Cover Page for Statistical Analysis Plan

Sponsor name:	Novo Nordisk A/S
NCT number	NCT04788511
Sponsor trial ID:	EX9536-4665
Official title of study:	Effect of semaglutide s.c. 2.4 mg once weekly on function and symptoms in subjects with obesity-related heart failure with preserved ejection fraction
Document date*	01 December 2022

*Document date refers to the date on which the document was most recently updated.

EX9536-4665

Statistical Analysis Plan

Effect of semaglutide s.c. 2.4 mg once weekly on function and symptoms in subjects with obesity-related heart failure with preserved ejection fraction

Biostatistics, GLP-1 & CV 1

Table of contents

Page

Version history

This Statistical Analysis Plan (SAP) for trial EX9536-4665 is based on the protocol version 6.0 dated 9th of September 2022.

Table 1-1 SAP Version History Summary

SAP Version	Approval Date	Change	Rationale
1.0	See approval date in the electronic document management system	Not Applicable	New document
2.0	See approval date in the electronic document management system	Added details on the classification of events of death in section 5.1 and pharmacokinetic modelling in section 5.7.3.	The SAP is updated to reflect changes in the protocol.
3.0	See approval date in the electronic document management system	Added previous heart failure events as intercurrent events in the statistical analyses.	Updated due to feed-back from regulatory authorities addressing heart failure events (urgent visit or hospitalisations) added as intercurrent events.
4.0	See approval date in the electronic document management system	Added the hierarchical composite confirmatory endpoint and changed supplementary analysis. More details provided for anchor based supportive secondary endpoints. Update to imputation model for intercurrent event.	The SAP is updated to reflect changes in the protocol in regards to the inclusion of the hierarchical composite endpoint and to address the updated supplementary analysis. Also new exploratory endpoints with a change score of 15 for KCCQ are added. More details and explanation text are added for the thresholds based on PGIs in the supportive secondary endpoints section 5.4.2. Furthermore, the endpoints defined by these thresholds are added in more detail in the endpoint tables. Furthermore, there has been an update to the imputation model for intercurrent events where timing of last available observation during the on-treatment period (LAO-OT) now is a covariate in

			the imputation model instead of a factor.

1 Introduction

This Statistical Analysis Plan (SAP) describes in detail the analyses of efficacy, other endpoints and assessments in trial EX9536-4665. Endpoints and assessments in the trial are listed in Appendix 6.3.

Specifications of tables, figures and listings (TFL) and other specifications not included in this SAP will be described in the mock TFL.

1.1 Objectives and endpoints

The objectives and estimands in the trial are described below.

1.1.1 Primary objective and estimands

The primary objective is to investigate the effects of semaglutide s.c. 2.4 mg once-weekly on physical function, symptoms and body weight compared with placebo, both added to standard of care, in subjects with obesity-related HFpEF.

The primary and secondary estimands for the primary objective are listed in Table 1-1.

Table 1-1 Primary and secondary estimands for the primary objective

Objective	Estimand category	Estimand attributes				
		Treatment condition	Population of interest	Variable/Endpoint	Intercurrent event strategy	Population-Level Summary Measure
To investigate the effects of semaglutide s.c. 2.4 mg once-weekly on physical function, symptoms and body weight compared with placebo, both added to standard of care, in subjects with obesity-related HFpEF	**Primary**	Semaglutide 2.4 mg or placebo, both added to standard-of-care, regardless of adherence to randomised treatment and initiation of other anti-obesity therapies	All randomised	Change in KCCQ clinical summary score from baseline (week 0) to end of treatment (week 52)	• Discontinuation of trial treatment: Data collected after intercurrent events used in analysis (treatment policy) • Initiation of other weight management drugs or bariatric surgery: Data collected after intercurrent events used in analysis (treatment policy) • Events of CV-death[a] and previous heart failure event[a] (if data are *not* collected): Will be incorporated into the KCCQ clinical summary score by ascribing the outcome an unfavourable value (composite)	Mean difference of semaglutide 2.4 mg versus placebo
				Change in body weight from baseline (week 0) to end of treatment (week 52)	• Discontinuation of trial treatment: Data collected after intercurrent events used in analysis (treatment policy) • Initiation of other weight management drugs or bariatric surgery: Data collected after intercurrent events used in analysis (treatment policy)	Mean difference of semaglutide 2.4 mg versus placebo
	Secondary [b]	Semaglutide 2.4 mg or placebo, both added to standard-of-care, had all	All randomised	Change in KCCQ clinical summary score from baseline (week 0) to end of treatment (week 52)	• Discontinuation of trial treatment: "had subjects adhered to treatment" (hypothetical) • Initiation of other weight management drugs or bariatric surgery: "had the subject not initiated weight management drugs or bariatric surgery" (hypothetical) • Events of death: "Had the subject not died" (hypothetical)	Mean difference of semaglutide 2.4 mg versus placebo

		subjects adhered to randomised treatment and not initiated other anti-obesity therapies		Change in body weight from baseline (week 0) to end of treatment (week 52)	• Discontinuation of trial treatment: "had subjects adhered to treatment" (hypothetical) • Initiation of other weight management drugs or bariatric surgery: "had the subject not initiated weight management drugs or bariatric surgery" (hypothetical) • Events of death: "Had the subject not died" (hypothetical)	Mean difference of semaglutide 2.4 mg versus placebo

[(a)] As determined by the event adjudication committee: heart failure is either heart failure hospitalization or urgent heart failure visit.

[(b)] Not related to any confirmatory hypothesis

1.1.2 Secondary objectives and estimands

The secondary objective is to investigate the effects of semaglutide s.c. 2.4 mg once-weekly on walking distance, biomarker of inflammation, body composition, disease specific aspects, social limitation, and health-related quality of life compared with placebo, both added to standard of care, in subjects with obesity-related HFpEF.

To investigate the effects of semaglutide s.c. 2.4 mg once-weekly in improving the overall clinical benefit compared with placebo, both added to standard of care, in subjects with obesity-related HFpEF.

The estimands for the secondary objective are listed in Table 1-2.

Table 1-2 Estimands for the secondary objective

Objective	Estimand category	Estimand attributes				
		Treatment condition	Population of interest	Variable/Endpoint	Intercurrent event strategy	Population-Level Summary Measure
Secondary objective: To investigate the effects of semaglutide s.c. 2.4 mg once-weekly on walking distance, biomarker of inflammation, body	**Primary**	Semaglutide 2.4 mg or placebo, both added to standard-of-care, regardless of adherence to randomised treatment and initiation of other anti-	All randomised	Change in six-minute walking distance from baseline (week 0) to end of treatment (week 52)	• Discontinuation of trial treatment: Data collected after intercurrent events used in analysis (treatment policy) • Initiation of other weight management drugs or bariatric surgery: Data collected after intercurrent events used in analysis (treatment policy) • Events of CV-death[(a)] and previous heart failure event[(a)] (if data are *not* collected): Will be incorporated into the six-minute walking distance by ascribing the outcome an unfavourable value (composite)	Mean difference of semaglutide 2.4 mg versus placebo

composition, disease specific aspects, social limitation, and health-related quality of life compared with placebo, both added to standard of care, in subjects with obesity-related HFpEF		obesity therapies		Hierarchical composite of:[1] Time to all-cause death, Number of heart failure events requiring hospitalisation or urgent heart failure visit, Time to first heart failure event requiring hospitalisation or urgent heart failure visit, Difference at least 15 in KCCQ clinical summary score change from baseline to 52 weeks, Difference at least 10 in KCCQ clinical summary score change from baseline to 52 weeks, Difference at least 5 in KCCQ clinical summary score change from baseline to 52 weeks, Difference at least 30 metres in six-minute walking distance change from baseline to 52 weeks (assessed by the win ratio)	• Discontinuation of trial treatment: Data collected after intercurrent events used in analysis (treatment policy) • Initiation of other weight management drugs or bariatric surgery: Data collected after intercurrent events used in analysis (treatment policy)	Total wins for each treatment group
				Change in C-Reactive Protein from baseline (week -2) to end of treatment (week 52)	• Discontinuation of trial treatment: Data collected after intercurrent events used in analysis (treatment policy) • Initiation of other weight management drugs or bariatric surgery: Data collected after intercurrent events used in analysis (treatment policy)	Geometric mean ratio of semaglutide 2.4 mg vs placebo
	Secondary (b)	Semaglutide 2.4 mg or placebo, both added to standard-of-care, had all subjects adhered to	All randomised	Change in six-minute walking distance from baseline (week 0) to end of treatment (week 52)	• Discontinuation of trial treatment: "had subjects adhered to treatment" (hypothetical) • Initiation of other weight management drugs or bariatric surgery: "had the subject not initiated weight management drugs or bariatric surgery" (hypothetical) • Events of death: "Had the subject not died" (hypothetical)	Mean difference of semaglutide 2.4 mg versus placebo

[1] Note that the different thresholds of KCCQ-CSS do not change the overall result of the win ratio, but it included to differentiate the individual levels

		randomised treatment and not initiated other anti-obesity therapies		Hierarchical composite of: Time to all-cause death, Number of heart failure events requiring hospitalisation or urgent heart failure visit, Time to first heart failure event requiring hospitalisation or urgent heart failure visit, Difference at least 15 in KCCQ clinical summary score change from baseline to 52 weeks, Difference at least 10 in KCCQ clinical summary score change from baseline to 52 weeks, Difference at least 5 in KCCQ clinical summary score change from baseline to 52 weeks, Difference at least 30 metres in six-minute walking distance change from baseline to 52 weeks (assessed by the win ratio)	• Discontinuation of trial treatment: "had subjects adhered to treatment" (hypothetical) • Initiation of other weight management drugs or bariatric surgery: "had the subject not initiated weight management drugs or bariatric surgery" (hypothetical)	Total wins for each treatment group
				Change in C-Reactive Protein from baseline (week -2) to end of treatment (week 52)	• Discontinuation of trial treatment: "had subjects adhered to treatment" (hypothetical) • Initiation of other weight management drugs or bariatric surgery: "had the subject not initiated weight management drugs or bariatric surgery" (hypothetical) • Events of death: "Had the subject not died" (hypothetical)	Geometric mean ratio of semaglutide 2.4 mg vs placebo

[(a)] As determined by the event adjudication committee: heart failure is either heart failure hospitalization or urgent heart failure visit.
[(b)] Not related to any confirmatory hypothesis

1.1.3 Exploratory objective

The exploratory objective is to investigate the effects of semaglutide s.c. 2.4 mg once-weekly versus placebo, both added to standard of care, in subjects with obesity-related HFpEF regarding:

- Heart failure outcomes
- Change in medications
- Change in biomarkers of myocardial strain
- Echocardiographic parameters of heart failure (in a subset of subjects)

1.2 Trial design

This is a 52-week, randomised, placebo-controlled, double blind, multi-centre clinical trial comparing semaglutide s.c. 2.4 mg with placebo in subjects with obesity-related heart failure with preserved ejection fraction.

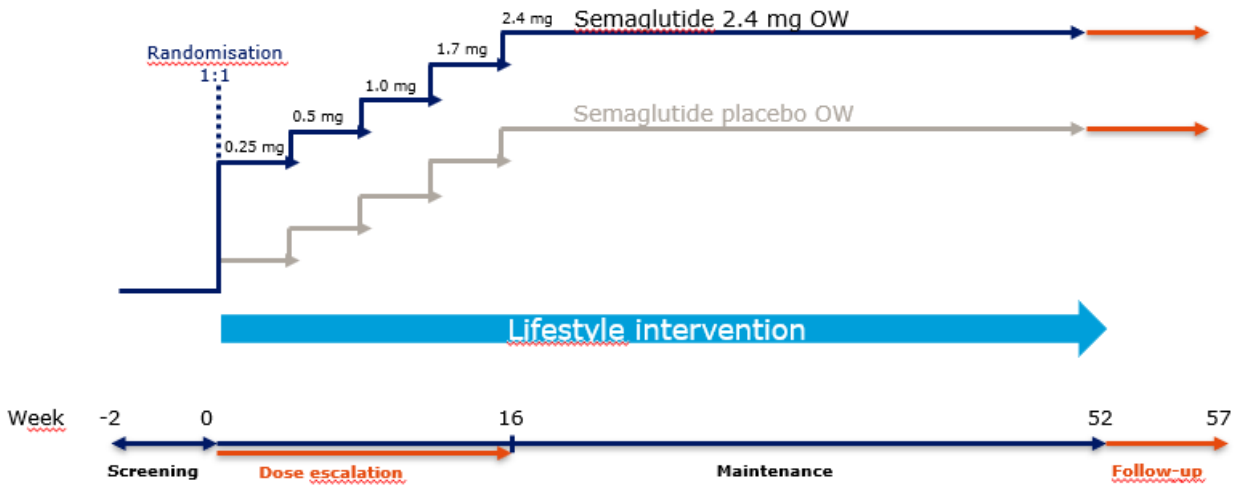
Eligible subjects will be randomised in a 1:1 manner to receive either semaglutide s.c. 2.4 mg or placebo once weekly as add-on to standard of care (Figure 1).

The trial includes a screening visit to assess the subject's eligibility followed by a randomisation visit. A period of 16 weeks of dose escalation is planned to minimise gastrointestinal adverse events with a dose increase every 4th week. Hereafter a visit will take place every 8th week until end of treatment (week 52). Follow up period is 5 weeks after end of treatment.

A subset of 240 randomised subjects will undergo echocardiography assessment at randomisation to ensure at least 180 subjects undergoing echocardiography at the end of treatment.

Randomisation will be stratified by BMI into two subgroups (BMI <35.0 kg/m^2 and BMI ≥35.0 kg/m^2).

Figure 1 A schematic diagram of the trial design



2 Statistical hypotheses

For the primary endpoints, change from baseline in the KCCQ clinical summary score (CSS) and change from baseline in body weight, the following confirmatory one-sided hypotheses are planned to be tested. Let $\mu_{semaglutide}$ and $\mu_{placebo}$ denote the true mean of change from baseline for semaglutide 2.4 mg and placebo group, respectively.

- Superiority of semaglutide 2.4 mg versus placebo in change in KCCQ-CSS from baseline (week 0) to end of treatment (week 52) will be tested as:

$$H: \mu_{semaglutide} \leq \mu_{placebo} \text{ vs}$$
$$H_A: \mu_{semaglutide} > \mu_{placebo}$$

- Superiority of semaglutide 2.4 mg versus placebo in change in body weight from baseline (week 0) to end of treatment (week 52) will be tested as:

$$H: \mu_{semaglutide} \geq \mu_{placebo} \text{ vs}$$
$$H_A: \mu_{semaglutide} < \mu_{placebo}$$

Operationally the hypotheses will be evaluated by two-sided tests.

The hypotheses for the confirmatory secondary endpoint, change in six-minute walking distance (6MWD) from baseline (week 0) to end of treatment (week 52), are the same as for KCCQ-CSS. Similarly, the hypotheses for the confirmatory secondary endpoint, ratio to baseline (week -2) at end of treatment (week 52) in C-Reactive Protein (CRP) are the same as for body weight.

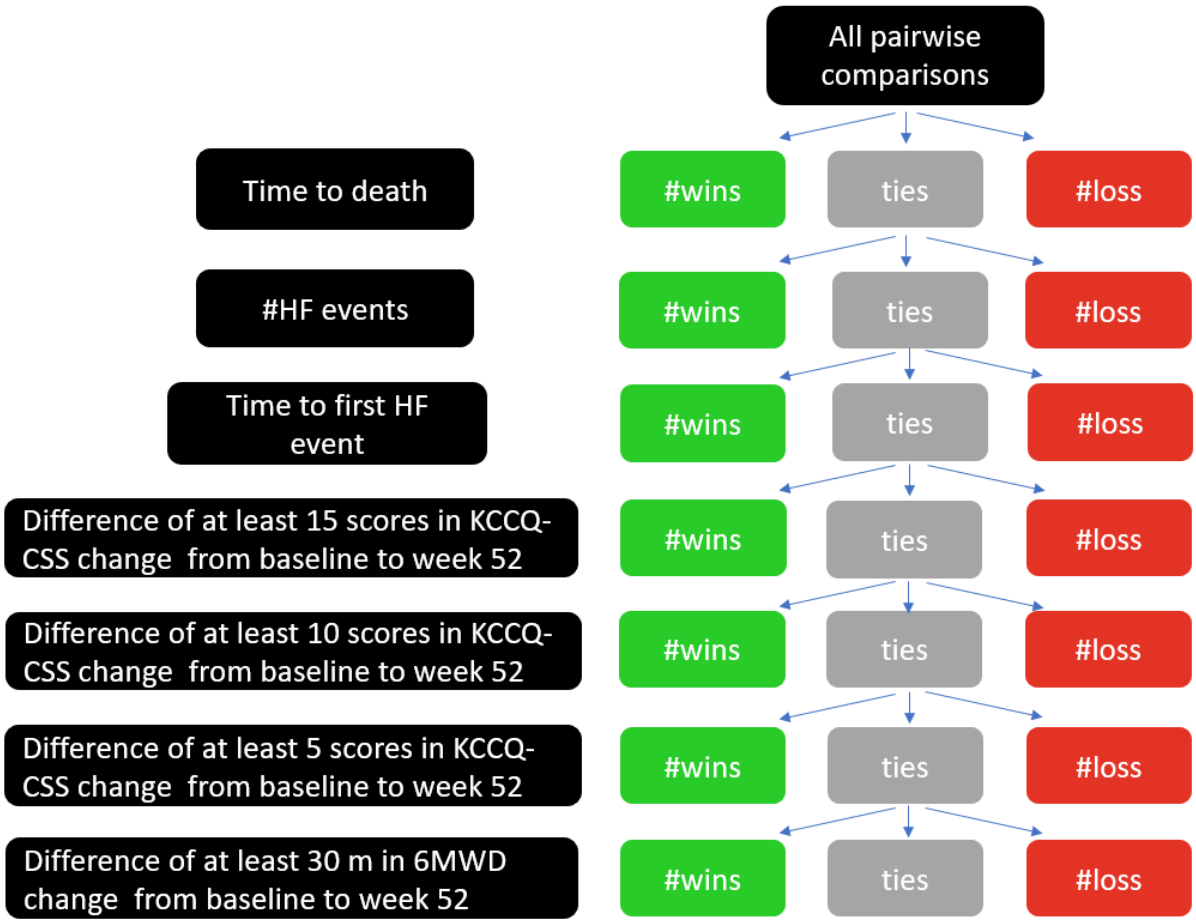
The hierarchical endpoint is a composite of death, number of heart failure events, time to first heart failure event, a difference of at least 15, 10 and 5 scores in KCCQ-CSS change from baseline to week 52 and a difference of at least 30 metres in 6MWD change from baseline to 52 weeks (Figure 2). This will be assessed using a win ratio approach. All patients randomised to semaglutide 2.4 mg are compared to all patient randomised to placebo within each stratum.

The stratified win ratio is calculated as the total number of wins in the semaglutide 2.4 mg group across all strata divided by the total number of losses. A value above one indicates a favourable outcome for semaglutide 2.4 mg as compared to placebo. Formally, we consider the following hypothesis testing for the win ratio statistic:

$$H: WR \leq 1$$

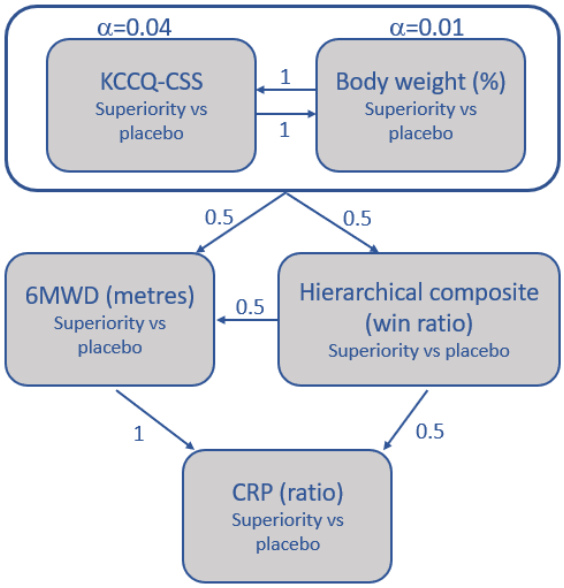
$$H_A: WR > 1$$

Figure 2 The composite hierarchical endpoint



Multiplicity adjustment

The tests of superiority of semaglutide s.c. 2.4 mg once weekly versus placebo for the primary and confirmatory secondary endpoints are performed using a graphical testing system (Figure 3) that preserves the overall study-wise type I error of 5%.



KCCQ-CSS: Kansas City Cardiomyopathy Questionnaire; CRP: C-Reactive Protein; 6MWD: six-minute walking distance. Hierarchical composite endpoint as defined in Section 1.1.2.

Figure 3 Tests of superiority of semaglutide s.c. 2.4 mg once weekly versus placebo for the primary and confirmatory secondary endpoints

First, the multiple primary endpoints are tested, where for these two endpoints the alpha is split in 1% for weight change and 4% for change in KCCQ-CSS. The tests for the multiple endpoint will follow the weighted Holm-Bonferroni procedure[6] (with weight one) such that if one of the two endpoints is superior then the full alpha can be recycled for the other endpoint and hence the remaining primary endpoint will be tested at the 5% significance level (two-sided).

If both hypotheses are rejected and superiority is confirmed for both primary endpoints, then the confirmatory secondary endpoints will be tested according to the flow and weights as specified in Figure 3. The 5% alpha will be 50/50 split between 6MWD and the hierarchical composite endpoint, such that the hierarchical composite endpoint will be tested at 2.5% significance level, and if superiority is confirmed, the 6MWD will be tested at 3.75% significance level. If superiority is not confirmed for the hierarchical composite endpoint, the 6MWD will be tested at 2.5% significance level. If superiority is confirmed for both 6MWD and the hierarchical composite endpoint, CRP will be tested at 5% significance level, but only if 6MWD is confirmed CRP will be tested at 2.5% significance level, or if only the hierarchical composite endpoint is confirmed, CRP will be tested at 1.25% significance level.

3 Sample size determination

See protocol Section 9.2.

4 Analysis sets

The following populations are defined:

Population	Description
Randomised	All subjects randomised
Full analysis set (FAS)	All subjects randomised. Exclusion of data from analyses should be used restrictively, and normally no data should be excluded from the FAS. Subjects will be analysed according to the randomised treatment
Safety analysis set (SAS)	All subjects randomly assigned to trial treatment and who take at least one dose of trial product. Subjects will be analysed according to the treatment they received for the majority of the period they were on treatment
Echocardiographic analysis set (EAS)	All subjects in the sub-population of FAS that are from the sites where the echocardiographic substudy is conducted

Three observation periods are defined for each subject:

- In-trial: The *in-trial period* is defined as the time interval from date of randomisation to
  - Date of last contact with trial site
  - Withdrawal date for subjects who withdraw informed consent
  - Date of the last subject-investigator/site contact as defined by investigator for subjects who are lost to follow up
  - Date of death for subjects who die before any of the above
- On-treatment (with trial product): A time-point is considered as 'on-treatment' if any dose of trial product has been administered within the prior 2 weeks (14 days). The *on-treatment period* is defined as all times which are considered on-treatment.
  - In general, the on-treatment period will therefore be from the date of first trial product administration to date of last trial product administration excluding potential off-treatment time intervals triggered by at least two consecutive missed doses.
  - For the evaluation of AEs, the lag time for each on-treatment time interval is 5 weeks (35 days).
- On-treatment without other anti-obesity therapies: The time period where subjects are considered to be treated with trial product and have not initiated other weight management drugs or bariatric surgery (i.e. anti-obesity therapies). The period begins at the date of first trial product administration and ends at the earliest date of:
  - The end date of the on-treatment observation period
  - Initiation of other anti-obesity therapies

The in-trial and on-treatment periods define the patient years of observation (PYO) and patient years of exposure (PYE), respectively, as the total time duration in the periods.

5 Statistical analyses

5.1 General considerations

Novo Nordisk will be responsible for the statistical analysis and reporting. Analysis and reporting will be based on pooled data from all sites and will be performed on un-blinded data after database release.

Effect endpoints will be analysed using the FAS; safety endpoints will be analysed using the SAS.

The comparison presented from a statistical analysis will be semaglutide 2.4 mg versus placebo and results will be presented by the estimated treatment contrast with associated two-sided 95% confidence intervals (CIs) and p-values corresponding to two-sided tests of no difference.

Data from all countries and sites will be analysed and reported together.

The last available and eligible observation at or before randomisation is used as the baseline value. If no assessments are available, the mean of baseline values across all subjects is used as the baseline value.

The stratification factor is BMI subgroup (BMI <35.0 kg/m^2, BMI ≥35.0 kg/m^2). Wrong stratification, e.g. if a subject with a BMI <35.0 kg/m^2 was mistakenly randomised using the randomisation list for subjects with a BMI ≥35 kg/m^2, the actual value of the stratification factor should in general be used in the analysis (i.e. <35 kg/m^2 should be used as factor level in the example above).

Laboratory values below the lower limit of quantification (LLOQ) will be set to ½LLOQ. Laboratory values above the upper limit of quantification (ULOQ) will be set to ULOQ.

All continuous variables will be summarized with n, mean, standard deviation, median, geometric mean, CV, min and max. When relevant, number of values n<LLOQ and n>ULOQ will also be presented.

For each subject a given assessment at week 52 may be available or missing and Table 5-1 describes the taxonomy for this. Note, this is done per assessment and per subject; subjects may be a different type for different assessments (a subject may have “available on randomised treatment” for body weight but “missing on randomised treatment” for waist circumference).

In the specific case where cause of death is not determined by the event adjudication committee (e.g. if the subject was lost to follow-up and hereafter died) it will be assumed the cause of death is CV death.

Table 5-1 Taxonomy for subjects based on week 52 assessments

Assessment at week 52	Subjects on randomised treatment at week 52	Type description
Available	Yes	**Available on randomised treatment**: Subjects who complete the trial on randomised treatment with an assessment at week 52: Includes those that stop and restart trial product.
	No	**Available but discontinued** Subjects who discontinued randomised treatment prematurely but returned to have an assessment at week 52. These are also called *retrieved subjects.*
Missing	Yes	**Missing on randomised treatment**: Subjects who complete the trial on randomised treatment without an assessment at week 52: Includes those that stop and restart trial product.
	No	**Missing and discontinued**: Subjects who discontinued randomised treatment prematurely and did not return to have an assessment at week 52. These are also called *non-retrieved subjects.*

5.2 Subject disposition

See mock TFLs.

5.3 Primary endpoints analyses

5.3.1 Definition of endpoints

The primary endpoints are change in the KCCQ-CSS (see Appendix 6.4) and change in body weight in percentage from baseline (week 0) to end of treatment (week 52) as listed in Appendix 6.3.

Change from baseline (week 0) to end of treatment (week 52) in body weight (%) is defined as

$$\frac{(\textit{body weight at week 52} - \textit{body weight at baseline})}{\textit{body weight at baseline}} \times 100.$$

5.3.2 Main analytical approach

5.3.2.1 Primary estimand

In line with the primary estimand strategy (see Section 1.1.1), the primary analysis of the primary endpoints will be based on all randomised subjects (full analysis set) and the in-trial observation period.

The superiority tests of semaglutide 2.4 mg vs. placebo will be carried out as described in Section 2.

All available data at week 52 are used and missing values at week 52 will be imputed and the endpoint will be derived from the imputed values. Missing data will be handled differently based on whether it is missing due to an intercurrent event or not.

The strategy for the handling of intercurrent events for the primary estimand is outlined in Table 5-2.

Subjects who have missing KCCQ-CSS values at week 52 due to cardiovascular (CV) death or subjects who have missing KCCQ-CSS values at week 52 with a previous heart failure event, as determined by the event adjudication committee, will be imputed using a composite strategy. More specifically, a single imputation approach where the lowest KCCQ-CSS value across treatment groups and time points will be used. Based on this value the change from baseline will be calculated individually for the subjects in scope.

A tipping point analysis will address the sensitivity of assigning these unfavourable values, see Section 5.3.3.

Table 5-2 Handling of intercurrent events for the primary estimand

Intercurrent event (endpoints)	Data collection	Data analysis
Discontinuation of trial treatment (KCCQ-CSS and body weight)	Subjects will be followed, and data collected after intercurrent events	Data collected after intercurrent events used in analysis (treatment policy strategy)
Initiation of other weight management drugs or bariatric surgery (KCCQ-CSS and body weight)		
CV death[(c)] (KCCQ-CSS only)	-	CV deaths will be incorporated into the KCCQ-CSS by ascribing the outcome an unfavourable value (composite strategy)
Heart failure hospitalization[(c)] (KCCQ-CSS only) Urgent heart failure[(c)] (KCCQ-CSS only)	Subjects will be followed, and data collected after intercurrent events	If data are not collected heart failure events will be incorporated into the KCCQ-CSS by ascribing the outcome an unfavorable value (composite strategy)

[(c)]As determined by the event adjudication committee

Subjects who have missing KCCQ-CSS values at week 52 due to other reasons than CV related death or previous heart failure events will have their values imputed using a missing at random (MAR) assumption.

Missing body weight values due to both CV death, non-CV death and with previous heart failure events will be imputed using a MAR assumption to allow bridging to the results of the STEP programme.

Multiple imputation of KCCQ-CSS

Missing data for the primary endpoint that do not arise from intercurrent events (e.g. lost to follow-up or withdrawal from the trial) will be imputed by multiple imputation approach similar to the one described by McEvoy et al.[1]

For subjects in the semaglutide 2.4 mg and the placebo groups, missing primary endpoint measurements at week 52 for non-retrieved (see Table 5-1) subjects are imputed using assessments from retrieved subjects in each treatment group. This will be done according to the timing of last available observation during the on-treatment period (LAO-OT) of KCCQ-CSS.

Missing measurements for the primary endpoint at week 52 for subjects on randomised treatment (at week 52) are imputed by sampling from available measurements at week 52 from subjects on randomised treatment in the relevant randomised treatment group.

The multiple imputation for KCCQ-CSS is done by defining an imputation model using retrieved subjects from the full analysis set and done within groups defined by randomised treatment and the timing of the LAO-OT of the endpoint. The model will be a linear regression of KCCQ-CSS at week 52 with gender (male/female) and baseline BMI (kg/m^2) (in categories <35 kg/m^2, ≥ 35 kg/m^2) as factors and baseline KCCQ-CSS, baseline body weight (kg), LAO-OT of KCCQ-CSS and timing of LAO-OT of KCCQ-CSS as covariates. No interactions will be included. If the imputation model cannot be fit then the model will be reduced until the model can fit. Reduction will be done in a fixed order by first removing the gender and then the baseline BMI. If the imputation model still cannot be fit, the imputation will be done regardless of the randomised treatment arm. If no LAO-OT exists post-baseline then the LAO-OT will be the baseline KCCQ-CSS. If any subjects are on-treatment and have missing values for KCCQ-CSS at week 52, an imputation model for missing KCCQ-CSS measurements at week 52 for these subjects will also be defined in a similar way using subjects that are on-treatment and have available values for KCCQ-CSS. The estimated posterior distribution for the parameters (regression coefficients and variances) in the imputation models are then used to impute missing week 52 KCCQ-CSS values for each randomised treatment arm. This will be done 1,000 times resulting in 1,000 complete data sets.

Analysis of KCCQ-CSS

The change from baseline to week 52 will be analysed using an analysis of covariance (ANCOVA) with randomised treatment and stratification (BMI <35.0 kg/m^2, BMI ≥ 35.0 kg/m^2) as factors, and baseline KCCQ-CSS as covariate for each of the 1000 complete data sets. The 1000 sets of analysis results will be combined using Rubin's rule to draw inference.

Based on NN9536-4153 phase 2 results 1,000 copies should be sufficient to establish stable results. If 1,000 copies are insufficient, 10,000 copies will be used. The multiple imputations will be generated using Novo Nordisk trial number 95364665 as seed number. In addition to the seed number, it is specified that the dataset is sorted by subject ID.

The estimated treatment difference between semaglutide 2.4 mg and placebo will be reported together with the associated two-sided 95% CI and corresponding p-value.

Multiple imputation and analysis of body weight

Imputation of missing data due to intercurrent events will be performed as described previously in this section. Multiple imputation of body weight will be done in a similar manner as for KCCQ-CSS, with the change that baseline KCCQ-CSS will not be included in the imputation model.

In the analysis of change from baseline to week 52 in body weight (%), baseline body weight (kg) will be used instead of baseline KCCQ-CSS as covariate.

5.3.2.2 Secondary estimand

In line with the secondary estimand strategy described in Section 1.1.1, the secondary estimand for the primary endpoints will be based on all randomised subjects (full analysis set) and the 'on-treatment without other anti-obesity therapies' observation period (see Section 4). In the secondary estimand observations after a heart failure event will be used if collected.

The secondary estimand for the primary endpoints addresses the efficacy of semaglutide 2.4 mg and will be assessed using a mixed model for repeated measurements (MMRM). Week 52 assessments from retrieved subjects are not used in this analysis. The MMRM will use assessments only from subjects who are taking the randomised treatment until end of treatment or until first discontinuing of randomised treatment.

The derived date of the second consecutive missed dose will be used as the latest date for using assessments in the MMRM. The assessment closest in time and before the derived date of the second consecutive missed dose will be used as last assessment on randomised treatment. For subjects who initiate other anti-obesity therapies, as defined in Section 4, before completion or first discontinuing of randomised treatment, the date of starting the other anti-obesity therapies will be used as latest date for using assessments in this MMRM.

Similarly, the assessment closest in time and before the date of starting other anti-obesity therapies will be used as last assessment on randomised treatment.

The MMRM will be fitted using KCCQ-CSS and the same factors and covariate as for the primary analysis model (see Section 5.3.2.1, "Analysis of KCCQ-CSS") all nested within visit. An unstructured covariance matrix for measurements within the same subject will be used to describe the variability for the repeated measurements for a subject.

From the MMRM model the estimated treatment difference between semaglutide 2.4 mg and placebo will be reported together with the associated two-sided 95% CI and corresponding p-value.

Analysis of body weight will be done in the same manner as for KCCQ-CSS.

5.3.3 Sensitivity analyses

Tipping point sensitivity analyses

As a sensitivity analysis for the primary estimand, a two-dimensional tipping point analysis will be performed. First, missing data are imputed according to the primary multiple imputation approach. Then, a penalty is added to the imputed values at end of treatment. The approach is to explore a range of penalties for both treatment groups, and the impact these would have on the trial conclusions. This sensitivity analysis evaluates the robustness of the superiority conclusions to departures from the observed change in KCCQ-CSS and body weight in both treatment groups.

Additionally, a tipping point analysis for KCCQ-CSS will be performed on imputed values due to CV death or previous heart failure events in a similar manner as for the tipping point analysis of the

multiple imputed values. This sensitivity analysis evaluates the robustness of the superiority conclusions in a setting where imputed values are added penalties using increasingly unfavourable values (including zero in the worst-case scenario).

5.3.4 Supplementary analyses

Supplementary analyses using rank ANCOVA

To access the composite strategy in relation to missing data due to CV death, non-CV death and heart failure events a non-parametric approach will be performed using the rank ANCOVA as a supplementary analysis for KCCQ-CSS.

The clinical importance of change in KCCQ-CSS will be applied for each subject where values within each order will be ranked according to Table 5-3.

Table 5-3 Ranking of KCCQ-CSS according to intercurrent and fatal events

Order	End of treatment (week 52) classification	Sub-ordering
1	Observed/imputed following a MAR[a] approach	Change from baseline ranked from highest to lowest (no priority between observed/imputed values)
2	Previous heart failure event in trial (if data at week 52 are not collected)	Ranking by number of events (more events ranked lowest) and then ranking by time to *first* event (shortest time ranked lowest)
3	Non-CV deaths	Ranking by time to event (shortest time ranked lowest)
4	CV deaths	Ranking by time to event (shortest time ranked lowest)

[a] The 1000 complete data sets generated for the primary imputation approach will be reused and pooled together summarising by taking the mean of the response values by subject. Order represents ranking from highest to lowest.

After the ranking of values, the rank ANCOVA will be performed, first adjusting for baseline KCCQ-CSS in a linear regression, and then use the Van Elteren test (stratified Wilcoxon) using stratification (BMI <35 kg/m^2, ≥35 kg/m^2) on the corresponding residuals. A (two-sided) p-value for the test of treatment difference between semaglutide 2.4 mg and placebo will be reported.

5.4 Secondary endpoints analyses

5.4.1 Confirmatory secondary endpoints

5.4.1.1 Definition of endpoints

The confirmatory secondary endpoints are:

- Change in 6MWD from baseline (week 0) to end of treatment (week 52)
- The hierarchical composite endpoint from baseline (week 0) to end of study (week 57)
- Change in CRP from baseline (week -2) to end of treatment (week 52)

A schematic overview of the endpoints is included in Appendix 6.3.

5.4.1.2 Main analytical approach

The superiority tests of semaglutide 2.4 mg vs. placebo will be carried out in a hierarchical order as described in Section 2.

Analyses addressing the primary estimand

Overall, the attributes of the primary estimand described in Section 5.3.2.1 is carried over to the confirmatory analysis of the secondary endpoints 6MWD and CRP, and specifically the intercurrent event strategy for 6MWD will mimic what was done for KCCQ-CSS, whereas the intercurrent event strategy for CRP will mimic body weight, see Table 5-2. This means that missing values for 6MWD due to CV death or previous heart failure event will be imputed using the lowest 6MWD value across treatment groups and time points. Based on this value the change from baseline will be calculated individually for the subjects in scope.

The analysis will be based on all randomised subjects (full analysis set) and the in-trial observation period.

The confirmatory secondary endpoints 6MWD and CRP will be analysed using the same imputation approach as used for the primary endpoints addressing the primary estimand. The imputation model is the same as described in Section 5.3.2.1 (multiple imputation of KCCQ-CSS) with KCCQ-CSS replaced by assessments of the endpoint to be analysed.

Similarly, the statistical model will be as described in Section 5.3.2.1 (bullet 2: analysis), i.e. including randomised treatment and stratification as factors, and associated baseline as covariate. For CRP both the dependent variable and covariate will be log-transformed and the ratio to baseline (week -2) to end of treatment (week 52) will be calculated instead of differences.

The analysis of the hierarchical composite endpoint will be based on direct comparisons of each subject randomised to semaglutide s.c. 2.4 mg versus each subject randomised to placebo within each stratum. For each of these pairs, a 'treatment winner' based on similar observation time will be declared based on the endpoint hierarchy. The win ratio (i.e. the number of winners randomised to semaglutide s.c. 2.4 mg divided by the number of winners randomised to placebo) will be reported together with the associated two-sided 95% CI and corresponding p-value. Furthermore, the contribution of wins from each individual component of the hierarchical composite endpoint will be summarised.

Missing KCCQ-CSS and 6MWD values due to other reasons than CV related death or previous heart failure events will be imputed using the MAR assumption as described in Section 5.3.2.1.

The analysis of the hierarchical composite endpoint will be done using a stratified win ratio approach, see Gaohong Dong, J. Q. (2018). The stratified win ratio. *Journal of Biopharmaceutical Statistics*, 778-796. The method can briefly be described as:

Consider a clinical trial with patients randomised into two groups with M strata. Let $N_t^{(m)}$ and $N_c^{(m)}$ denote the number of patients in the treatment group and the control group, respectively, and $N^{(m)} = N_t^{(m)} + N_c^{(m)}$ the total sample size in the mth stratum $(m = 1,2, \dots, M)$. We define the weighted stratified win ratio as

$$WR = \frac{\sum_{m=1}^{M} w^{(m)} n_t^{(m)}}{\sum_{m=1}^{M} w^{(m)} n_c^{(m)}}$$

In the unstratified situation ($M = 1$), this reduces to $n_t^{(1)}/n_c^{(1)}$, the original win ratio. With two strata and the weight defined as $w^{(m)} = 1/N^{(m)}$ (Mantel-Haenszel), the weighted stratified win ratio is given by

$$WR = \frac{\sum_{m=1}^{2} n_t^{(m)}/N^{(m)}}{\sum_{m=1}^{2} n_c^{(m)}/N^{(m)}}$$

The variance is calculated by the asymptotic normal U statistic approach[5].

Analyses addressing the secondary estimand

Analyses of change in 6MWD and change in CRP will be done following the same approach as described in Section 5.3.2.2, where the analysis for change in 6MWD will mimic the analysis for change in KCCQ-CSS and the analysis for change in CRP will mimic the analysis for change in body weight. CRP will be analysed on log-scale in line with the strategy of the main analytical approach for CRP. The analysis will be based on all randomised subjects (full analysis set) and the 'on-treatment without other anti-obesity therapies' observation period

Similarly the hierarchical composite endpoint will be based on all randomised subjects (full analysis set) and the 'on-treatment without other anti-obesity therapies' observation period.

5.4.1.3 Sensitivity analysis

Sensitivity analysis for 6MWD will follow the approach described for KCCQ-CSS in Section 5.3.3 where the tipping point analyses will be applied using the same methodology as for KCCQ-CSS. CRP will follow the approach described for body weight.

5.4.1.4 Supplementary analyses

The supplementary analyses for 6MWD and CRP will follow the approach described in Section 5.3.4.

5.4.2 Supportive secondary endpoints

A schematic overview of the supportive secondary endpoints is included in Appendix 6.3 and a brief overview is given in Table 5-4.

The supportive secondary endpoints will be analysed using the same imputation approach as used for the confirmatory secondary endpoints and to address the primary estimand. The intercurrent event strategy for systolic blood pressure and waist circumference will mimic what was done for body weight, whereas the intercurrent event strategy for KCCQ overall summary score (KCCQ-OSS) will mimic KCCQ-CSS, see Table 5-2. The analysis will be based on all randomised subjects (full analysis set) and the in-trial observation period.

For the continuous endpoints, the analysis model will be with randomised treatment and stratification as factors, and associated baseline as covariate. The analyses will be on original scale

(i.e. no log-transformation). The responder (binary) endpoints will be derived from the 1000 imputed data sets described in Section 5.3.2.1. For the responder endpoints, each of the complete data sets will be analysed using a logistic regression (LR) with randomised treatment and stratification as factors, and associated baseline value as covariate. Estimated odds ratios will be log-transformed and inference will be drawn using Rubin's rule. The results will be back-transformed and described by the odds ratio between treatments and the associated 95% CI and p-value for no treatment difference.

5.4.2.1 Clinically meaningful within-subject change endpoints

The KCCQ-CSS scale measures symptoms and physical function for participants with heart failure. Superiority with respect to KCCQ-CSS is tested on the continuous scale in the confirmatory testing hierarchy, see Section 2. If superiority is demonstrated, the clinical relevance of the treatment effect is evaluated based on the proportions of participants in each treatment group that have experienced clinically relevant improvements with respect to symptoms and physical function as measured by KCCQ-CSS.

Likewise, superiority with respect to 6MWT is tested as a confirmatory endpoint. If superiority is demonstrated, the clinical relevance of the treatment effect is evaluated based on the proportions of participants in each treatment group that have experienced clinically relevant improvements in walking distance.

These clinically meaningful improvements can be characterised as the smallest difference in an endpoint that is perceived by the subject to be beneficial[3]. In order to characterise a meaningful within-subject change in KCCQ-CSS and 6MWD the patient global impression of severity (PGI-S) will be used in an anchor based approach, where the PGI-S is the so-called anchors. The responder definition values will be calculated for KCCQ-CSS and 6MWD by an external vendor.

The proportion of subjects improving equal or more than these thresholds based on PGI-S will be considered as supportive secondary endpoints (see Table 5-4) and will be analysed following the methods in Section 5.4.2.

Furthermore, to be able to visually evaluate the robustness of the results based on the PGI-S 1 point improvement threshold, the distribution of the change in KCCQ-CSS and change in 6MWD from baseline will be illustrated in an empirical Cumulative Distribution Function plot by treatment arms. To further aid this evaluation, the proportion of subjects improving equal or more than these thresholds based on the PGI-S 0 point improvement and PGI-S 2 point improvement will be analysed following the methods in Section 5.4.2.

Table 5-4 Statistical analysis of supportive secondary endpoints

Title	Statistical model
Subject achieving 10 % weight loss or more (Yes/No)	LR
Subject achieving 15 % weight loss or more (Yes/No)	LR
Subject achieving 20 % weight loss or more (Yes/No)	LR
Subject improving 5 points or more in KCCQ clinical summary score (Yes/No)	LR
Subject improving 10 points or more in KCCQ clinical summary score (Yes/No)	LR
Change in KCCQ overall summary score	ANCOVA
Change in systolic blood pressure (SBP) (mmHg)	ANCOVA

Change in waist circumference (cm)		ANCOVA
Subjects achieving CMWSC or more in KCCQ-CSS (Yes/No)	(PGI-S)	LR
Subjects achieving CMWSC or more in 6MWD (Yes/No)	(PGI-S)	LR

CMWSC**:** Clinically meaningful within-subject change

5.5 Exploratory endpoints analysis

The exploratory endpoints in the trial are listed in Section 6.3. The exploratory endpoints to be statistically analysed are:

- Change in NT-proBNP from baseline (week -2) to end of treatment (week 52)
- Time to first heart failure event (hospitalisation or urgent visit)
- Echocardiographic substudy endpoints
- Change in KCCQ subscales from baseline (week 0) to end of treatment (week 52)
- Endpoints of KCCQ

Remaining exploratory endpoints will be summarised as described in mock TFLs.

NT-proBNP

Analysis of NT-proBNP will be performed as described in Section 5.3.2.2 using the strategy of the secondary estimand. The MMRM will be fitted using the same factors and covariate, substituting KCCQ-CSS at baseline with (log-transformed) NT-proBNP as baseline, as for the primary analysis all nested within visit. As indicated, both the dependent variable and covariate will be log-transformed and the ratio to baseline (week -2) to end of treatment (week 52) will be calculated instead of differences.

Time to first heart failure event

Time to first occurrence of a heart failure event from randomisation will be analysed using a Cox proportional hazards model stratified by baseline BMI (kg/m^2) (in categories <35 kg/m^2, ≥ 35 kg/m^2) with treatment as a covariate. Subjects without an event will be censored at either end of treatment visit (week 52), permanent discontinuation or all-cause death, whichever comes first.

A cumulative incidence plot will be produced using the Nelson-Aalen estimator adjusting for competing risk (all-cause death) for the two treatment groups. The difference between the cumulative incidences at week 52 for the treatment groups will be tested using a log-rank test.

Echocardiographic substudy

The echocardiographic substudy endpoints are:

- Change in left atrial volume from baseline (week 0) to end of treatment (week 52)
- Change in left ventricular (LV) filling pressure (diastolic function) (E/e´) from baseline (week 0) to end of treatment (week 52)
- Change in global longitudinal strain from baseline (week 0) to end of treatment (week 52)

Analysis of the echocardiographic substudy endpoints will be performed on all subjects with observed values at baseline and at week 52 in the echocardiographic analysis set using an ANCOVA with randomised treatment and stratification (BMI <35 kg/m^2, ≥35 kg/m^2) as factor and baseline value as covariate. Left atrial volume and LV filling pressure will be analysed on log-scale i.e. the ratio between end of treatment and baseline will be calculated instead of differences, and both the dependent variable and covariate will be log-transformed.

Endpoints of KCCQ

A schematic overview of the exploratory KCCQ endpoints is included in Appendix 6.3 and a brief overview is given in Table 5-5.

Table 5-5 Explorative KCCQ endpoints

Title	Statistical model
Change in KCCQ total symptom score	ANOVA
Change in KCCQ physical limitations score	ANOVA
Change in KCCQ social limitations score	ANOVA
Change in KCCQ health-related quality of life score	ANOVA
Subject worsening 5 points or more in KCCQ clinical summary score (Yes/No)	LR
Subject worsening 10 points or more in KCCQ clinical summary score (Yes/No)	LR
Subject improving 15 points or more in KCCQ clinical summary score (Yes/No)	LR
Subject worsening 15 points or more in KCCQ clinical summary score (Yes/No)	LR
Subject improving 5 points or more in KCCQ overall summary score (Yes/No)	LR
Subject improving 10 points or more in KCCQ overall summary score (Yes/No)	LR
Subject improving 15 points or more in KCCQ overall summary score (Yes/No)	LR
Subject worsening 5 points or more in KCCQ overall summary score (Yes/No)	LR
Subject worsening 10 points or more in KCCQ overall summary score (Yes/No)	LR
Subject worsening 15 points or more in KCCQ overall summary score (Yes/No)	LR

Continuous endpoints will be analysed using the same imputation approach as used for the confirmatory secondary endpoints and to address the primary estimand. The intercurrent event strategy will mimic KCCQ-CSS, see Table 5-2. The analysis will be based on all randomised subjects (full analysis set) and the in-trial observation period.

The responder endpoints of improving and worsening will be analysed in the same manner as described for responder endpoints in Section 5.4.2.

5.6 Safety analyses

There are no safety endpoints in the trial. Safety assessments will be summarised descriptively by treatment group and visit. Categorical safety assessments will be summarised as counts and relative frequencies.

5.7 Other analyses

At the end of treatment visit, subjects are asked which treatment (semaglutide or placebo) that they believe they have received. The results will be summarised.

Treatment differences at other time points before end of treatment (week 52) for the confirmatory endpoints could be analysed using the secondary estimand.

COVID-19

A supplementary analysis using the primary estimand will be performed for all confirmatory endpoints excluding subjects who experience a COVID-19 related adverse event. If less than 5 subjects have experienced a COVID-19 related adverse event the analyses will be excluded.

A supplementary analysis will be performed using the primary estimand for all confirmatory endpoints excluding subjects who co-participates in COVID-19 vaccination trials. If less than 5 subjects are co-participating in COVID-19 vaccination trials the analyses will be excluded.

5.7.1 Subgroup analyses

No subgroup analyses are planned to be reported in the trial specific study report.

5.7.2 Pharmacokinetic modelling

Exposure-response analyses will be used as supportive evidence for the evaluation of efficacy and safety and further to support the investigated dose of semaglutide in subjects with HFpEF and obesity. Plasma semaglutide concentrations will be used to derive model-based estimates of steady-state average concentrations for each subject, utilizing a population pharmacokinetic modelling approach that leverages information from the STEP programme. Relevant efficacy and safety endpoints will be related to steady-state average concentrations and subjected to model-based exposure-response analysis. A modelling analysis plan will be prepared prior to database lock outlining details of the analyses. The results from the exposure-response analysis will be reported separately from the CTR.

5.8 Interim analyses

No interim analyses are planned.

5.8.1 Data monitoring committee

There is no data monitoring committee for this trial.

6 Supporting documentation

6.1 Appendix 1: List of abbreviations

6MWD	six-minute walking distance
AE	adverse event
ANCOVA	analysis of covariance
BMI	body mass index
CI	confidence interval
COVID	coronavirus disease
CRP	C-Reactive Protein
CSS	clinical summary score
CV	cardiovascular
EAS	echocardiographic analysis set
FAS	full analysis set
HFpEF	heart failure with preserved ejection fraction
KCCQ	Kansas City Cardiomyopathy Questionnaire
LAO-OT	last available observation during the on-treatment period
LLOQ	lower limit of quantification
LR	logistic regression
LV	left ventricular
MAR	missing at random
MMRM	mixed model for repeated measurements
NT-proBNP	N-terminal pro-brain natriuretic peptide
NYHA	New York Heart Association
OSS	overall summary score
PGI-S	patient global impression of severity
PGI-C	patient global impression of change

PYE	patient years of exposure
PYO	patient years of observation
SAP	statistical analysis plan
SBP	systolic blood pressure
SAS	safety analysis set
s.c.	subcutaneous
TFL	tables, figures and listings
ULOQ	upper limit of quantification

6.2 Appendix 2: Changes to protocol-planned analyses

The changes from the protocol of EX9536-4665 are summarised below.

- The following exploratory endpoints were added
  - Subject improving 15 points or more in KCCQ clinical summary score
  - Subject worsening 15 points or more in KCCQ clinical summary score
  - Subject improving 15 points or more in KCCQ overall summary score
  - Subject worsening 15 points or more in KCCQ overall summary score

6.3 Appendix 3: Definition and calculation of endpoints, assessments and derivations

Type	Title	Time frame	Unit	Details
Primary endpoint	Change in KCCQ clinical summary score	From baseline (week 0) to end of treatment (week 52)	Score (no unit, range; 0-100)	
Primary endpoint	Change in body weight	From baseline (week 0) to end of treatment (week 52)	%	
Confirmatory secondary endpoint	Change in six-minute walking distance (6MWD)	From baseline (week 0) to end of treatment (week 52)	Metres	
Confirmatory secondary endpoint	Hierarchical composite of: Time to all-cause death, number of heart failure events requiring hospitalisation or urgent heart failure visit, time to first heart failure event requiring hospitalisation or urgent heart failure visit, difference of at least 15 in KCCQ clinical summary score change from baseline to 52 weeks, difference of at least 10 in KCCQ clinical summary score change from baseline to 52 weeks, difference of at least 5 in KCCQ clinical summary score change from baseline to 52 weeks, difference of at least 30 metres in sex-minute walking distance change from baseline to 52 weeks (assessed by the win ratio)	From baseline (week 0) to end of study (week 57)	Total wins for each treatment group	
Confirmatory secondary endpoint	Change in C-Reactive Protein (CRP)	From baseline (week -2) to end of treatment (week 52)	Ratio to baseline (no unit)	Log transformed
Supportive secondary endpoint	Subject achieving 10 % weight loss or more (Yes/No)	From baseline (week 0) to end of treatment (week 52)	Count of subjects	
Supportive secondary endpoint	Subject achieving 15 % weight loss or more (Yes/No)	From baseline (week 0) to end of treatment (week 52)	Count of subjects	
Supportive secondary endpoint	Subject achieving 20 % weight loss or more (Yes/No)	From baseline (week 0) to end of treatment (week 52)	Count of subjects	

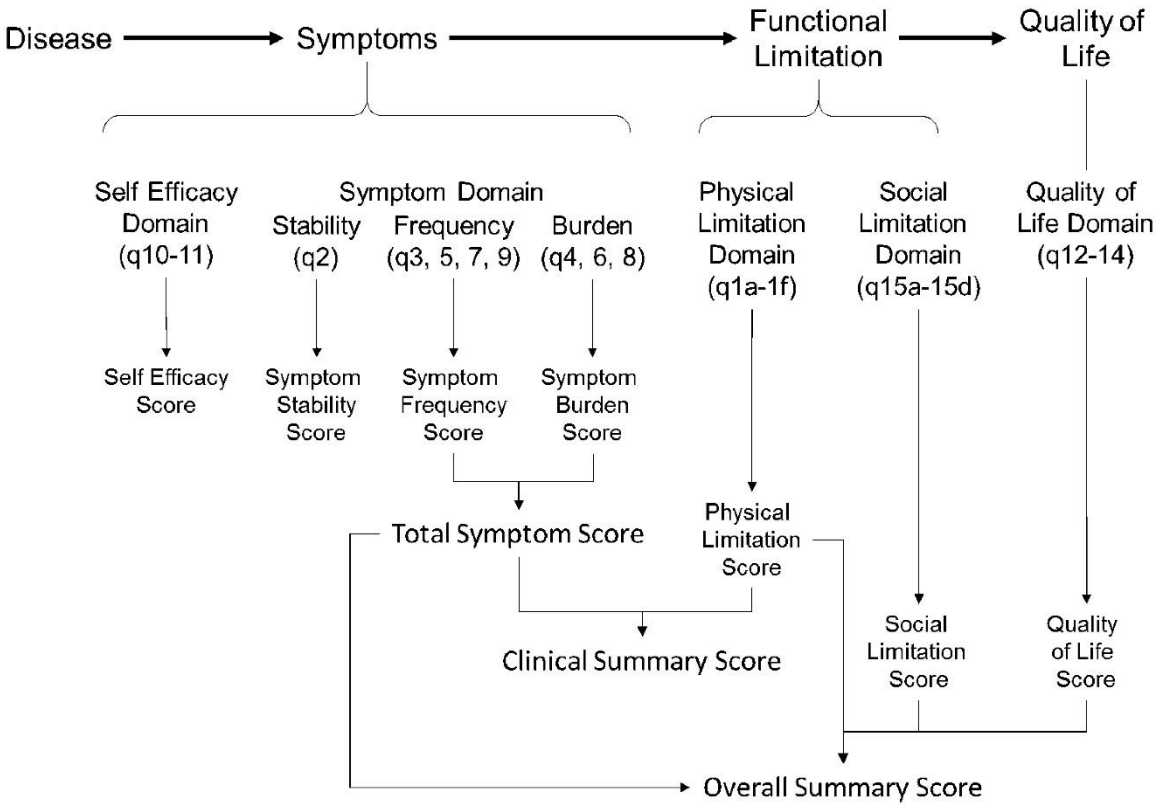
Type	Title	Time frame	Unit	Details
Supportive secondary endpoint	Subject improving 5 points or more in KCCQ clinical summary score (Yes/No)	From baseline (week 0) to end of treatment (week 52)	Count of subjects	
Supportive secondary endpoint	Subject improving 10 points or more in KCCQ clinical summary score (Yes/No)	From baseline (week 0) to end of treatment (week 52)	Count of subjects	
Supportive secondary endpoint	Change in KCCQ overall summary score	From baseline (week 0) to end of treatment (week 52)	Score (no unit, range 0-100)	
Supportive secondary endpoint	Change in systolic blood pressure (SBP)	From baseline (week -2) to end of treatment (week 52)	mmHg	
Supportive secondary endpoint	Change in waist circumference	From baseline (week 0) to end of treatment (week 52)	cm	
Supportive secondary endpoint	Subject achieving the threshold for clinically meaningful within-subject change or more in KCCQ-CSS (PGI-S)	From baseline (week 0) to end of treatment (week 52)	Count of subjects	
Supportive secondary endpoint	Subject achieving the threshold for clinically meaningful within-subject change or more in 6MWD (PGI-S)	From baseline (week 0) to end of treatment (week 52)	Count of subjects	
Exploratory endpoint	Change in antihypertensive medication	From baseline (week 0) to end of treatment (week 52)	Category (no unit; decrease / no change / increase)	
Exploratory endpoint	Change in loop diuretic medication	From baseline (week 0) to end of treatment (week 52)	Category (no unit, decrease / no change / increase)	
Exploratory endpoint	Change in NT-proBNP	From baseline (week -2) to end of treatment (week 52)	Ratio to baseline (no unit)	Log-transformed
Exploratory endpoint	Change in EQ-5D-5L score	From baseline (week 0) to end of treatment (week 52)	Score (no unit, range per item; 1-5)	
Exploratory endpoint	Subject worsening 5 points or more in KCCQ clinical summary score (Yes/No)	From baseline (week 0) to end of treatment (week 52)	Count of subjects	
Exploratory endpoint	Subject worsening 10 points or more in KCCQ clinical summary score (Yes/No)	From baseline (week 0) to end of treatment (week 52)	Count of subjects	
Exploratory endpoint	Subject improving 15 points or more in KCCQ clinical summary score (Yes/No)	From baseline (week 0) to end of treatment (week 52)	Count of subjects	
Exploratory endpoint	Subject worsening 15 points or more in KCCQ clinical summary score (Yes/No)	From baseline (week 0) to end of treatment (week 52)	Count of subjects	
Exploratory endpoint	Subject improving 5 points or more in KCCQ overall summary score (Yes/No)	From baseline (week 0) to end of treatment (week 52)	Count of subjects	
Exploratory endpoint	Subject improving 10 points or more in KCCQ overall summary score (Yes/No)	From baseline (week 0) to end of treatment (week 52)	Count of subjects	
Exploratory endpoint	Subject improving 15 points or more in KCCQ overall summary score (Yes/No)	From baseline (week 0) to end of treatment (week 52)	Count of subjects	

Type	Title	Time frame	Unit	Details
Exploratory endpoint	Subject worsening 5 points or more in KCCQ overall summary score (Yes/No)	From baseline (week 0) to end of treatment (week 52)	Count of subjects	
Exploratory endpoint	Subject worsening 10 points or more in KCCQ overall summary score (Yes/No)	From baseline (week 0) to end of treatment (week 52)	Count of subjects	
Exploratory endpoint	Subject worsening 15 points or more in KCCQ overall summary score (Yes/No)	From baseline (week 0) to end of treatment (week 52)	Count of subjects	
Exploratory endpoint	Change in subscales of KCCQ: Total symptom score Physical limitations score Social limitations score Health-related quality of life	From baseline (week 0) to end of treatment (week 52)	Score (no unit, range 0-100)	
Exploratory endpoint	Subject experiencing improvement in NYHA Class (Yes/No)	From baseline (week -2) to end of treatment (week 52)	Count of subjects	
Exploratory endpoint	Subject experiencing deterioration in NYHA Class (Yes/No)	From baseline (week -2) to end of treatment (week 52)	Count of subjects	
Exploratory endpoint	Time to first heart failure event (hospitalisation or urgent visit)	From baseline (week 0) to end of treatment (week 52)	Days	
Echocardiographic sub-study				
Exploratory endpoint	Change in left atrial volume	From baseline (week 0) to end of treatment (week 52)	Ratio to baseline (no unit)	
Exploratory endpoint	Change in left ventricular (LV) filling pressure (diastolic function) (E/e´)	From baseline (week 0) to end of treatment (week 52)	Ratio to baseline (no unit)	
Exploratory endpoint	Change in global longitudinal strain	From baseline (week 0) to end of treatment (week 52)	%	

6.4 Appendix 4: Kansas City Cardiomyopathy Questionnaire and scoring

The KCCQ consists of 23 questions that are mapped into 10 summary scores that are based on the as shown in Figure 4. Scale scores range from 0 to 100 where higher scores indicates a better health status with regards to heart failure[4]. KCCQ-CSS is an average of the "physical limitation score" and "total symptom score", and KCCQ-OSS is an average of the "physical limitation score," "total symptom score," "quality of life score," and "social limitation score".

Figure 4 Mapping of KCCQ



Item No.	Item text	Response scale
1a	How much are you limited by heart failure in dressing yourself over the past 2 weeks?	1: Extremely limited 2: Quite a bit limited 3: Moderately limited 4: Slightly limited 5: Not at all limited Missing: Limited for other reasons or did not do the activity
1b	How much are you limited by heart failure in showering/bathing over the past 2 weeks	
1c	How much are you limited by heart failure in walking 1 block on level ground over the past 2 weeks	
1d	How much are you limited by heart failure in doing yard work, housework or carrying groceries over the past 2	

	weeks	
1d	How much are you limited by heart failure in climbing flight of stairs without stopping over the past 2 weeks	
1f	How much are you limited by heart failure in hurrying or jogging over the past 2 weeks	
2	Compared with 2 weeks ago, have your symptoms of heart failure (shortness of breath, fatigue, or ankle swelling) changed?	1: Much worse 2: Slightly worse 3: Not changed 4: Slightly better 5: Much better 3: I 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?	1: Every morning 2: 3 or more times a week, bur not every day 3: 1-2 times a week 4: Less than once a week 5: Never over the past 2 weeks
4	Over the past 2 weeks, how much has swelling in your feet, ankles or legs bothered you?	1: Extremely bothersome 2: Quite a bit bothersome 3: Moderately bothersome 4: Slightly bothersome 5: Not at all bothersome 5: 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?	1: All of the time, 2: Several times per day 3: At least once a day 4: 3 or more times per week but not every day 5: 1-2 times per week 6: Less than once a week 7: Never over the past 2 weeks
6	Over the past 2 weeks, how much has your fatigue bothered you?	1: Extremely bothersome 2: Quite a bit bothersome 3: Moderately bothersome 4: Slightly bothersome 5: Not at all bothersome 5: 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?	1: All of the time, 2: Several times per day 3: At least once a day 4: 3 or more times per week but not every day 5: 1-2 times per week 6: Less than once a week

		7: Never over the past 2 weeks
8	Over the past 2 weeks, how much has your shortness of breath bothered you?	1: Extremely bothersome 2: Quite a bit bothersome 3: Moderately bothersome 4: Slightly bothersome 5: Not at all bothersome 5: 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?	1: Every night 2: 3 or more times a week, but not every day 3: 1-2 times a week 4: Less than once a week 5: Never over the past 2 weeks
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?	1: Not at all sure 2: Not very sure 3: Somewhat sure 4: Mostly sure 5: Completely sure
11	How well do you understand what things you are able to do to keep your heart failure symptoms from getting worse?	1: Do not understand at all 2: Do not understand very well 3: Somewhat understand 4: Mostly understand 5: Completely understand
12	Over the past 2 weeks, how much has your heart failure limited your enjoyment of life?	1: It has extremely limited my enjoyment of life 2: It has limited my enjoyment of life quite a bit 3: It has moderately limited my enjoyment of life 4: It has slightly limited my enjoyment of life 5: 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?	1: Not at all satisfied 2: Mostly dissatisfied 3: Somewhat satisfied 4: Mostly satisfied 5: Completely satisfied
14	Over the past 2 weeks, how often have you felt discouraged or down in the dumps because of your heart failure?	1: I felt that way all of the time 2: I felt that way most of the time 3: I occasionally felt that way 4: I rarely felt that way 5: I never felt that way
15a	How much have your heart failure limited your participation in hobbies, recreational activities over the past 2 weeks?	1: Severely limited 2: Limited quite a bit 3: Moderately limited 4: Slightly limited

15b	How much have your heart failure limited your participation in working or doing household chores over the past 2 weeks?	5: Did not limited at all Missing: Does not apply or did not do for other reasons
15c	How much have your heart failure limited your participation in visiting family or friends out of your home over the past 2 weeks?	
15d	How much have your heart failure limited your participation in intimate relationships with loved ones over the past 2 weeks?	

- If at least three of Questions 1a-1f are not missing, then the Physical Limitation score is
  - $\text{Physical Limitation Score} = 100[(\text{Mean of Question 1a} - \text{1f actually answered}) - 1]/4$
- If Question 2 is not missing, then the Symptom Stability Score is calculated as follows
  - $\text{Symptom Stability Score} = 100[(\text{Question 2}) - 1]/4$
- If at least two of Questions 3, 5, 7, and 9 is not missing, then the Symptom Frequency Score is calculated as
  - $S3 = [(\text{Question 3}) - 1]/4$
  - $S5 = [(\text{Question 5}) - 1]/6$
  - $S7 = [(\text{Question 7}) - 1]/6$
  - $S9 = [(\text{Question 9}) - 1]/4$
  - $\text{Symptom Frequency Score} = 100[\text{mean of S3, S5, S7 and S9}]$
- If at least one of Questions 4, 6 and 8 is not missing, then the Symptom Burden Score is defined as
  - $\text{Symptom Burden Score} = 100[(\text{mean of Questions 4, 6 and 8 actually answered}) - 1]/4$
- The Total Symptom Score is defined as the mean of the following available summary scores:
  - Symptom Frequency Score and Symptom Burden Score
- If at least one of Questions 10 and 11 is not missing, then the Self-Efficacy Score is calculated as
  - $\text{Self} - \text{Efficacy Score} = 100[(\text{mean of Questions 10 and 11 actually answered}) - 1]/4$
- If at least one of Question 12, 13 and 14 is not missing, then the Quality of Life Score is calculated as

- Quality of Life Score $= 100[(\text{mean of Questions } 12, 13 \text{ and } 14 \text{ actually answered}) - 1]/4$

- If at least two of the Questions 15a-15d are not missing then the Social Limitation Score is calculated as
  - Social Limitation Score $= 100[(\text{mean of Questions } 15a - 15d \text{ actually answered}) - 1]/4$

- The Overall Summary Score is defined as the mean of the following available summary scores:
  - Physical Limitation Score, Total Symptom Score, Quality of Life Score and Social Limitation Score

- The Clinical Summary Score is defined as the mean of the following available summary scores:
  - Physical Limitation Score and Total Symptom Score

7 References

1. McEvoy BW. Missing data in clinical trials for weight management. J Biopharm Stat. 2016;26(1):30-6.
2. Wang D, Pocock S. A win ratio approach to comparing continuous non-normal outcomes in clinical trials. Pharm Stat. 2016;15(3):238-45.
3. Byrom B, Breedon P, Tulkki-Wilke R, Platko J. Meaningful change: Defining the interpretability of changes in endpoints derived from interactive and mHealth technologies in healthcare and clinical research. Journal of Rehabilitation and Assistive Technologies Engineering. 2020;7:2055668319892778.
4. 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. 2
5. Gaohong Dong, J. Q. (2018). The stratified win ratio. *Journal of Biopharmaceutical Statistics*, 778-796.
6. Burman CF, S. C. (2009). A recycling framework for the construction of Bonferroni-based multiple tests. *Stat Med*, 28(5):739-61.