Cover Page for Statistical analysis plan

Sponsor name:	Novo Nordisk A/S
NCT number	NCT02863328
Sponsor trial ID:	NN9924-4223
Official title of study:	PIONEER 2 – vs. SGLT-2 Inhibitor A 52-week, randomised, open label, active-controlled trial to investigate the efficacy and safety of oral semaglutide versus empagliflozin in subjects with type 2 diabetes mellitus inadequately controlled on metformin
Document date:	02 October 2018

16.1.9 Documentation of statistical methods

List of contents

Statistical Analysis Plan

Trial ID: NN9924-4223

PIONEER 2 – vs. SGLT-2 Inhibitor

Efficacy and Safety of Oral Semaglutide versus Empagliflozin in Subjects with Type 2 Diabetes Mellitus

A 52-week randomised, Open-label, Active-controlled Trial

Redacted statistical analysis plan
Includes redaction of personal identifiable information only.

.

Author:

[redacted]

Table of contents

List of abbreviations

ADA	American Diabetes Association
AE	adverse event
ANCOVA	analysis of covariance
BG	blood glucose
BMI	body mass index
CI	confidence interval
CRP	c-reactive protein
CoEQ	Control of Eating Questionnaire
CTR	clinical trial report
EAC	event adjudication committee
ECG	electrocardiogram
eCRF	electronic case report form
EMA	European Medicines Agency
FAS	full analysis set
FPG	fasting plasma glucose
GLP-1	glucagon-like peptide-1
HbA_{1c}	glycosylated haemoglobin
HDL	high-density lipoprotein
HOMA-B	homeostatic assessment index of beta-cell function
HOMA-IR	homeostatic model assessment index of insulin resistance
HRQoL	health-related quality of life
IWRS	interactive web response system
LDL	low-density lipoprotein
LLoQ	lower limit of quantification
LOCF	last observation carried forward
MAR	missing at random
MCMC	Markov Chain Monte Carlo
MCS	mental component score
MedDRA	Medical Dictionary for Regulatory Activities
MI	multiple imputation
MMRM	mixed model for repeated measurements
NBS	norm-based score
PCS	physical component score
PG	plasma glucose
PD	pharmacodynamics
PK	pharmacokinetics
PP	per protocol
PRO	patient reported outcomes

SAP	statistical analysis plan
SAS	safety analysis set
s.c.	subcutaneous
SD	standard deviation
SF-36v2 (acute version)	SF-36v2® Health Survey (acute version)
SMPG	self-measured plasma glucose
SNAC	sodium N-[8-(2-hydroxybenzoyl)amino]caprylate
TE	treatment effect
TEAE	treatment-emergent adverse events
US	United states
VLDL	very low-density lipoprotein

1 Introduction

1.1 Trial information

This is a 52-week, randomised, open-label, active-controlled, parallel-group, multicentre, multinational trial with 2 arms comparing the efficacy and safety of oral semaglutide with empagliflozin in subjects with type 2 diabetes mellitus inadequately controlled on metformin.

Primary objective

To compare the effect of once-daily dosing of 14 mg oral semaglutide versus 25 mg empagliflozin, both in combination with metformin, on glycaemic control in subjects with type 2 diabetes mellitus.

Secondary objectives

To compare the effect of once-daily dosing of 14 mg oral semaglutide versus 25 mg empagliflozin, both in combination with metformin, on body weight in subjects with type 2 diabetes mellitus.

To compare the safety and tolerability of once-daily dosing of 14 mg oral semaglutide versus 25 mg empagliflozin, both in combination with metformin, in subjects with type 2 diabetes mellitus.

Trial design

Subjects with type 2 diabetes mellitus treated with metformin will after a 2-week screening period be randomised 1:1 to receive a dose of either 14 mg oral semaglutide once-daily or 25 mg empagliflozin once-daily. After a 52-week randomised treatment period, all subjects enter a follow-up period of 5 weeks ending with a follow-up visit. The total trial duration for the individual subject will be approximately 59 weeks. For further details, see the trial protocol.

1.2 Scope of the statistical analysis plan

This statistical analysis plan (SAP) is based on the protocol for trial NN9924-4223 "Efficacy and safety of oral semaglutide versus empagliflozin in subjects with type 2 diabetes mellitus", version 3.0 (17 November 2016) as well as the Protocol amendment no. 1, version 1.0 (15 November 2016), and includes more detailed procedures for executing the statistical analyses of the primary and secondary endpoints. Statistical analyses and a number of clarifications additional to those specified in the trial protocol are pre-specified with this SAP. All changes to the statistical analyses planned in the trial protocol are documented in Section 3.

Novo Nordisk will be responsible for the statistical analyses and reporting.

2 Statistical considerations

General considerations

Data from all sites will be analysed and reported together.

The latest available measurement, at or prior to the randomisation visit, will be used as the baseline measurement. If no measurement(s) have been obtained, at or prior to randomisation, the baseline value will be left missing.

Laboratory values below the lower limit of quantification (LLoQ) will be set to ½LLoQ. Number of values below LLoQ by treatment and visit will be summarised if deemed relevant.

The primary and confirmatory efficacy endpoints will be evaluated at week 26. This approach is expected to result in a lower proportion of missing data, use of rescue medication and premature treatment discontinuation, compared to the expected proportion of missing data, use of rescue medication and premature treatment discontinuation at week 52, and therefore considered a meaningful representation and confirmation of the effect of oral semaglutide.

Results from a statistical analysis will as a minimum be presented by the estimated treatment contrasts for oral semaglutide 14 mg vs. empagliflozin 25 mg with associated two-sided 95% confidence intervals and p-values corresponding to two-sided tests of no difference.

If no statistical analysis is specified, data will be presented using relevant summary statistics.

Primary and secondary estimands

Two estimands addressing different aspects of the primary trial objective will be defined as follows:

- Primary estimand – 'Treatment policy'
  - treatment difference at week 26 for all randomised subjects regardless of adherence to randomised treatment and initiation of rescue medication

The treatment policy estimand assesses the expected glycaemic benefit in a future population that results from subjects initiating treatment with oral semaglutide including potential rescue medication(s) as compared to initiating treatment with empagliflozin including potential rescue medication(s). Generalisation of this estimand depends among other things on the extent to which the use of rescue medication in this trial reflects clinical practice and the adherence to trial product administration in this trial reflects the behaviour of the target population. Accordingly, data collected regardless of discontinuation of trial product or initiation of rescue medication(s) will be used to draw inference.

- Secondary estimand – 'Hypothetical'
  - treatment difference at week 26 for all randomised subjects if all subjects adhered to treatment and did not initiate rescue medication

The hypothetical estimand assesses the glycaemic benefit a future subject is expected to achieve if initiating and continuing treatment with oral semaglutide as compared to empagliflozin. It is considered a clinically relevant estimand as it provides information to treating clinicians about the expected glycaemic efficacy of oral semaglutide compared to empagliflozin for purposes of treating individual subjects. Generalisation of this estimand depends among other things on the extent to which the adherence to trial product administration in this trial reflects the behaviour of the target population. Accordingly, only data collected prior to discontinuation of trial product or initiation of rescue medication will be used to draw inference. This will avoid confounding from rescue medication.

Analogously, two estimands will be pre-defined for the remaining secondary endpoints.

Missing data considerations at week 26

When estimating the primary estimand, the proportion of missing data, i.e. data that do not exist even though subjects are intended to stay in the trial regardless of treatment status and initiation of rescue medication(s), is expected to be maximum 10% based on the oral semaglutide phase 2 trial (NN9924-3790). Thus, missing data will be due to withdrawal from trial or lost to follow-up.

When estimating the secondary estimand, the proportion of missing data is expected to be higher (20%) since data collected after discontinuation of trial product or initiation of rescue medication(s) will be set to missing. The 10% of missing data, which is due to discontinuation of trial product or initiation of rescue medication(s), is based on the empagliflozin assessment report[1] and the oral semaglutide phase 2 trial (NN9924-3790) indicates that a low starting dose with gradual dose escalation diminishes gastrointestinal AEs compared with more aggressive dosing regimens. The main reasons for missing data in the two treatment arms are expected to be early treatment discontinuation due to AEs (particular gastrointestinal AEs for the oral semaglutide arm) and initiation of rescue medication. A higher proportion of subjects is expected to discontinue treatment due to AEs in the oral semaglutide arm compared to empagliflozin whereas initiation of rescue medication is expected to be more frequent in the empagliflozin arm. So overall the frequency of missing data is expected to be similar across treatment arms.

Descriptive summaries and graphical representation of extent, reason(s) for and pattern of missing data will be presented by treatment arm.

2.1 Sample size calculation

The primary endpoint is change from baseline to week 26 in HbA_{1c}. For HbA_{1c}, both non-inferiority and superiority of oral semaglutide versus empagliflozin are planned to be tested. The confirmatory secondary endpoint, change from baseline to week 26 in body weight, is planned to be tested for superiority of oral semaglutide versus empagliflozin.

The sample size calculation is made to ensure a power of at least 90% for testing HbA_{1c} superiority of oral semaglutide versus empagliflozin out of the three pre-specified confirmatory hypotheses shown in Figure 2–1. The closed testing procedure described in Bretz et al 2011[2] is used to control the overall type-I error at a nominal two-sided 5% level.

The statistical testing strategy is based on the principle that glycaemic effect must be established in terms of HbA_{1c} non-inferiority before testing for added benefits in terms of HbA_{1c} superiority and/or body weight superiority.

The sample size is calculated using the calcPower function in the R package, gMCP[3] using 10000 simulations. All of the three pre-specified confirmatory tests are assumed to be independent. Since some of the tests are positively correlated, the assumption of independence is viewed as conservative.

The sample size assumptions for treatment effects (TE), adjusted treatment effects and the standard deviations (SD) are given in Table 2–1. These assumptions are primarily based on the oral semaglutide phase 2 results (NN9924-3790), empagliflozin assessment report[1] and supported by results from the s.c. semaglutide phase 2 trial (NN9535-1821).

To assess the effect of empagliflozin on glycaemic effect, a similar trial (NCT01159600) where empagliflozin was used as add on to metformin was reviewed. Based on this trial, the chosen margin of 0.4 provides assurance that oral semaglutide has a clinically relevant effect greater than zero. With regards to the constancy assumption, controlled clinical trials have consistently established that empagliflozin is an effective anti-diabetic drug. Therefore, lack of trial sensitivity with empagliflozin as comparator is not anticipated to be an issue in this trial.

With regards to preserving an acceptable proportion of the effect of empagliflozin, the broader margin of 0.4, has been chosen instead of 0.3 because of the anticipated body weight advantage of oral semaglutide compared to empagliflozin. The trial has been powered to meet HbA_{1c} superiority of oral semaglutide versus empagliflozin. With the anticipated added benefit on body weight, it is considered acceptable to use a non-inferiority margin of 0.4.

Since the equalising effect of rescue medication will be included in the primary analysis as well as a conservative approach for handling of missing data will be performed, an adjustment in treatment effect will be implemented for the 10% of subjects who are expected to discontinue trial product or

initiate rescue medication and for the 10% of subjects who are expected to have actual missing data. The treatment effects used in the sample size calculation will be adjusted according to a 75% smaller effect in these subjects. For the 10% of the subjects with missing data, the non-inferiority margin of 0.4% for HbA_{1c} is added to the imputed values, when testing for non-inferiority. The adjusted treatment effects for testing non-inferiority (HbA_{1c} only) and superiority are as described below:

- Non-inferiority
  - 0.8×TE + 0.2×TE×0.25 + non-inferiority margin×0.1
- Superiority
  - 0.8×TE + 0.2×TE×0.25

Table 2–1 Assumptions used in the for sample size calculation

Parameter	Treatment effect (TE)	Adjusted TE, non-inferiority	Adjusted TE, superiority	Standard deviation	Non-inferiority margin
HbA_{1c}	-0.3%	-0.215%	-0.255%	1.1%	0.4%
Body weight	-1 kg		-0.85 kg	4 kg	

With the above assumptions, allocating 408 subjects to each of the two arms provides 90% power to confirm HbA_{1c} superiority of oral semaglutide versus empagliflozin. In total 2×408 = 816 subjects are planned to be randomised. Calculated powers for individual hypotheses are presented in Table 2–2.

Table 2–2 Calculated powers for individual hypotheses

	HbA_{1c} non-inferiority	HbA_{1c} superiority	Body weight superiority
Power	> 99%	90%	85%

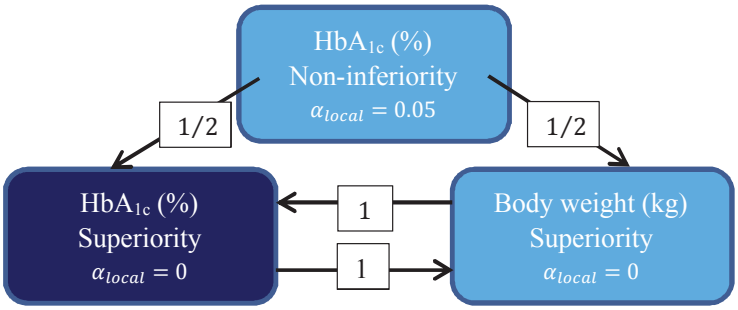


Figure 2–1 Graphical illustration of the closed testing procedure

The overall significance level of α = 0.05 (two-sided) is initially allocated to the HbA_{1c} non-inferiority test of oral semaglutide vs. empagliflozin. The local significance level (α-local) will be reallocated if a hypothesis is confirmed according to the weight given by the directed edges between nodes (hypotheses). The sample size is based on the hypothesis in the dark box.

2.2 Definition of analysis sets

The following analysis sets will be defined:

Full analysis set (FAS): Includes all randomised subjects. Subjects in the FAS will contribute to the evaluation "as randomised".

Safety analysis set (SAS): Includes all subjects exposed to at least one dose of trial product. Subjects in the SAS will contribute to the evaluation "as treated".

Per protocol (PP) analysis set: Includes all subjects in the FAS who fulfils the following criteria

- have not violated any inclusion criteria
- have not fulfilled any exclusion criteria
- have a valid baseline HbA_{1c} measurement
- is exposed to trial product and have at least one valid HbA_{1c} measurement while on treatment without rescue medication at or after week 14

Subjects in the PP analysis set will, as in the SAS, contribute to the analysis "as treated".

Data selections and observation periods

Unless subjects withdraw their informed consent, data collection will continue for the full duration of the trial. The full duration of the trial is defined as up to and including

- the follow-up visit (V13) for subjects on trial product
- the latest occurring visit of the end-of-treatment visit (V12) or the follow-up premature discontinuation visit (V13A), for subjects who have discontinued trial product prematurely.

Subjects and data to be used in an analysis will be selected in a two-step manner.

- Firstly, subjects will be selected based on the specified analysis set
- Secondly, data points on the selected subjects from the first step will be selected based on the specified observation period

Definition of the observation periods:

In-trial: This observation period represents the time period where subjects are considered to be in the trial, regardless of discontinuation of trial product or initiation of rescue medication. The in -trial observation period starts at randomisation (as registered in the IWRS) and ends at the date of

- the last direct subject-site contact, which is scheduled to take place 5 week after planned last dose of trial product at the follow-up visit.
- withdrawal for subjects who withdraw their informed consent.
- the last subject-investigator contact as defined by the investigator for subjects who are lost to follow-up.
- death for subjects who dies before any of the above.

On-treatment: This observation period represents the time period where subjects are considered treated with trial product. The observation period is a subset of the in-trial observation period. It starts at the date of first dose of trial product. Two slightly different end dates will be needed to cover all assessments appropriately.

For adjudicated events, ECGs, eye examination category, anti-semaglutide antibodies, and AEs including hypoglycaemic episodes, the observation period ends at the first date of any of the following:

- the follow-up visit (V13)
- the follow-up prematurely discontinuation visit (V13A)
- the last date on trial product +38 days
- the end-date for the in-trial observation period

The follow-up visit is scheduled to take place 5 weeks after the last date on trial product corresponding to approximately five half-lives of oral semaglutide. The visit window for the follow-up visit is +3 days.

For efficacy and other safety assessments (laboratory assessments, physical examination and vital signs) the observation period ends at the last date on trial product +3 days. This will be used in order to ensure specificity to reversible effects of treatment.

On-treatment without rescue medication: This observation period is a subset of the on-treatment observation period, where subjects are considered treated with trial product, but have not initiated

any rescue medications. The on-treatment without rescue medication observation period starts at first date on trial product and the observation period ends at the first date of any of the following:

- the last dose of trial product +3 days
- the date of initiation of rescue medication

The in-trial observation period will be the primary observation period w hen estimating the primary estimand. The on-treatment without rescue medication observation period will be the primary observation period when estimating the secondary estimand. Safety will be evaluated based on the in-trial and the on-treatment observation periods.

Data points collected outside an observation period will be treated as missing in the analysis. Baseline data will always be included in an observation period. For adjudicated events, the onset date will be the EAC adjudicated onset date.

Before data are locked for statistical analysis, a review of all data will take place. Any decision to exclude either a subject or single observations from the statistical analysis is the joint responsibility of the members of the Novo Nordisk study group. Exclusion of data from analyses should be used restrictively, and normally no data should be excluded from the FAS. The subjects or observations to be excluded, and the reasons for their exclusion must be documented and signed by those responsible before database lock. The subjects and observations excluded from analysis sets, and the reason for this, will be described in the clinical trial report.

Confirmatory hypotheses

For the primary HbA_{1c} endpoint and the confirmatory secondary body weight endpoint the following confirmatory one-sided hypotheses are planned to be tested for oral semaglutide versus empagliflozin. Let the mean treatment difference be defined as μ = (oral semaglutide minus empagliflozin):

- HbA_{1c} non-inferiority, using a non-inferiority margin of 0.4%
  - H_0: $\mu \geq 0.4\%$ against H_a: $\mu < 0.4\%$

- HbA_{1c} superiority
  - H_0: $\mu \geq 0.0\%$ against H_a: $\mu < 0.0\%$

- HbA_{1c} body weight superiority
  - H_0: $\mu \geq 0.0$kg against H_a: $\mu < 0.0$kg

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

Multiplicity and criteria for confirming hypotheses

The type I error for testing the three confirmatory hypotheses related to the HbA_{1c} and body weight endpoints will be preserved in the strong sense at 5% (two-sided) using the weighted Bonferroni-based closed testing procedure described in Bretz et al 2011[2] and outlined in Figure 2–1.

The first hypothesis to be tested is non-inferiority of HbA_{1c}. It will be tested at the overall significance level (5%) while allocating 0% local significance level to the remaining of the hypotheses. For this hypothesis, and in general, if a hypothesis is confirmed the significance level will be reallocated according to the weight and the direction of the edges going from the confirmed hypothesis to the next hypotheses as specified in Figure 2–1. Each of the following hypotheses will be tested at their local significance level (α-local). This process will be repeated until no further hypotheses can be confirmed.

Non-inferiority and/or superiority will be considered confirmed if the mean treatment difference is supporting the corresponding alternative hypothesis and the two-sided p-value from the primary analysis of the primary estimand is strictly below its local two-sided significance level as defined by the closed testing procedure in Figure 2–1. This is equivalent to using a one-sided p-value (nominal $\alpha = 0.025$) and a one-sided 2.5% overall significance level in the closed testing procedure.

2.3 Primary endpoint

The primary endpoint is change from baseline to week 26 in HbA_{1c}.

2.3.1 Primary analysis for the primary estimand

The primary estimand will be estimated based on the FAS using week 26 measurements from the in-trial observation period. The primary statistical analysis will be a pattern mixture model using multiple imputation to handle missing data assuming that the missing data mechanism is missing at random (MAR) within the groups used for imputation. Imputation of missing data at week 26 will be done within 4 groups of subjects defined by randomised treatment arm, and whether subjects at week 26; (i) have discontinued treatment or initiated rescue medication or (ii) are still on treatment and have not initiated rescue medication. It is hereby assumed that the likely values of what the missing data would have been if available are best described by information from subjects who at week 26 are similar in terms of randomised treatment arm and treatment adherence/rescue medication status.

Missing values for each group will be imputed as follows:

- An analysis of covariance (ANCOVA) with region as a categorical fixed effect and baseline HbA_{1c} measurement as a covariate will be fitted to observed values of the change from baseline in HbA_{1c} at week 26.

- The estimated parameters for location and dispersion will be used to impute 1000 values for each subject with missing week 26 data based on region and baseline HbA_{1c}. Thus, 1000 complete data sets will be generated including observed and imputed values.

In the statistical analysis models the variable region is included as a categorical fixed effect. The regions to be used in the statistical analyses are defined as Europe, North America, South America and Asia. When addressing the treatment policy estimand, the imputation is to be done within groups defined by randomised treatment and treatment adherence at time of evaluation. The number of subjects in the groups who at time of evaluation (week 26 and 52) have discontinued trial product or initiated rescue medication are expected to be relatively low. Therefore, the region variable included in the imputation model will be reduced in levels avoiding estimation problems due to sparse data. The regions to be used in these imputations are defined as North America and Other regions.

Analysis used for confirming superiority versus empagliflozin at week 26:

For each of the 1000 (now complete) imputed data sets, the change in HbA_{1c} from baseline to week 26 will be analysed using an ANCOVA with treatment and region as categorical fixed effects, and baseline HbA_{1c} as covariate. The results obtained from analysing the datasets will be combined using Rubin's rule[4] to draw inference.

Analysis used for confirming non-inferiority versus empagliflozin at week 26:

Prior to analysing the data using the same model and approach as used for confirming superiority (see above), a value of 0.4% (the non-inferiority margin) will be added to imputed values at week 26 for the oral semaglutide treatment arms only[5]. For evaluating non-inferiority versus empagliflozin unadjusted two sided p-value for testing no difference from the non-inferiority margin will be presented.

2.3.2 Primary analysis for the secondary estimand

The secondary estimand will be estimated based on the FAS using post-baseline measurements up to and including week 26 from the on-treatment without rescue observation period. The primary analysis for the secondary estimand will be a Mixed Model for Repeated Measurements (MMRM). A restricted maximum likelihood will be used. The model will include all post baseline HbA_{1c} measurements collected at scheduled visits up to and including week 26 as dependent variables. The independent effects included in the model will be treatment and region as categorical fixed effects and baseline HbA_{1c} as a covariate, all nested within visit. An unstructured covariance matrix for HbA_{1c} measurements within the same subject will be employed, assuming measurements from different subjects are independent.

The MMRM is a well-established method that accounts for the uncertainty pertaining to missing data. This analysis assumes that the missing data mechanism is MAR. Under this assumption the

statistical behaviour of the missing data (given the observed responses and model fixed effects and covariates) is assumed to be same as the observed data.

For subjects who do not have post-baseline assessments for planned visits available in the on-treatment without rescue medication period, the baseline value will be carried forward to the first planned visit to ensure that all randomised subjects will contribute to the statistical analysis.

2.3.3 Sensitivity analyses

To investigate the sensitivity of the primary analysis results, complementary and separate analyses will be performed for the primary and secondary estimand. In line with European Medicines Agency recommendations[6] and with a report from the US National Research Council[7], these analyses will primarily evaluate the sensitivity of the results due to the impact of missing data. Since conservatism, i.e. avoiding bias in favour of oral semaglutide, depends on the context, separate sensitivity analyses will be made for non-inferiority and superiority testing.

The evaluation of the robustness of the primary analysis results will primarily be based on a pattern mixture model approach using multiple imputation (MI). An overview of the sensitivity analyses for each of the estimands are specified below followed by a more detailed description of the three different pattern mixture models used. Finally, one additional sensitivity analysis for the primary analysis will be described that is not based on the pattern mixture model approach (see Section 2.3.3.1).

Sensitivity analyses for the primary estimand

The estimation of the primary estimand will be repeated using the following sensitivity analyses:

- A comparator multiple imputation analysis based on FAS using the in-trial observation period (superiority).
- A comparator multiple imputation analysis differentiating between reasons for discontinuing treatment prematurely based on FAS using the in-trial observation period (superiority).
- A tipping-point multiple imputation analysis based on FAS using the in-trial observation period (non-inferiority and superiority).

Sensitivity analyses for the secondary estimand

The estimation of the secondary estimand will be repeated using the following sensitivity analyses:

- A tipping-point multiple imputation analysis based on FAS using the on-treatment without rescue medication observation period (non-inferiority and superiority).

2.3.3.1 Pattern mixture models

Common for the three pattern mixture model sensitivity analyses is that they all aim to stress-test the primary HbA_{1c} results by changing the assumptions for part or all missing data in the oral

semaglutide treatment arm, while maintaining the missing data assumption for the empagliflozin arm.

- *Comparator multiple imputation analysis:* In this sensitivity analysis missing data at week 26 for all subjects will be imputed to resemble the distribution of the week 26 values observed in the empagliflozin treatment arm. In effect, this imputation approach removes the treatment difference between oral semaglutide and empagliflozin for all subjects randomised to oral semaglutide, given that oral semaglutide is better than empagliflozin. Due to the potential lack of sensitivity for testing non-inferiority this sensitivity analysis will only be used to evaluate the robustness of HbA_{1c} superiority conclusions.
- *Comparator multiple imputation analysis differentiating between reasons for discontinuing treatment prematurely:* In this sensitivity analysis missing data at week 26 for subjects who discontinue oral semaglutide treatment due to treatment related AE(s) will be imputed to resemble the distribution of the week 26 values observed in the empagliflozin treatment arm. Treatment related AEs are defined as AEs classified as possible or probable related to trial product as reported by the investigator. In effect this imputation approach removes the treatment difference between oral semaglutide and empagliflozin for this selected group of subjects randomised to oral semaglutide. This sensitivity analysis is less conservative as compared to the first sensitivity analysis. Due to the potential lack of sensitivity for testing non-inferiority this sensitivity analysis will only be used to evaluate the robustness of HbA_{1c} superiority conclusions.
- *Tipping-point multiple imputation analysis:* In this sensitivity analysis, firstly, missing data will be imputed according to the primary analysis for the treatment policy estimand, whereas for the hypothetical estimand imputation will be done as described below for the binary endpoints (see section 2.4.2.1). Secondly, for the oral semaglutide treatment arm a penalty will be added to the imputed values at week 26. The approach is to gradually increase this penalty until the HbA_{1c} conclusion from the primary analysis is changed. The specific value of the penalty that changes the conclusion will be used to evaluate the robustness of the primary analysis result. This sensitivity analysis will be used for evaluating the robustness of the HbA_{1c} non-inferiority and superiority conclusions.

2.3.3.2 Other sensitivity analyses

The following additional sensitivity analysis will be specified

- *Per-protocol analysis:* This sensitivity analysis will be based on the per-protocol analysis set. Data from the on-treatment without rescue medication observation period will be analysed using the primary analysis approach for the secondary estimand. This sensitivity analysis will be used to evaluate the robustness of the HbA_{1c} non-inferiority conclusions.

2.3.3.3 Assessment of sensitivity analyses

The results from the sensitivity analyses will be collectively used to interpret the robustness of the trial results for HbA_{1c}. Due to the large number of sensitivity analyses and their inherent

conservative nature, it will not be a requirement that all confirmatory hypotheses are consistently confirmed across the sensitivity analyses. Thus, no absolute success criteria will be pre-defined for each sensitivity analysis. The sensitivity results in totality will be used to substantiate the credibility of the trial results.

2.4 Secondary endpoints

2.4.1 Confirmatory secondary endpoints

Change from baseline to week 26 in body weight (kg) will be a confirmatory secondary endpoint.

The primary and secondary estimands will be estimated using the same approaches as described for the primary HbA_{1c} endpoint. Body weight will only be tested for superiority. Baseline body weight will be used as a covariate instead of baseline HbA_{1c} in both the multiple imputation and analysis model.

Superiority will be considered confirmed if the mean treatment difference is supporting the corresponding hypothesis and the two-sided p-value from the analysis of the primary estimand is strictly below its local two-sided significance level resulting from the closed testing procedure in Figure 2–1. Sensitivity analyses similar to the ones pre-specified for testing superiority for the primary HbA_{1c} endpoint will be made to evaluate the robustness of the body weight results.

2.4.2 Supportive secondary endpoints

2.4.2.1 Efficacy endpoints

The below supportive secondary efficacy endpoints will be evaluated for

- the primary estimand based on FAS using the in-trial observation period
- the secondary estimand based on FAS using the on-treatment without rescue medication observation period

No sensitivity analyses are planned for these.

Continuous efficacy endpoints

Change from baseline to week 52 in:

- HbA_{1c}
- Body weight (kg)

Change from baseline to week 26 and week 52 in:

- Body weight (%)
- FPG
- Fasting C-peptide

- Fasting insulin and proinsulin
- Fasting glucagon
- Insulin resistance (homeostatic model assessment index of insulin resistance [HOMA-IR]) and beta-cell function (homeostatic assessment index of beta-cell function [HOMA-B])
- BMI
- Waist circumference
- Fasting lipid profiles (total cholesterol, LDL cholesterol, VLDL cholesterol, HDL cholesterol, triglycerides, free fatty acids)
- C-reactive protein (CRP)

BMI will be calculated based on body weight and height based on the formulae:

$$\text{BMI kg/m}^2 = \text{body weight (kg)/(Height (m) x Height (m)) or (kg/m}^2 = [\text{lb/in}^2 \text{ x } 703])$$

Change from baseline to week 26 and week 52 in 7-point SMPG profile:

- Mean 7-point profile; defined as the area under the profile, calculated using the trapezoidal method, divided by the measurement time
- Mean postprandial increment (over all meals)

The above continuous endpoints will be analysed separately using similar model approaches as for the primary endpoint with the associated baseline response as a covariate. All endpoints, except HbA_{1c}, body weight, FPG, BMI, waist circumference and endpoints related to 7-point SMPG profile, will be log-transformed prior to analysis with the associated log-transformed baseline value as a covariate.

For evaluation of the primary estimand, the analysis will be performed separately for week 26 and week 52. For the analysis at week 52, the imputation of missing data will be further differentiated by whether subjects have discontinued trial product or initiated rescue medication prior to week 26 or at/after week 26. This will result in imputation of missing data within 6 groups of subjects instead of the 4 groups as described for the week 26 evaluation in Section 2.3.1. If less than 5 subjects have available data in one of the 6 groups, the imputation will be made within 4 groups without differentiating by time of discontinuation of trial product or initiation of rescue medication in the same way as specified for the primary evaluation. The frequency of missing data is expected to be slightly larger at week 52 compared to week 26. The rate of missing data is expected to decline over time.

For evaluation of the secondary estimand, the MMRM based primary analysis will include all scheduled post-baseline measurement up to and including week 52. From this model the estimated treatment differences (ratios) will be presented at week 26 (except for HbA_{1c} and body weight), and week 52 with 95% confidence intervals and two-sided p-values for test of no difference. The

baseline will not be carried forward to first planned visit if the first planned visit falls later than 8 weeks after randomisation.

Binary efficacy endpoints

If a subject after week 26 achieves (yes/no):

- $HbA_{1c} < 7.0\%$ (53 mmol/mol) (ADA) target
- $HbA_{1c} \leq 6.5\%$ (48 mmol/mol) (AACE) target
- Weight loss $\geq 5\%$
- Weight loss $\geq 10\%$
- $HbA_{1c} < 7.0\%$ (53 mmol/mol) without hypoglycaemia (severe or BG-confirmed symptomatic hypoglycaemic episodes) and no weight gain
- HbA_{1c} reduction $\geq$ 1%-point (10.9 mmol/mol) and weight loss $\geq 3\%$

When addressing the treatment policy estimand the 'no hypoglycaemia' component of the composite endpoint will also include non-treatment-emergent events of severe or BG-confirmed symptomatic hypoglycaemia as data collected regardless of discontinuation of trial product or initiation of rescue medication(s) is used. The above six binary endpoints will also be evaluated after week 52.

Missing data for the above six binary endpoints will be accounted for using multiple imputation techniques. For the treatment policy estimand the binary endpoints will be calculated as dichotomisations of the 1000 multiple imputations underlying the primary MI analysis. For the hypothetical estimand the model will be implemented using a sequential imputation approach assuming MAR. The imputation will be done as described below:

- Intermittent missing values in the on-treatment without rescue observation period will be imputed using a Markov Chain Monte Carlo (MCMC) method, in order to obtain a monotone missing data pattern. This imputation is done for each treatment group separately and a 1000 copies of the data set will be generated.
- A sequential regression approach for imputing monotone missing values at planned visits will be implemented starting with the first visit after baseline and sequentially continuing to the planned end of treatment visit. For each treatment group an analysis of covariance model will be used to impute missing values at each planned visit. The model will include region as categorical effect and baseline and post-baseline values prior to the visit in question as covariates.

The binary endpoints will be derived as dichotomisations of the 1000 multiple imputations from the sequential imputation.

For both estimands, each of the 1000 data sets will be analysed using a logistic regression model with treatment and region as fixed effects and baseline value as covariate (i.e. baseline Hb A_{1c} for binary HbA_{1c} endpoints, baseline body weight for body weight endpoints and both HbA_{1c} and baseline body weight for the composite endpoints that comprises both parameters). The results will be combined using Rubin's rule[4] to draw inference.

For the composite endpoints involving both HbA_{1c} and body weight the imputed data sets will be combined by imputation number.

Time to event endpoints

- Time to additional anti-diabetic medication (to support the treatment policy estimand)
- Time to rescue medication (to support the hypothetical estimand)

Definition of additional anti-diabetic medication: New anti-diabetic medication and/or Intensification of anti-diabetic medication initiated at or after randomisation and before (planned) end-of-treatment.

Definition of rescue medication: New anti-diabetic medication and/or Intensification of anti-diabetic medication initiated at or aft er randomisation and before last date on trial product. This is a subset of the additional anti-diabetic medication.

The following rules will be applied based on the concomitant medication data reported by the investigator, to determine whether or not the recorded anti-diabetic medication is 1. *New anti-diabetic medication* or 2. *Intensification of anti-diabetic medication*

1. ***New anti-diabetic medication:*** Anti-diabetic medication (4th-level ATC code) that is initiated after randomisation and is new compared to the anti-diabetic background medication at randomisation (see above) and with a dosing duration of more than 21 days
2. ***Intensification of anti-diabetic medication:*** A more than 20% increase in the dose of anti-diabetic medication after randomisation as compared to the anti-diabetic medication dose at randomisation (5th-level ATC code not changed) and with a dosing duration of more than 21 days.

More than 21 days is chosen as a minimum duration for the medication to be considered as 'anti-diabetic medication'. This threshold is set to ensure that the short-term durations (i.e. $\leq$ 21 days) of anti-diabetic medication (e.g. in connection with concurrent illnesses) are not included because such intensifications are not likely to affect the effect endpoints.

Treatment policy estimand: Time to additional anti-diabetic medication

The treatment policy estimand is addressed for the FAS using the in-trial observation period and additional anti-diabetic medication will be considered an event regardless of treatment adherence.

Time from randomisation to additional anti-diabetic medication will be analysed using a Cox proportional hazards model with treatment and region as categorical fixed effects and baseline HbA_{1c} as a covariate. From this analysis the estimated Hazard ratios between oral semaglutide and empagliflozin together with associated two-sided 95% CIs and unadjusted two-sided p-values will be presented. The analysis aims to address the need of additional anti-diabetic medication regardless of this is due to lack of effect or tolerability. Switch to other anti-diabetic treatment is therefore also considered an event and withdrawn subjects or subject lost to follow-up will be considered as having an event on the day of withdrawal. Subjects will be censored on the day before planned end of treatment visit.

Hypothetical estimand: Time to rescue medication

The hypothetical estimand is addressed for the FAS using the on-treatment without rescue medication observation period. Time from first dose of trial product to initiation of rescue medication will be analysed using the same model as described above. The analysis aims to address lack of effect and only initiation of rescue medication as add-on to randomised treatment is considered an event. Switch to other anti-diabetic treatment is not considered an event and as a consequence subjects will be censored on the day before date of last trial product. Potential events occurring between randomisation and first date on trial product will be included in the analysis as events at day 0, in order to count all events of rescue medication.

Pharmacokinetic endpoints

- SNAC plasma concentrations
- Semaglutide plasma concentrations for population PK analyses

The SNAC plasma concentrations and semaglutide plasma concentrations collected in this trial will be evaluated using relevant summary statistics. In addition, the semaglutide plasma concentration will be part of a meta-analysis across the oral semaglutide phase 3a trials, see Section 2.6.

2.4.2.2 Safety endpoints

The safety endpoints will be evaluated based on SAS using the on-treatment observation period and the in-trial observation period unless otherwise stated. The following endpoints are used to support the safety objective:

Adverse events

- Number of treatment emergent adverse events (TEAEs) during exposure to trial product, assessed up to approximately 57 weeks

All AEs will be coded using version 20.1 of the Medical Dictionary for Regulatory Activities (MedDRA) coding.

A treatment-emergent AE is defined as an AE with onset in the on-treatment observation period (see definition of observation periods in Section 2.2).

TEAEs will be summarised in terms of the number of subjects with at least one event (N), the percentage of subjects with at least one event (%), the number of events (E) and the event rate per 100 patient years of observation time (R) for the on-treatment observation period. Supportive summaries of AEs will be made for the in-trial observation period. The development over time in gastrointestinal AEs will be presented graphically.

Other safety endpoints

Change from baseline to week 26 and week 52 in:

- Amylase
- Lipase
- Pulse
- Systolic blood pressure
- Diastolic blood pressure

The above safety endpoints will be evaluated using the primary analysis for the primary estimand based on SAS using the in-trial observation period and using the primary analysis for the secondary estimand based on SAS using the on-treatment observation period. Endpoints will be analysed separately as described above for continuous efficacy endpoints. Results will be presented at week 26 and at week 52. Amylase and lipase endpoints will be log-transformed prior to analysis with the associated log-transformed baseline value as a covariate.

Change from baseline to week 26 and week 52 in:

- ECG evaluation
- Physical examination (week 52 only)
- Eye examination category (week 52 only)

Any occurrence of anti-semaglutide antibodies (yes/no) up to approximately 57 weeks:

- Anti-semaglutide binding antibodies
- Anti-semaglutide neutralising antibodies
- Anti-semaglutide binding antibodies cross reacting with native GLP-1
- Anti-semaglutide neutralising antibodies cross reacting with native GLP-1

Anti-semaglutide binding antibodies up to approximately 57 weeks:

- Anti-semaglutide binding antibody levels

Other safety assessments

Change from baseline to week 26 and week 52 in:

- Haematology
- Biochemistry (except for amylase and lipase)
- Calcitonin

Change from pre-dose to post-dose (25 and 40 min) at week 4, 26, and 52 in:

- Lactate

The above safety endpoints and assessments will be summarised descriptively by treatment arm and visit. Categorical safety endpoints and assessments will be summarised as counts and relative frequencies. Calcitonin will also be presented by gender.

Hypoglycaemia

- Number of treatment-emergent severe or BG-confirmed symptomatic hypoglycaemic episodes during exposure to trial product, assessed up to approximately 57 weeks
- Treatment-emergent severe or BG-confirmed symptomatic hypoglycaemic episodes during exposure to trial product, assessed up to approximately 57 weeks (yes/no)

Classification of hypoglycaemia:

Hypoglycaemic episodes will be summarised for the SAS and the on-treatment observation period only.

Treatment-emergent: hypoglycaemic episodes will be defined as treatment-emergent if the onset of the episode occurs within the on-treatment observation period (see definition of observation periods in Section 2.2).

Nocturnal hypoglycaemic episodes: episodes occurring between 00:01 and 05.59 both inclusive.

Hypoglycaemic episodes are classified according to the Novo Nordisk classification of hypoglycaemia and the ADA classification of hypoglycaemia (see Figure 2–2).

Novo Nordisk classification of hypoglycaemia

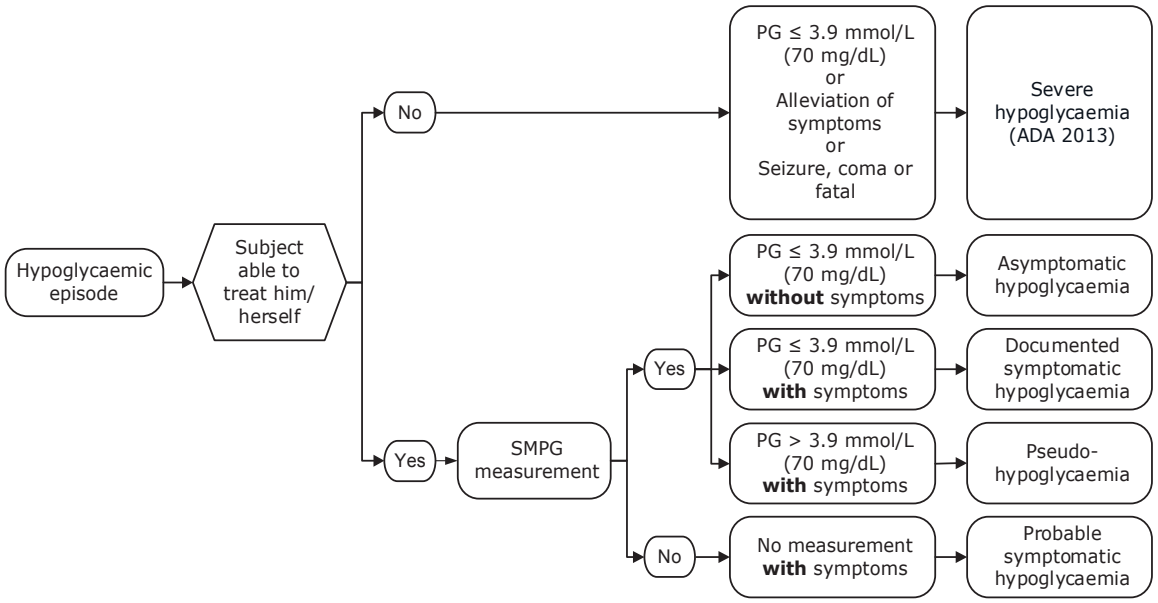
In normal physiology, symptoms of hypoglycaemia occur below a PG level of 3.1 mmol/L (56 mg/dL)[8]. Therefore, Novo Nordisk has included hypoglycaemia with PG levels below this cut-off point in the definition of BG-confirmed hypoglycaemia.

Novo Nordisk uses the following classification in addition to the ADA classification:

Severe or BG-confirmed symptomatic hypoglycaemia: An episode that is severe according to the ADA classification[9] or BG-confirmed by a PG value < 3.1 mmol/L (56 mg/dL) with symptoms consistent with hypoglycaemia.

ADA classification of hypoglycaemia[9]

- Severe hypoglycaemia: An episode requiring assistance of another person to actively administer carbohydrate, glucagon, or take other corrective actions. PG concentrations may not be available during an event, but neurological recovery following the return of PG to normal is considered sufficient evidence that the event was induced by a low PG concentration.
- Asymptomatic hypoglycaemia: An episode not accompanied by typical symptoms of hypoglycaemia, but with a measured PG concentration ≤ 3.9 mmol/L (70 mg/dL).
- Documented symptomatic hypoglycaemia: An episode during which typical symptoms of hypoglycaemia are accompanied by a measured PG concentration ≤ 3.9 mmol/L (70 mg/dL).
- Pseudo-hypoglycaemia: An episode during which the person with diabetes reports any of the typical symptoms of hypoglycaemia with a measured PG concentration > 3.9 mmol/L (70 mg/dL) but approaching that level.
- Probable symptomatic hypoglycaemia: An episode during which symptoms of hypoglycaemia are not accompanied by a PG determination but that was presumably caused by a PG concentration ≤ 3.9 mmol/L (70 mg/dL).



Note: Glucose measurements are performed with capillary blood calibrated to plasma equivalent glucose values

Figure 2–2 ADA classification of hypoglycaemia

PG: plasma glucose. SMPG: Self-measured plasma glucose

Data on treatment-emergent hypoglycaemic episodes will be presented in terms of the number of subjects with at least one episode, the percentage of subjects with at least one episode (%), the total number of episodes and the episode rate per 100 patient years of observation time.

Analysis of severe or BG-confirmed symptomatic hypoglycaemic endpoints

The number of treatment-emergent severe or BG-confirmed symptomatic hypoglycaemic episodes will be evaluated for the on-treatment period using a negative binomial regression model with a log-link function and the logarithm of the duration of the subject's on-treatment observation period as offset. The model will include treatment and region as fixed factors and baseline HbA_{1c} as covariate.

The binary endpoint showing whether a subject has at least one treatment-emergent severe or BG-confirmed symptomatic hypoglycaemic episode will be analysed using a logistic regression model with treatment and region as fixed factors and baseline HbA_{1c} as covariate.

2.5 Interim analysis

No interim analyses will be performed before the database is locked.

2.6 Pharmacokinetic and/or pharmacodynamic modelling

Data from this trial will be evaluated using population pharmacokinetic analysis and exposure-response for semaglutide. The purpose of the population pharmacokinetic analysis will be 1) to describe the covariate factors (such as weight, age, gender, race and ethnicity) that influence semaglutide exposure, 2) to estimate a steady-state exposure level for each subject with pharmacokinetic data in order to facilitate subsequent exposure-response analyses. The purpose of the exposure-response analyses will be to support the recommended doses by investigating response and potentially side effects across the exposure range.

The population pharmacokinetic and exposure-response analyses will be conducted as a meta-analysis, including all relevant Phase 3a trials with PK assessment relevant oral semaglutide phase 3 trials. A separate modelling analysis plan will be prepared before database lock, outlining details of the analyses. The modelling will be performed by Quantitative Clinical Pharmacology at Novo Nordisk A/S and will be reported separately from the CTR.

2.7 Patient reported outcomes

Change from baseline to week 26 and week 52 in:

- SF-36v2® Health Survey (acute version) (SF-36v2 (acute version)) health survey: Scores from the 8 domains and the physical component score and mental component score summary scores
- CoEQ: Scores from the 4 domains and scores from 19 individual items

A more detailed description of the handling of the two patient reported outcomes (PRO) questionnaires used in this trial is provided in the following sections.

No mulitplicity adjustments will be done for the PRO questionnaires (SF-36v2 (acute version) and CoEQ).

The PRO endpoints will be evaluated using the primary analysis for the primary estimand based on FAS using the in-trial observation period. All of the above scores will be analysed separately as the other continuous efficacy endpoints with the associated baseline response as a covariate.

2.7.1 SF-36v2 (acute version)

The SF-36v2 (acute version) instrument is a commonly used generic instrument measuring health-related quality of life (HRQoL)/general health status across disease areas including diabetes[10]. The SF-36v2 (acute version) is a PRO questionnaire for adults and contains 36 items (see Table 2–3).

A total of 35 items measure eight domains of functional health and well-being as well as two summary domains with a 1-week recall period: physical functioning (10 items), role limitation due to physical health problems (4 items), bodily pain (2 items), general health perceptions (5 items), vitality (4 items), social functioning (2 items), role limitations due to emotional problems (3 items)

and general mental health (5 items), mental component score (MCS), physical component score (PCS). There is an additional single item giving information on health change over the past week.

Table 2–3 Overview of items in SF-36v2 (acute version) questionnaire

Item No.	Item text	Response scale
1	In general, would you say your health is:	Excellent; Very good; Good; Fair; Poor
2	Compared to one week ago, how would you rate your health in general now	Much better now than one week ago; Somewhat better now than one week ago; About the same as one week ago; Somewhat worse now than one week ago; Much worse now than one week ago
Question 3: The following questions are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?		
3a	Vigorous activities, such as running, lifting heavy objects, participating in strenuous sports	Yes, limited a lot; Yes, limited a little; No, not limited at all
3b	Moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf	as Item 3a
3c	Lifting or carrying groceries	as Item 3a
3d	Climbing several flights of stairs	as Item 3a
3e	Climbing one flight of stairs	as Item 3a
3f	Bending, kneeling, or stooping	as Item 3a
3g	Walking more than a mile	as Item 3a
3h	Walking several hundred yards	as Item 3a
3i	Walking one hundred yards	as Item 3a
3j	Bathing or dressing yourself	as Item 3a
Question 4: During the past week, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of your physical health?		
4a	Cut down on the amount of time you spent on work or other activities	All of the time; Most of the time; Some of the time; A little of the time; None of the time
4b	Accomplished less than you would like	as Item 4a
4c	Were limited in the kind of work or other activities	as Item 4a
4d	Had difficulty performing the work or other activities (for example, it took extra effort)	as Item 4a
Question 5: During the past week, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?		
5a	Cut down on the amount of time you spent on work or other activities.	as Item 4a
5b	Accomplished less than you would like	as Item 4a
5c	Did work or other activities less carefully than usual	as Item 4a
6	During the past week, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbours, or groups?	Not at all; Slightly; Moderately; Quite a bit; Extremely
7	How much bodily pain have you had during the past week?	None; Very mild; Mild; Moderate; Severe Very severe
8	During the past week, how much did pain interfere with your normal work (including both work outside the home and housework)?	Not at all; A little bit; Moderately; Quite a bit; Extremely
Question 9: These questions are about how you feel and how things have been with you during the past week. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past week…		
9a	Did you feel full of life?	All of the time; Most of the time; Some of the time; A little of the time; None of the time
9b	Have you been very nervous?	as Item 9a

9c	Have you felt so down in the dumps that nothing could cheer you up?	as Item 9a
9d	Have you felt calm and peaceful?	as Item 9a
9e	Did you have a lot of energy	as Item 9a
9f	Have you felt downhearted and depressed?	as Item 9a
9g	Did you feel worn out?	as Item 9a
9h	Have you been happy?	as Item 9a
9i	Did you feel tired?	as Item 9a
10	During the past week, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting with friends, relatives, etc.)?	as Item 9a
Question 11: How TRUE or FALSE is each of the following statements for you?		
11a	I seem to get sick a little easier than other people	Definitely true; Mostly true; Don't know; Mostly false; Definitely false
11b	I am as healthy as anybody I know	as Item 11a
11c	I expect my health to get worse	as Item 11a
11d	My health is excellent	as Item 11a

Domain scores

Norm-based scores (NBS) will be derived using the QualityMetric Health Outcomes™ Scoring Software[11] including the 2009 US general population norm used for all the oral semaglutide phase 3a programme trials. Version 4.5 of the QualityMetric Health Outcomes™ Scoring Software available at time of licensing will be used. Table 2–4 provides an overview of the domains. NBS standardises domain and component scores into T-scores using the means and standard deviations from the US general population. Higher scores on all domains and component summary measures (PCS and MCS) indicate better HRQoL/general health status. Item 2 (i.e. Question 2 in eCRF) is not included in any score.

Table 2–4 Overview of domains for SF-36v2 (acute version)

Domain	Items numbers of items included in domain	Comment
Physical Functioning (PF)	Items 3a-j	
Role Limitations Due to Physical Health (Role-Physical; RP)	Items 4a-d	
Bodily Pain (BP)	Items 7, 8	Both item scores reversed
General Health Perceptions (General Health; GH)	Items 1, 11a-d	Item scores 1, 11b and 11d reversed
Vitality (VT)	Items 9a, 9e, 9g, 9i	Item scores 9a and 9e reversed
Social Functioning (SF)	Items 6, 10	Item score 6 reversed
Role Limitations Due To Emotional Problems (Role-Emotional; RE)	Items 5a-c	
Mental Health (MH)	Items 9b, 9c, 9d, 9f, 9h	Item scores 9d and 9h reversed
Physical component summary (PCS)	NA	The PCS score is a weighted average of the 8 domain scores.
Mental component summary (MCS)	NA	The MCS score is also a weighted average of the 8 domain scores. Weights differ from PCS to MCS.

Missing data at instrument level will be handled using the Maximum Data Recovery method: The method applies a value to a domain item rendered missing if at least one of the items in that domain has valid data. A domain score is considered missing if all item values in the domain are missing.

PCS and MCS are calculated when at least seven of the eight domains have valid data, either actual or estimated. However, to calculate PCS, the Physical Functioning domain must be one of the seven domains having valid data. Also, to calculate MCS, the Mental Health domain must be one of the seven domains having valid data.

Responder threshold values

The responder threshold values, in terms of T-score points for change from baseline are defined in Table 2–5 [11].

Table 2–5 Responder thresholds for SF-36v2 (acute version)

Domain	Responder threshold
Physical Functioning (PF)	4.3
Role Limitations Due to Physical Health (Role-Physical; RP)	4.0
Bodily Pain (BP)	5.5
General Health Perceptions (General Health; GH)	7.0
Vitality (VT)	6.7
Social Functioning (SF)	6.2
Role Limitations Due To Emotional Problems (Role-Emotional; RE)	4.6
Mental Health (MH)	6.7
Physical component summary (PCS)	3.8
Mental component summary (MCS)	4.6

2.7.2 Control of Eating Questionnaire (CoEQ)

The CoEQ comprises 21-items designed to assess the intensity and type of food cravings, as well as subjective sensations of appetite and mood[11]. In the oral semaglutide phase 3a programme trials a version with 19 items has been used (see Table 2–6). One of the two excluded items is open-ended and addresses specific foods, and the other excluded item concerns how difficult it has been to resist this specific food; and the items are therefore not part of any of the four domains.

Table 2–6 Overview of items in CoEQ

Item No.	Item text	Response scale
1	How hungry have you felt?	10 = Extremely hungry, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0 = Not at all hungry
2	How full have you felt?	10 = Extremely full, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0 = Not at all full
3	How strong was your desire to eat sweet foods?	10 = Extremely strong, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0 = Not at all strong
4	How strong was your desire to eat savoury foods?	10 = Extremely strong, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0 = Not at all strong
5	How happy have you felt?	10 = Extremely happy, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0 = Not at all happy
6	How anxious have you felt?	10 = Extremely anxious, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0 = Not at all anxious

7	How alert have you felt?	10 = Extremely alert, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0 = Not at all alert
8	How contented have you felt?	10 = Extremely contended, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0 = Not at all contended
9	During the last 7 days how often have you had food cravings?	10 = Very often, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0 = Not at all
10	How strong have any food cravings been?	10 = Extremely strong, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0 = Not at all strong
11	How difficult has it been to resist any food cravings?	10 = Extremely difficult, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0 = Not at all difficult
12	How often have you eaten in response to food cravings?	10 = After every one, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0 = Not at all
13	How often have you had cravings for chocolate and chocolate flavoured foods?	10 = Extremely often, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0 = Not at all
14	How often have you had cravings for other sweet foods (cakes, pastries, biscuits, etc)?	10 = Extremely often, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0 = Not at all
15	How often have you had cravings for fruit or fruit juice?	10 = Extremely often, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0 = Not at all
16	How often have you had cravings for dairy foods (cheese, yoghurt)?	10 = Extremely often, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0 = Not at all
17	How often have you had cravings for starchy foods (bread, pasta)?	10 = Extremely often, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0 = Not at all
18	How often have you had cravings for savoury foods (fries, crisps, burgers etc)?	10 = Extremely often, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0 = Not at all
19	Generally, how difficult has it been to control your eating?	10 = Extremely difficult, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0 = Not at all difficult

Item scores

The CoEQ items are scored on an 11-point graded response scale ranging from 10 to 0. If data are missing for an item, the item score is treated as missing. No reversal of item scores will be done.

Domain scores

The sum of the items in each domain is calculated, and divided by the number of items in the domain in order to obtain a domain score. Items 1 and 2 are not included in any domain score. Details are given in Table 2–7.

Table 2–7 Overview of domains for CoEQ

Domain	Items numbers of items included in domain	Comment
Craving Control	9-12, 19	The domain score is reversed so that a greater score represents a greater level of Craving Control (i.e. 10 to 0, 9 to 1, …, 0 to 10)
Positive Mood	5-8	Scores from item 6 (“How anxious have you felt?’’) are reversed (i.e. 10 to 0, 9 to1, …, 0 to 10)
Craving for Savoury	4, 16-18	
Craving for Sweet	3, 13-15	

Missing data at instrument level will be handled in the following way. To score a domain it is required that at least 50% of the items need to be answered. Then, the domain is scored based on the average of the items answered. If less than 50% of the items of a domain are answered no score will be derived.

Responder threshold values

Half of a standard deviation (SD) of the baseline CoEQ item and domain scores are used as distribution-based approach defining the responder thresholds. The thresholds are derived from baseline CoEQ data across Oral sema 14 mg and Empa 25 mg.

2.7.3 Responder analyses for PRO to be reported in a separate report

Responder analyses will be reported in a separate report from the CTR for SF-36v2® and CoEQ. These additional analyses were not in scope during the development process of the protocols for the oral semaglutide phase 3a programme trials. Responder analyses will be conducted for both estimands, for the same time points that are defined for the analyses of PRO endpoints and separately for each domain.

For descriptive statistics the following subject responder categorization is applied for all relevant time points and domain:

- Responder (improvement): Individual change from baseline in score ≥ positive responder threshold
- Non-responder (no change): Individual change from baseline in score > negative responder threshold value and < positive responder threshold value
- Non-responder (worsening): Individual change from baseline in score ≤ negative responder threshold value

The following binary subject responder definition is applied for all relevant time points and scores:

- Responder: Individual change from baseline in score ≥ positive responder threshold
- Non-responder: Individual change from baseline in score < positive responder threshold

The binary responder endpoints will be analysed as the other supportive secondary binary efficacy endpoints (see section 2.4.2.1). Estimated proportions and differences in proportions will be reported in addition to odds and odds ratios.

3 Changes to the statistical analyses planned in the protocol

The main analyses were described in the protocol for the trial NN9924-4223. However, clarifications, more detailed descriptions of endpoints and analyses are provided in this SAP. The changes from the protocol of NN9924-4223 are summarised below:

- It has been specified which countries belong to which regions.

- The primary and secondary estimands have changed names from de-facto and de-jure to treatment policy and hypothetical, respectively.
- The MMRM sensitivity analysis of the primary estima nd has been omitted in section 2.3.3. It is considered sufficient to keep the two current sensitivity analyses to stress test the primary results.
- For the MMRM analyses, it is specified that for subjects who do not have post-baseline assessments for planned visits available in the on-treatment without rescue medication period, the baseline value will be carried forward to the first planned visit, if the first planned visit do not fall later than 8 weeks after randomisation, to ensure that all randomised subjects will contribute to the statistical analyses.
- The three MI sensitivity analyses of the secondary estimand have been omitted in section 0. It is considered sufficient to keep the tipping point sensitivity analysis for the secondary hypothetical estimand as it can be considered as a progressive stress-testing to assess how severe departures from MAR must be in order to reverse the conclusions from the primary MMRM analysis used to address the hypothetical estimand.
- The LOCF sensitivity analysis specified in the trial protocol (section 17.3.3.2) has been omitted, as it is not realistic that subjects with missing data would have had stable results from the point of drop out to trial completion.
- The statistical analyses of the two binary effect endpoints (HbA_{1c} reduction ≥ 1%-point (10.9 mmol/mol) and body weight loss ≥ 3%) have been omitted, because they are being analysed as a part of the two composite binary effect endpoints.
- For the binary efficacy endpoints, it has been specified how missing data in the analyses for the hypothetical estimand will be imputed using a sequential imputation approach assuming MAR.
- A clarification of the 'no hypoglycaemia' component in composite binary endpoints has been added.
- It has been specified which assessments will be analysed on logarithmic scale.
- The definitions of initiation of rescue medication and additional anti-diabetic medication used for the time-to-event endpoints as well as the accompanying statistical analyses have been further clarified.
- It has been specified that all safety laboratory results (except amylase and lipase) are safety assessments and not safety endpoints as written in the trial protocol.
- Both PROs, SF-36v2 (acute version) and CoEQ, will be analysed statistically, however only using the primary analysis of the primary estimand.
- The responder analyses and the primary analysis for the secondary estimand of both PROs, SF-36v2 (acute version) and CoEQ, will be presented in a report separate from and after finalisation of the CTR.

4 References

1. European Medicines Agency. Assessment report for Jardiance (INN name: empagliflozin). EMA/CHMP/137741/2014. 20 March 2014.
2. Bretz F, Posch M, Glimm E, Klinglmueller F, Maurer W, Rohmeyer K. Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. Biometrical Journal. 2011;53(6):894-913.
3. Rohmeyer K, Klinglmueller F. gMCP: Graph Based Multiple Test Procedures. R package version 0.8-8. 3 Oct 2014.
4. Little RJA, Rubin DB. Statistical analysis with missing data: New York: John Wiley & Sons. 1987.
5. Koch GG. Comments on 'Current issues in non-inferiority trials' by Thomas R. Fleming, Statistics in Medicine, DOI: 10.1002/sim.2855. Stat Med. 2008;27(3):333-42.
6. Committee for Medicinal Products for Human Use (CHMP). Guideline on Missing Data in Confirmatory Clinical Trials. EMA/CPMP/EWP/1776/99 Rev. 1. 2 Jul 2010.
7. National Academy of Sciences (NAS). The Prevention and Treatment of Missing Data in Clinical Trials. Washington D.C.: The National Academies Press. 2010.
8. Schwartz NS, Clutter WE, Shah SD, Cryer PE. Glycemic thresholds for activation of glucose counterregulatory systems are higher than the threshold for symptoms. J Clin Invest. 1987;79(3):777-81.
9. Seaquist ER, Anderson J, Childs B, Cryer P, Dagogo-Jack S, Fish L, et al. Hypoglycemia and diabetes: a report of a workgroup of the American Diabetes Association and the Endocrine Society. Diabetes Care. 2013;36(5):1384-95.
10. Maruish, M. E. (2011). *User's manual for the SF-36v2 Health Survey (3rd ed.).* Lincoln, RI: QualityMetric Incorporated.
11. Dalton, M., Finlayson, G., Hill, A., & Blundell, J. (2015). Preliminary validation and principal components analysis of the Control of Eating Questionnaire (CoEQ) for the experience of food craving. *European Journal of Clinical Nutrition*, Vol. 69, p. 1313-1317.

Section removed: 16.1.9.1 Pre-defined MedDRA search – list of preferred terms - (not part of statistical analysis plan).