Statistical Analysis Plan	
Drug Substance	Verinurad
Study Code	D5496C00005
Edition Number	4.0
Date	02 June 2022

A Phase 2, Multicentre, Double-Blind, Placebo and Active Control Efficacy and Safety Study to Evaluate Verinurad combined with Allopurinol in Heart Failure with Preserved Ejection Fraction (AMETHYST)

SIGNATURE PAGE

A Phase 2, Multicentre, Double-Blind, Placebo and Active Control Efficacy and Safety Study to Evaluate Verinurad combined with Allopurinol in Heart Failure with Preserved Ejection Fraction (AMETHYST)

AstraZeneca Global
Product Statistician

_____ _____

Redacted Date

ClinChoice Inc.
Study Statistician

_____ _____

Redacted Date

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
AE	Adverse event
ANCOVA	Analysis of covariance
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical
AZ	AstraZeneca
CC	Complete case
CFB	Change from baseline
CI	Confidence interval
CPET	Cardiopulmonary exercise test
CSP	Clinical study protocol
CSR	Clinical study report
CSS	Clinical summary score
CV	Cardiovascular
DAE	Discontinuation of investigational product due to adverse event
DBP	Diastolic blood pressure
DRMI	Drop-out reason-based multiple imputation
ECG	Electrocardiogram
eCRF	Electronic case report form
PRO	patient reported outcome
FAS	Full analysis set
HF	Heart failure
HFpEF	Heart failure with preserved ejection fraction
ICF	Informed consent form
IDMC	Independent data monitoring committee
IP	Investigational product
IPD	Important protocol deviation
ITT	Intent-to-treat
KCCQ	Kansas City Cardiomyopathy Questionnaire
LLOQ	Lower limit of quantification
LSMD	Least squares means difference
LSMEANS	Least squares means

Abbreviation or special term	Explanation
MAR	Missing at random
MCP	Multiple comparisons procedure
MedDRA	Medical dictionary for regulatory activities
MI	Multiple imputation
MMRM	Mixed model for repeated measures
OSS	Overall summary score
PD	Protocol deviation
PGIC	Patient Global Impression of Change
PGIS	Patient Global Impression of Severity
C[redacted]	[redacted]
PRO	Patient reported outcome
PT	Preferred term
SAE	Serious adverse event
SAP	Statistical analysis plan
SBP	Systolic blood pressure
SD	Standard deviation
SI	System international (units)
SOC	System organ class
sUA	Serum uric acid
TBL	Total bilirubin
TSS	Total symptom score
CCI [redacted]	[redacted]
ULN	Upper limit of normal
VO_2	Oxygen volume

AMENDMENT HISTORY

Category*: Change refers to	Date	Description of change	In line with the CSP?	Rationale
Other	08 July 2021	Removed 'Three Arm' from title.	Yes	Implementation of Protocol Version 3.0. Patients randomised after implementation are assigned to one of 4 treatment arms as part of modified study design.
Other	08 July 2021	Updated protocol version number.	Yes	New Protocol Version 3.0.
Statistical analysis method for the primary or secondary endpoints	08 July 2021	Updated study design to include new planned analysis of pooled data from the 24 mg verinurad + allopurinol and 12 mg verinurad + allopurinol treatment arms.	Yes	Protocol Version 3.0 includes an additional treatment arm.
Other	08 July 2021	Re-labelled study schema figure as Part (A) and inserted new study schema as Part (B).	Yes	Schema dependent on patient randomisation prior to or after implementation of Protocol Version 3.0.
Other	08 July 2021	Added additional dosing schedule for the 24 mg verinurad + allopurinol arm.	Yes	Protocol Version 3.0 includes an additional treatment arm.
Other	08 July 2021	Updated to include additional treatment arm (24 mg verinurad in combination with allopurinol) and 1:1:1:1 randomisation ratio.	Yes	Protocol Version 3.0 includes an additional treatment arm.
Statistical analysis method for the primary or secondary endpoints	08 July 2021	Added new sensitivity analyses of the primary and secondary analyses.	Yes	Protocol Version 3.0 includes an additional sensitivity analysis.
Statistical analysis method for the primary or secondary endpoints	08 July 2021	Added distribution-based analysis for estimating group-level minimal clinically important difference analyses for primary and secondary endpoints.	Yes	Protocol Version 3.0 includes additional distribution-based analyses.
Other	08 July 2021	Removed safety analyses text in Section 7 (Changes of Analysis from Protocol)	Yes	SAP and Protocol Version 3.0 are in alignment with the definition of the Safety Analysis Set.

Other	08 July 2021	Removed DECT/PET endpoints	No	Substudy cancelled.
Primary or secondary endpoints	08 July 2021	Changed two of the sensitivity analyses to supplemental analyses	Yes	ICH9 addendum clarification of differences between sensitivity and supplemental analyses.
Statistical analysis method for the primary or secondary endpoints	08 July 2021	Added protocol version to analysis covariates	Yes	Accounting for potential differences due to protocol versions.
CCI	[redacted]	[redacted]	[redacted]	[redacted]
Other	19 November 2021	Removed 24 mg verinurad arm, interim analysis, and references to DECT and PET/CT.	Yes	Updated based on CSP version 4.0
Other	02 June 2022	Removed AEs of special interest and MACE. Updated for edits and clarifications. Update treatment compliance definition.	Yes	AEs of special interest and MACE not needed. Minor edits. Treatment compliance dependent on visit.

* Pre-specified categories are
Primary or secondary endpoints; Statistical analysis method for the primary or secondary endpoints; Derivation of primary or secondary endpoints; Multiple Testing Procedure; Data presentations; Other

1 STUDY DETAILS

This statistical analysis plan (SAP) has been developed based on the AstraZeneca clinical study protocol (CSP) D5496C00005 version 4.0 (September 20, 2021) and outlines the analyses to be generated for the global clinical study report (CSR). Additional analyses required for regional submissions will be prespecified in a separate analysis plan and will be submitted to the appropriate authorities.

1.1 Study objectives

1.1.1 Primary objective

Primary objective:	Endpoint/variable:
To assess effect of verinurad + allopurinol compared to placebo on exercise capacity	Change from baseline at Week 32 in peak VO_2 consumption

VO_2 Oxygen volume.

1.1.2 Secondary objectives

Secondary objective:	Endpoint/variable:
To assess effect of verinurad + allopurinol compared to allopurinol monotherapy on exercise capacity	Change from baseline at Week 32 in peak VO_2 consumption
To assess effect of verinurad + allopurinol compared to placebo and compared to allopurinol monotherapy on Kansas City Cardiomyopathy Questionnaire (KCCQ)-Total Symptom Score (TSS)	Change from baseline at Week 32 in KCCQ-TSS

VO_2 Oxygen volume.

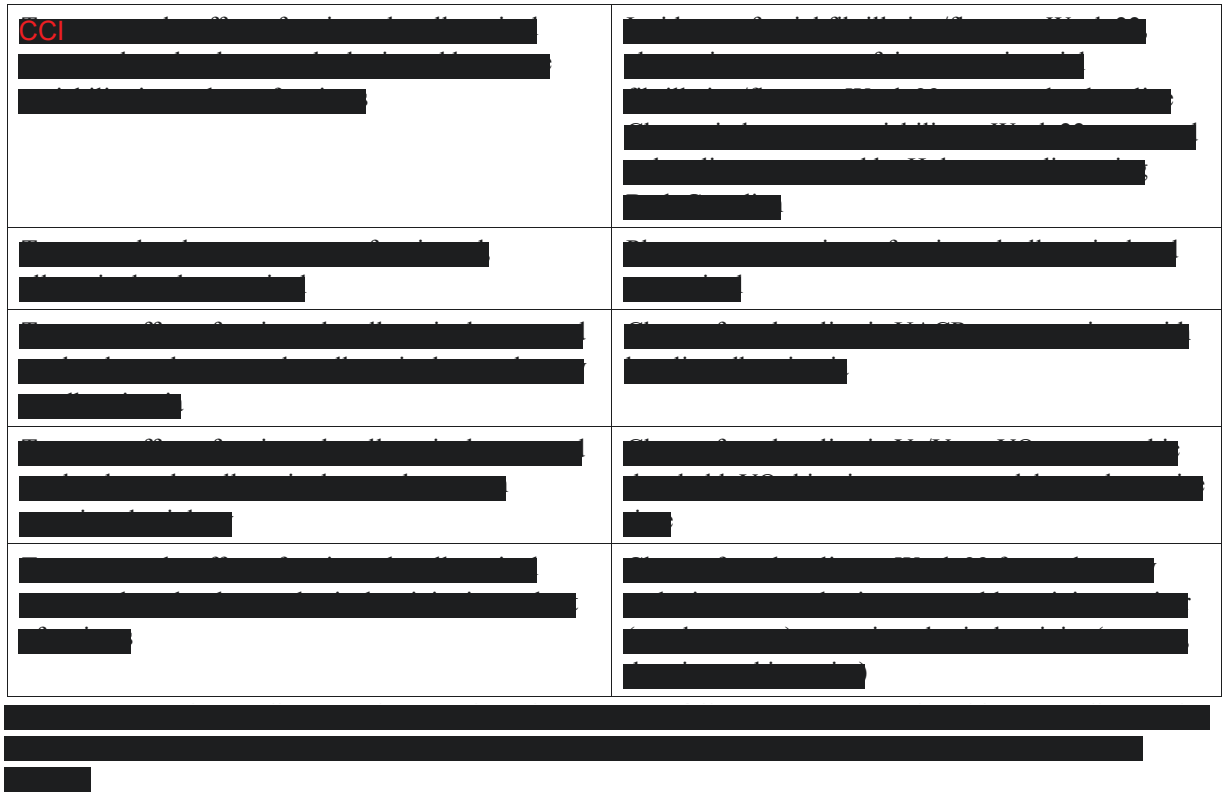
1.1.3 Safety objective

Safety Objective:	Endpoint/Variable:
To assess the safety and tolerability of verinurad + allopurinol as compared to placebo and to allopurinol in patients with HFpEF.	Safety and tolerability will be evaluated in terms of AEs, vital signs, clinical laboratory, and ECG. Assessments related to AEs cover • Occurrence • Relationship to study treatment as assessed by investigator • Intensity • Seriousness • Death • AEs leading to discontinuation of study treatment • Other action taken related to study treatment, including dose interruptions Vital signs parameters include systolic and diastolic blood pressure, pulse, temperature.

AE Adverse event. ECG Electrocardiogram. HFpEF Heart failure with preserved ejection fraction.

CCI [redacted]

[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]



1.2 Study design

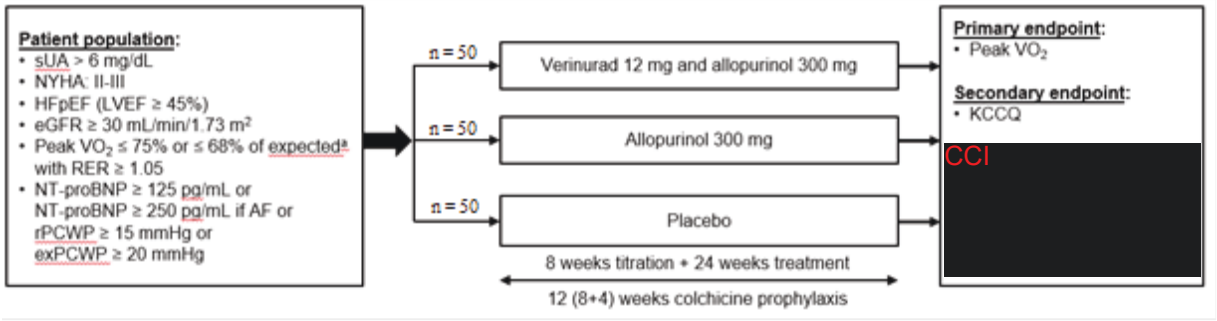
This is a randomised, multicentre, international, double-blind, parallel arm, active- and placebo-controlled, Phase 2 study to assess safety and efficacy of 12 mg verinurad in combination with 300 mg allopurinol on functional endpoints in patients with hyperuricaemia (sUA > 6 mg/dL) and HFpEF (left ventricular ejection fraction (LVEF) ≥ 45%).

Refer to the CSP sections 5.1 and 5.2 for inclusion/exclusion criteria.

Approximately 150 patients (approximately 50 per arm) will be randomised at around 70 sites in approximately 12 countries. With an estimated screen-failure rate of 65%, up to 500 patients should be enrolled.

The general study schema is summarised in Figure 1. See the CSP Section 1.1 (SoA) for a detailed list of visits and assessments.

Figure 1 Study Schema



[a] Based on Fletcher et al 1995

AF Atrial fibrillation. eGRF Estimated glomerular filtration rate. exPCWP Pulmonary capillary wedge pressure during exercise. HFpEF Heart failure with preserved ejection fraction. KCCQ Kansas City Cardiomyopathy Questionnaire. LVEF Left ventricular ejection fraction. NT-proBNP N-terminal pro b-type natriuretic peptide. NYHA New York Heart Association. RER Respiratory exchange ratio. rPCWP Pulmonary capillary wedge pressure during rest. sUA Serum uric acid. VO_2 Volume oxygen.

Patients who meet the eligibility criteria will be randomised in a 1:1:1 ratio to 12 mg verinurad plus 300 mg allopurinol, 300 mg allopurinol, or placebo. There will be matching placebo for both the verinurad and allopurinol. Titration is done in discrete steps of 4 weeks to reduce the risk of adverse reactions due to allopurinol. Patients will be followed until they complete 8 weeks of titration followed by 24 weeks of treatment at target dose plus 4 weeks of follow up. The allopurinol/placebo tablet and verinurad/placebo capsule will be taken orally in combination once a day. The primary endpoint will be analysed at Week 32.

Patients unable to tolerate the stepped dosage may be down-titrated only by reversing the assigned steps within treatment group. Verinurad and allopurinol cannot be unpaired from the titration schedule. Patients who cannot tolerate the step-1 dose will be discontinued from study treatment and be followed for the remainder of the study.

The dosing schedule is summarised in Table 1 below.

Table 1 Dosing Schedule

Treatment Arm	Week 0-3 Step 1 - Titration	Week 4-7 Step 2 - Titration	Week 8-12 Step 3 – Target Dose	Week 13-32 Target Dose
12 mg verinurad + allopurinol	Colchicine prophylaxis + 3 mg verinurad + 100 mg allopurinol	Colchicine prophylaxis + 7.5 mg verinurad + 200 mg allopurinol	Colchicine prophylaxis + 12 mg verinurad + 300 mg allopurinol	12 mg verinurad + 300 mg allopurinol
Allopurinol	Colchicine prophylaxis + 100 mg allopurinol	Colchicine prophylaxis + 200 mg allopurinol	Colchicine prophylaxis + 300 mg allopurinol	300 mg allopurinol
Placebo	Colchicine prophylaxis + placebo	Colchicine prophylaxis + placebo	Colchicine prophylaxis + placebo	Placebo

1.3 Number of subjects

The study will employ a 1:1:1 randomisation ratio between 12 mg verinurad + allopurinol, allopurinol, and placebo. Based on this design with 50 patients randomised to the verinurad + allopurinol group and to the placebo group respectively, and assuming a standard deviation for the primary endpoint of 2 mL/kg/min, the width of 95% confidence intervals for estimated differences between treatments will be about 1.57 ml/kg/min. The minimum detectable treatment difference for statistical significance in a two-group t-test with a two-sided significance level of 5% given the assumptions above is 0.794 mL/kg/min.

The assumed standard deviation of 2 mL/kg/min is derived from Borlaug (Borlaug et al 2018) and from Redfield (Redfield et al 2013).

2 ANALYSIS SETS

2.1 Definition of analysis sets

Four analysis sets are defined below: Enrolled, full analysis set (FAS), safety analysis set, CCI [redacted]. If no signed informed consent is collected (important protocol deviation), then the patient will be excluded from all analysis sets defined below.

2.1.1 Enrolled set

This analysis set comprises all patients who sign the main informed consent form (ICF) and will be used for reporting of disposition and screening failures.

2.1.2 Full analysis set

All patients who have been randomised to study treatment will be included in the FAS, irrespective of their protocol adherence and continued participation in the study. Patients will be analysed according to their randomised study treatment assignment, irrespective of the treatment actually received. Data from patients who withdraw consent will be included up to the date of their study termination.

The FAS will be considered the primary analysis set for the primary and secondary variables and for the exploratory efficacy variables.

All efficacy analyses will be performed using an intent-to-treat (ITT) approach based on the FAS. For consistency, demographic and baseline characteristics will be presented using the FAS.

2.1.3 Safety analysis set

The Safety analysis set consists of all patients who have received at least 1 dose of study treatment. Patients will be analysed according to their randomised study treatment assignment, irrespective of the treatment actually received.

All safety summaries will be based on this analysis set.

CCI [redacted]

2.2 Violations and deviations

2.2.1 Important protocol deviations

Important protocol deviations (IPDs) are a subset of protocol deviations (PDs) that may significantly impact the completeness, accuracy, and/or reliability of the study data or that may significantly affect a patient's rights, safety, or well-being.

The final list of IPDs will be documented prior to unblinding the study data and will include but may not be limited to the following deviations:

- Patient was randomised but did not meet inclusion criteria
- Patient was randomised but met exclusion criteria
- Patient received/took the wrong study treatment or incorrect dose or took medication from only one bottle at any time of the study period

- Patient received any prohibited medication from those listed in Table 6 of the CSP
- Patient developed discontinuation criteria but continued on investigational product (IP)

IPDs will be summarised by the highest level deviation category and will be based on the FAS. IPDs or PDs related to the ongoing and emerging novel coronavirus (COVID-19) will be summarised for the FAS population overall and by randomised treatment group. Patients with IPDs or PDs related to COVID-19 will be listed.

IPDs will not be used to exclude any patient from any analysis set, nor to exclude any data from patients included in an analysis set.

3 PRIMARY AND SECONDARY VARIABLES

3.1 General definitions

3.1.1 Visit window definitions

For endpoints that present visit-based data, the variables will be summarised based on the scheduled days with analysis visit windows. The analysis visit windows will be based on the collection schedule listed in the protocol and variables will be windowed to the closest scheduled visit for that variable.

Analysis visit windows have been constructed so that every observation collected can be allocated to a particular visit, including unscheduled assessments. No analysis visit windows will be defined for screening visits.

The window for the analysis visits following baseline will be constructed in such a way that the upper limit of the interval falls halfway between the two visits (the lower limit of the first post-baseline visit will be Day 2). Analysis visit will be mapped based on study day windows in Table 2 below.

Table 2 Visit Windows

Visit Name	Visit Study Day	Study Days (Visit Window)
Week 4 (Visit 4)	Day 28	2 to 42
Week 8 (Visit 5)	Day 56	43 to 70
Week 12 (Visit 6)	Day 84	71 to 119
Week 22 (Visit7)	Day 154	120 to 189
Week 32 (Visit 8)	Day 224	190 to 230
Week 36 (Visit 9)	Day 252	231 to end of study

If multiple readings are recorded within a single visit window, the rules below will be applied.

- If there are 2 or more observations within the same visit window, then the non-missing observation closest to the scheduled visit will be used in the analysis.
- If 2 observations are equidistant from the scheduled visit, then the non-missing observation with the earlier collection date will be used in the analysis.
- If 2 or more observations are collected on the same day, then the non-missing observation with the earlier collection time will be included in the analysis. If any observations' collection time is missing the mean of the observations will be used in the analysis.

If a visit window does not contain any observations, then the data will remain missing.

For the patient-report questionnaires collected by electronic patient reported outcomes (ePRO), the following rules will be applied:

For overall analyses not based on any particular study visit (eg, maximum post-baseline value), all data will be listed and/or analysed, including any repeat or unscheduled visits, unless otherwise specified. For safety endpoints, all post-baseline results will be included in these overall analyses.

3.1.2 Baseline

In general, the last recorded value on or not after the date of randomisation will serve as the baseline measurement for efficacy endpoints while the last recorded value prior to first dose of study treatment will serve as the baseline measurement for safety endpoints. When time of assessment is not recorded or missing, it is assumed that assessments recorded on the date of first dose of study treatment were performed prior to dosing, except in cases of protocol-specified post-dose assessments. If there is no value prior to randomisation (or the first dose of study treatment, depending on the endpoint), then the baseline value will not be imputed and will be set to missing. No data known to be collected post first dose will be used in determining the baseline value, unless otherwise specified.

Baseline serum creatinine will be the highest of all creatinine measurements collected before randomisation.

3.1.3 Missing values

A treatment policy estimand will be applied to the analysis of the primary and secondary endpoints whereby all data up to Week 32 are included, regardless of whether the patient has discontinued study drug or received other medications.

Missing values for the primary and secondary endpoints at post-baseline visits up to and including Week 32 (Visit 8) will be imputed using a dropout reason-based multiple imputation (DRMI) approach. Additionally, sensitivity analyses to assess the robustness of the primary analysis with respect to the handling of missing data will be performed. Missing values at baseline will not be imputed and change from baseline will be set to missing if the baseline assessment is missing.

Details for the imputation and sensitivity analyses can be found in Section 4.2.5. Missing data will not be imputed for the safety and exploratory analyses.

3.1.4 Prior/concomitant medications

A medication will be regarded as prior if it was stopped on or before the date of randomisation (medication stop date ≤ date of randomisation). A medication will be regarded as concomitant if the start date is after the date of randomisation, or if it started on or prior to the date of randomisation and ongoing after the date of randomisation.

The handling of partial/missing dates for AEs and prior/concomitant medications is detailed in Appendix 8.1.

3.2 Endpoints

3.2.1 Primary efficacy endpoint

The primary efficacy endpoint is the absolute change from baseline (Visit 2) in peak VO_2 at Week 32 (Visit 8), calculated as the value at Week 32 minus the baseline value, as assessed by a cardiopulmonary exercise test (CPET).

The primary modality for a CPET will be either a motor-driven treadmill or cycle ergometer. Patients must use the same testing modality and the same exercise equipment for all exercise tests during the study. The modality for each CPET will be recorded. Peak VO_2 is the maximum rate of oxygen consumption measured during the exercise test, and will be expressed in millilitres oxygen per kilogram of body mass per minute (mL/kg/min).

3.2.2 Secondary efficacy endpoint

The secondary efficacy endpoint is the absolute change from baseline (Visit 3) in KCCQ Total Symptom Score (TSS) to Week 32 (Visit 8), calculated as the value at Week 32 minus the baseline value.

The KCCQ is a 23-item, self-administered ePRO instrument that quantifies physical limitations, symptoms, self-efficacy, social interference, and quality of life, and has been shown to be a valid, reliable and responsive measure for patients with HF (Green et al 2000). The KCCQ was developed to measure the patient's perception of their health status independently, which includes HF-related symptoms (frequency, burden and recent change),

impact on physical and social function, self-efficacy and knowledge, and how the patient's HF affects their quality of life.

The KCCQ-TSS is a summary score measuring symptom experience in the symptom frequency domain and the symptom burden domain. The score ranges from 0-100, where higher scores reflect better health status.

See Appendix 8.2 for the full KCCQ. Domains and scoring algorithms are described in Appendix 8.3.

KCCQ-TSS is assessed using the ePRO instrument during clinic visits at the time points indicated in Section 1.1 (SoA) of the CSP. It is not possible to leave assessment questions within the 23-item KCCQ questionnaire unanswered on the ePRO instrument and it is checked at each visit to ensure completion.

CCI [redacted]

CCI

[redacted]

CCI [redacted]

[redacted]

CCI

[redacted]

3.3 Safety endpoints

The following safety data will be collected: reported AEs (including serious AEs and AEs leading to discontinuation of IP), laboratory endpoints, 12-lead electrocardiogram (ECG), physical examination, and vital signs.

All safety measurements will use all available data for analyses, including data from unscheduled visits and repeated measurements.

Missing safety data will not be imputed. The handling of partial/missing dates for AEs and prior/concomitant medications is detailed in Appendix 9.1. Duration of AEs will not be calculated using imputed dates and will instead be set to missing.

3.3.1 Adverse events

Adverse events experienced by the patients will be collected throughout the study and will be coded using the latest version of the Medical Dictionary for Regulatory Activities (MedDRA) per the Data Management Plan. Adverse events will be collected from the time of the first dose of study treatment throughout the treatment period and including the follow-up period and last contact. Serious AEs will be recorded from the time of signing of ICF throughout the treatment period and including the follow-up period and last contact. All AEs will be coded

using the latest version of the Medical Dictionary for Regulatory Activities (MedDRA) per the Data Management Plan.

The following events are considered treatment emergent:

- Adverse events with an onset date on or after first dose of IP
- Worsening of pre-existing events on or after first dose of IP

Pre-treatment (prior) adverse events are defined as those with a start prior to the day of first dose of study treatment.

If an AE has a missing onset date then unless the stop date of the AE indicates otherwise, this will be considered an on-treatment AE. Similarly, if an AE has a partial onset date, then unless the partial onset date or the stop date indicates otherwise, this will be considered an on-treatment AE.

Adverse events that have missing causality (after data querying) will be assumed to be related to study drug.

3.3.2 Laboratory variables

Blood and urine samples for determination of clinical chemistry, haematology and urinalysis parameters will be taken at the times detailed in Section 1.1 (SoA) of the CSP and will be assessed in a central laboratory. The parameters outlined in Section 8.2.1 (Table 7) of the CSP will be collected.

In summaries, figures, and listings, lab results and normal ranges will be presented in System International (SI) units.

Changes in haematology and clinical chemistry variables between baseline and each post-baseline assessment will be calculated. For values recorded with a reading greater than or less than ('>', '<') symbol, the reported numeric value will be used for analysis and the value with the symbol will be included in the listings, unless otherwise specified. For example, a value of <0.01 will be analysed as 0.01 and listed as <0.01.

Absolute values will be compared to the relevant reference range and classified as low (below range), normal (within range or on limits) or high (above range). The central reference ranges will be used for laboratory variables. All absolute values falling outside the reference ranges will be flagged.

Urinalysis data will be categorised as negative, trace, positive (+), or strongly positive (++, +++, or ++++) at each timepoint.

For the liver function tests: aspartate Aminotransferase (AST), alanine Aminotransferase (ALT), alkaline phosphatase (ALP), and total bilirubin (TBL), the multiple of the central laboratory upper limit of the normal (ULN) range will be calculated for each data point.

Multiple = Value / ULN

That is, if the ALT value was 72 IU/L (ULN=36) then the multiple would be 2.

Patients who meet any of the following criteria at any point during the study will be flagged:

- AST ≥ 3x ULN
- ALT ≥ 3x ULN
- TBL ≥ 2xULN

3.3.3 Twelve-lead ECGs

ECG measurements will be assessed at the times detailed in Section 1.1 (Table SoA) of the CSP, with the baseline visit being defined as the last available non-missing measurement(s) prior to first dose of study treatment.

The ECG assessment (normal, abnormal but not clinically significant, or abnormal and clinically significant; and QT interval, QTcF) will be recorded in the eCRF. Any clinically significant findings will be reported as AEs or in the Medical History, as appropriate.

3.3.4 Vital signs

Temperature, pulse rate, weight, and blood pressure will be assessed at times detailed in Section 1.1 (Table SoA) of the CSP.

Changes in vital signs variables between baseline and each subsequent scheduled assessment will be calculated. The change from baseline is defined as the post-baseline visit value minus the baseline visit value. There will be no imputation for missing values.

Absolute values will be compared to the reference ranges in Table 3 below and classified as low (below range), normal (within range or on limits) or high (above range). All values falling outside the reference ranges will be flagged.

Table 3 Vital Signs Reference Ranges

Parameter	Standard Units	Lower Limit	Upper Limit
Diastolic Blood Pressure (DBP)	mmHg	60	120
Systolic Blood Pressure (SBP)	mmHg	100	160
Pulse Rate	Beats/min	50	120

Table 3 Vital Signs Reference Ranges

Parameter	Standard Units	Lower Limit	Upper Limit
Body Temperature	Celsius	36.5	38

3.3.5 Physical examination

Complete and brief physical examinations will be performed at time points detailed in Section 1.1 (Table SoA) of the CSP. What is included in the assessment will be dependent on whether the examination is complete or brief, as described in Section 8.2.2 of the CSP.

Any new or aggravated clinically relevant abnormal medical finding at a physical examination as compared with the baseline assessment will be reported as an AE unless unequivocally related to the disease under study.

4 ANALYSIS METHODS

4.1 General principles

All efficacy CCI endpoints will be analysed using the full analysis set (FAS). Patients will be analysed according to their randomised treatment. The analysis of safety endpoints will be based on the safety analysis set. Pharmacokinetic endpoints will be analysed using the pharmacokinetic analysis set. The enrolled set will be used for disposition and screen failure reporting. Data from unscheduled visits will not be used in the efficacy analyses. Missing values at baseline will not be imputed.

The data analyses will be conducted using the SAS® System (SAS Institute Inc., Cary, NC). All SAS® programs used to generate analytical results will be developed and validated according to AstraZeneca SAS® programming standards and validation procedures.

Summary data will be presented in tabular format by treatment group and visit when applicable. Categorical data will be summarised by the number and percentage of patients in each category. Continuous variables for parametric data will be summarised by descriptive statistics including N, mean, SD, median, minimum and maximum. Data listings will be sorted by treatment and patient number.

Continuous endpoints assessed at several visits will be displayed graphically as arithmetic mean, geometric mean, or median curves over time (visit), by treatment and visit. Suitable error bars will be included in the graphical displays.

Continuous endpoints that are derived change-from-baseline variables will be summarised in

tables and graphs as above for both the measurements at each visit and the change-from-baseline values. CCI [redacted]

All hypothesis testing will be reported using 2-sided tests. All p-values will be nominal and will be displayed in SAS pvalue6.4 format.

4.1.1 Estimand of the primary and secondary efficacy endpoints

A treatment policy estimand will be applied to the analysis of the primary and secondary endpoints whereby all data up to Week 32 (Visit 8) are included, regardless of whether a patient remains on blinded study treatment or not. Details regarding primary and secondary estimands are provided in Table 4 below, with additional details including sensitivity analyses provided in Section 4.2.5.

Table 4 Primary and Secondary Efficacy Endpoints Estimands

Statistical Category	Estimand			Section
	Endpoint (Population)	**Intercurrent Event Strategy**	**Population Level Summary (Analysis)**	
Primary Objective: To assess effect of verinurad + allopurinol compared to placebo and compared to allopurinol monotherapy on exercise capacity				
Primary/MCP	Change from baseline at Week 32 in peak VO_2 consumption (FAS)	Included in analysis regardless of having discontinued study drug or received other medications (treatment policy) using DRMI approach	Mean difference between treatments (LSMD from CFB ANCOVA at Week 32)	4.2.5
Sensitivity	Change from baseline at Week 32 in peak VO_2 consumption (FAS)	Included in analysis regardless of having discontinued study drug or received other medications (treatment policy) using MI approach	Mean difference between treatments (LSMD from CFB ANCOVA at Week 32)	4.2.5
Secondary Objective: To assess effect of verinurad + allopurinol compared to placebo and compared to allopurinol monotherapy on KCCQ-TSS				
Secondary/MCP	Change from baseline at Week 32 in KCCQ-TSS (FAS)	Included in analysis regardless of having discontinued study drug or received other medications (treatment policy) using DRMI approach	Mean difference between treatments (LSMD from CFB MMRM to Week 32)	4.2.7
Sensitivity	Change from baseline at Week 32 in KCCQ-TSS (FAS)	Sensitivity analyses as described for the primary objective	Mean difference between treatments (LSMD from CFB MMRM to Week 32)	4.2.5

MCP Multiple comparisons procedure. FAS Full analysis set. Dropout reason-based multiple imputation. MI Multiple imputation. LSMD Least squares means difference. CFB Change from baseline. ANCOVA Analysis of covariance. MMRM Mixed model for repeated measures.

4.1.2 Methods for multiplicity control

A hierarchical test sequence will be used for the confirmatory analysis of the primary and secondary objectives in order to address the issue of multiple testing and control the Type I error rate at an overall two-sided 0.05 level. Statistical significance will be assessed in the following sequence:

- Comparison between verinurad + allopurinol and placebo in change from baseline in peak VO_2 at Week 32
- Comparison between verinurad + allopurinol and allopurinol monotherapy in change from baseline in peak VO_2 at Week 32
- Comparison between verinurad + allopurinol and placebo in change from baseline in KCCQ-TSS at Week 32
- Comparison between verinurad + allopurinol and allopurinol monotherapy in change from baseline in KCCQ-TSS at Week 32

The testing procedure will continue down the hierarchy if the preceding endpoint is rejected at a two-sided 0.05 level and will stop if the preceding endpoint is not rejected at a two-sided 0.05 level.

CCI [redacted]

4.2 Analysis methods

4.2.1 Subject disposition

Patient disposition will be summarised using the enrolled patients set. The total number of patients will be summarised for those who enrolled and those who were not randomised (and reason). The number and percentage of patients within each treatment group will be presented by the following categories: randomised, received treatment with study drug, did not receive treatment with study drug (and reason), completed treatment with study drug, discontinued treatment with study drug (and reason), completed study, and withdrawn from study (and reason). Subjects are considered to have completed the study when he or she has completed the last scheduled visit (Visit 9).

CCI [redacted]

The number of patients randomised by region, country and site will also be summarised by treatment group in the full analysis set.

4.2.2 Impact of COVID-19 on study visits

Impact of the COVID-19 pandemic on study visits will be evaluated by summarising the number and percentage of patients at each visit by treatment and contact mode (remote-audio, remote-video, home visit, or on-site visit).

4.2.3 Demography data and subject characteristics

Demography and patient characteristics will be summarised by treatment group and for total in the FAS.

Age will be derived from the date of informed consent minus date of birth, rounded down to the nearest integer. For patients where date of birth is not recorded, the age as recorded in the electronic case report form (eCRF) will be used.

Medical history will be summarised separately for past and current conditions by MedDRA preferred term (PT) within MedDRA system organ class (SOC).

4.2.4 Study treatment administration

Duration of each study IP administration will be summarised by treatment group for the safety analysis set and will be calculated in days as:

Duration of exposure = min (last dose date of IP, date of death)-first dose date of IP + 1.

Study treatment compliance in percent will be summarised by treatment group for the safety analysis set and calculated as:

Study treatment compliance = (total doses administered/total doses expected) x 100.

Total doses administered will also be summarised by treatment group for the safety analysis set. The total number of doses administered is based on dispensed doses minus returned unused doses and will be calculated as:

Total doses administered = dispensed doses – returned doses + 1 (excluding records with missing returned doses information).

The total number of doses expected includes all visits with protocol scheduled IP administration on or before a patient's IP discontinuation or treatment completion date and will be calculated as:

*Total doses expected = {min(treatment/study discontinuation, last treatment date) - min(treatment start, dispensed doses date, Visit 3) + 1} * 2 + extra dose (where extra dose is the number of days between visit 4 and visit 5 as patients took 1 more tablet than other periods).*

Patients who didn't receive a study IP will have zero compliance for this study IP.

Tabular summaries for the percentage of patients by the reason for discontinuation of randomised treatment as well as for withdrawal from the study will be presented by treatment to describe why patients discontinue from randomised treatment or withdraw from the study.

Tabular summaries for the percentage of patients with treatment interruptions and resumptions due to sCr elevations will be presented by treatment.

Kaplan-Meier (KM) estimates of the cumulative proportion of patients which have discontinued treatment during the treatment period will be calculated and plotted, with the number of patients at risk of discontinuation indicated at each specific time point in the plot. Time will be defined as number of days from randomisation until discontinuation (discontinuation date - randomisation date + 1), or for patients who do not discontinue, from randomisation to censoring (censor date - randomisation date + 1). Censor date will be the earliest date of date of withdrawal or death when applicable, otherwise the date of the last treatment visit (visit 8), whichever occurs first.

4.2.5 Primary efficacy endpoint: verinurad + allopurinol vs placebo

The efficacy variable for the analysis of the primary objective to assess the effect of verinurad + allopurinol compared to placebo on exercise capacity is the absolute change from baseline in peak VO_2 at Week 32. Patients will be analysed using the FAS according to randomised treatment.

The null hypothesis is that there is no difference in mean change from baseline at Week 32 in peak VO_2 consumption between verinurad + allopurinol and placebo. The alternative hypothesis is that there is a difference in mean change from baseline between verinurad + allopurinol and placebo, ie,

H_0: Difference in mean CFB (verinurad + allopurinol vs placebo) = 0
H_a: Difference in mean CFB (verinurad + allopurinol vs placebo) ≠ 0

Summary statistics for peak VO_2 at each visit and change from baseline VO_2 at Week 32 will be presented by treatment group and visit.

Change from baseline in peak VO_2 at Week 32 between the treatment groups will be compared using ANCOVA analysis, with change from baseline as the dependent variable, treatment as the independent variable and baseline peak VO_2 included as covariate.

Results will be presented in terms of least square means (LSMEANS), treatment differences in LSMEANS, 95% confidence intervals (CI), and p-values.

The analysis will include data from all patients at Week 32 irrespective of whether the patient has died, discontinued study drug, or received other medications. Missing peak VO_2 values at Week 32 will be imputed using a dropout reason-based multiple imputation approach (DRMI); missing data in the verinurad + allopurinol group (12 mg verinurad + allopurinol treatment arms) due to potentially treatment-related reasons namely following the intercurrent events death, withdrawal of consent, loss to follow-up or premature discontinuation of study drug due to an AE will be imputed based on the peak VO_2 values in the placebo group whereas missing values for all other patients will be imputed assuming missing at random (MAR), ie, based on the peak VO_2 values in their own respective treatment group. Missing values at baseline will not be imputed.

One mitigation to minimise the amount of missing data this study is to allow patients to down-titrate dosage by reversing the assigned steps within treatment group. Patients who cannot tolerate the step-1 dose will be discontinued from the study treatment but will still be followed for the remainder of the study.

The normality assumption of the ANCOVA models will be checked graphically by producing histograms and normal probability plots of the residuals from the model estimated excluding patients with missing peak VO_2 values from the analysis.

Primary peak VO_2 analysis under the treatment policy estimand using a dropout reason-based multiple imputation (DRMI) approach

The predictive mean matching imputation algorithm in the SAS PROC MI procedure will be used to ensure that imputed values fall within permissible ranges. The algorithm will use placebo arm patient data to impute missing values due to potentially treatment-related reasons in the verinurad + allopurinol group, and use data within each respective treatment arm to impute missing values for all other reasons in remaining patients. The algorithm starts by fitting a linear regression model to peak VO_2 values at Week 32 in patients without missing values, including baseline peak VO_2 as covariate. A set of regression coefficients is drawn from the posterior predictive distribution of the estimated coefficients. The set of drawn coefficients is then used to calculate predicted values for all patients, including those with missing values. For each patient with missing values, the algorithm finds a set of K closest matching predicted values among patients without missing values. Finally, an imputed replacement value is drawn at random from the observed values of these closest matching patients. The algorithm is repeated, creating M imputation datasets. A seed of 230185 will be used with K=5 closest values and M=100 datasets will be imputed. ANCOVA analyses of each of the imputed datasets will be performed as described for the primary analysis in

Section 4.2.5. Results from each fitted model will be combined using Rubin's rules implemented with the PROC MIANALYZE SAS procedure.

Sensitivity analyses under the treatment policy estimand using multiple imputation (MI)

To examine the sensitivity of the results of the primary analysis to departures from the underlying assumptions, a sensitivity analysis will be performed. The sensitivity analysis will include data from all patients at Week 32 irrespective of whether the patient has discontinued study drug or received other medications. The following sensitivity analysis will be performed:

- Missing peak VO_2 values at Week 32 will be imputed using multiple imputation as described above but all imputations will be based on values in the placebo group.

Supplementary analyses for the primary endpoint

Two supplementary analyses will be performed to gain additional insights into the understanding of the treatment effect on the primary endpoint. All supplementary analyses will include data from all patients up to Week 32 irrespective of whether the patient has discontinued study drug or received other medications.

- Supplementary analysis 1 – Complete case (CC): Missing peak VO_2 values at Week 32 will not be imputed and patients with missing values will be excluded from the analysis.

- Supplementary analysis 2– Rank ANCOVA: Change from baseline at Week 32 in peak VO_2 will be transformed into ranks incorporating deaths using fractional rank with the mean method to handle ties.

The rank ANCOVA supplementary analysis 2 is to assess the potential effect of non-normality of the primary endpoint and different handling of the intercurrent event death. In this analysis, change from baseline at Week 32 in peak VO_2 as well as the baseline peak VO_2 values will be transformed into ranks using fractional ranking with the mean method to handle ties. The ANCOVA model will have ranked change from baseline as the dependent variable, treatment as the independent variable, ranked baseline values included as covariate. Patients who die before Week 32 will be given the worst ranks. Among patients who die, patients will be ranked according to the time from randomisation until death, with shorter times given worse ranks. Missing peak VO_2 data following intercurrent events other than death will be imputed using the same multiple imputation procedure as described for the primary analysis, prior to ranking, and combined with Rubin’s rules.

As the point estimate from the rank ANCOVA analysis lack a clear clinical interpretation, a supplementary Hodges–Lehmann (HL) estimate of the median difference between treatment groups in change from baseline at Week 32 in peak VO_2, together with a 95% CI, will be estimated. The HL estimate is the median of all possible pairwise differences between verinurad + allopurinol and placebo patients. Deaths will be handled in a similar way as in the rank ANCOVA analysis, where change from baseline values for patients who died are set to a temporary negative change value lower than all observed change values, with lower negative change values given to patients with shorter time from randomisation to death. The estimation will be made for each imputed data set, and estimates will be combined using Rubin's rules.

Subgroup analyses for the primary endpoint

To explore the uniformity of the detected overall treatment effect on the primary efficacy endpoint, subgroup analyses will be performed for the following factors:

- CPET exercise mode (treadmill, cycle ergometer).
- LVEF at baseline (<50%, 50-60%, >60%)
- History of atrial fibrillation at baseline (yes, no)
- Dispensed prophylactic colchicine at randomisation (yes, no)

For each of the subgroup variables, a separate ANCOVA model will be fitted using the same terms as used in the primary analysis but additionally including the subgroup variable and the subgroup variable-by-treatment interaction term. Missing values in peak VO_2 will be handled in the same way as in the primary analysis. Patients with missing values in the subgroup variables is expected to be minimal, and will not be imputed but excluded before analysis. Similar output will be presented for each subgroup as for the primary analysis. The p-value for the interaction term will be presented, when applicable. Resulting treatment differences per subgroup in LS means and their 95% CIs will be presented in a forest plot. These subgroup analyses are considered as exploratory, and p-values for subgroup estimates and interaction terms will therefore not be adjusted for multiplicity but interpreted descriptively.

If any category of a subgroup variable contains less than 20% of the patients, an ANCOVA model similar to the primary analysis will instead be fitted excluding patients from that category. Descriptive data will be summarised for patients in the excluded category.

Anchor-based analysis for estimating thresholds for clinically meaningful within-patient change

Anchor-based analyses will be conducted to derive thresholds for meaningful within-patient change from baseline in the primary endpoint, i.e the level of change that an individual patient will perceive as meaningful. Anchor-based analysis is used to estimate a threshold for an endpoint which translates to a meaningful within-patient change by utilising anchor variables as the reference. Missing values will not be imputed, instead only patients with complete data at baseline and Week 32 will be included. This analysis will be made using the FAS, using data from all treatment arms combined.

For the primary endpoint, change from baseline at Week 32 in peak VO_2, the following anchor variables will be used (described in Appendix 8.4):

- Change from baseline at Week 32 in KCCQ item 1F (limitation in patient's ability of "Hurrying or jogging (as if to catch a bus)"),
- Change from baseline at Week 32 in KCCQ item 1E (limitation in patient's ability of "Climbing a flight of stairs without stopping").

Further details on the anchor analyses are described in Appendix 8.4.

Distribution-based analysis for estimating group-level minimal clinically important difference

To support the responder analyses by thresholds for clinically meaningful within-patient change from baseline on an individual level, distribution-based methods will be used to derive a value to be considered a minimal clinically important difference when comparing group-level averages in the primary endpoint. The analysis of group-level averages is not directly related to the analysis of responders and responder thresholds (described above). The group-level averages (means or medians) represent a mixture of some patients who improved and other patients who deteriorated. The anchor-based thresholds for individual-level within-patient change, which generally only look at one direction of change at a time (improvement or deterioration), are expected to be larger than what constitutes the magnitude of minimal clinically meaningful difference between group-level averages. Essentially, the two concepts are fundamentally different and should not be confused and as described in the FDA public workshop guidance from 2018 (FDA 2018), we note the following key discrepancies between the concepts of meaningful "within-patient change" and meaningful "between-group mean differences":

Individual within-patient change is different than between-group mean difference or treatment effect. From a regulatory standpoint, FDA is more interested in what constitutes a meaningful within-patient change in scores from the patient perspective (i.e., individual patient level). The between-group mean difference is the difference between the average score change between two study arms that is commonly used to evaluate treatment difference, but it does not address the individual within-patient change that is used to evaluate whether a meaningful score

change is observed. A treatment effect is different than a meaningful within-patient change. The terms minimally clinically important difference (MCID) and minimum important difference (MID) do not define meaningful within-patient change if derived from group-level data.

Missing values will not be imputed, instead only patients with complete data at baseline and Week 32 will be included. This analysis will be made using the FAS, using data from all treatment arms combined. The distribution-based method used will be the one-half SD method. The one-half SD method is based on the effect size of a change score. It is suggested that 0.5 corresponds to a "medium" effect size, relative to the SD (Cohen 1977). The minimal important difference is thus calculated by dividing the SD of the primary endpoint at baseline by 2 in all patients combined. In addition, the distribution-based method will be supported by anchor-based analyses of meaningful within-patient change described below, using the smallest levels of the anchors, to triangulate a single value to be considered the minimal clinically important group-level difference. The difference between average change in peak VO_2 in each treatment group will then be compared to this value.

4.2.6 Primary efficacy endpoint: verinurad + allopurinol vs allopurinol

The efficacy variable for the analysis of the secondary objective to assess the effect of verinurad + allopurinol compared to allopurinol monotherapy on exercise capacity is the absolute change from baseline in peak VO_2 at Week 32. Patients will be analysed using the full analysis set according to randomised treatment.

Mean change from baseline at Week 32 in peak VO_2 consumption between verinurad + allopurinol and allopurinol monotherapy will be analysed using a similar ANCOVA model as outlined for the primary efficacy endpoint in Section 4.2.5.

The analysis will include data from all patients irrespective of whether the patient has discontinued study drug or received other medications. Missing peak VO_2 at Week 32 values will be handled using the same method as described for the analysis of the primary objective in Section 4.2.5.

The estimate of the treatment effect of verinurad + allopurinol compared to allopurinol monotherapy will be extracted from the analysis described for the analysis of the primary objective as described in Section 4.2.5.

All sensitivity, supplementary, and subgroup analyses will be performed as described in Section 4.2.5.

Testing strategy to account for multiplicity considerations is addressed in Section 4.1.2.

4.2.7 Secondary efficacy endpoint

The efficacy variable for the analysis of the secondary objective to assess the effect of verinurad + allopurinol compared to placebo and compared to allopurinol monotherapy on KCC-TSS is the absolute change from baseline in KCCQ-TSS to Week 32. The analysis set will be the FAS.

Summary statistics for KCCQ-TSS at each visit and change from baseline KCCQ-TSS at week 22 and 32 will be presented by treatment group and visit. KCCQ-TSS ePRO questionnaire completion status and reasons for not completing will be summarised by treatment group and visit.

Mean change from baseline in KCCQ-TSS to Week 32 between the treatment groups will be compared using a mixed model repeated measure (MMRM) model, with change from baseline to week 22 and 32 (visit 7 and 8) as the dependent variable, treatment as the independent variable, and visit, visit and treatment interaction, and baseline KCCQ-TSS included as covariates. Baseline KCCQ-TSS will be the KCCQ-TSS value at Week 0 (Visit 3). Visit will be fitted as a categorical variable. An unstructured covariance matrix will be used with degrees of freedom calculated using the Kenward-Roger approach. If the estimation does not converge, the following covariance structures will be used in order until convergence is met: Toeplitz, compound symmetric, variance components.

Normality assumptions will be checked graphically by producing histograms and normal probability plots of the residuals from the model estimated excluding patients with missing KCCQ-TSS values from the analysis.

Results will be presented in terms of least square means (LSMEANS), treatment differences in LSMEANS, 95% confidence intervals (CI), and p-values for Week 22 (Visit 7) and Week 32 (Visit 8). The results for change from baseline to Week 32 will be of primary interest and will be the basis for the confirmatory testing. Graphical displays of KCCQ-TSS LS mean change from baseline by treatment and visit will also be presented.

Testing strategy to account for multiplicity considerations is addressed in Section 4.1.1.

The analysis will include data from all patients irrespective of whether the patient has discontinued study drug or received other medications. Missing KCCQ-TSS values at Week 22 and Week 32 will be imputed in a sequential manner with previous measures as covariates using a DRMI approach as described below. Missing values at baseline will not be imputed.

Secondary KCCQ-TSS analysis under the treatment policy estimand using a dropout reason-based multiple imputation (DRMI) approach

The imputation procedure for missing KCCQ-TSS values will follow the same procedure as in the primary analysis described in Section 4.2.5. Missing values at Week 22 and Week 32 will be imputed in a sequential manner. Missing Week 22 values will be imputed first and will be included as a covariate in imputing Week 32.

Sensitivity analyses under the treatment policy estimand using multiple imputation (MI)

To examine the sensitivity of the results of the secondary analysis to departures from the underlying assumptions, a sensitivity analysis will be performed. The sensitivity analyses will include data from all patients irrespective of whether the patient has discontinued study drug or received other medications. The sensitivity analyses will be performed in a similar manner as in Section 4.2.5.

- Missing KCCQ-TSS values at Week 22 and Week 32 will be imputed using multiple imputation as described above but all imputations will be based on values in the placebo group.

Supplementary analyses

Three supplementary analyses will be performed to gain additional insights into the understanding of the treatment effect on the secondary endpoint. The supplementary analyses will include data from all patients irrespective of whether the patient has discontinued study drug or received other medications. The supplementary analyses will be performed in a similar manner as in Section 4.2.5.

- Supplementary analysis 1 - CC: Missing KCCQ-TSS values at Week 22 and Week 32 will not be imputed and patients with missing values will be excluded from the analysis.

- Supplementary analysis 2 – Rank ANCOVA: Change from baseline at Week 32 in KCCQ-TSS will be transformed into ranks using fractional ranking with mean method to handle ties. Week 22 assessments will be disregarded and not used in the analysis.

Anchor-based analysis for estimating thresholds for clinically meaningful within-patient change

Anchor-based analyses will be conducted to derive thresholds for meaningful within-patient change in the secondary endpoint in a similar manner as for the primary endpoint. Missing values will not be imputed, instead only patients with complete data at baseline and Week 32 will be included. This analysis will be made using the FAS population, using data from all treatment arms combined.

For the secondary endpoint, change from baseline in KCCQ-TSS, the following anchor variables will be used (described in Appendix 8.4):

- Change from baseline at Week 32 in PGIS,
- The Week 32 assessment of PGIC.

Further details on the anchor analyses are described in Appendix 8.4.

Distribution-based analysis for estimating group-level minimal clinically important difference

As described above, for the primary endpoint, distribution-based methods will be used to interpret what constitutes a minimal clinically important difference between group-level averages (means or medians). The main distribution-based method used for the secondary endpoint will be the standard error of measurement (SEM) method. The SEM will be calculated using baseline data as the standard deviation of secondary endpoint at baseline multiplied by the square root of 1 minus the reliability (internal consistency) of the KCCQ-TSS instrument at baseline. A difference of 1 SEM has been suggested as a minimal clinically important difference in a score (Wyrwich 1999). In other words:

$$1\,SEM = SD * \sqrt{(1 - r_X)},$$

where r_X represents the reliability coefficient of the score, calculated as Cronbach's alpha, as a measure of internal consistency. Cronbach's alpha will be calculated as:

$$r_X = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k} \sigma_i^2}{\sigma_X^2}\right),$$

where k equals the number of items in KCCQ-TSS, σ_i^2 is the variance in score of item i, and σ_X^2 is the variance of the sum of the scores of all k items, using the numerical values assigned to the verbal response options in the KCCQ scoring algorithm in Appendix 9.3. The SEM method will be supplemented by values calculated using the one-half standard deviation method described for the primary endpoint, as well as the anchor-based analyses of meaningful within-patient change described below in order to triangulate a single value to be considered the minimal clinically important group-level difference in the secondary endpoint. The difference between group average change scores by treatment group will then be compared to this value.

Missing values will not be imputed, instead only patients with complete data at baseline and Week 32 will be included. This analysis will be made using the FAS, using data from all treatment arms combined.

CCI [redacted]

[redacted]

CCI

[redacted]

4.2.9 Safety endpoints

All safety variables will be summarised using the safety analysis set. Patients who receive study treatment which is not consistent with the treatment they were randomised to will be listed. All summaries will be presented by treatment group.

4.2.9.1 Adverse events

All AEs occurring from the randomisation visit and onwards will be summarised. All AEs will be listed, including SAEs in the pre-treatment period (with start date prior to the first dose of IP). All summaries will be presented by treatment group.

An overall summary table will be produced showing the total number of events and the number, percentage, and exposure-adjusted rate of patients with at least 1 AE in any of the following categories; AEs, serious adverse events (SAEs), AEs with outcome of death, AEs leading to discontinuation of investigational product (DAEs).

AEs, AEs with outcome of death, SAEs and DAEs will be summarised by System Organ Class (SOC) and Preferred Term (PT) assigned to the event by MedDRA. For each PT, the number, percentage and exposure-adjusted rate of patients reporting at least one occurrence will be presented (ie, multiple occurrences of an AE for a patient will only be counted once). Summaries by SOC and PT will be sorted in descending order by frequency of PT in the verinurad + allopurinol arm. A summary of the most common (frequency of >3%) AEs will be presented by PT. Additionally, a summary of non-serious AEs occurring in >5% of patients in any treatment group will be presented by PT. AEs and SAEs causing discontinuation of the

study treatment and AEs and SAEs causing discontinuation from the study will also be summarised.

AEs and SAEs will be summarised by preferred term and investigator's causality assessment and maximum intensity. If a patient reports multiple occurrences of the same AE within the same study period, the maximum intensity will be taken as the highest recorded maximum intensity (the order being mild, moderate, and severe).

Key patient information will be presented for patients with adverse events with outcome of death, serious adverse events, and adverse events leading to discontinuation of treatment.

The latest version of the MedDRA dictionary will be used for coding of the AEs.

CCI

4.2.9.3 Laboratory data

All protocol-specified continuous laboratory parameters will be summarised descriptively by absolute value at each visit by treatment group, together with the corresponding changes from baseline. All parameters will be summarised in SI units. Results reported by the central laboratory in conventional units will be converted to SI units for reporting.

Key patient information will be presented for patients with post-baseline values outside of laboratory reference ranges.

Maximum post-baseline total bilirubin (TBL) elevations by maximum post-baseline alanine aminotransferase (ALT) and aspartate aminotransferase (AST) will be presented, expressed as multiples of the upper limit of normal (ULN). TBL will be presented in multiples of the following ULN: ≤1.5, >1.5-2, >2. AST and ALT will be presented in multiples of the following ULN: ≤1, >1-3, >3-5, >5-10, >10.

Maximum post-baseline TBL will be presented (<2 and ≥2 x ULN) and plotted against maximum post-baseline ALT (<3, ≥3-<5, ≥5-<10, and ≥10 x ULN), expressed as multiples of ULN. This will be repeated to show maximum post-baseline TBL against maximum post-baseline AST.

Data for patients who meet the biochemical criteria for Hy's law (ALT or AST ≥3 x ULN together with TBL≥2 x ULN) will be presented, which will include all visits for this subset of patients. A line plot of liver biochemistry test results (including ALP, ALT, AST, TBL, and GGT) over time will also be presented for this subset of patients. For all patients who meet the biochemical criteria for Hy's law definition, a SAE narrative will be produced.

For urinalysis data, a shift table will be generated to present changes from baseline to maximum post-baseline value for selected parameters and will include patients with both baseline and post-baseline data.

Any data outside the central laboratory reference ranges will be explicitly noted on the listings that are produced.

CCI [redacted]

4.2.9.5 ECGs

A shift table will be produced for the ECG overall interpretation to display normal, abnormal – not clinically significant, abnormal – clinically significant, and not done. The shift table will present baseline and last observation post-baseline value.

QTcF intervals will be summarised by treatment group. This will consist of a summary of the number and percentage of patients with a post-baseline QTcF interval greater than 450 ms, 480 ms, and 500 ms and the number and percentage of patients with an increase from baseline in QTcF interval of greater than 30 ms and 60 msec. A shift plot showing maximum post-dose QTcF change from baseline (with reference lines at 30 ms and 60 ms) will also be produced.

4.2.9.6 Vital signs

Descriptive statistics and change from baseline for vital signs data will be presented for each treatment group by visit. Baseline to maximum post-baseline and baseline to minimum postbaseline value shift tables will be generated, as applicable for each parameter and will include patients with both baseline and post-baseline data.

5 INDEPENDENT DATA MONITORING COMMITTEE (IDMC) ANALYSES

Refer to the IDMC charter

6 CHANGES OF ANALYSIS FROM PROTOCOL

Patients who completed treatment at Week 28 under CSP version 1.0 will have their Week 28 assessment counted as their Week 32 assessment. The scheduled Week 28 visit falls within the Week 32 assessment window as defined in CSP version 2.0.

Sensitivity analyses added for the secondary endpoint.

7 REFERENCES

Borlaug et al 2018
Borlaug BA, Anstrom KJ, Lewis GD, Shah SJ, Levine JA, Koepp GA, et al. Effect of inorganic nitrite vs placebo on exercise capacity among patients with heart failure with preserved ejection fraction: The INDIE-HFpEF randomised clinical trial. JAMA 2018;320(17):1764-1773.

Cohen J 1977
Cohen J. Statistical power analysis for the behavioral sciences. Academic Press. 1977

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

Fletcher et al 1995
Fletcher GF, Balady G, Froelicher VF, Hartley LH, Haskell WL, Pollock ML. Exercise standards. Circulation 1995;91(2):580-615.

Green et al 2000
Green CP, Porter CB, Bresnahan DR, Spurtus JA. Development and evaluation of the Kansas City Cardiomyopathy Questionnaire: A new health status measure for heart failure. Journal of the American College of Cardiology 2000;35(5):1245-1255.

Levey et al 2009
Levey AS, Stevens LA, Schmid CH, Zhang Y, Castro AF, Feldman HI, et al. A new equation to estimate glomular filtration rate. Annals of Internal Medicine 2009;150(9):604-12

Redfield et al 2013
Redfield MM, Chen HH, Borlaug BA, Semigran MJ, Lee KL, Lewis G, et al. Effect of phosphodiesterase-5 inhibition on exercise capacity and clinical status in heart failure with preserved ejection fraction: a randomised clinical trial. JAMA 2013;309(12):1268-1277.

FDA 2018
Patient-focused drug development guidance public workshop: Methods to identify what is important to patients & select, develop or modify fit-for-purpose clinical outcomes assessments. Retrieved July 05, 2021, from https://www.fda.gov/media/116277/download

Wyrwich 1999
Wyrwich K, Tierney W, Wolinsky F. Further evidence supporting a SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. J Clin Epidemiol 1999;52:861–873.

8 APPENDIX

8.1 Partial dates for adverse events and prior/concomitant medication

Dates missing the day, or both the day and month of the year will adhere to the following conventions in order to classify treatment-emergent AEs and to classify prior/concomitant medications:

Adverse Events

- The missing day of onset of an AE will be set to:

  - the first day of the month that the event occurred if the onset YYYY-MM is after the YYYY-MM of the first study treatment.
  - the day of the first study treatment if the onset YYYY-MM is the same as YYYY-MM of the first study treatment.
  - the date of informed consent if the onset YYYY-MM is before the YYYY-MM of the first treatment.

- The missing day of resolution of an AE will be set to the last day of the month of the occurrence. If the patient died in the same month, then set the imputed date to the death date.
- If the onset date of an AE is missing both the day and month, the onset date will be set to:
  - January 1 of the year of onset if the onset year is after the year of the first study treatment.
  - the date of the first treatment if the onset year is the same as the year of the first study treatment.
  - the date of informed consent if the onset year is before the year of the first study treatment.
- If the resolution date of an AE is missing both the day and month the date will be set to December 31 of the year of occurrence. If the patient died in the same year, then set the imputed date to the death date.

Prior/concomitant medication

- The missing day of start date of a therapy will be set to the first day of the month that the event occurred.
- The missing day of end date of a therapy will be set to the last day of the month of the occurrence.
- If the start date of a therapy is missing both the day and month, the onset date will be set to January 1 of the year of onset.

- If the end date of a therapy is missing both the day and month, the date will be set to December 31 of the year of occurrence.

- If the start date of a therapy is null and the end date is not a complete date, the start date will be set to the date of the first study visit.

- If the start date of a therapy is null and the end date is a complete date and the end date is after the date of the first study visit, the start date will be set to the date of the first study visit. Otherwise, the start date will be set to the end date of the therapy.

- If the end date of a therapy is null and the start date is not a complete date, the end date will be set to the study end date.

- If the end date of a therapy is null and the start date is a complete date and the start date is prior to the study end date, the end date will be set to the study end date. Otherwise, the end date will be set to the start date of the therapy.

8.2 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
A. Dressing yourself	☐	☐	☐	☐	☐	☐
B. Showering/ Bathing	☐	☐	☐	☐	☐	☐
C. Walking 1 block on level ground	☐	☐	☐	☐	☐	☐
D. Doing yardwork, housework or carrying groceries	☐	☐	☐	☐	☐	☐
E. Climbing a flight of stairs without stopping	☐	☐	☐	☐	☐	☐
F. 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
☐	☐	☐	☐	☐	☐

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
☐	☐	☐	☐	☐	☐	☐

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
□	□	□	□	□

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?

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
A. Hobbies, recreational activities	☐	☐	☐	☐	☐	☐
B. Working or doing household chores	☐	☐	☐	☐	☐	☐
C. Visiting family or friends out of your home	☐	☐	☐	☐	☐	☐
D. Intimate relationship with loved ones	☐	☐	☐	☐	☐	☐

8.3 KCCQ Scoring Instructions

The 23 items and corresponding 15 questions are listed in Appendix 8.2. The 6 items in question 1 constitute the physical limitations score. Question 2 is for the symptom stability domain. Questions 3, 5, 7, and 9 constitute the symptom frequency domain, and questions 4, 6, and 8 constitute the symptom burden domain. Questions 10 and 11 constitute the self-efficacy domain. Questions 12, 13, and 14 constitute quality of life domain. Question 15 is for the social limitation domain.

It is not possible to leave individual assessment questions unanswered on the ePRO instrument and it is checked at each visit to ensure completion. However, for physical limitation and social limitation domain, the response to the items/questions could be "Limited for other reasons or did not do" and "Does not apply or did not do for other reasons", respectively. The response represents the scenario that the question doesn't apply and the score is not calculable (NC). For example, if a patient stays at home, the item "Hurrying or jogging (as if to catch a bus)" is not answerable, and the corresponding scale score will be not calculable. If at a time point, at least 4 items in physical limitation score selected "Limited for other reasons or did not do", or at least 3 items in social limitation selected "Does not apply or did not do for other reasons", the corresponding PLS or social limitation domain at the time point will be considered as NC.Each KCCQ item or question is scored by assigning each response an ordinal value, beginning with 1 for the response that implies the lowest level of functioning. If at least half of the components within the domain are not NC (for physical limitation and social limitation), then the domain score can be calculated by summing the responses of the questions actually answered within the domain. Scale scores are transformed to a 0 to 100 range by subtracting the lowest possible scale score, dividing by the range of the scale, and multiplying by 100. If the domain has more than one component, the domain score will be the mean value of the transformed score over the actually answered components. Higher scores represent a better outcome. The scoring algorithm of each domain and summary score is described in detail below.

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(Questions 1a-f actually answered) – 1]/4

2. Symptom Stability

- Code the response to Question 2 as follows:

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

- 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

- 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(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

- Compute

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

5. Total Symptom Score

- 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

- Compute

  Self-Efficacy Score = 100*[mean(Questions 10 and 11) – 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

- Compute

  Quality of Life Score = 100*[mean(Questions 12, 13, and 14) – 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(Questions 15a-d actually answered) – 1]/4

9. Overall Summary Score

- 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

- Clinical Summary Score = mean of the available summary scores:

Physical Limitation Score
Total Symptom Score

8.4 Anchor-based analysis for estimating thresholds for clinically meaningful within-patient change

Anchor-based analysis is used to estimate a threshold for an endpoint which constitutes a clinically meaningful within-patient change by utilising anchor variables as a reference. Anchors will be relevant ePRO responses from the patients measuring the same concept as the endpoint in question.

Note that, in the tables below, the order of responses in the PGIS and PGIC scales have been reversed from the way they are presented in the CSP (Very severe to No symptoms instead of the other way around). This is only a visual re-ordering so that the order of the (arbitrary) numerical coding values assigned to the response levels go from low to high, so that a change > 0 signifies "improvement" for all anchors.

Categorisation of anchors

The ordinal responses of KCCQ 1F, KCCQ 1E, PGIS, and PGIC at baseline and/or Week 32 will be assigned numerical values starting at 1 indicating the lowest level function/symptom/change. The numerical coding for each variable is shown in Table 5. Patients are allowed to respond to KCCQ 1F and KCCQ 1E using the option “Limited for other reasons or did not do” (in accordance with the KCCQ-23). Such responses will not be coded numerically and those responses are excluded from the anchor-based analyses.

Table 5 Ordinal response coding of anchor variables at baseline and/or Week 32.

KCCQ 1F		KCCQ 1E		PGIS		PGIC	
1	"Extremely limited"	1	"Extremely limited"	1	“Very severe”	-3	"Much worse"
2	"Quite a bit limited"	2	"Quite a bit limited"	2	“Severe”	-2	"Moderately worse"
3	"Moderately limited"	3	"Moderately limited"	3	“Moderate”	-1	"A little worse"
4	"Slightly limited"	4	"Slightly limited"	4	“Mild”	0	"About the same"
5	"Not at all limited"	5	"Not at all limited"	5	“Very mild”	1	"A little better"
				6	"No symptoms"	2	"Moderately better"
						3	"Much better"

Using the coded numerical values variables, change from baseline anchors at Week 32 in KCCQ 1F, KCCQ 1E, and PGIS will be calculated as the difference in numerical value (Week 32 – baseline). Change from baseline anchors in KCCQ 1F and 1E ranges from -4 to 4, while change from baseline in PGIS ranges from -5 to 5. For these change anchor variables, negative values indicate a decrease from baseline to Week 32 and thus a deterioration of function or symptoms.

Change from baseline in KCCQ 1F, KCCQ 1E, and PGIS will be categorised into 6-category variables as:

- "Large improvement": ≥+3 points
- "Moderate improvement": +2 points,
- "Small improvement": +1 point,
- "Stable": 0,
- "Small deterioration": -1 points,
- "Moderate deterioration": -2 point.
- "Large deterioration": ≤-3 points.

PGIC will also be analysed as a 6-category variable with categories defined as:

- "Large improvement": +3 points
- "Moderate improvement": +2 points,
- "Small improvement": +1 point,
- "Stable": 0,
- "Small deterioration": -1 points,
- "Moderate deterioration": -2 point.
- "Large deterioration": -3 points.

Correlations between change from baseline in peak VO_2 and KCCQ-TSS and their corresponding anchors, both uncategorised (raw score) and categorised anchor variables, will be estimated using Spearman's correlation coefficient, recognizing that the anchor variables are ordinal variables (even when assigned numerical values for the calculations). An anchor is generally considered adequate if the estimated correlation coefficient greater than or equal to 0.3 (Coon et al 2018).

Descriptive statistics (N, mean, SD, median, quartiles, minimum, and maximum) will be presented for the primary and secondary endpoint across categories of their corresponding categorised anchor variables. To assist the interpretation of the anchor-based analysis, empirical cumulative distribution (CDF) function and probability density function (PDF) plots of the primary and secondary endpoint by anchor categories will be constructed. The results from these analyses will be used in an effort to find agreement across different types of evidence for a single threshold which represents clinically meaningful within-patient change for each endpoint, for both improvement and deterioration.

SIGNATURE PAGE

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

Document Name: d5496c00005-sap-ed-4		
Document Title:	Statistical Analysis Plan Edition 4	
Document ID:	Doc ID-004360069	
Version Label:	5.0 CURRENT LATEST APPROVED	
Server Date **(dd-MMM-yyyy HH:mm 'UTC'Z)**	**Signed by**	**Meaning of Signature**
07-Jun-2022 15:28 UTC	Redacted	Content Approval
07-Jun-2022 15:44 UTC	Redacted	Author Approval

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