (procedure dates, office visits, hospital admissions, etc.) but not actual start/end dates of events prior to claims being filed. Therefore, no imputation of AE start/end dates will be feasible based on the claims data. Dates of service(s) rendered to participants during the study (i.e., starting from the time informed consent is signed until the end of double-blind treatment plus 30 days) will be taken into account. The claims data will utilize ICD codes to identify the SAEs and AEs of interest which are defined in the protocol and for the final analysis. The AEs of interest include all severe hypoglycemic events, acute kidney injury, fractures, diabetic ketoacidosis, and lower limb amputation. In addition, hypotension, Fournier's gangrene, genital mycotic infections, urinary tract infections (including urosepsis and pyelonephritis), increased urination and hypersensitivity reactions (including angioedema and anaphylaxis) will be considered for coding and final analysis.

For the prohibited medications, analysis period will include data from the date of informed consent until the end of double-blind treatment period. For all other concomitant mediations, analysis period will include data from the date of informed consent until the end of the study. For the HRU data analysis, information collected from date of informed consent until the end of study will be considered in the final analysis.

For both the double-blind treatment period + 30 days after last dose and the 9-month entire study period, descriptive summaries (numbers and percentages) of participants having an SAE/AE of interest will be produced, including by participant's diabetes status (T2DM and non-T2DM) and by randomized intervention group. Discontinuations (and reported reasons) will be summarized by randomized intervention group based on available claims data.

There will be no reconciliation/analysis of SAE/AE data that are self-reported to the Call Center with the data obtained from the claims database.

A listing will be presented for treatment emergent hospitalizations and deaths for each randomized intervention group. AEs occurring between informed consent and the day prior to the first dose of study intervention will also be listed. This listing will be based on participants with informed consent into the study. Depending on the claims/pharmacy data, AEs occurring in participants who used prohibited (commercially available SGLT2is other than the study intervention, canagliflozin) will also be listed.

Based on pharmacy profiles of the participants after-consent in this study, the total number of participants concomitantly receiving UGT inducers during the double-blind treatment period will be summarized combined across both study intervention groups. If the number of such participants is >2% of the total 1,900 targeted for enrollment in this study, the users and non-users of UGT inducers will be summarized by each study intervention group as well and examined further (as

necessary) for any relationship to key study outcomes (eg, PRO endpoints and clinical outcomes reported in claims).

6.2. Clinical Laboratory Tests

Not Applicable.

6.3. Vital Signs and Physical Examination Findings

Not Applicable.

6.4. Electrocardiogram

Not Applicable.

6.5. Other Safety Parameters

Not Applicable.

7. PHARMACOKINETICS/PHARMACODYNAMICS

Not Applicable.

8. BIOMARKERS

Not Applicable.

9. HEALTH ECONOMICS

As treatment of participants with symptomatic HF with JNJ-28431754 (canagliflozin)JNJ-28431754 (canagliflozin) versus placebo may result in lower utilization of hospitalization/readmissions due to HF or other reason, ED visits, the duration of stay at each hospitalization/readmission, and discharge destination will be summarized by each intervention group. Comparison will be made between groups by analyzing differences in hospitalizations/readmissions and ED visits using the Fisher exact test. Descriptive statistics (eg, mean, median, minimum, and maximum) will be provided for the total length of stay in hospitals for each intervention group. Cumulative distribution function of the time to first occurrence of hospitalization/readmission will be estimated by the Kaplan-Meier method and tested using a log-rank test, stratified by the type of HF ejection fraction (ie, HFrEF or HFpEF). In addition, a Cox proportional hazards mode with intervention and stratification factor as covariates will be fitted. A point-estimate along with 2-sided 95% CI for HR of the intervention effect (in terms of relative risk reduction (RRR) = 100 x [1-HR] %) will be provided. The cumulative event rate derived from Kaplan-Meier estimate will also be displayed graphically to evaluate the timing of event occurrence and the consistency of the treatment effect over time.

REFERENCES

1. Berglund P and Heeringa S. Multiple Imputation of Missing Data Using SAS®. 2014. Cary, NC: SAS Institute Inc. CDISC. “Analysis Data.
2. Dreyer RP, Jones PG, Kutty S, Spertus JA. Quantifying clinical change: discrepancies between patients' and providers' perspectives. Qual Life Res. 2016 Sep;25(9): 2213-2220.
3. Green, C., Porter C., Bresnahan D., and Spertus, J. 2000. Development and evaluation of the Kansas City Cardiomyopathy Questionnaire: a new health status measure for heart failure. Journal of American College of Cardiology. 2000 Apr; 35(5):1245-1255.
4. International Council For Harmonisation (ICH) of Technical Requirement for Pharmaceuticals for Human Use (final version, 2019): ICH Harmonised Guideline – Addendum on estimands and sensitivity analysis in clinical trials (to the guideline on Statistical Principles for Clinical Trials) - E9 (R1).
5. Khariton Y. et al. Association Between Sacubitril/Valsartan Initiation and Health Status Outcomes in Heart Failure with Reduced Ejection Fraction: Findings from the CHAMP-HF Registry. Data presented at the European Society of Cardiology Heart Failure (ESC-HF); 2018 May 26-29; Vienna, Austria.
6. Luo N, et al. Relationship between changing patient-reported outcomes and subsequent clinical events in patients with chronic heart failure: insights from HF-ACTION, European Journal of Heart Failure (2019) 21, 63-70.
7. McMurray JJV, Solomon SD, Inzucchi SE, et al; DAPA-HF Trial Committees and Investigators. Dapagliflozin in patients with heart failure and reduced ejection fraction. N Engl J Med. 2019 Sep 19.
8. Nassif ME, Windsor SL, Tang F, et al. Dapagliflozin Effects on Biomarkers, Symptoms, and Functional Status in Patients With Heart Failure With Reduced Ejection Fraction: The DEFINE-HF Trial. Circulation. 2019 Oct 29;140(18):1463-1476.
9. Pokharel Y, et al. Association of serial Kansas City Cardiomyopathy Questionnaire assessments with death and hospitalization in patients with heart failure with preserved and reduced ejection fraction. A secondary analysis of 2 randomized clinical trials. JAMA Cardiology. 2017 Dec 1; 2912):1315-1321.
10. Shoemaker M, Curtis AB, Vangsnes E, Dickinson, MG. Clinically Meaningful Change Estimates for the Six-Minute Walk Test and Daily Activity in Individuals With Chronic Heart Failure. Cardiopulm Phys Ther J, Sep 2013;24(3), 21-29.
11. Spertus, J, Jones P. Development and validation of a short version of the Kansas City Cardiomyopathy Questionnaire. Circulation: Cardiovascular Quality and Outcomes, 2015; 8:469-476.
12. Spertus J, Peterson E, Conard MW, et al. Monitoring clinical changes in patients with heart failure: a comparison of methods. Am Heart J 2005;150(4):707-715.
13. Thorup C, Hansen J, Gronkjaer M, et al. Cardiac Patients’ Walking Activity Determined by a Step Counter in Cardiac Telerehabilitation: Data From the Intervention Arm of A Randomized Controlled Trial. J Med Internet Res 2016 Apr; 18(4); e69.
14. Vetrovsky T, Siranec M, Parenica J, et al. Effect of a 6-month pedometer-based walking intervention on functional capacity in patients with chronic heart failure with reduced (HFrEF) and with preserved (HFpEF) ejection fraction: study protocol for two multicenter randomized controlled trials, J Transl Med, 2017; 15(1):153.
15. Gail, M. H., and Simon, R. (1985). “Tests for Qualitative Interactions between Treatment Effects and Patient Subsets.” *Biometrics* 41:361–372.
16. Mehta, L. et al. (2015). Identifying an important change estimate for the multiple sclerosis walking scale-12 (MSWS-12v1) for interpreting clinical trial results, Multiple Sclerosis Journal – Experimental, Translational, Clinical” 1:1-9

17. Nassif, M. et al. (2019). Dapagliflozin Effects on Biomarkers, Symptoms, and Functional Status in Patients With Heart Failure With Reduced Ejection Fraction - The DEFINE-HF Trial. *Circulation*, Vol. 140, No. 18, 1463-1476

18. Bhatt, D. et al. (2020). Sotagliflozin in Patients with Diabetes and Recent Worsening Heart Failure. *N Engl J Med*, 2020; Nov 16

19. Gasparyan, S. et al. (2019). Adjusted Win Ratio with Stratification: Calculation Methods and Interpretation, *Stat.ME*

20. Wang, D. and Pocock, S. (2016). A win-ratio approach to comparing continuous non-normal outcomes in clinical trials, *Pharmaceutical Statistics*, online 11 March 2016

APPENDICES

Appendix 1: The Kansas City Cardiomyopathy Questionnaire (KCCQ)

The KCCQ (Appendix 1: The Kansas City Cardiomyopathy Questionnaire (KCCQ)) is a well-established PRO for use in patients with HF, with published validity and reliability and was recently endorsed by the Center for Devices and Radiological Health (CDRH, US FDA) as a Certified Outcome Assessment through its Medical Device Development Tools Program and is currently under review at the Center for Drug Evaluation and Research (CDER, US FDA).

The Kansas City Cardiomyopathy Questionnaire (KCCQ) is a 23-item self-administered questionnaire developed to independently measure the patient's perception of their health status, which includes heart failure symptoms, impact on physical and social function, and how their heart failure impacts their quality of life (QOL) within a 2-week recall period. This questionnaire requires, on average, 4 to 6 minutes to complete, with a resulting score of 0 to 100, and higher scores indicating better health.

The intent of the study is to enroll participants with NYHA class II and III. As it would be difficult to verify NYHA class with claims and EHR data, the sponsor will use a KCCQ overall summary score at baseline of >40 and <80, which is highly correlated to class II and III in HFrEF and HFpEF based on the research published by Joseph, et al.[6] In addition to the KCCQ overall summary score, a baseline KCCQ TSS score will also be calculated for each participant.

The 23-item KCCQ quantifies 7 domains of patients' HF-related health status:

- Physical Limitation (6 items)
- Symptom Stability (1 item)
- Symptom Frequency (4 items)
- Symptom Burden (3 items)
- Self-Efficacy (2 items)
- Quality of Life (3 items)
- Social Limitations (4 items)

Item responses are coded sequentially (1, 2, 3, etc.) from worst to best status. Scores are generated for each domain and scaled from 0 to 100, with 0 denoting the worst and 100 the best possible status. In addition, three summary scores are calculated: **Total Symptom score** (average of Symptom Frequency and Symptom Burden), **Clinical Summary score** (average of Physical Limitation and Total Symptoms), and **Overall Summary score** (average of Physical Limitation, Total Symptoms, Quality of Life, and Social Limitation).

Among the 7 domains and 3 summary scores of KCCQ described above, this study will focus on analyzing only two individual domains (physical limitation and quality of life) and three summary scores (clinical summary, overall summary, and total symptom).

Below is a representative example of the scale questions that will be used in this study.

The KC Cardiomyopathy Questionnaire

The following questions refer to your **heart failure** and how it may affect your life. Please read and complete the following questions. There are no right or wrong answers. Please mark the answer that best applies to you.

1. **Heart failure** affects different people in different ways. Some feel shortness of breath while others feel fatigue. Please indicate how much you are limited by **heart failure** (shortness of breath or fatigue) in your ability to do the following activities over the past 2 weeks.

Place an **X** in one box on each line

Activity	Extremely Limited	Quite a bit Limited	Moderately Limited	Slightly Limited	Not at all Limited	Limited for other reasons or did not do the activity
Dressing yourself	❑	❑	❑	❑	❑	❑
Showering/Bathing	❑	❑	❑	❑	❑	❑
Walking 1 block on level ground	❑	❑	❑	❑	❑	❑
Doing yardwork, housework or carrying groceries	❑	❑	❑	❑	❑	❑
Climbing a flight of stairs without stopping	❑	❑	❑	❑	❑	❑
Hurrying or jogging (as if to catch a bus)	❑	❑	❑	❑	❑	❑

2. Compared with 2 weeks ago, have your symptoms of **heart failure** (shortness of breath, fatigue, or ankle swelling) changed?

My symptoms of **heart failure** have become…

Much worse	Slightly worse	Not changed	Slightly better	Much better	I've had no symptoms over the last 2 weeks
❑	❑	❑	❑	❑	❑

Original US English

3. Over the past 2 weeks, how many times did you have **swelling** in your feet, ankles or legs when you woke up in the morning?

Every morning	3 or more times a week, but not every day	1-2 times a week	Less than once a week	Never over the past 2 weeks
❑	❑	❑	❑	❑

4. Over the past 2 weeks, how much has **swelling** in your feet, ankles or legs bothered you?

It has been ...

Extremely bothersome	**Quite a bit** bothersome	**Moderately** bothersome	**Slightly** bothersome	**Not at all** bothersome	I've had **no swelling**
❑	❑	❑	❑	❑	❑

5. Over the past 2 weeks, on average, how many times has **fatigue** limited your ability to do what you want?

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

6. Over the past 2 weeks, how much has your **fatigue** bothered you?

It has been ...

Extremely bothersome	**Quite a bit** bothersome	**Moderately** bothersome	**Slightly** bothersome	**Not at all** bothersome	I've had **no fatigue**
❑	❑	❑	❑	❑	❑

7. Over the past 2 weeks, on average, how many times has **shortness of breath** limited your ability to do what you wanted?

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

Original US English

8. Over the past 2 weeks, how much has your **shortness of breath** bothered you?

It has been ...

Extremely bothersome	**Quite a bit** bothersome	**Moderately** bothersome	**Slightly** bothersome	**Not at all** bothersome	I've had **no shortness of breath**
❑	❑	❑	❑	❑	❑

9. Over the past 2 weeks, on average, how many times have you been forced to sleep sitting up in a chair or with at least 3 pillows to prop you up because of **shortness of breath**?

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

10. **Heart failure** symptoms can worsen for a number of reasons. How sure are you that you know what to do, or whom to call, if your **heart failure** gets worse?

Not at all sure	**Not very** sure	**Somewhat** sure	**Mostly** sure	**Completely** sure
❑	❑	❑	❑	❑

11. How well do you understand what things you are able to do to keep your **heart failure** symptoms from getting worse? (for example, weighing yourself, eating a low salt diet etc.)

Do not understand at all	Do not understand very well	Somewhat understand	Mostly understand	Completely understand
❑	❑	❑	❑	❑

12. Over the past 2 weeks, how much has your **heart failure** limited your enjoyment of life?

It has **extremely** limited my enjoyment of life	It has limited my enjoyment of life **quite a bit**	It has **moderately** limited my enjoyment of life	It has **slightly** limited my enjoyment of life	It has **not limited** my enjoyment of life at all
❑	❑	❑	❑	❑

13. If you had to spend the rest of your life with your **heart failure** the way it is right now, how would you feel about this?

Not at all satisfied	Mostly dissatisfied	Somewhat satisfied	Mostly satisfied	Completely satisfied
❑	❑	❑	❑	❑

Original US English

14. Over the past 2 weeks, how often have you felt discouraged or down in the dumps because of your **heart failure**?

I felt that way **all of the time**	I felt that way **most of the time**	I **occasionally** felt that way	I **rarely** felt that way	I **never** felt that way
❑	❑	❑	❑	❑

15. How much does your **heart failure** affect your lifestyle? Please indicate how your **heart failure** may have limited your participation in the following activities over the past 2 weeks.

Please place an **X** in one box on each line

Activity	**Severely** limited	Limited **quite a bit**	**Moderately** limited	**Slightly** limited	**Did not** limit at all	Does not apply or did not do for other reasons
Hobbies, recreational activities	❑	❑	❑	❑	❑	❑
Working or doing household chores	❑	❑	❑	❑	❑	❑
Visiting family or friends out of your home	❑	❑	❑	❑	❑	❑
Intimate relationships with loved ones	❑	❑	❑	❑	❑	❑

Original US English

Appendix 2: Participant Satisfaction Survey

Appendix 2 provides a representative example of the questions that will be provided to the participant.

For each of the statements below, please indicate your level of agreement:	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Trying something new to possibly treat my heart failure was important to me.					
The virtual study was easy for me to sign up for and get started.					
The Fitbit was easy to set up and use.					
The study app was easy to set up on my smartphone.					
The questionnaires were easy to complete on my smartphone.					
The reminders to complete the questionnaires were helpful.					
The study Fitbit provided helpful information to me.					
It was easy to reach the study team when I had questions or concerns.					
I would recommend participating in a study like this to my friends and family.					

Appendix 3: Schedule of Activities

Period	Screening			Double-Blind Treatment				No Treatment		Notes
	Screening	Baseline								
Day	Within 21 days before Day 1	Within 7 days before Day 1	1 (first dose)	14±3	28±3	42±3	84±3 End of Treatment	85-251-±7	252-±7 End of Study	Day -21 to 0=Screening; Day 1 = first dose, Day 84= end of treatment; Day 252=End of Study
Week				2	4	6	12	13-36		Week 12 is equivalent to Day 84 of treatment
Month			0				3	6	9	Month 3 is equivalent to Day 84 of treatment
Study Procedure										
Screening/ Administrative										
Electronic Informed consent[a]	X									
Review medical history	X									
Inclusion/ exclusion criteria[b]	X		X							All inclusion/exclusion verification should be completed within the 21-day screening period and before randomization
KCCQ baseline overall summary score <80			X							
Study Intervention Administration										
Randomization			X							
Dispense/administer study intervention			←	—	—	—	→			
Participant-reported adherence via app - weekly			←	—	—	—	→			
Assessments										
KCCQ				X	X	X	X	X[c]	X	Kansas City Cardiomyopathy Questionnaire
PGIC				X	X	X	X	X[c]	X	Patient Global Impression of Change
PGI-S			X	X	X	X	X	X[c]	X	Patient Global Impression of Severity
Step count and floors climbed			←	—	—	—	→			
Claims data collection[d]			←	—	—	—	→			

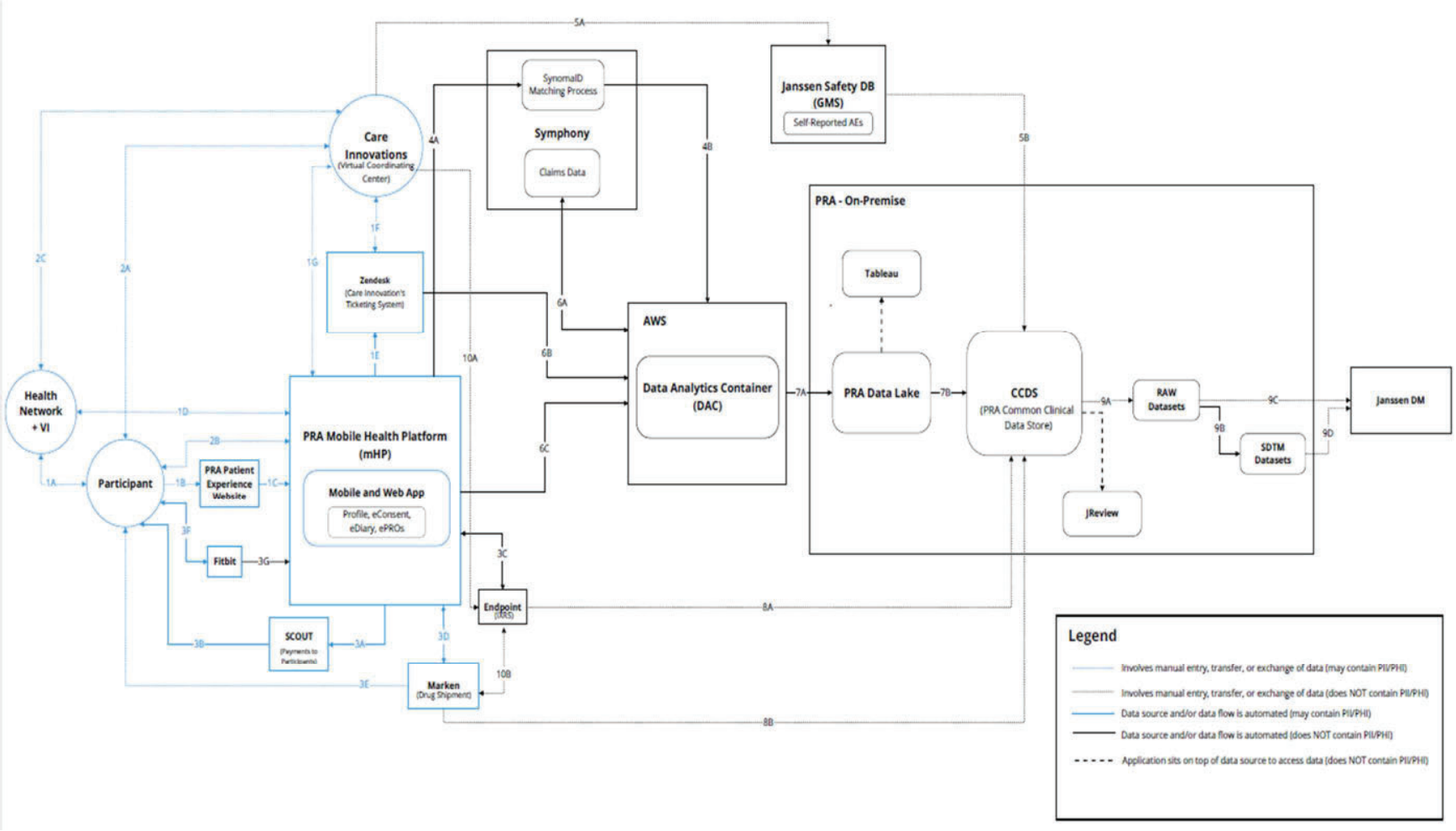
Period	Screening			Double-Blind Treatment				No Treatment		Notes
	Screening	Baseline								
Day	Within 21 days before Day 1	Within 7 days before Day 1	1 (first dose)	14±3	28±3	42±3	84±3 End of Treatment	85-251-±7	252-±7 End of Study	Day -21 to 0=Screening; Day 1 = first dose, Day 84= end of treatment; Day 252=End of Study
Week				2	4	6	12	13-36		Week 12 is equivalent to Day 84 of treatment
Month			0				3	6	9	Month 3 is equivalent to Day 84 of treatment
Participant Satisfaction Survey (optional)									X	
Health Resource Utilization and Health Economics										
HHF for any reason and duration			←				→			Hospitalizations for heart failure
Emergency department visits			←				→			
Readmissions and duration			←				→			
Ongoing Review of Participant Data										
Prohibited medications and new diagnoses			←				→			
Reporting of adverse events[e]	←						→			

Appendix 4: Usage of the Analysis Sets

Appendix 4 presents an overview of the different analysis sets and their usage.

Analyses/data displays	All randomized analysis set	Full analysis set (FAS)	Safety analysis set	Per protocol set
Patient disposition	✓	✓	✓	✓
Demographics	✓	✓	✓	✓
Baseline characteristics		✓		✓
Medical history		✓		
Treatment exposure			✓	✓
Effectiveness: Primary endpoint		✓		✓
Effectiveness: Secondary endpoints		✓		✓
Effectiveness; Exploratory endpoints		✓		
Safety endpoints			✓	
All other patient listings	✓			

Appendix 5: CHIEF-HF Study Data Flow Diagram



Step	Description
1A	Health Network sends email to prospective study participants.
1B	Pre-consent Participant clicks link to PRA Patient Experience Website and completes Technology Screening form and contact info submission to "opt in".
1C	PRA Patient Experience opt in triggers Patient ID creation in PRA mHP.
1D	Health Network + VI utilize the PRA mHP Web Application to update Pre-Consent Participant demographic info and assess eligibility.
1E	Care Innovations Zendesk Ticketing System automatically pulls information from PRA mHP for Pre-Consent Participants assessed as "Eligible" and creates an assigned open ticket.
1F	Care Innovations interacts with Zendesk Ticketing System to pick up assigned open tickets.
1G	Care Innovations interacts with PRA mHP Web Application to access Pre-Consent Participant contact information to schedule eConsent call.
2A	Care Innovations initiates communication with Pre-consent Participant to schedule eConsent call and verify contact information (mobile number and address).
2B	Pre-consent Participant interacts with PRA mHP Participant Web Application and Mobile Application to complete pre-consent review.
2C	Care Innovations interacts with VI to coordinate and facilitate completion of eConsent call. Post-consent Participant moves to ePRO completion post eConsent call.
3A	PRA mHP sends questionnaire completion details to SCOUT for payment to participants.
3B	SCOUT sends reloadable payment card and payment to consented participants upon completion of study questionnaires.
3C	PRA mHP interacts with Endpoint IRT at Randomization via a two-way database integration.
3D	PRA mHP interacts with Marken's system via a two-way database integration.
3E	Marken schedules directly with Participant to deliver drug to Participant's home. Care Innovations sends study supplies Welcome Kit to Participant.
3F	During Welcome Call with Care Innovations, Participant is guided through the supplies in the Welcome Kit, which includes the Fitbit watch. Participant interacts with Fitbit Mobile Application for Fitbit set-up and wears Fitbit to collect Biodata / ADL data.
3G	Once authorized within the PRA mHP Mobile Application, Fitbit data is retrieved by PRA mHP automatically through Fitbit API.
4A	PRA mHP sends Symphony encrypted, tokenized PII to match consented Participant in PRA mHP with their available medical claims data in Symphony.
4B	Synoma ID Matches identified in Symphony are sent to DAC
5A	Care Innovations manually submits participant-reported AEs directly to Janssen Safety (GMS).
5B	Janssen Safety (GMS) extracts participant-reported AEs and transfers data to PRA for upload into CCDS.
6A	Symphony Claims data is pulled into DAC.
6B	Care Innovations Zendesk ticketing data is sent to DAC via JSON payloads.
6C	PRA mHP data is pulled into DAC.
7A	DAC data is staged and moved into consolidation (CNS) layer, aka "PRA Data Lake".
7B	Consolidated DAC data tables are loaded into CCDS.
8A	Data Transfer from Endpoint is manually loaded into CCDS.
8B	Data Transfer from Marken is manually loaded into CCDS.
9A	RAW SAS datasets are extracted from CCDS for each data delivery.
9B	SDTM Datasets are generated from the RAW data tables for each data delivery.
9C	RAW SAS datasets are manually posted to LSAF for delivery to Janssen.
9D	SDTM SAS datasets are manually posted to LSAF for delivery to Janssen.
10A	Care Innovations manually enters drug resupply request into Endpoint IRT (if applicable)
10B	Endpoint notifies Marken of need for drug resupply (if applicable)

Appendix 6: Patient Global Impression of Change (PGIC)

Appendix 6 provides a representative example of the scale questions that will be used in this study.

Since the start of the treatment you've received in this study, your heart failure symptoms are: (select one response)

1. Very much improved
2. Somewhat improved
3. A little improved
4. No change
5. A little worse
6. Somewhat worse
7. Very much worse

Appendix 7: Patient Global Impression of Severity (PGI-S)

Appendix 7 provides a representative example of the scale question that will be used in this study.

Considering all aspects of your heart failure symptoms right now, would you say your heart failure symptoms are: (select one response)

1. None
2. Mild
3. Moderate
4. Severe
5. Very Severe

Appendix 8: KCCQ Score Instructions

The Kansas City Cardiomyopathy Questionnaire Scoring Instructions

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

1. PHYSICAL LIMITATION

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

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

- If at least three of Questions 1a-f are not missing, then compute

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

  (see footnote at end of this document for explanation of meaning of "actually answered")

2. SYMPTOM STABILITY

- Code the response to Question 2 as follows:

  Much worse = 1

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

- If Question 2 is not missing, then compute

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

3. SYMPTOM FREQUENCY

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

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

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

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

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

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

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

4. SYMPTOM BURDEN

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

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

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

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

5. TOTAL SYMPTOM SCORE

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

6. SELF-EFFICACY

- Code responses to Questions 10 and 11 as follows:

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

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

- If at least one of Questions 10 and 11 is not missing, then compute

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

7. QUALITY OF LIFE

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

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

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

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

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

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

8. SOCIAL LIMITATION

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

  Severely limited = 1
  Limited quite a bit = 2
  Moderately limited = 3
  Slightly limited = 4
  Did not limit at all = 5

Does not apply or did not do for other reasons = <*missing value*>

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

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

9. OVERALL SUMMARY SCORE

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

10. CLINICAL SUMMARY SCORE

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

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

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

If doing these calculations seems like too much trouble, consider using one of our tools – available at www.cvoutcomes.org:

- SAS or SPSS code
- Excel spreadsheets
- Web data services