Document Type:	Study Analysis Plan
Official Title:	A randomized parallel-group, placebo-controlled, double-blind, multi-center trial to evaluate the efficacy and safety of the oral sGC stimulator vericiguat to improve physical functioning in activities of daily living in patients with heart failure and preserved ejection fraction (VITALITY-HFpEF)
NCT Number:	NCT03547583
Document Date:	21-Nov-2019

Title Page

A randomized parallel-group, placebo-controlled, double-blind, multi-center trial to eValuate the effIcacy and safeTy of the orAL sGC stImulator vericiguaT to improve phYsical functioning in activities of daily living in patients with heart failure and preserved ejection fraction (VITALITY-HFpEF)

Bayer study drug	BAY 1021189 / Vericiguat / sGC stimulator vericiguat		
Study purpose:	efficacy, safety, pharmacokinetics		
Clinical study phase:	2b	**Date:**	21 NOV 2019
Study No.:	19334	**Version:**	2.0
Author:	PPD		

This Statistical Analysis Plan is produced on a word-processing system and bears no signatures.

The approval of the Statistical Analysis Plan is documented in a separate Signature Document.

Table of Contents

List of Tables

List of Figures

Abbreviations

ACE	Angiotensin converting enzyme
ACIR	Assessment Criteria Identification Requirement
AE	Adverse event
ALT	Alanine aminotransferase
AN(C)OVA	Analysis of (Co-)Variance
ARB	Angiotensin receptor blocker
ATC	Anatomical Therapeutic Chemical
BMI	Body mass index
CAD	Coronary artery disease
CDF	Cumulative distribution function
cGMP	Cyclic guanosine monophosphate
CKD	Chronic kidney disease
COPD	Chronic obstructive pulmonary disease
CSR	Clinical Study Report
CSS	Clinical Summary Score
CV	Cardiovascular
DSMC	Data Safety Monitoring Committee
DBP	Diastolic blood pressure
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
EF	Ejection fraction
eGFR	Estimated glomerular filtration rate
e.g.	*exempli gratia*, for example
FAS	Full analysis set
GCP	Good Clinical Practice
HEOR	Health Economics, Outcomes & Reimbursement
HF	Heart failure
HFrEF	Heart failure with reduced ejection fraction
HFpEF	Heart failure with preserved ejection fraction
HR	Heart rate
ICH	International Committee on Harmonization
i.e.	*id est*, that is
IV	Intravenous
KCCQ	Kansas City Cardiomyopathy Questionnaire
LAV	Left atrial volume
LLOQ	Lower limit of quantification
LVEF	Left ventricular ejection fraction
MACE	Major adverse cardiovascular event
MAP	Mean arterial pressure
MAR	Missing at random
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed-effects model for repeated measures
MNAR	Missing not at random
MRA	Mineralocorticoid receptor antagonist
NO	Nitric oxide
NT-proBNP	N-terminal pro-brain natriuretic peptide
NYHA	New York Heart Association
o.d.	*omni die*, once daily
OSS	Overall Summary Score
PD	Pharmacodynamics

PDF	Probability Density Function
PLS	Physical Limitation Score
PK	Pharmacokinetics
PKS	Pharmacokinetic analysis set
PP	Pulse pressure
PPS	Per protocol set
PT	Preferred term
QoL	Quality of Life
ROC	Receiver operating characteristics
SAC	Statistical Analysis Center
SAE	Serious adverse event
SAF	Safety analysis set
SAP	Statistical Analysis Plan
SBP	Systolic blood pressure
SFS	Symptom Frequency Score
sGC	Soluble guanylate cyclase
SMQ	Standard MedDRA query
SOC	System organ class
TEAE	Treatment emergent adverse event
TLF	Tables, listings, and figures
TSS	Total Symptom Score
ULN	Upper limit of normal
VAS	Visual analogue scale
WCHF	Worsening chronic heart failure
WHO-DD	World Health Organization Drug Dictionary

1. Introduction

Patients hospitalized with heart failure (HF) with preserved ejection fraction (HFpEF) are a large and growing population. The relative proportion of HFpEF has increased to more than 50% of all HF hospitalizations [1]. Event rates after hospitalization in patients with HFpEF are as high as in patients with HF with reduced ejection fraction (HFrEF). Since patients with HFpEF are highly symptomatic with a poor quality of life, it is important to find new therapies that can alleviate symptoms and improve patient well-being [2].

No specific therapy has been currently established for HFpEF patients, and recommendations are limited to symptomatic treatment of congestive symptoms by diuretics, and to treating causes and comorbidities including hypertension, coronary artery disease, and atrial fibrillation [3]. Trials that established the clinical benefits of angiotensin converting enzyme (ACE) inhibitors, angiotensin receptor antagonists, and β-blockers have all selected patients on the basis of reduced EF. Subsequent trials in patients with HFpEF did not confirm equal effectiveness in this subgroup [4]. Therefore, there is need for new drug targets such as cyclic guanosine monophosphate (cGMP).

Vericiguat has been evaluated in approximately 65 preclinical studies and has been administered to over 500 healthy patients in 19 completed Phase I studies with single- and multiple-dose administration, and in approximately 740 heart failure patients in two Phase IIb studies (SOCRATES- REDUCED and SOCRATES- PRESERVED) that studied vericiguat in worsening chronic heart failure patients with HFrEF and with HFpEF. In the ongoing phase III event-driven outcome trial VICTORIA (NCT02861534), as of Aug 31 2018, 3791 patients with HFrEF out of a planned total sample size of 4872 have been randomized to vericiguat 10 mg target dose or placebo with a median study drug treatment duration at that time of 5 months. The study treatment duration is planned to continue until 2020 for a median follow-up duration of 18 months.

In SOCRATES-PRESERVED 4 different dose regimens of vericiguat were studied in patients with chronic HFpEF enrolled within 1 month after a qualifying heart failure decompensation event [5]. Data from SOCRATES-PRESERVED established 2.5 mg vericiguat as the safe and well tolerated starting dose. Uptitration by two dose doublings with 14 day intervals each to a target dose of 10 mg was tolerated in patients with HFpEF without an increase in adverse events compared to placebo, with low discontinuation rates in all groups, and no changes in blood pressure at 10 mg compared to placebo. Although analyses of the primary efficacy variables in SOCRATES-PRESERVED did not demonstrate changes in NT-proBNP and LAV at 12 weeks in the vericiguat dose groups compared to placebo, patients receiving vericiguat in the 10 mg target dose arm experienced a more pronounced improvement in health related QoL, physical limitations, NYHA class, signs and symptoms of congestion and trends towards improved diastolic function at 12 weeks [5,6].

The Kansas City Cardiomyopathy Questionnaire (KCCQ) has been widely used in HF trials. In SOCRATES-PRESERVED, improvements in the KCCQ overall summary score (OSS) as well as the clinical summary score (CSS) were observed. Consistent changes in other domains of the KCCQ were also noted. However, the improvement appeared to be driven primarily by improvements in the physical limitation score (PLS). KCCQ PLS has previously been

validated against the distance covered in a 6MWT, NYHA class, and physical limitation domains within other quality of life (QoL) instruments including the 36-Item Short Form Survey in patients with HF [7]. Improvements in physical functioning and reduction in symptom frequency are considered to be direct measures of treatment benefit in patients with HFpEF, and have been shown to be associated with subsequent reductions in mortality and hospitalizations [8]. These data generate the hypothesis that vericiguat improves physical functioning and clinically relevant patient-reported outcomes in patients with HFpEF. Based on these data, the VITALITY phase 2 trial (study 19334) has been designed to study whether in patients with HFpEF vericiguat 10 mg or 15 mg improves the KCCQ PLS compared to placebo after 24 weeks of treatment.

This statistical analysis plan (SAP) is based on the study protocol (Version 1.0 dated 22 FEB 2018) and describes the interim and final analyses of study 19334.

2. Study Objectives

The primary hypothesis is treatment with vericiguat 10 mg or 15 mg in patients with HFpEF improves physical functioning measured by the KCCQ PLS compared to placebo after 24 weeks of treatment. The secondary hypothesis is treatment with vericiguat 10 mg or 15 mg in patients with HFpEF improves walking distance as measured by the 6MWT compared to placebo after 24 weeks of treatment. The study is designed to have enough power for the primary and secondary hypotheses testing. Other objectives are listed only for descriptive purposes.

Primary objectives:

- To evaluate the efficacy of vericiguat 10 mg in comparison to placebo on improving physical functioning as measured by KCCQ PLS from baseline to week 24.
- To evaluate the efficacy of vericiguat 15 mg in comparison to placebo on improving physical functioning as measured by KCCQ PLS from baseline to week 24.
- To evaluate the safety and tolerability of vericiguat.

Secondary objectives:

- To evaluate the efficacy of vericiguat 10 mg in comparison to placebo on improving distance traveled on a 6MWT from baseline to week 24.
- To evaluate the efficacy of vericiguat 15 mg in comparison to placebo on improving distance traveled on a 6MWT from baseline to week 24.
- To evaluate the efficacy of vericiguat 10 mg in comparison to placebo in increasing the proportion of patients with KCCQ PLS improvement from baseline by >5 points at 24 weeks and other thresholds (e.g. >3, >7, >10, >15, >20), and the proportions with these improvements in the other KCCQ domains OSS, CSS, TSS and SFS.

- To evaluate the efficacy of vericiguat 15 mg in comparison to placebo in increasing the proportion of patients with KCCQ PLS improvement from baseline by >5 points at 24 weeks and other thresholds (e.g. >3, >7, >10, >15, >20), and the proportions with these improvements in the other KCCQ domains OSS, CSS, TSS and SFS.
- To evaluate the efficacy of vericiguat 10 mg in comparison to placebo in decreasing the proportion of patients with KCCQ PLS decline from baseline by >5 points at 24 weeks and the proportions with decline in the other KCCQ domains OSS, CSS, TSS and SFS.
- To evaluate the efficacy of vericiguat 15 mg in comparison to placebo in decreasing the proportion of patients with KCCQ PLS decline from baseline by >5 points at 24 weeks and the proportions with decline in the other KCCQ domains OSS, CSS, TSS and SFS.

Other objectives:

- To evaluate the efficacy of vericiguat 10 mg and the efficacy of vericiguat 15 mg in comparison to placebo on improving:
    - Symptom frequency as measured by the KCCQ SFS
    - Perceived exertion experienced by patients during 6MWT as measured by Borg CR 10.
    - Other patient-reported outcomes such as the generic health-related quality of life measure EQ-5D-5L and the Fried-based frailty score
    - NYHA class
    - Laboratory variables such as NT-proBNP from baseline to 24 weeks
- To collect and summarize death including cardiovascular death and CV hospitalizations including heart failure hospitalizations, MI and stroke, and outpatient HF events.
- To create the cumulative distribution function of KCCQ PLS, OSS, CSS, TSS and SFS change from baseline to 24 weeks.
- To evaluate PK of vericiguat in patients with HFpEF, and
- To evaluate the concentration-QTc relationship of vericiguat at the 15 mg dose level.
- Optional pharmacogenetic research: To explore the relationship between genetic variation and clinical characteristics of patients, independent and dependent of the treatment administered. Variations across the human genome may be analyzed for association with clinical data collected in this study.
- To evaluate further biomarkers to investigate the drug (i.e. mode-of-action-related effect and / or safety) and / or the mechanisms of the disease

- Optional accelerometry substudy: to collect exploratory data by activity tracking

Note: Analyses for PK, biomarker, and optional objectives (including accelerometry) will be specified in a separate supplementary SAP.

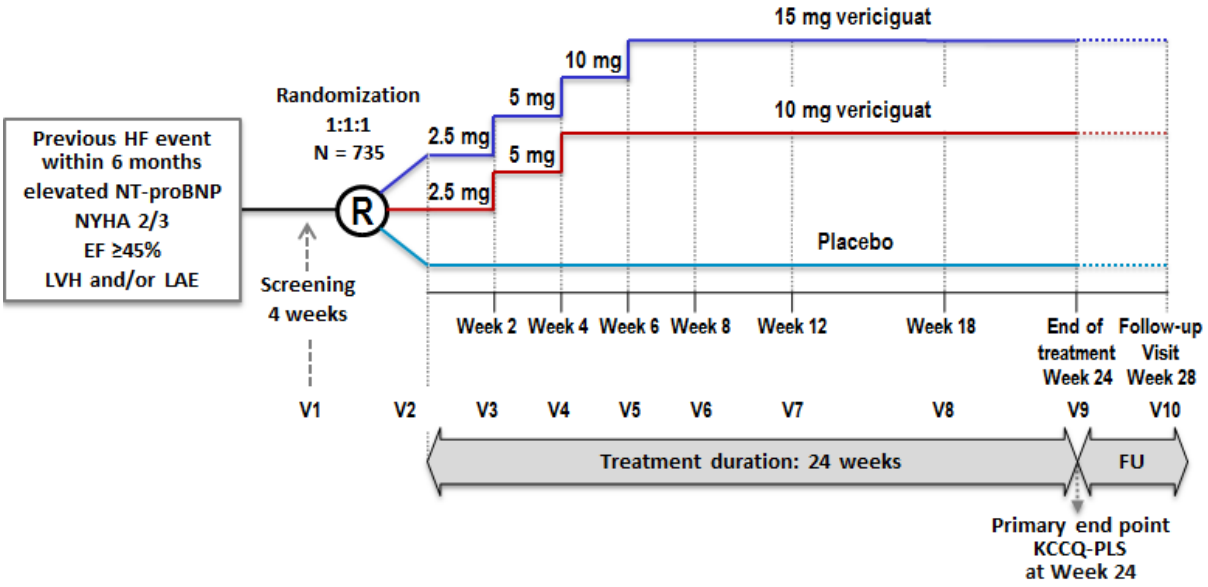
3. Study Design

This is a randomized, parallel-group, placebo-controlled, double-blind, multi-center trial of vericiguat in patients with HFpEF to be conducted in conformance with GCP. Approximately 735 patients will be randomized as described in Figure 3-1 to evaluate the efficacy of vericiguat 10 mg and 15 mg in comparison to placebo on improving physical functioning from baseline to week 24. To reduce heterogeneity in the studied population patients younger than 45 years of age are not eligible. Patients are expected to be on a background treatment for concurrent conditions including hypertension (anti-hypertensives) and volume overload (diuretics).

Screening may be initiated any time after admission of a patient to the hospital for HF, before or after discharge, and up to 6 months after hospitalization for HF. Use of IV diuretic treatment for HF, even without hospitalization, is indicative of HF decompensation and accepted as a qualifying HF event equivalent to hospitalization. The most recent decompensation should be considered for qualifying the patient for the study. In either scenario, HF must be the primary reason for hospitalization or IV diuretic treatment. Patients will be randomized within up to 4 weeks after the screening visit and within up to 6 months after the onset of the qualifying event of hospitalization for HF or treatment with IV diuretic for HF (see Figure 3-1).

Primary endpoint assessment is at week 24. All patients will be followed until study completion to assess for vital status and all study endpoints until week 28.

Figure 3-1 Study design



EF, ejection fraction; FU, follow-up; HF, heart failure; KCCQ, Kansas City Cardiomyopathy Questionnaire; LAE, left atrial enlargement; LVH, left ventricular hypertrophy; NT-proBNP, N-terminal pro-brain natriuretic peptide; NYHA, New York Heart Association; PLS, physical limitation score; V, Visit. KCCQ as the entire questionnaire is administered, not just the PLS domain.

Stratification

Patients will be randomized within 4 weeks of the screening visit and within up to 6 months after the date of hospitalization for HF or treatment with IV diuretic for HF, and randomization will be stratified according to region and heart rhythm at baseline ECG.

Region

- Americas
- Europe (incl. Israel and South Africa)
- Asia/Pacific

Heart rhythm

- Atrial fibrillation (defined as atrial fibrillation or atrial flutter in baseline ECG)
- Sinus rhythm (defined as no atrial fibrillation and no atrial flutter in baseline ECG)

Treatment

Patients will be randomized 1:1:1 within 4 weeks after the Screening Visit to either:

1. Placebo arm: placebo and sham up-titration at weeks 2, 4, and 6
2. 10 mg arm: vericiguat, which will be started at 2.5 mg at randomization and up-titrated to 5 mg at week 2, and to 10 mg at week 4, with sham titration at week 6.
3. 15 mg arm: vericiguat, which will be started at 2.5 mg at randomization and up-titrated to 5 mg at week 2, to 10 mg at week 4, and to 15 mg at week 6.

Withdrawal from study

Unnecessary withdrawal of patients from the study follow-up should be avoided and all efforts should be taken to motivate patients to adhere to all study procedures and to be followed until the end of the trial. All patients will be followed until study completion to assess for vital status and all study endpoints until week 28.

The investigator should explore all possible options to reach a patient who fails to return to a visit or to respond to the site contact attempts. The site must document all attempts to try to contact the patient in the medical records / source documents. In order to avoid loss-to-follow-up, the investigator should ask the patient at the study start for the contact details of a relative or friend who can be contacted in case the patient cannot be reached. Patients should not be withdrawn from follow-up unless the patient explicitly withdraws consent to be contacted. All

efforts should therefore be made to discuss solutions with the patient that would enable the patient to continue with follow-up in order to minimize the number of patients who withdraw such consent. The vital status will be collected for all randomized patients who have not specifically withdrawn consent for further study follow-up by any method (telephone call, email etc.), irrespective of completion of study procedures.

End of study

The end of the study as a whole will be reached as soon as the last visit of the last patient has been reached in all centers in all participating countries (EU and non-EU).

4. General Statistical Considerations

4.1 General Principles

The statistical analyses will be performed using the software package SAS release 9.2 or higher (SAS Institute Inc., Cary, NC, USA) [9].

All variables will be analyzed by descriptive statistical methods. For continuous data, the summary statistics including mean, standard deviation, minimum, median, quartiles, and maximum, and the number of missing data will be provided. For categorical data, frequency count and percentages will be generated.

All patients will be analyzed according to the assigned treatment group if not specified otherwise. Each treatment group will comprise all patients within one randomized titration scheme regardless of the individual subject's actual titration, i.e. in case a subject could not be uptitrated or was down-titrated during the study.

On treatment

A measurement is considered to be "on-treatment," if it was taken during treatment or up to 5 days after end of treatment with study medication.

Analysis time point and visit window

Based on the randomization date, each scheduled visit date will be computed and a midpoint between two adjacent scheduled visit dates will be identified as the boundary of two visit windows. For example, Day 28 will be used to label the boundary between week 2 and week 6 windows. Table 4-1 shows an example of the visit window for each analysis time point for KCCQ. Each visit date will be mapped to one visit window. For any particular window, if there are multiple measurements within the window then the one that is closest to the scheduled visit date will be used. In case there are multiple measurements on the same day, the latest one will be selected. Also note that the baseline is defined as the last non-missing measurement before first intake of study treatment (Section 4.4.2). In case KCCQ assessments may simultaneously be considered as baseline and follow up, for example if a patient takes first study treatment 14 days after randomization, under such scenario these measurements will be assigned as baseline only.

Table 4-1: Visit windows for KCCQ

Analysis time point	Scheduled visit date	Scheduled visit window
Baseline	Day 1	Day 1
Week 2	Day 14	Day 2 – Day 27
Week 6	Day 42	Day 28 – Day 62
Week 12	Day 84	Day 63 – Day 104
Week 18	Day 126	Day 105 – Day 146
Week 24	Day 168	Day 147 – Day 181
Safety follow-up*	Day 196	Day 182 – Day 209

*If patients have completed all scheduled doses. In case of premature discontinuation, safety follow-up will be scheduled at 28 ± 7 days after the last dose, and the visit window will span from two weeks before the scheduled safety follow-up visit to two weeks after.

4.2 Handling of Missing Data

All missing or partial data will be presented in the subject data listings as they are recorded on the electronic Case Report Form (eCRF) or electronic PRO diary (ePRO).

For missing components in KCCQ domains, the algorithm for score calculation will follow the KCCQ scoring instructions [10]. For example in the PLS domain, if a patient answers "Limited for other reasons or did not do the activity", the response will be considered as missing. KCCQ scoring instructions specify that at least 3 out of 6 non-missing components are needed for the domain to be evaluable.

Non-evaluable domains will be treated the same way as missing domains and will be handled based on missing data imputation rules specified in Appendix 9.1.

4.3 Interim Analyses and Data Monitoring

An independent Data Safety Monitoring Committee (DSMC) has been established for this study. The DSMC will review safety and tolerability regularly, and conduct the futility interim analysis during the course of the study. Members of the DSMC will receive unblinded safety data. The involvement of an external Statistical Analysis Center in this process will ensure that unblinded information is not available to anyone except the DSMC. The study team and study committees (steering committee and the PRO sub-committee, see below) will be kept blinded in this process. There are no predefined stopping conditions for the ongoing safety monitoring of this trial. However, the DSMC may recommend termination, temporary suspension of the study, and intervention of treatment arm or modification of the study. Details on DSMC responsibilities and decisions, as well as a detailed plan for DSMC meetings will be described in the DSMC Charter.

A sub-committee of the study steering committee will be focused on PROs (called PRO sub-committee The PRO sub-committee is made up of an additional panel of experts in the field acting in support of the steering committee. Steering committee and PRO sub-committee will review and monitor the blinded PRO data regularly to ensure data quality and completeness. The PRO sub-committee will also contribute to the interpretation of the analyses to evaluate important within patient change which will be used to determine the responder definitions for

the KCCQ-PLS. Details on PRO sub-committee membership, roles, responsibilities and operational aspects will be described in the PRO sub-committee charter.

4.3.1 Futility interim analysis

A single interim assessment by the DSMC to evaluate futility is planned for when 50% of the patients complete Visit 7 (week 12) and their KCCQ data become available. Although the effects of 24 weeks of treatment with vericiguat have not been evaluated in patients with HFpEF, prior data in this population suggest that the 12 week timepoint will be sufficiently predictive of the 24 week data to allow for an adequate interim assessment of efficacy. Furthermore, in view of the expected timeline for patient recruitment, a sufficiently large proportion of data at the 24 week timepoint will only become available shortly before completion of the trial. Based on the operating characteristics of a range of futility criteria (Table 4-2), an observed mean treatment effect of 0.6 point difference in KCCQ PLS (vericiguat vs placebo) was chosen to serve as the boundary for futility for both dose groups; this is approximately equivalent to a joint conditional power (CP) of 23% at week 24 (under the assumption of a true difference of 5 points and a SD of 21 points).

Table 4-2: Operating characteristics for futility boundaries

Equivalent treatment effect boundary (points)	Conditional power [1]	Joint conditional power [2]	Probability of meeting futility criterion (Observed treatment difference < boundary for both arms)		Joint probability [3] of observing a >7/5/3 point treatment effect on any dose at final analysis
			True effect = 0	True effect = 5	
1.2	20%	31%	53%	2.5%	0.004 / 0.134 / 0.696
0.7	15%	24%	44%	1.4%	0 / 0.095 / 0.621
0.6	**14%**	**23%**	**43%**	**1.3%**	**0 / 0.089 / 0.605**
0.5	13%	21%	40%	1.1%	0 / 0.082 / 0.590
0.4	12%	20%	38%	1.0%	0 / 0.077 / 0.574
0	10%	16%	34%	0.7%	0 / 0.056 / 0.509

[1] Probability of rejecting any individual hypothesis at the end of the study based on observed interim value (under the assumption of a true difference of 5 points).
[2] Joint probability of rejecting at least one of the hypotheses at the end of the study based on observed interim value (under the assumption of a true difference of 5 points).
[3] Joint probability of observing a treatment effect larger than certain threshold in either of the comparisons at the end of the study based on observed interim value (under the assumption of a true difference of 5 points).

An observed mean treatment effect <0.6 in both comparisons (vericiguat 10 mg vs. placebo and vericiguat 15 mg vs. placebo) would therefore signal to the DSMC to consider their non-binding recommendation that the trial should be terminated early. Assuming the true treatment effect is a 5-point difference, as expected, the probability of stopping at the interim is approximately 1.3%. Under the null hypothesis, i.e., assuming 0 treatment effect for both vericiguat vs. placebo comparisons, the probability of stopping at the interim for futility is approximately 43%. If only one arm is futile then it will not be dropped as enrollment is expected to be almost completed at the time of the futility analysis. Stopping for futility will only be done if both arms meet the futility criterion. Under this observed interim value (0.6 point difference) and assuming the alternative hypothesis is true, the trial has a probability of 8.9% to observe a >5 point improvement and a probability of 60.5% to observe a >3 point mean improvement in either of the comparisons in KCCQ PLS at week 24. Details of the futility interim analysis will be specified in the DSMC charter. No formal interim report will be written.

4.3.2 Anchor-based analysis and responder definitions

The main analyses to aid the interpretation and definition of a responder threshold for the KCCQ PLS change scores will be based on an anchor-based approach by categorizing patients according to their responses from baseline to Visit 7 (week 12) to the Patient Global Impression of Change (PGIC) (see Section 6.2.1). A responder definition represents the threshold for meaningful within-patient change, and it can be used to categorize patients as responders or non-responders based on the KCCQ PLS scores.

Supportive anchor-based analyses will be conducted by categorizing patients according to their change in Patient Global Impression of Severity (PGIS) (see Section 6.2.1) responses from baseline to Visit 7 (week 12) (e.g. unchanged, improved by 1 response category, deteriorated by 1 response category, etc.) and additionally, change in PGIS response between week 18 and 24 (when more subjects are expected to rate themselves in the same severity category). In addition to anchor-based approaches, supportive distribution-based approaches will also be used, i.e. 0.5 standard deviation and standard error of measurement (SEM). These methods provide an indication of the smallest score that would exceed measurement error.

Graphical displays via empirical cumulative distribution function (CDF) and probability density function (PDF), receiver operating characteristics (ROC) and non-parametric discriminant analyses will also be created to support the above estimations.

These analyses will serve to support the interpretation of study results relative to thresholds for meaningful within-patient change in addition to the established thresholds of clinically meaningful between group changes for the KCCQ PLS [11,12,13]. They will be performed on blinded data, with all treatments combined, from an interim dataset prior to study unblinding. These analyses are planned for when 100% of the patients complete Visit 7 (week 12) and their KCCQ data become available. Only patients with actual data (not imputed data) on the KCCQ PLS, PGIC, PGIS at baseline and the relevant post-baseline date (collected on the same day) will be included. The blinded analysis results will be reviewed by the PRO sub-committee which will propose an appropriate responder definition estimate for within-patient

improvement and/or worsening in KCCQ PLS scores. Correlations between KCCQ-PLS and anchors will be calculated to assess the level of confidence in the interpretation of results. Correlations >0.35 are generally considered acceptable, and confidence in an individual analysis is increased between KCCQ-PLS and the anchor if the correlation is moderate to strong (0.4 – 0.6) [11].

Anchor-based approach

The anchor-based approach will be applied by categorizing patients according to their responses to the PGIC questions at Visit 7 (week 12).

The PGIC contains two questions (i.e. Q1 and Q1a or Q1b):

1. Compared to the start of your treatment, how would you describe your physical limitations due to heart failure today?

    - Response options: “Much better”, “Better”, “A little better”, “The same”, “A little worse”, “Worse”, or “Much worse”;

- If a subject responds “Much better”, “Better”, or “A little better”, then he/she is asked:
    - 1a. Was this improvement in your physical limitations an important change for you?
    - Response options: “Yes”, “No”.
- If a subject responds “Much worse”, “Worse”, or “A little worse”, then he/she is asked:
    - 1b. Was this worsening in your physical limitations an important change for you (by important we mean did it bother you)?
    - Response options: “Yes”, “No”.

Patients will be categorized according to their responses to the PGIC as follows:

- Group 1: Patients who report “A little better” to PGIC #1 as well as “Yes” to 1a;
- Group 2: Patients who report “A little worse” to PGIC #1 as well as ‘Yes’ to 1b

KCCQ PLS Mean (SD) change scores from baseline to Visit 7 (week 12) will be calculated for each group above to estimate a meaningful within-patient threshold.

KCCQ PLS mean (SD) change scores from baseline to Visit 7 (week 12) will also be calculated for the all PGIC response categories separately, to assess trends in change scores. In addition, KCCQ PLS mean (SD) change scores from baseline to Visit 7 (week 12) will be calculated by overall improvement or worsening in PGIC response categories.

Distribution-based approaches

Distribution-based approaches will also be performed to supplement the anchor-based approaches. These will be based on 0.5 SD and standard error of measurement (SEM) of the KCCQ-PLS at baseline. The SEM is calculated as the standard deviation at baseline multiplied by the square root of one minus the reliability of the instrument [SD * (2(1-r))1/2].

The reliability coefficient will be estimated by calculating correlation (Intraclass Correlation Coefficient; ICC [14, 15]) between the baseline and Week 6 KCCQ-PLS scores among stable patients. Stable patients will be defined as patients who respond 'the same' to the PGIC at Week 6. KCCQ-PLS mean scores for baseline and week 6, and differences between the two assessments and respective standard deviations will be reported for this stable subgroup.

Empirical Cumulative Distribution Function (CDF) and Probability Density Function (PDF)

Empirical cumulative distribution functions of the change from baseline in KCCQ-PLS scores for each PGIC anchor category will be created to show the distribution of change scores among subjects who experienced different levels of change. Examining the distribution of change scores via empirical cumulative distribution function (CDF) and probability density function (PDF) plots provides supporting evidence for the chosen responder threshold [16].

Receiver Operating Curves (ROC)

ROC curves will be used to assess the ability of change in KCCQ PLS to discriminate between patients responding "a little better" with "yes" to PCIG question 1a, "better" or "much better" on the PGIC question at Visit 7 (week 12) and those with other responses. The point(s) on the curve that optimizes sensitivity and/or specificity will be considered the optimal threshold in differentiating between the two groups.

Non-Parametric Discriminant Analysis

Non-parametric discriminant analysis will be conducted to explore the responder definitions [17] further. Given sufficient sample size (at least 20 patients in each of the response groups: Patients who report "A little better" to PGIC #1 as well as "Yes" to 1a; or "A little worse" to PGIC #1 as well as "Yes" to 1b; or "The same" to PGIC #1), the KCCQ PLS score density curve will be estimated for each group and the intersection points of these density curves will be considered as the discrimination points. The discrimination point between the two groups "A little better" with response "Yes" to 1a and "The same" is considered to be an estimate of the responder definition for improvement. Similarly, the estimate of the responder definition for worsening is calculated based on those who report "A little worse" with response "Yes" to 1b and "The same" to PGIC #1. In case there is not sufficient sample size in each group, patients who respond "A little better", "Better" or "Much better" will be aggregated to the new category "Improved". Patients who respond "A little worse", "Worse" or "Much worse" will be combined into the new category "Worsened". Discriminant point analysis will be performed based on these new categories in a similar way.

4.3.2.1 Updates from FDA Type C meeting

The proposed KCCQ-PLS thresholds for meaningful within-patient changes determined during the interim and blinded dataset analysis of 8.33 and 12.5 for improvement and -4.17 for worsening will be added to the pre-specified thresholds (e.g. >3, >7, >10, >15, >20) for exploratory responder analysis already included in the SAP version 1.0. Following an FDA request made during the Type C meeting held on 6th November 2019, the anchor-based analysis will be additionally run on the full trial dataset prior to unblinding to further inform interpretation of the primary endpoint of KCCQ-PLS change from baseline to week 24.

Supportive analyses, including distributional-based approaches, receiver-operating characteristic curves and non-parametric discriminant methods will not be re-evaluated on the full trial dataset prior to unblinding.

4.4 Data Rules

4.4.1 Compliance

The patient drug intake compliance is calculated as percent of planned tablet intake:

100 * *Number of taken tablets* / *Number of planned tablets*.

The *number of planned tablets* is calculated as: *Treatment duration* (days) * *Number of planned tablets per day*. The *number of planned tablets per day* is two.

To derive *treatment duration*, the start and stop dates of tablet intake are derived from the collected study treatment exposure (dataset EX). The *number of taken tablets* is derived from the drug accountability (dataset DA).

For patients who withdraw prematurely from the study treatment, compliance will be calculated up to the time of last dose.

4.4.1.1 Imputation rules for partially missing end dates for study medication

In case of (partially) missing end dates for study medication intake the following approach will be applied to impute end of study medication intake:

- In case the subject stopped treatment due to death: impute partially missing end dates by the 'maximum study medication end date' defined as the maximal possible date (i.e. last month of the year, and last day of the month, respectively). Take the minimum of death date and, if available, imputed maximum study medication end date as the imputed study medication end date.
- In case the subject completed treatment or stopped due to reasons other than death: impute partially missing end dates by the 'minimum study medication end date' defined as the minimal possible date (i.e. first day of month and first month of year, respectively). Take the maximum of the last available study medication intake date (i.e. largest non-missing study medication start or stop date) and, if available, imputed minimum study medication end date as the imputed study medication end date.

4.4.2 Baseline and Change from Baseline

Baseline is defined as the last non-missing measurement before or on randomization date. If more than one measurement was taken for a scheduled time point (i.e. blood pressure measurements and heart rate), the average value is used

Change from baseline will be displayed as the difference to baseline defined as:

Change = Post baseline value – baseline value.

4.4.2.1 Vital Signs

Three measurements of vital signs parameters will be taken at time intervals of about 2 minutes. Averages of non-missing values of these three measurements will be calculated and used for the statistical analysis. If only one of the planned measurements is available, this value will be used.

4.4.3 Laboratory Data Handling

Only the data provided by the central laboratory will be used for analysis, values from local laboratories will not be used.

For values which are below the lower limit of quantification (LLOQ), half the value of the LLOQ will be used for analysis. Differences between two values below the LLOQ will be assigned values of 0.

4.4.4 Subgroup analyses

The following subgroups will be considered for descriptive and exploratory analyses. To determine whether the treatment effect is consistent across various subgroups, the estimate of the between-group treatment effect (with a nominal 95% CI) for the primary endpoint will be estimated and plotted within each category of the following classification variables in all subjects:

- Gender: Male, Female
- Region:
  - Americas
  - Europe (incl. Israel and South Africa)
  - Asia/Pacific
- Heart rhythm: Atrial fibrillation (defined as atrial fibrillation or atrial flutter in baseline ECG), Sinus rhythm (defined as no atrial fibrillation and no atrial flutter in baseline ECG)
- Randomization time:
  - Before discharge from HF hospitalization
  - After discharge from HF hospitalization
  - Post IV diuretic treatment without hospitalization
- Time from qualifying event (HF hospitalization or IV diuretic) to randomization: <3 month, ≥3 month
- Baseline Body Mass Index (BMI; kg/m^2): ≤30, >30
- Baseline Estimated Glomerular Filtration Rate (eGFR; $mL/min/1.73m^2$): ≤60, >60
- Diabetes: Yes, No

- COPD: Yes, No
- Anemia: Yes, No
- NYHA class: II, III
- Baseline KCCQ-PLS: <45, 45 – <75, ≥75
- Baseline subgroups defined by combined KCCQ-PLS and NYHA class criteria, as follows: KCCQ-PLS <65 and NYHA II; KCCQ-PLS <65 and NYHA III; KCCQ-PLS ≥65 and NYHA II; KCCQ-PLS ≥65 and NYHA III

5. Analysis Sets

5.1 Assignment of analysis sets

Final decisions regarding the assignment of patients to analysis sets will be made during the validity meeting (see Section 5.1.8).

Data from all patients who signed informed consent, regardless of their assignment to certain analysis sets, will be presented in individual subject data listings. In addition, the following analysis sets will be used for this study:

5.1.1 Full analysis set (FAS)

All patients randomized will be valid for the FAS. The patients will be analyzed based on the planned treatment. The population characteristics (such as disposition, demographic and baseline characteristics etc.) will be based on FAS.

5.1.2 FAS (KCCQ)

All patients randomized and treated (at least one dose of the study treatment), and have at least one observed KCCQ-PLS assessment at both baseline and post-baseline (excluding safety follow-up), will be valid for this analysis set. The primary analysis for the primary efficacy endpoint will be based on FAS (KCCQ).

5.1.3 FAS (6MWT)

All patients randomized and treated (at least one dose of the study treatment), and who are able to perform at least one 6MWT assessment at both baseline and post-baseline (excluding safety follow-up), will be valid for this analysis set. The primary analysis for the secondary efficacy endpoint will be based on FAS (6MWT).

5.1.4 Pre-specified supportive efficacy analysis set (PEAS)

All patients randomized and treated (at least one dose of the study treatment), and have at least one observed KCCQ-PLS assessment at both baseline and post-baseline (excluding safety follow-up), excluding those with no physical limitation at baseline (baseline KCCQ PLS = 100 or baseline PGIS as “no limitation”), will be valid for this supportive efficacy analysis set. The primary efficacy analysis will be repeated on PEAS as a supportive analysis.

5.1.5 Safety analysis set (SAF)

All patients that received at least one dose of study treatment (vericiguat or placebo). Analyses will be based on actual treatment received, and patients who received both vericiguat and placebo by mistake will be assigned to the planned treatment. All safety analyses will be based on the SAF.

5.1.6 Per protocol set (PPS)

All patients who meet the major inclusion and exclusion criteria at randomization that may affect efficacy, who are not taking excluded concomitant medications during the study that could have an effect on efficacy, have the KCCQ assessed at baseline and at least once during the treatment phase and who are at least 80% compliant with study medication and show no major protocol deviations. The PPS will only be used in the supportive analysis.

Major protocol deviations leading to exclusion from PPS include:

- Overall compliance with study treatment intake of <80% or >120%.
- Study medication was interrupted for more than 10 consecutive days
- No history of chronic heart failure.

A complete list of all validity findings and important protocol deviations leading to exclusion from the PPS will be specified in the Assessment Criteria Identification Requirement (ACIR).

5.1.7 Pharmacokinetics analysis set (PKS)

All patients with at least one valid pharmacokinetic concentration will be included in the statistical evaluation of pharmacokinetics.

5.1.8 Validity Review

A validity meeting will be held before final database closure where subject validity for the individual analysis sets will be decided. The results of the validity meeting will be documented and may comprise decisions and details relevant for statistical evaluation. Any changes to the statistical analysis prompted by the results of the validity review meeting will be documented in an amendment or in a supplement to this SAP, as applicable.

6. Statistical Methodology

6.1 Population characteristics

Population characteristics will be summarized overall and by treatment group. Analyses will be performed in the FAS unless otherwise noted.

6.1.1 Disposition

The number of patients enrolled, randomized, and valid for the FAS, FAS-KCCQ, FAS-6MWT, SAF, PPS and PKS will be summarized overall and by treatment groups. The number of patients discontinuing the screening phase together with the primary reason for discontinuation will be presented. The number of patients discontinuing the treatment and

follow-up phases together with the primary reason for discontinuation will be presented by treatment groups and overall in separate tables.

6.1.2 Demography and Baseline Characteristics

Demographic variables and baseline characteristics will be summarized by treatment group and overall. Summary statistics will be presented for metric variables. Frequency tables will be presented for categorical variables.

Demography includes age, gender, race, ethnicity, region, body height, body weight, body mass index (BMI), smoking history, and alcohol consumption. Age and BMI will each be given as continuous variable and categorized with the following categories:

- Age (years): <65, 65-75, >75
- BMI (kg/m^2): ≤30, >30.

The following additional baseline characteristics will be analyzed:

- Time from the qualifying HF event to randomization (days, continuous)
- Qualifying HF event: Hospitalization, IV diuretic
- LVEF (%) at baseline: 45% - <50%, 50% - <60%, ≥60%
- NT-proBNP (pg/mL) at baseline (continuous)
- eGFR (mL/min/1.73m^2) at baseline (continuous)
- eGFR (mL/min/1.73m^2) at baseline: ≤60, >60
- Heart rhythm: Atrial fibrillation, Sinus rhythm
- SBP (mmHg) at baseline (continuous)
- SBP (mmHg) at baseline: <120, ≥120
- Diastolic blood pressure (DBP; mmHg) at baseline (continuous)
- Heart rate (HR; beats/min) at baseline (continuous)
- NYHA class at baseline: Class II, III
- Medical history of coronary artery disease (CAD): present, not present
- Medical history of diabetes mellitus: diabetes, no diabetes
- Medical history of atrial fibrillation: yes, no
- Medical history of arterial hypertension: present, not present
- Medical history of chronic kidney disease (CKD): present, not present
- Medical history of chronic obstructive pulmonary disease (COPD): present, not present
- Medical history of anemia: present, not present

Demographics and baseline characteristics tables will also be presented by the subgroups defined in Section 4.4.4.

6.1.3 Protocol deviations

The number of patients with validity findings and important deviations will be presented overall and by investigator and country for each treatment group and in total. The frequencies of each major protocol deviation will be presented by treatment group and total.

6.1.4 Medical History

Medical history findings will be coded by Medical Dictionary for Regulatory Activities (MedDRA) codes. Medical history will be presented for each MedDRA Primary System Organ Class (SOC) and Preferred Term (PT) by treatment group and overall.

6.1.5 Concomitant Medication

Prior and concomitant medications will be coded by Anatomical Therapeutic Chemical (ATC) classification system according to the World Health Organization Drug Dictionary (WHO-DD).

6.1.5.1 Selected Concomitant Medication of Special Interest

Concomitant medication of special interest, derived via ATC codes, Bayer Drug Grouping (BDG), or combinations of both ATCs and BDGs, will be summarized by class and corrected generic name. Of special interest are the drug groups:

- diuretics
- β-blockers
- ACE inhibitors
- Angiotensin receptor blockers (ARB)
- Mineralocorticoid receptor antagonists (MRA)
- SGLT-2 inhibitors
- Calcium channel blockers

Diuretics will be categorized into the subcategories loop diuretics, thiazides, and potassium sparing diuretics. In addition the combined drug group ACE inhibitor and/or ARB will be analyzed.

The number of patients taking a medication in a drug group of special interest or combination of concomitant medications of special interest at any time during the study will be given. In addition, the number of patients taking a medication in a drug group of special interest at baseline and week 24 will be given. Medication start and stop date information will be used to assess, if a medication was taken at the respective visits. In addition, the tables with number of patients taking a medication in a drug group of special interest at baseline (Day 1) will be provided for the subgroups specified in Section 4.4.4.

A cross-tabulation comparing the number of patients taking a combination of concomitant medications of special interest at baseline 1 and the number of patients taking individual drug groups of interest at week 24 will be presented by treatment group.

Analysis of diuretics

An analysis of percentage of equivalence dose taken will be performed. Percentage of equivalence dose is calculated individually per subject and medication as:

Percentage of equivalence dose = actual dose / equivalence dose * 100

In case of missing dose and/or frequency information, the respective medication will be omitted from the analysis. Percentage of equivalence dose will be calculated for medications taken at baseline and week 24.

In case a subject takes several medications from one diuretics subcategory at a visit, the sum of the percentages of equivalence dose of the different medications will be used for the subject for the respective visit.

For each diuretics subcategory, summary statistics will be provided for percentages of equivalence dose as well as changes from baseline in percentages of equivalence dose by treatment group and overall.

In addition, patients will be grouped into 'below 50% of equivalence dose', '50% to 100% of equivalence dose', 'more than 100% of equivalence dose' and 'no respective medication intake'. For each diuretics subcategory, number and percentage of patients within each percentage group will be given by treatment group and overall at baseline and week 24.

Changes from baseline to week 24 will be categorized into the categories 'dose increased', 'dose not changed', 'dose reduced', 'new drug started', 'drug intake stopped'.

Table 6-1: Equivalence doses for different diuretics

Diuretic (ingredient)	Daily Equivalence* Dose (mg)
Loop diuretics	
Bumetanide	1
Furosemide	40
Torasemide	10
Thiazides	
Bendroflumethiazide	2.5
Hydrochlorothiazide	12.5
Indapamide	2.5
Metolazone	2.5
Potassium-sparing diuretics	
Amiloride	5# or 10†
Spironolactone/eplerenone	50# or 100†
Triamterene	100# or 200†

#in combination with ACEi and/or ARB. †without ACEi and ARB.
*Equivalence does not mean bio-equivalence.
Daily equivalence doses are taken from table 16 in the ESC HF Guideline 2012 [18].

6.1.6 Treatment Duration and Exposure

Treatment duration (number of days with study treatment intake, calculated based on the start and stop dates of tablet intake, not including interruptions of drug) will be summarized using descriptive statistics by treatment group and overall. In addition, treatment duration will be categorized to ≤7 days, >7-42 days, >42-84 days, >84-126 days, >126-168 days, and presented with the corresponding number and percentage of patients by treatment group and overall. A table will be presented with the absolute and relative frequencies of patients still in the study at each visit. Kaplan-Meier plots for time-to end of study treatment will be provided for each treatment group. The extent of exposure to study treatment (total amount of intake in mg) will be summarized using descriptive statistics and by 200-mg intervals by treatment group and overall.

Gaps as well as interruptions will not be counted towards duration calculations. Overlaps will be handled by setting the first record with overlap as the day before the start of the following record. The exposure will be computed as the total dose assumed record by multiplying the treatment days by the relative dose and summing them across days on treatment for each individual patient.

6.1.7 Treatment Compliance

The compliance will be summarized descriptively by treatment group and overall. In addition, compliance will be categorized into three groups (<80%, 80-120%, >120%) and summarized by treatment group and overall.

Compliance will be calculated as the number of tablets taken divided by the number of tablets dispensed * 100. In addition:

1) If the patient is missing return tablets or missing dispensing tablets then those tablets are not counted toward the number of tablets taken for the compliance calculation.
2) If the dispensing or return dates are missing but the number of return and dispensing tablets are present, the number of tablets taken will be considered for the compliance calculation.
3) Missing compliance will be noted as missing instead of 0.

6.1.8 Completer definition

A patient is defined as a completer if the following criteria is met:

- On the "End of Treatment" form page of the eCRF, the box corresponding to the question of "Did the subject complete treatment?" is checked off as "Yes."

6.2 Efficacy

All analyses will be performed according to ITT principle, i.e., including all patients randomized, if not otherwise specified (Section 5.1).

6.2.1 Patient Reported Outcomes

Kansas City Cardiomyopathy Questionnaire (KCCQ)

The Kansas City Cardiomyopathy Questionnaire (KCCQ) is a self-administered disease-specific measure comprised of 23 items grouped into seven domains (Physical Limitation, Symptom Stability, Symptom Frequency, Symptom Burden, Self Efficacy, Quality of Life and Social Limitations). The following five domains will be considered measures of treatment efficacy:

- Physical Limitation (PLS) (questions 1a to 1f),
- Symptom Frequency (SFS) (questions 3,5,7,9),
- Symptom Burden (question 4,6,8),
- Quality of Life (questions 12,13,14),
- Social limitations (questions 15a to 15d).

The following two KCCQ domains are not considered measures of treatment efficacy and are not included in algorithms for summary scores, these domains will be reported for completeness;

- Symptom Stability (question 2), referred to as 'change in symptoms' in the protocol, assesses patient perceived change in symptoms over previous two weeks,
- Self-Efficacy (question 10,11), referred to as 'self-efficacy and knowledge' in the protocol, is a quality of care metric assessing patient perceived knowledge of disease management.

For each item, options are on a 5- to 7-point Likert-type scale with varying response options depending on the question. Domain scores are calculated by summing item scores and then

transforming scores to a 0-100 unit scale with higher scores indicating better health status, as described in the KCCQ scoring instruction [10], reproduced in protocol appendix 8.

The individual domain scores will additionally be summarized to three summary scores:

- The Total Symptom Score (TSS) is the mean of the Symptom Burden and the Symptom Frequency scores.
- The Clinical Summary Score (CSS) is the mean of the Total Symptom score and the Physical Limitation score.
- The Overall Summary Score (OSS) is the mean of the Total Symptom score, and the Physical Limitation, Quality of Life, and Social Limitation scores.

The KCCQ will be administered at baseline (visit 2), 2 weeks (visit 3), 6 weeks (visit 5), 12 weeks (visit 7), 18 weeks (visit 8), 24 weeks (visit 9), and safety follow-up (week 28, visit 10); and following premature discontinuation: at time of discontinuation (visit 11) and at 28 +/-7 days following premature discontinuation (visit 12).

EQ-5D-5L

The EQ-5D-5L is a self-administered generic measure of health-related quality of life which includes five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Patients provide a rating for each question on a five-level Likert scale: having no problems, having slight problems, having moderate problems, having severe problems and being unable to do/having extreme problems. In addition, patients are asked to self-rate their own health today on a vertical 0-100 unit visual analogue scale (VAS), with 0 corresponding to "the worst health you can imagine", and 100 corresponding to "the best health you can imagine." For the EQ-5D-5L, summary scores will be calculated out of the five dimensions using the preference weighting derived by Craig et al from a representative US general population[19].

The EQ-5D-5L will be administered at baseline (visit 2), 2 weeks (visit 3), 6 weeks (visit 5), 12 weeks (visit 7), 18 weeks (visit 8), 24 weeks (visit 9), and at safety follow-up (week 28, visit 10); and following premature discontinuation: at time of discontinuation (visit 11) and at 28 +/-7 days following premature discontinuation (visit 12).

Patient Global Impression of Change (PGIC) and Patient Global Impression of Severity (PGIS) Questions

The PGIC contains two questions (i.e. Q1 and Q1a or Q1b):

1. Compared to the start of your treatment, how would you describe your physical limitations due to heart failure today?
    - Response options: “much better”, “better”, “a little better”, “the same”, “a little worse”, “worse”, or “much worse”;
- If a subject responds “much better”, “better”, or “a little better”, then he/she is asked:

- 1a. Was this improvement in your physical limitations an important change for you?
- Response options: “Yes”, “No”.

- If a subject responds “much worse”, “worse”, or “a little worse”, then he/she is asked:
  - 1b. Was this worsening in your physical limitations an important change for you (by important we mean did it bother you)?
  - Response options: “Yes”, “No”.

The PGIS contains 1 question:

- How would you rate your **physical limitations** due to heart failure today?
  - Response options: “No Limitations”, “Mild”, “Moderate”, “Severe”, “Very Severe”

These questions will be used to support interpretation of the KCCQ PLS mean change scores.

The PGIC will be administered at 2 weeks (visit 3), 6 weeks (visit 5), 12 weeks (visit 7), 18 weeks (visit 8), 24 weeks (visit 9), and also at safety follow-up (visit 10) and following premature discontinuation; at time of discontinuation (visit 11) and at 28 +/-7 days following premature discontinuation (visit 12).

The PGIS will be administered at the same timepoints as the PGIC but will be additionally administered at baseline (visit 2).

6.2.2 Clinician Reported Outcomes

Fried frailty score

The Fried frailty phenotype defines 5 key domains for frailty: weakness, low energy, slowed walking speed, decreased physical activity, and weight loss.

The questions patterned from the Fried Frailty score domains are assessed at baseline and at 24 weeks. The questions address five domains of the frailty phenotype: (a) shrinking (weight loss of $\geqslant$10 pounds in the prior year); (b) weakness (decreased or weakened grip strength); (c) exhaustion (fatigue or declining endurance); (d) slowness (slower walking pace); and (e) low activity (decline in physical activity). The 5 questions are as follows:

Within the last 12 months has the patient experienced any of the following?

- Unintentional weight loss ($\geq$ 5 kg/10 lbs).
- Developed decreased grip strength.
- Developed increasing fatigue/lethargy or declining endurance.
- Walks a distance of 5 m/15 feet at a slower pace.

- Decline in typical activity level.

The Fried frailty score will be completed at baseline (visit 2), 24 weeks (visit 9) and at time of premature discontinuation (visit 11).

Assessment of NYHA class

NYHA class will be assessed according to the classification below:

- Class I: No limitation of physical activity
- Class II: Slight limitation of physical activity in which ordinary physical activity leads to fatigue, palpitation, dyspnea, or anginal pain; the person is comfortable at rest
- Class III: Marked limitation of physical activity in which less-than-ordinary activity results in fatigue, palpitation, dyspnea, or anginal pain; the person is comfortable at rest
- Class IV: Inability to carry on any physical activity without discomfort but also symptoms of heart failure or the anginal syndrome even at rest, with increased discomfort if any physical activity is undertaken

The NYHA class will be completed at baseline and all following visits.

6.2.3 Primary efficacy variable

6.2.3.1 Primary Analysis of the Primary Efficacy Variable

The primary endpoint will be change in KCCQ PLS from baseline to week 24.

The estimand of interest is the de facto or treatment policy estimand.

Population: Defined through study inclusion/exclusion criteria (see protocol Section 6.1 and 6.2).

Variable: Change from baseline in KCCQ PLS from baseline to Week 24.

Approach of intervention effect: Regardless of stopping study treatment or adherence to study treatment.

Summary measure: Difference of LS means at Week 24.

The primary analysis will be performed on the FAS (KCCQ) (Section 5.1.2). Missing data will be imputed based on a pattern-mixture model with control-based pattern imputation (Appendix 9.1). The main idea of this method was introduced by Little and Yau (1996) [20] and later refined by Ratitch and O'Kelly (2011) [21]. In this control-based pattern imputation, PROC MI is called in such a way that it builds its imputation model only on data from the control arm, while it imputes missing data in both control and experimental treatment arms using a single control-based imputation model. In other words, the methods will be implemented using copy reference approach. For those patients who had died during the study, they will be imputed with the worst case outcome, i.e., a zero score from the time of death onward.

The analysis of each imputed dataset will be performed using mixed-effects model for repeated measures (MMRM), including treatment, region, heart rhythm, study visit as fixed effects, interaction between study visit and treatment group, and adjustment for the baseline PLS values as covariates. The model will assume an unstructured covariance matrix for within-patient variability.

The model statement is:

$$Y_{ijklm} = \mu + \beta x_i + t_k + r_l + h_m + v_j + (tv)_{jk} + \varepsilon_{ij}$$

where Y_{ijklm} is the change from baseline in PLS to visit j for patient i; μ is the intercept, β is the baseline covariate effect, x_i is the baseline PLS for patient i, t_k is the fixed effect of treatment k, r_l is the fixed effect of region l, h_m is the fixed effect of heart rhythm m (atrial fibrillation or sinus rhythm from ECG baseline), v_j is the fixed effect of visit j, $(tv)_{jk}$ is the interaction effect of treatment k by visit j, and $\varepsilon_{ij} \sim$ Normal(0, σ^2) represents the residual variance component with corr($\varepsilon_{ij}, \varepsilon_{ij'}$) = $\rho_{jj'}$, $j \neq j'$.

The core SAS-code to be used is the following:

```
proc mixed data=imputed_pls_data;
     by _imputation_;
     class subjidn  trt01pn cntygr1n ATRFIBN avisitn;
     model chg= trt01pn  cntygr1n ATRFIBN base avisitn avisitn*trt01pn /ddfm=kr outp=resid s
covb;
     repeated avisitn/ subject=subjidn type=un group=trt01pn;
          lsmeans  avisitn*trt01pn/e  cl diff;
     estimate 'Vericiguat10-Placebo'  trt01pn -1 1 0 ;
     estimate 'Vericiguat10-Placebo at Week 2' trt01pn -1 1 0 avisitn*trt01pn -1 0 0 0 0 1 0 0
0 0/e cl;
     estimate 'Vericiguat10-Placebo at Week 6' trt01pn -1 1 0 avisitn*trt01pn  0 -1 0 0 0 0 1
0 0 0/cl;
     estimate 'Vericiguat10-Placebo at Week 12' trt01pn -1 1 0 avisitn*trt01pn 0 0 -1 0 0 0 0
1 0 0/cl;
     estimate 'Vericiguat10-Placebo at Week 18' trt01pn -1 1 0 avisitn*trt01pn 0 0 0 -1 0 0 0
0 1 0/cl;
     estimate 'Vericiguat10-Placebo at Week 24' trt01pn -1 1 0 avisitn*trt01pn 0 0 0 0 -1 0 0
0 0 1/cl;

     estimate 'Vericiguat15-Placebo'  trt01pn -1 0 1;
     estimate 'Vericiguat15-Placebo at Week 2' trt01pn -1 0 1 avisitn*trt01pn -1 0 0 0 0 0 0 0
0 0 1 0 0 0 0/e cl;
     estimate 'Vericiguat15-Placebo at Week 6' trt01pn -1 0 1 avisitn*trt01pn  0 -1 0 0 0 0 0
0 0 0 0 1 0 0 0/cl;
     estimate 'Vericiguat15-Placebo at Week 12' trt01pn -1 0 1 avisitn*trt01pn 0 0 -1 0 0 0 0
0 0 0 0 0 1 0 0/cl;
     estimate 'Vericiguat15-Placebo at Week 18' trt01pn -1 0 1 avisitn*trt01pn 0 0 0 -1 0 0 0
0 0 0 0  0 0 1 0/cl;
     estimate 'Vericiguat15-Placebo at Week 24' trt01pn -1 0 1 avisitn*trt01pn 0 0 0 0 -1 0 0
0 0 0 0 0 0 0 1/cl;

RUN;
```

Results of the MMRM analysis on 100 separately imputed datasets can now be combined to derive an overall result. This is done by applying PROC MIANALYZE with the following generic code:

```
ods output parameterestimates=mi_lsmean;
proc mianalyze parms=lsmean;
    by trt01pn avisitn;
    modeleffects trt01pn*avisitn;
RUN;
```

Based on the responder definitions obtained from anchor-based analysis (Section 4.3.2), a responder is defined as a patient whose improvement from baseline in KCCQ PLS is greater than or equal to responder definition. The number of responders at each visit will be summarized and tested using a chi-square test. Note that missing data will not be imputed in the responder analysis. In addition, the number of patients with KCCQ PLS improvement from baseline by >5 points at 24 weeks and other thresholds (e.g. >3, >7, >10, >15, >20) will be summarized for each arm and assessed using a chi-square test. The number of patients with KCCQ PLS decline from baseline by >5 points and other thresholds (>3, >7, >10, >15, >20) at 24 weeks will be summarized by each arm and examined using a chi-square test as well. Further, PDF and CDF of change from baseline at 24 weeks for KCCQ PLS domain will be plotted to compare the proportion of patients who are meeting certain threshold from vericiguat and placebo arms. The Kolmogorov–Smirnov test will be performed to determine if the curves are significantly different.

6.2.3.2 Supportive Analysis of the Primary Efficacy Variables

A tipping point analysis will be performed as the sensitivity analysis under missing not at random (MNAR) assumption. Based on the primary analysis model, we will vary assumptions about the missing outcomes to assess how severe departures from the control-based imputation model it can be to overturn the statistical significance conclusions. The procedure will impute the missing data with observations only from placebo arm, and adjust the imputed values for observations in the vericiguat treatment arm for a specified sequence of shift parameters. Multiple imputation will be done using SAS PROC MI under missing not at random (MNAR) assumption (Appendix 9.2).

Also, the same primary efficacy model will be repeated on PEAS (Section 5.1.4) as a supportive efficacy analysis.

In addition, the primary analysis model will be applied to all assessments post-baseline prior to study treatment discontinuation on FAS and PPS.

Furthermore, the same primary analysis model will be repeated under missing at random (MAR) assumption, which means no missing data imputations will be performed.

6.2.4 Secondary efficacy variable

The secondary endpoint is change from baseline in the 6 MWT at 24 weeks. The estimand of interest is the de facto or treatment policy estimand.

Population: Defined through study inclusion/exclusion criteria (see protocol Section 6.1 and 6.2).

Variable: Change from baseline in 6 MWT from baseline to Week 24.

Approach of intervention effect: Regardless of stopping study treatment or adherence to study treatment.

Summary measure: Difference of LS means at Week 24.

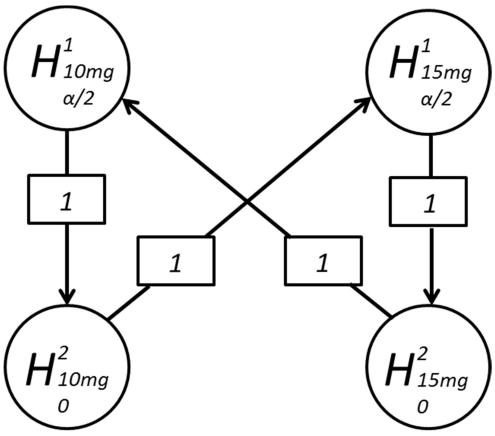
The secondary endpoint will be analyzed in the same manner as the primary endpoint and on the KCCQ (6MWT) (Section 5.1.3). Missing data will be imputed based on a pattern-mixture model with control-based pattern imputation (Appendix 9.1). For those patients who had died during the study, a score of 0 will be assumed from the time of death onward.

A supportive analysis will be repeated using the MMRM model without any missing data imputations. Also, the same model will be applied to all assessments post-baseline prior to study treatment termination.

6.2.5 Multiplicity adjustment

The study uses the graphical method of Maurer and Bretz [22] to provide strong multiplicity control across the primary KCCQ PLS hypotheses and the secondary 6 MWT hypotheses. The overall two-sided alpha level will be controlled at 0.05. The two primary hypotheses will be tested using the Bonferroni procedure, each at a splitted level of alpha 0.025. The two secondary hypotheses will be tested based on a parallel gatekeeping principle. As shown in Figure 6-1, if the test of both primary and secondary endpoints at one of the dose groups is significant at the alpha level of 0.025, the primary and secondary endpoint at the other dose group can be tested at the alpha level of 0.05, in an ordered manner. Figure 6-1 shows the α-allocation for each hypothesis in the circle representing the hypothesis.

Figure 6-1 α-allocation for the primary and secondary hypotheses



Overall, the primary analysis will be claimed to have a positive outcome if the null hypothesis for at least one of the two primary hypotheses can be rejected.

6.2.6 Other KCCQ domains

Change from baseline in other KCCQ domains (Symptom Frequency, Symptom Burden, Social limitation, Quality of Life), KCCQ summary scores (TSS, CSS and OSS) will be summarized by visit. Cumulative distribution functions will be plotted to compare the proportion of patients who are meeting certain thresholds from vericiguat and placebo arms. The variables will be analyzed using the MMRM model with no missing data imputation. The proportion of patients with changes from baseline by >5 points at 24 weeks and other thresholds (e.g. >3, >7, >10, >15, >20) in the other KCCQ domains as well as OSS, CSS, TSS and SFS will be summarized. The proportion of patients with decline from baseline by >5 points at 24 weeks, as well as other thresholds (e.g. >3, >7, >10, >15, >20) in the other KCCQ domains and summary scores will also be summarized.

6.2.7 EQ-5D-5L

Frequencies of answers to individual questions will be displayed by treatment group and overall by visit. In addition, changes from baseline for the single questions will be classified into the following categories: Improvement by two or greater Likert points, Improvement by one Likert point, No change, Worsening by one Likert point, Worsening by two or greater Likert points. Frequencies of the different categories will be displayed by assigned treatment group and overall by visit.

The EQ-5D Visual Analogue Scale (VAS), the preference weighted summary score and the corresponding changes from baseline will be described by treatment group and visit.

The summary scores of EQ-5D-5L (UK and US) as well as the EQ Visual Analogue Scale (VAS) and the corresponding changes from baseline will be described by treatment group and visit.

6.2.8 Fried frailty score

Frequencies of patients with each response option for each question will be presented by treatment group and overall by visit. Proportions of patients with one or more responses characteristic of frailty and those with three or more characteristics will be summarized.

6.2.9 Clinical outcome events

All patients will be followed until study completion to assess for vital status. Potential pre-specified clinical outcome events will be submitted for adjudication to an independent CEC. The following events will be collected and analyzed to explore the efficacy of study drug treatments on clinical outcomes:

- All deaths (CV and non-CV, including outpatient death)
- CV hospitalizations including HF and non-HF hospitalizations (i.e. myocardial infarction and stroke)
- Outpatient HF events, such as IV diuretic use for HF or urgent HF visits

Incidences of events, including all deaths, CV deaths and non-CV deaths, total CV hospitalizations, HF hospitalizations, non-HF hospitalizations (i.e. myocardial infarction and

stroke), as well as outpatient HF events will be summarized. Patients without an event will be censored at their last contact date (the latest date among all known assessments or follow-up dates).

In addition, Kaplan Meier curves of the composite endpoints will be presented by treatment:

- Time to first occurrence of the composite of CV death or heart failure hospitalization.
- Time to first occurrence of the composite of CV death, HF hospitalization, worsening from baseline in NYHA class, or worsening from baseline in PGIC
- Time to first occurrence of the composite of CV death, heart failure hospitalization, worsening from baseline in NYHA class, or worsening from baseline in KCCQ CSS

6.2.10 NYHA Class

Frequencies of patients in the different NYHA classes will be presented by treatment group and overall by visit.

Frequencies of patients with change in NYHA class from baseline (e.g. unchanged, improved by 1 category, deteriorated by 1 category, etc.) will be presented by treatment group and visit.

The number of patients with transitions from baseline with respect to categories (class I, class II, class III, class IV) will be provided by baseline value, treatment group and visit.

6.2.11 PGIC and PGIS

Frequencies of patients in the different PGIC/S categories will be presented by treatment group and overall by visit.

Frequencies of patients with change in PGIC/S class from baseline (e.g. unchanged, improved by x categories, deteriorated by x categories, etc.) will be presented by treatment group and visit.

6.3 Safety

Safety analyses will be performed in the SAF.

6.3.1 Adverse events

Adverse events (AEs) will be coded using MedDRA terminology. The version number of MedDRA used for the analyses will be stored in the clinical database. A listing will be provided linking the original investigator terms and the coded terms.

AEs are considered to be treatment-emergent if they have started or worsened after first application of study medication up to 5 calendar days after end of treatment with study medication.

An overall summary of AEs and treatment-emergent (TE) AEs will be generated by treatment group and overall.

Incidences of patients with TEAEs, drug-related and/or serious TEAEs, TEAEs causing discontinuation of study treatment will be summarized by treatment group and overall

grouped by MedDRA Primary System Organ Class (SOC) and Preferred Term (PT). In addition, the incidence of pre-treatment AEs and AEs during the follow-up (more than 5 calendar days after end of treatment with study medication) will be tabulated. Serious adverse events (SAEs), deaths, and AEs leading to discontinuation will be listed. The date, relative day (to study medication) and phase of the study (pre-treatment, during treatment, post-treatment) will be included.

6.3.2 Adverse events of special safety interest

For this trial, the following adverse events are considered of special interest:

- Symptomatic hypotension events (Bayer SMQ "Hypotension (Riociguat)" (code=SMQ_1388)). Hypotension events will be categorized in symptomatic/asymptomatic as captured in the eCRF.
- Syncope (Bayer SMQ "Syncope Riociguat"(code=SMQ_1389))

For each of the SMQs, the following tables will be provided: overall summary of TEAEs, incidences of patients with TEAEs, and with serious TEAEs grouped by SOC and PT.

A listing of all hypotension and syncope events will be provided, including date, relative day (to study medication), dose group, preferred term and phase of the study (pre-treatment, during treatment, post-treatment).

6.3.3 Further safety parameters

6.3.3.1 Laboratory parameters

Summary statistics including changes from baseline will be calculated by treatment group and visit for all quantitative laboratory parameters. The summary statistics for troponin T and eGFR will also be provided by treatment group.

The number of patients with transitions from baseline with respect to reference range categories (low, normal, high) will be provided by treatment group and visit. In addition, the number of patients with treatment-emergent abnormal laboratory values above or below the normal range will be tabulated by treatment group and overall.

To assess acute renal failure, the number and incidence of patients with increase in serum creatinine >0.3 mg/dL, >0.5 mg/dL, and ≥ ×2 from baseline, as well as the number and incidence of patients with decrease in eGFR >50% or to <15 mL/min/1.73m² will be presented by visit and treatment group.

The number and incidence of patients with ALT > 3× upper limit of normal (ULN), and total bilirubin > 2×ULN will be presented by visit and treatment group.

6.3.3.2 Vital signs

The absolute values and the changes from baseline values at each visit will be summarized by treatment group using descriptive statistics for systolic blood pressure (SBP), diastolic blood pressure (DBP), heart rate (HR).

The change in vital signs parameters from baseline to week 24 will be analyzed by two-sided two-sample t-tests comparing each individual vericiguat treatment group with the placebo treatment group at the 5% level in an explorative setting.

The number of patients with systolic blood pressure below 90 mmHg, based on the individual measurements rather than the mean value, will be displayed by visit and by treatment group. A frequency listing of all patients with systolic blood pressure below 90 mmHg post baseline will be provided.

6.3.3.3 Electrocardiogram (ECG)

The incidence rates of treatment-emergent ECG abnormalities will be tabulated by treatment group. A descriptive analysis of continuous ECG parameters and their changes from baseline by visit and treatment group will also be presented.

Prolongations of QT interval will be summarized by visit and treatment group based on the PKS. At every time point the 90% Confidence Intervals (CI) for the mean change from baseline in QTcF will be calculated.

7. Document history and changes in the planned statistical analysis

7.1 Document history

- Version 1.0
- Version 2.0
    - Added updates per the FDA Type C meeting held on 11/6/2019
    - Added definition of completer
    - Updated definition of baseline, treatment compliance calculation , and treatment duration calculation
    - Updated with latest SAS code
    - Included imputation rules for partially missing end dates for study medication

7.1.1 Changes from the protocol

None

8. References

1. Steinberg BA, Zhao X, Heidenreich PA, et al. Trends in patients hospitalized with heart failure and preserved left ventricular ejection fraction: prevalence, therapies, and outcomes. Circulation. 2012; 126 (1): 65-75.
2. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC)Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur Heart J. 2016 Jul 14;37(27):2129-200.
3. Shah SJ, Gheorghiade M. Heart failure with preserved ejection fraction: treat now by treating comorbidities. JAMA. 2008; 300 (4): 431-3.
4. Borlaug BA, Redfield MM. Diastolic and systolic heart failure are distinct phenotypes within the heart failure spectrum. Circulation. 2011; 123 (18): 2006-13; discussion 2014.
5. Pieske B, Maggioni AP, Lam CSP, Pieske-Kraigher E, Filippatos G, Butler J, et al. Vericiguat in patients with worsening chronic heart failure and preserved ejection fraction: results of the SOluble guanylate Cyclase stimulatoR in heArT failurE patientS with PRESERVED EF (SOCRATES-PRESERVED) study. Eur Heart J. 2017 Apr 14;38(15):1119-27.
6. Filippatos G, Maggioni AP, Lam CSP, Pieske-Kraigher E, Butler J, Spertus J, et al. Patient-reported outcomes in the SOluble guanylate Cyclase stimulatoR in heArT failurE patientS with PRESERVED ejection fraction (SOCRATES-PRESERVED) study. Eur J Heart Fail. 2017 Jun;19(6):782-91.
7. Green CP, Porter CB, Bresnahan DR, Spertus JA. Development and evaluation of the Kansas City Cardiomyopathy Questionnaire: a new health status measure for heart failure. J Am Coll Cardiol. 2000 Apr;35(5):1245-55.
8. Butler J, Hamo CE, Udelson JE, Pitt B, Yancy C, Shah SJ, et al. Exploring New Endpoints for Patients With Heart Failure With Preserved Ejection Fraction. Circ Heart Fail. 2016; 9(11).
9. SAS Institute Inc. SAS/STAT® 9.2 User's Guide. Cary, NC: SAS Institute Inc. 2008.
10. The Kansas City Cardiomyopathy Questionnaire Scoring Instructions (2000-2003), John Spertus MD MPH [http://cvoutcomes.org/ (6th September 2013)]
11. Spertus J, Peterson E, Conard MW, et al. Monitoring clinical changes in patients with heart failure: a comparison of methods. Am Heart J 2005;150:707 - 15.
12. Dreyer RP, Jones PG, Kutty S, Spertus J. Quantifying clinical change: discrepancies between patients' and providers' perspectives. Qual Life Res. 2016; 25: 2213-2220.

13. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol. 2008; 61(2), 102–109.
14. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin. 1979; 86 (2), 420-428.
15. Haywood KL, Garratt AM, Dziedzic K, Dawes PT. Generic measures of health-related quality of life in ankylosing spondylitis: reliability, validity and responsiveness. Rheumatology. 2002; 41: 1380-1387.
16. Coon CD, Cook KF. Moving from significance to real-world meaning: methods for interpreting change in clinical outcome assessment scores. Qual Life Res. 2018; 27: 33-40.
17. Gerlinger C, Schmelter T. Determining the non-inferiority margin for patient reported outcomes. Pharm Stat. 2011;10(5):410-3.
18. McMurray JJV, Adamopoulos S, Anker SD, et al. ESC Guideline for the diagnosis and treatment of acute and chronic heart failure 2012. The Task Force for the Diagnosis and Treament of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. European Heart Journal. 2012; 33:1787-1847.
19. Craig BM, Rand K. Choice Defines QALYs: A US Valuation of the EQ-5D-5L. Medical Care. 2018; 56(6):529-536
20. Little, R, Yau.L. Intent-to-Treat Analysis for Longitudinal Studies with Drop-Outs. Biometrics. 1996; 52: 1324-1333.
21. Ratitch, B, O'Kelly, M. Implementation of Pattern-Mixture Models Using Standard SAS/STAT Procedures. Proceedings of PharmaSUG2011 Conference – Paper SP04. http://www.pharmasug.org/proceedings/2011/SP/PharmaSUG-2011-SP04.pdf
22. Maurer W, Bretz F. Multiple Testing in Group Sequential Trials Using Graphical Approaches. Statistics in Biopharmaceutical Research. 2013;5(4):311-20.

9. Appendix

9.1 Pattern-mixture modeling approach

The main idea of this method was introduced by Little and Yau (1996) [20] and later refined by Ratitch and O'Kelly (2011) [21]. The Ratitch and O'Kelly (2011) approach suggests using sequential regression and multiple imputation methodology to impute missing values after a subject's discontinuation from the trial based on "as treated" model, using actual dose after drop-out if it is known, or based on some plausible assumptions if unknown. In most clinical trials, patients stop taking experimental medication after discontinuation. In this case, the "as treated" model for discontinued patients would be based on the idea that patients are taking a zero dose of the experimental treatment.

In this control-based pattern imputation, the intent is to make no direct use of observed data from the experimental treatment arm for estimating the imputation model. Hence, PROC MI is called in such a way that it builds its imputation model only on data from the control arm, while it imputes missing data in both control and experimental treatment arms using a single control-based imputation model. This is achieved with a sequence of calls to PROC MI.

The general procedure for control-based pattern imputation is the following:

(a) Impute the non-monotone data with the MCMC method and store the result of this partial imputation. In this study, the non-monotone data would represent those patients that discontinued study treatment prematurely, but was able to comeback at a later visit for assessment (e.g., dropped out at week 2 and missed assessments for week 6, then came back and completed week 12 visit).

(b) Next, impute the monotone data with control based pattern imputation. Here it implies that the patients who were in the treatment group and no longer received the treatment will assume the same conditional distribution as those in the control group. We start with imputing missing data at the first time-point that has some missing data (defined as time-point 1).

(c) Repeat (b) for all other time-points sequentially. Patients whose missing values were imputed in the last call to PROC MI will be included in the input dataset for the next call to PROC MI. Thus data for time-point t, filled in during the last call, will be used for predictor variables in the next call to PROC MI (for time point t + 1), which is consistent with the internal workings of a single call to PROC MI to impute all time-points automatically.

This procedure is illustrated below by the following example code:

```
proc mi data = DATAIN out = DATAIN_MONO nimpute =100  seed = 123;
    var SCORE_0 SCORE_1 SCORE_2 SCORE_3 SCORE_4 SCORE_5;
    mcmc chain = multiple impute = monotone;
run;

proc mi data=DATAIN_Mono out = DATAIN_REG_IMP1 seed=123 nimpute=1;
       class Trt;
       monotone reg ;
```

```
       mnar model( SCORE_1 SCORE_2 SCORE_3 SCORE_4 SCORE_5 / modelobs= (Trt='0'));
       var SCORE_0 SCORE_1 SCORE_2 SCORE_3 SCORE_4 SCORE_5 ;
run;
```

where SCORE_0 , SCORE_1, SCORE_2, SCORE_3, SCORE_4, SCORE_5 represent KCCQ PLS scores at baseline and post-baseline visits week 2, 6, 12, 18 and 24.

The same procedure is then repeated until all time-points are imputed.

9.2 Tipping point analysis approach

We intend to perform multiple imputation analysis for a specified sequence of shift parameters, which only adjust the imputed values for observations in the treatment group (TRT=1). Multiple imputation will be done using SAS PROC MI under MNAR assumption using the following generic code:

```
proc mi data=PLS_data seed=19334 nimpute=1 out=PLS_mi;
       class Trt;
       monotone reg (/details);
       mnar model( SCORE_1 SCORE_2 SCORE_3 SCORE_4 SCORE_5 / modelobs= (Trt='0'))
            adjust (SCORE_1 SCORE_2 SCORE_3 SCORE_4 SCORE_5 /shift=sj adjustobs=(Trt='1'));
       var SCORE_0 SCORE_1 SCORE_2 SCORE_3 SCORE_4 SCORE_5;
run;
```

Each imputed observation (missing KCCQ PLS scores) will be modified according to the required delta shift. Assuming there is no penalty in KCCQ PLS score in the placebo group, then for those with missing outcomes in vericiguat treatment arm, we decrease the score gradually until a p-value of 0.025 or higher is produced (beyond the tipping point). The actual penalty (to 1 decimal place) that gives a p-value above and closest to 0.025 will then be identified.

9.3 Region Definition

CNTYGR1 **Pooled Country/Geographic Region 1**	**COUNTRY** **Country**
Americas	ARGENTINA
Americas	CANADA
Americas	COLOMBIA
Americas	UNITED STATES
Asia/Pacific	JAPAN

Asia/Pacific	MALAYSIA
Asia/Pacific	SINGAPORE
Asia/Pacific	TAIWAN (PROVINCE OF CHINA)
Europe (incl. Israel and South Africa)	AUSTRIA
Europe (incl. Israel and South Africa)	BELGIUM
Europe (incl. Israel and South Africa)	BULGARIA
Europe (incl. Israel and South Africa)	GERMANY
Europe (incl. Israel and South Africa)	SPAIN
Europe (incl. Israel and South Africa)	GREECE
Europe (incl. Israel and South Africa)	HUNGARY
Europe (incl. Israel and South Africa)	ISRAEL
Europe (incl. Israel and South Africa)	ITALY
Europe (incl. Israel and South Africa)	POLAND
Europe (incl. Israel and South Africa)	PORTUGAL
Europe (incl. Israel and South Africa)	RUSSIAN FEDERATION
Europe (incl. Israel and South Africa)	SOUTH AFRICA