16.1.9 Documentation of statistical methods

List of contents

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

.

Statistical Analysis Plan

Trial ID: NN9924-4222

PIONEER 3 - vs. DPP-4 inhibitor

Efficacy and long-term safety of oral semaglutide versus sitagliptin in subjects with type 2 diabetes

Author:

[redacted]

Table of Contents

List of abbreviations

ADA	American Diabetes Association
AE	adverse event
AACE	American Association of Clinical Endocrinologists
ANOVA	analysis of variance
ANCOVA	analysis of covariance
ATC	Anatomical Therapeutic Chemical
BG	blood glucose
BMI	body mass index
CI	confidence interval
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
GI	gastrointestinal
HbA1c	glycosylated haemoglobin
HDL	high-density lipoprotein
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
MCAR	missing completely at random
MCS	mental component score
MCMC	Markov Chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
MI	multiple imputation
MMRM	mixed model for repeated measurements
MNAR	missing not at random
NBS	norm-based score
PCS	physical component score
PG	plasma glucose
PK	pharmacokinetic
PP	per protocol analysis set
PRO	patient reported outcome
REML	restricted maximum likelihood
SAP	statistical analysis plan
SAS	safety analysis set
s.c.	subcutaneous
SD	standard deviation

SF-36v2 (acute version)	SF-36v2® Health Survey (acute version)
SNAC	sodium N-[8-(2-hydroxybenzoyl)amino]caprylate
SU	sulfonylurea
SMPG	self-measured plasme glucose
T2DM	type 2 diabetes mellitus
TE	treatment effect
TEAE	treatment emergent adverse event
VLDL	very low-density lipoprotein

1 Introduction

1.1 Trial information

1.1.1 Objective(s)

1.1.1.1 Primary objective

To compare the effect of once-daily dosing of three dose levels (3 mg, 7 mg and 14 mg) of oral semaglutide versus sitagliptin 100 mg once-daily, both in combination with metformin with or without SU, on glycaemic control in subjects with T2DM.

1.1.1.2 Secondary objectives

To compare the effect of once-daily dosing of three dose levels (3 mg, 7 mg and 14 mg) of oral semaglutide versus sitagliptin 100 mg once-daily, both in combination with metformin with or without SU, on body weight in subjects with T2DM.

To compare the long-term safety and tolerability of once-daily dosing of three dose levels (3 mg, 7 mg and 14 mg) of oral semaglutide versus sitagliptin 100 mg once-daily, both in combination with metformin with or without SU, in subjects with T2DM.

1.1.2 Type of trial

The trial is a 78-week, randomised, double-blind, double-dummy, active-controlled, parallel-group, multi-centre, multi-national trial with four arms comparing efficacy and safety of oral semaglutide 3 mg, 7 mg and 14 mg once-daily with sitagliptin 100 mg once-daily. Subjects with T2DM inadequately controlled on metformin alone or in combination with SU will be randomised in a 1:1:1:1-manner to receive either:

- oral semaglutide 3 mg and sitagliptin placebo
- oral semaglutide 7 mg and sitagliptin placebo
- oral semaglutide 14 mg and sitagliptin placebo
- sitagliptin 100 mg and oral semaglutide placebo

Randomisation will be stratified according to descent (Japanese subjects/non-Japanese subjects) and anti-diabetic pre-trial background medication (metformin or metformin + SU) to ensure even distribution of the four treatment arms within strata.

To maintain the blinding of the trial, the treatment will consist of two tablets daily as the sitagliptin tablet is not visually identical to the oral semaglutide tablets.

Total trial duration for the individual subject will be approximately 85 weeks. The trial includes a 2-week screening period, followed by a 78 week randomised treatment period and a follow-up period of 5 weeks.

A schematic diagram of the trial design is shown in Figure 1–1.

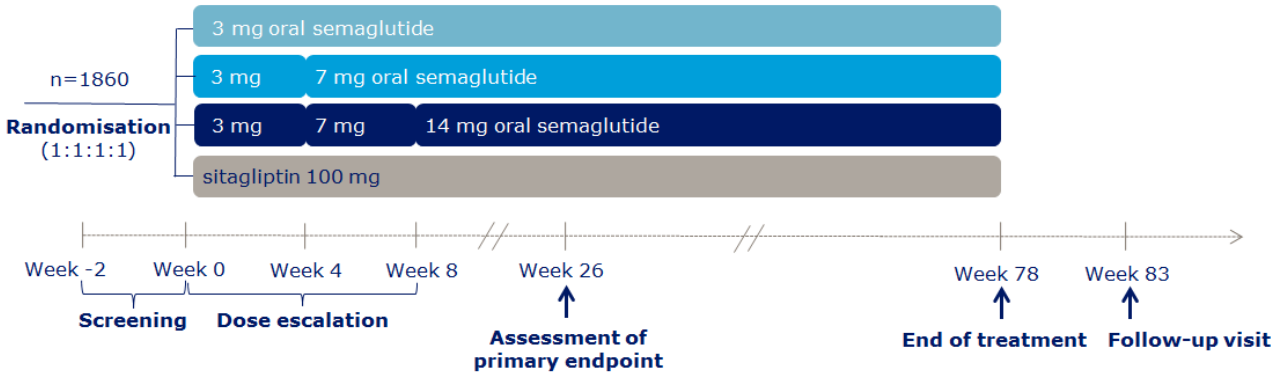


Figure 1–1 Trial design

1.2 Scope of the statistical analysis plan

This statistical analysis plan (SAP) is based on the protocol for trial NN9924-4222 "Efficacy and long-term safety of oral semaglutide versus sitagliptin", version 4.0 (23 November 2016) as well as the Protocol amendment no. 1, version 1.0 (10 November 2015) and the Protocol amendment no. 1, version 2.0 (14 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.

2 Statistical considerations

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

General considerations

The blinding of the randomised treatments will be maintained until the database has been released for statistical analysis. No interim analyses or other analyses of unblinded data will be performed before the database is locked.

Data from all sites will be analysed and reported together.

The randomisation is stratified based on pre-trial therapy at screening (metformin or metformin+SU) and descent (Japanese subjects/non-Japanese subjects). In the statistical analyses descent is a part of region and pre-trial therapy at screening is referred to as stratification factor. The two combinations (metformin; metformin+SU) in the stratification factor will be included based on the actual information collected through the eCRF. In case of missing eCRF information the information collected from IWRS system will be used.

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 confirmed at week 26. This approach will result in a lower proportion of missing data compared to the expected proportion of missing data at week 52 and week 78 and therefore a more reliable estimate of efficacy.

Results from a statistical analysis will as a minimum be presented by the estimated treatment contrasts for the below three comparisons at week 26, week 52 and week 78 with associated two-sided 95% confidence intervals and p-values corresponding to two-sided tests of no difference:

- oral semaglutide 14 mg vs. sitagliptin 100 mg
- oral semaglutide 7 mg vs. sitagliptin 100 mg
- oral semaglutide 3 mg vs. sitagliptin 100 mg

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

Primary and secondary estimands

Two estimands adressing different aspect of the primary trial objective will be defined as follows:

- Primary estimand – 'Treatment policy'
  - treatment difference at week 26 for all randomised subjects
- 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 treatment policy estimand assesses the average effect in a future population that results from initiating treatment with oral semaglutide including potential rescue medication(s) as compared to initiating treatment with sitagliptin 100 mg including potential rescue medication(s). Generalisation of this estimand depends among other things on the extent to which the use of rescue medication and treatment adherence in this trial reflects clinical practice. All post-baseline scheduled visit data will be included in the analysis, including data collected after discontinuation of trial product or initiation of rescue medication(s).

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 sitagliptin 100 mg. It is considered a clinically relevant estimand as it provides information to treating clinicians about the expected glycaemic efficacy of oral semaglutide compared to sitagliptin 100 mg for purposes of treating individual subjects. Generalisation of this estimand depends among other things on the extent to which the compliance to trial product administration in this trial reflects clinical practice. Only post-baseline scheduled visit data collected prior to discontinuation of trial product or initiation of rescue medication will be included in the analysis. This will avoid confounding from rescue medication.

Missing data considerations at week 26

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), when estimating the primary estimand, 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.

The proportion of missing data when estimating the secondary estimand 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 20% of missing data is based on the sitagliptin phase 3 trials[1] the oral semaglutide phase 2 trial (NN9924-3790) and the indication that a low starting dose with gradual dose escalation diminishes GI AEs compared with more aggressive dosing regimens. Across treatment arms the main reasons for missing data are expected to be early treatment discontinuation due to GI AEs and eventually initiation of rescue medication. Initiation of rescue medication is expected to be more frequent in the sitagliptin 100 mg arm and in the oral semaglutide 3 mg arm

than for the two highest dose levels of oral semaglutide. A higher proportion of subjects are expected to discontinue treatment due to AEs in the oral semaglutide 14 mg arm, compared to the other treatment arms. 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 versus sitagliptin 100 mg are planned to be tested at each dose level. The confirmatory secondary endpoint is change from baseline to week 26 in body weight (kg). For body weight, superiority versus sitagliptin 100 mg is planned to be tested at each dose level. The sample size calculation is based on jointly meeting the below three out of the nine pre-specified confirmatory hypotheses shown in Table 2–1. The closed testing procedure described in Bretz et al 2011[2] is used to control the overall type-1 error at a nominal two-sided 5% level. The three hypotheses are

- HbA_{1c} superiority of oral semaglutide 14 mg vs. sitagliptin 100 mg
- HbA_{1c} superiority of oral semaglutide 7 mg vs. sitagliptin 100 mg
- HbA_{1c} non-inferiority of oral semaglutide 3 mg vs. sitagliptin 100 mg (margin 0.3%)

The statistical testing strategy is based on the following two principles:

- Within a dose level, glycaemic efficacy must be established by HbA_{1c} non-inferiority before testing for added benefits in terms of superiority for HbA_{1c} and/or superiority of body weight.
- Glycaemic efficacy by HbA_{1c} non-inferiority must be established on all higher dose levels before continuing testing hypotheses on lower dose levels.

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

The sample size assumptions for treatment effects (TE), adjusted treatment effects and the common standard deviation (SD) used across dose levels are given in Table 2–1. These are based on the oral semaglutide phase 2 results (NN9924-3790), sitagliptin phase 3a trial results and supported by results from the s.c. semaglutide phase 2 trial (NN9535-1821).

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 discontinue trial product or initiate rescue medication and for the 10% of subjects with missing data. The treatment effects used in the sample size calculation will be adjusted according to a 75% smaller effect in these subjects. In addition, the non-inferiority margin of 0.3 % for HbA_{1c} is used, 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 for sample size calculation

Semaglutide vs. sitagliptin	HbA_{1c}			Body weight		
Treatment dose	14 mg	7 mg	3 mg	14 mg	7 mg	3 mg
Treatment effect (TE)	-0.5%	-0.3%	-0.1%	-3.0 kg	-2.0 kg	-1.0 kg
Adjusted TE, non-inferiority	-0.395%	-0.225%	-0.055%			
Adjusted TE, superiority	-0.425%	-0.255%	-0.085%	-2.55 kg	-1.70 kg	-0.85 kg
Standard deviation	1.1%	1.1%	1.1%	4.0 kg	4.0 kg	4.0 kg

With the above assumptions, allocating 465 subjects to each of the semaglutide treatment arms and sitagliptin 100 mg provides 90 % power to jointly confirm HbA_{1c} superiority of semaglutide 14 mg vs. sitagliptin 100 mg, HbA_{1c} superiority of semaglutide 7 mg vs. sitagliptin 100 mg and HbA_{1c} non-inferiority of semaglutide 3 mg vs. sitagliptin 100 mg. Calculated powers for selected individual hypotheses are presented in Table 2–2. In total 4×465 = 1860 subjects are planned to be randomised.

Table 2–2 Powers for individual hypotheses

Statistical test	HbA_{1c} superiority			Body weight superiority			HbA_{1c} non-inferiority (margin = 0.3 %)
Treatment dose	14 mg	7 mg	3 mg	14 mg	7 mg	3 mg	3 mg
Power (%)	> 99	90	19	> 99	> 99	89	> 99

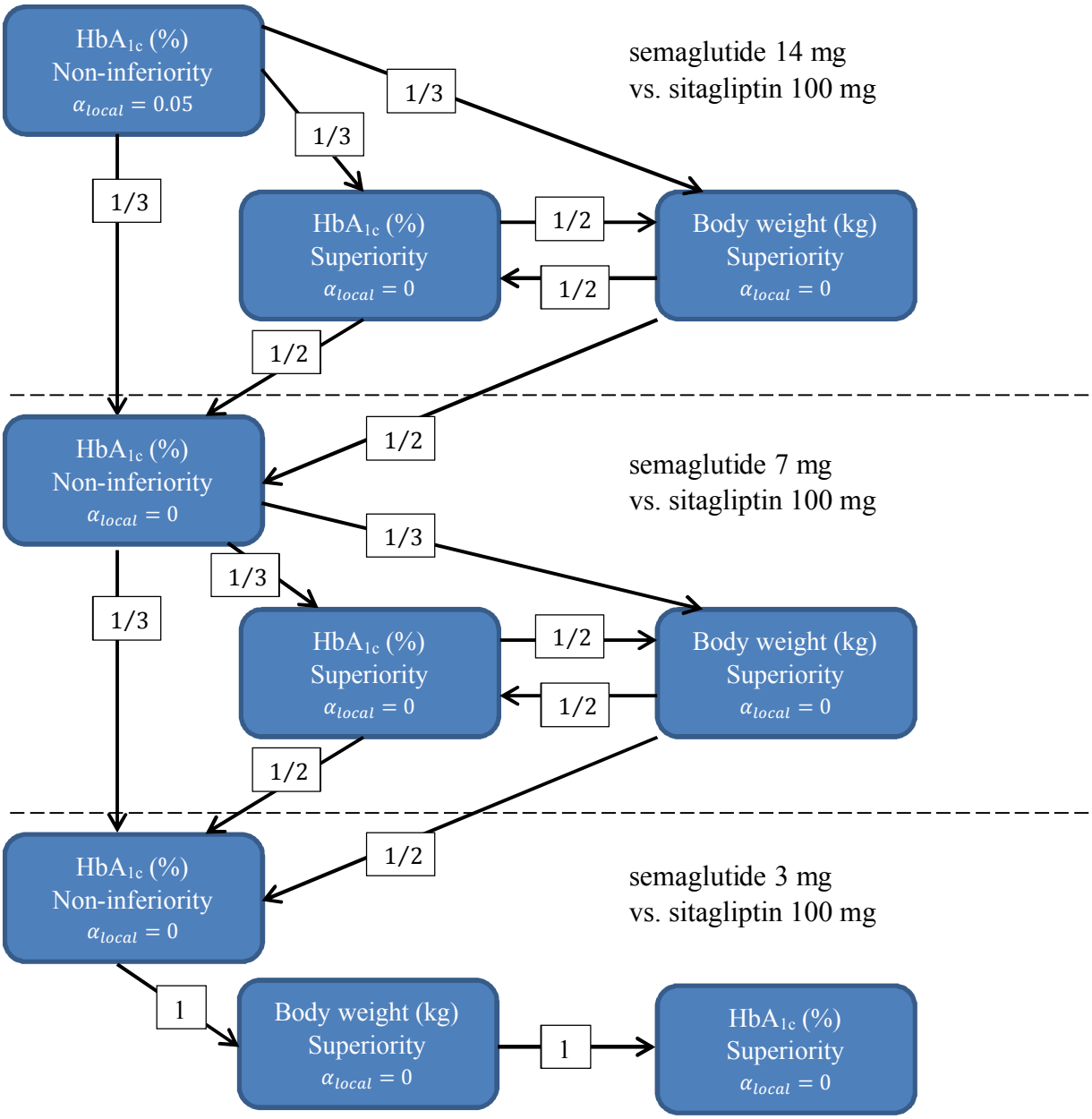


Figure 2–1 Graphical illustration of the closed testing procedure

The overall significance level of α = 0.05 (two-sided) is initially allocated to the HbA_{1c} non-inferiority test of semaglutide 14 mg vs. sitagliptin 100 mg. 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).

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 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 based on the trial product received for the majority of the period where they were on treatment. This will be referred to as contributing to the evaluation "as treated".

Per protocol (PP) analysis set: includes all subjects in the full analysis set 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 contribute to the analysis "as treated".

Before data are locked for statistical analysis and the randomisation code is broken, 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 will be used restrictively and normally no subjects will be excluded from the FAS. If any subjects or observations are to be excluded from the analysis sets or the observation periods defined below, this, together with the reasons for their exclusion, will be documented and signed by those responsible before database lock. A description of these exclusions will be included in the clinical trial report.

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 (V17) for subjects on trial product
- the End-of-treatment visit (V16) or the follow-up premature discontinuation visit (V17A), whichever comes last, 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.

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

In this trial, three observation periods will be defined, as described below. 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, onset date will be the EAC adjudicated onset date.

Definition of the observation periods:

- **In-trial:** This observation period will include information assessed at or after date of randomisation (as registered in the IWRS system) and up to and including the last direct subject-site contact, which is scheduled to take place 5 weeks (+3 days visit window) after the planned last dose on trial product at a follow-up visit. For subjects who withdraw their informed consent, the in-trial observation period ends at their date of withdrawal. If a subject is lost to follow-up, the end of the in-trial period is defined as the date of the last subject-investigator contact (site or phone visit). In the case a subject dies during the trial the date of death will be the end-date of the in-trial observation period regardless of the above defined end-dates. This observation period will be the primary observation period for estimating the primary estimand and used for evaluating safety.

- **On-treatment:** This observation period is a subset of the in-trial observation period and represents the time period in which a subject is considered exposed to trial product. For adjudicated events, ECGs, eye examination category, anti-semaglutide antibodies and AEs including hypoglycaemic episodes information collected on or after the first date of trial product up to and including the first date of (i) the follow-up visit V17, (ii) the follow-up prematurely discontinuation visit V17A, (iii) the last date on trial product +38 days or (iv) the end-date for the in-trial observation period will be used. 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. In addition, the ascertainment window includes the follow-up visit window of +3 days after last date on trial product. For efficacy and other safety assessments (laboratory assessments, physical examination and vital signs) information collected on or after the first date of trial product up to and including the last date on trial product +3 days will be used in order to ensure specificity to reversible effects of treatment. The end-date for all assessments in the on-treatment observation period will be before or the same date as the end-date for the in-trial observation period. This observation period will be used for evaluating safety.

- **On-treatment without rescue medication:** This observation period is a subset of the on-treatment observation period. To avoid potential confounding from rescue medications, information that is collected after initiation of rescue medication will be excluded from this observation period. Specifically, it includes information collected at or after date of first dose on trial product up to and including the first date of; (i) last date on trial product +3 days or (ii) initiation of rescue medication. Thus, an ascertainment window of 3 days for subjects not

initiating rescue medication. This observation period will be the primary observation period for estimating the secondary estimand.

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). Imputation of missing data at week 26 will be done within 8 groups of subjects defined by randomised treatment arm and whether subjects at week 26; (i) have discontinued treatment or initiated rescue medication or (ii) 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 status.

Missing values for each group will be imputed as follows:

- An analysis of covariance (ANCOVA) with region and stratification factor as categorical fixed effects 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, stratification factor 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, Africa 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 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 sitagliptin at week 26:

For each of the 1000 (now complete) imputed data sets the change from baseline to week 26 will be analysed using an ANCOVA with treatment, region and stratification factor 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.

From this analysis the estimated treatment differences between each of the oral semaglutide dose levels and sitagliptin 100 mg together with associated two-sided 95 % CIs and unadjusted two-sided p-values for testing no difference from zero will be presented.

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

Prior to analysing the data using the same model and approach as used for confirming superiority (see above) a value of 0.3 % (the non-inferiority margin) will be added to imputed values at week 26 for the oral semaglutide treatment arms only (Koch 2008)[5]. For evaluating non-inferiority versus sitagliptin 100 mg unadjusted two sided p-values 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 (REML) 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, stratification factor 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. For subjects who have no post-baseline scheduled assessments available, the baseline value will be carried forward to the first scheduled visit to ensure that all randomised subjects will contribute to the statistical analysis.

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. Even if the assumption of MAR is not satisfied, this analysis is not expected to bias the estimated HbA_{1c} treatment effect for the secondary estimand in favour of oral semaglutide to any important degree. This is supported by the oral semaglutide phase 2 results (NN9924-3790) that showed that subjects who discontinue oral semaglutide do not have better outcome on average than those who remain on treatment.

Primary hypotheses

For the primary HbA_{1c} endpoint the following two confirmatory one-sided hypotheses are planned to be tested at each dose level of oral semaglutide versus sitagliptin 100 mg. Let the mean treatment difference be defined as μ = (oral semaglutide minus sitagliptin 100 mg):

- Non-inferiority, using a non-inferiority margin of 0.3 %
  - H_0: $\mu \geq 0.3$ % against Ha: $\mu < 0.3$ %

- Superiority
  - H_0: $\mu \geq 0.0$ % against Ha: $\mu < 0.0$ %

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

Multiplicity and criteria for confirming hypotheses

The Type-I error for testing the nine confirmatory hypotheses related to the HbA_{1c} and body weight endpoints (see Section 2.1) 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 . The first hypothesis to be tested is non-inferiority of HbA_{1c} at the highest dose level. 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 . 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 . 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.

2.3.3 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 EMA recommendations[6] and with a report from the US National Research Council[7], these analysis will primary evaluate the sensitivity of the results due to the impact of missing data. Since conservatism, i.e. avoiding bias in favour of 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. 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 analyses for the primary analysis will be described that are not based on the pattern mixture model approach.

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 arms, while maintaining the missing data assumption for the sitagliptin 100 mg arm. The analyses will all be implemented using multiple imputation as described for the primary analysis of the primary estimand:

- *Comparator multiple imputation analysis:* In this sensitivity analysis missing data at week 26 for all subjects will be imputed to resemble in distribution the week 26 values observed in the sitagliptin 100 mg treatment arm. In effect this imputation approach removes the treatment difference between oral semaglutide and sitagliptin 100 mg for all subjects randomised to oral semaglutide, given that oral semaglutide is better than sitagliptin 100 mg. 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 in distribution the week 26 values observed in the sitagliptin 100 mg 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 sitagliptin 100 mg 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, missing data will first be imputed according to the primary analysis. Second, for all oral semaglutide treatment arms a penalty will be added to the imputed values at week 26. The approach is to gradually increase this penalty until all confirmed HbA_{1c} conclusions from the primary analysis are reversed. For each hypothesis tested the specific value of the penalty that reverses the conclusion will be used to evaluate the robustness of the primary analysis results. 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 analysis

The following additional sensitivity analysis will be specified

- *Per-protocol analysis*: This sensitivity will be based on the per-protocol analysis set. Data from the on-treatment without rescue 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} and body weight. Due to the 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 imputation and analysis model.

From the analyses the three estimated treatment differences between each of the oral semaglutide dose levels and sitagliptin 100 mg will be presented together with associated two-sided 95 % CIs and unadjusted two-sided p-values for testing no difference from zero.

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 and 78 in:

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

Change from baseline to week 26, 52 and 78 in:

- Body weight (%)
- FPG
- BMI
- Waist circumference
- Fasting lipid profiles (total cholesterol, LDL cholesterol, VLDL cholesterol, HDL cholesterol, triglycerides, free fatty acids)

Change from baseline to week 26, 52 and 78 in the below derived endpoints from the 7-point profile:

- Mean of the 7-point profile, defined as the area under the profile, calculated using the trapezoidal method, divided by the measurement time
- Mean 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. Fasting lipid profile endpoints 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, 52 and 78. For the analysis at week 52 and at week 78, 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 within12 groups of subjects instead of the 8 groups as described for the week 26 evaluation in Section 2.3.1. If less than five subjects have available data in one of the 12 groups, the imputation will be made within the 8 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 and week 78 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 78. From this model the estimated treatment differences (ratios) will be presented at week 26 (except for HbA_{1c} and body weight), 52 and 78 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

Subjects who after 26 weeks of treatment achieve (yes/no):

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

When addressing the treatment policy estimand the 'without 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 endpoints will be evaluated after week 52 and after week 78 as well.

Handling of missing data for binary endpoints

HbA_{1c} and body weight

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 HbA_{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.

Without hypoglycaemia

For both estimands missing data in the 'without hypoglycaemia' component will be accounted for using a multiple imputation approach[8] in which the number of events of severe or BG-confirmed hypoglycaemia in the missing data period will be imputed. The 'without hypoglycaemia' component will be calculated as a dichotomization of the sum of the observed and the imputed number of episodes covering the observed and missing data periods.

Treatment policy estimand

Analogously to the primary imputation model for the continuous endpoints the imputation of hypoglycaemic episodes will assume that the missing data is best described by information from subjects who are similar in terms of randomised treatment arm and whether subjects at time of evaluation; (i) have discontinued treatment or initiated rescue medication or (ii) are still on treatment and have not initiated rescue medication (treatment status). I.e. withdrawn subjects or subjects who are lost to follow-up are assumed to have the same conditional event rate after withdrawal conditional on the observed event rate before withdrawal and on randomised treatment

and treatment status and the other covariates. For evaluation at week 52 and 78 treatment status will be differentiated by whether subjects have discontinued trial product or initiated rescue medication prior to week 26 or at/after week 26. The imputation will be done as described below:

- A Bayesian log-linear negative binomial model is fitted to the observed data from the in-trial observation period to obtain the posterior distribution of the model parameters. The model will include randomised treatment, treatment status and stratification factor as fixed factors and baseline HbA_{1c} as a covariate. The logarithm of the duration of the subject's observed data period will be included as offset.
- Based on the estimated parameters in this model the number of events in the missing data period will be imputed conditional on the event rate in the observed data period. 1000 imputations will be done sampling from the posterior distribution of the model parameters.

Hypothetical estimand

Imputation of events when addressing the hypothetical estimand will be done in a similar way. Subjects who discontinue trial product or initiate recue medication are assumed to have the same conditional event rate after discontinuation/rescue initiation conditional on the observed event rate before discontinuation/rescue initiation and on randomised treatment and the other covariates. The imputation will be done as described below:

- A Bayesian log-linear negative binomial model is fitted to the observed data from the on-treatment without rescue observation period to obtain the posterior distribution of the model parameters. The model will include randomised treatment and stratification factors as fixed factors and baseline HbA_{1c} as a covariate. The logarithm of the duration of the subject's observed data period will be included as offset.
- Based on the estimated parameters in this model the number of events in the missing data period will be imputed conditional on the event rate in the observed data period. 1000 imputations will be done sampling from the posterior distribution of the model parameters.

For the composite endpoint the imputed datasets of the three components will be combined by imputation number and analysed in the same way as described above for the other binary endpoints.[4]

Time to event endpoint

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

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 after 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 sitagliptin 100 mg 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.

2.4.2.2 Safety endpoints

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

Adverse events

- Number of TEAEs during exposure to trial product, assessed up to approximately 83 weeks

All AEs will be coded using the most recent version 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.

Other safety endpoints

Change from baseline to week 26, 52 and 78 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, 52 and 78. 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, 52 and 78 in:

- Electrocardiogram (ECG) evaluation
- Physical examination (week 52 and 78 only)
- Eye examination category (week 52 and 78 only)

Occurrence of anti-semaglutide antibodies (positive/negative):

- 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 antibody levels

Safety assessments

Change from baseline to week 26, 52 and 78 in:

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

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

Pharmacokinetics

Please refer to Section 2.5.

Hypoglycaemia

- Number of treatment emergent severe or BG confirmed symptomatic hypoglycaemic episodes during exposure to trial product, assessed up to approximately 83 weeks
- Treatment emergent severe or BG confirmed symptomatic hypoglycaemic episodes during exposure to trial product, assessed up to approximately 83 weeks (yes/no)

Classification of Hypoglycaemia:

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

Treatment emergent: hypoglycaemic episode 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: are 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

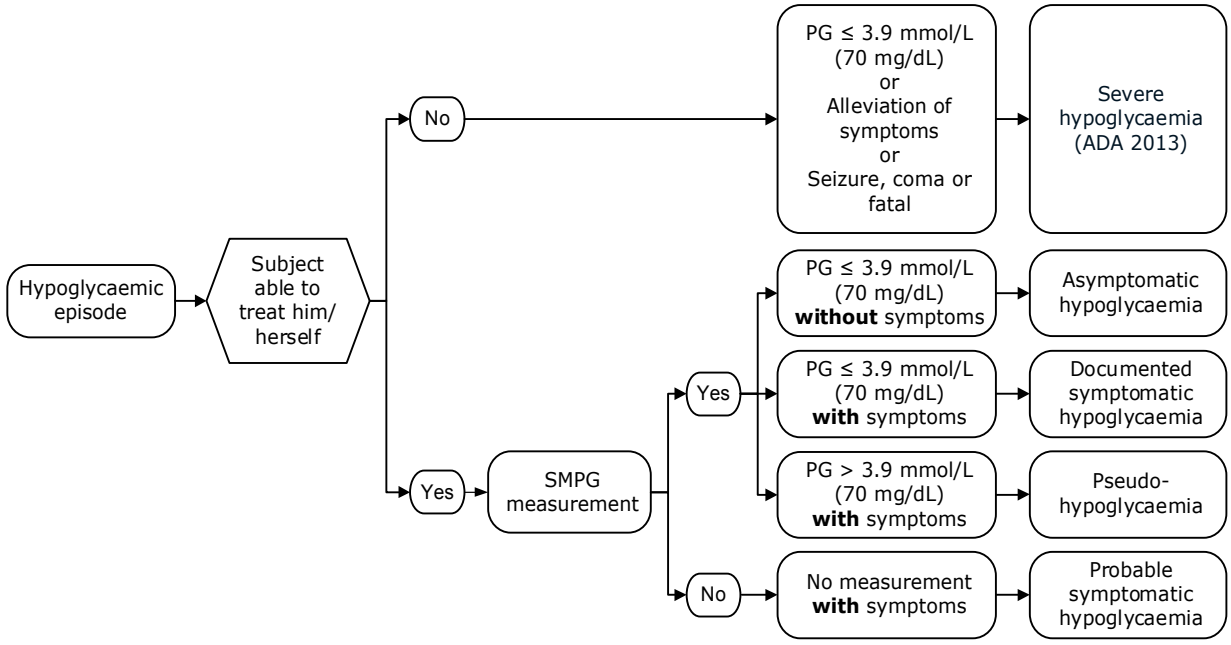
In normal physiology, symptoms of hypoglycaemia occur below a PG level of 3.1 mmol/L (56 mg/dL)[9]. Therefore, Novo Nordisk has included hypoglycaemia with PG levels below this cut-off point in the definition of BG confirmed hypoglycaemia.

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[10] or BG confirmed by a PG value < 3.1 mmol/L (56 mg/dL) **with** symptoms consistent with hypoglycaemia.

ADA classification[10] of hypoglycaemia

- 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

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 analysed 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 factors for treatment, stratification factor 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, stratification factor and region as fixed factors and baseline HbA_{1c} as covariate.

2.5 Pharmacokinetic modelling

Exploratory population PK modelling will be used to evaluate semaglutide exposure and the effects of pre-specified covariates on the exposure. As a minimum, the covariates sex, body weight at

baseline and age at baseline will be explored. The modelling will include data from approximately 50 % of randomised subjects that were exposed to semaglutide in this trial and will be performed as a meta-analysis across phase 3a trials. Actual time points for dose administration and PK sampling will be used. Results will be presented using criteria which will be pre-specified in a modelling analysis plan that is to be finalised before DBL. The modelling will be performed by Quantitative Clinical Pharmacology at Novo Nordisk A/S and will be reported separately from the CTR.

Covariate analysis

Covariates will be tested on CL/F. Covariates for inclusion:

- Body weight
- Age group (< 65, 65-75 and ≥ 75 years)
- Sex (Male/Female)
- Race (White/ Black or African American/ Asian/ Hawaiian or Pacific Islander/ Other)
- Ethnicity (Hispanic or Latino/ not Hispanic or Latino)
- Upper GI comorbidity (Yes/No)

2.6 Health economics and/or patient reported outcomes

PROs

Change from baseline to week 26, 52 and 78 in:

- SF-36v2™ (acute version) health survey: Physical component score, mental component score and scores from the 8 domains
- IWQoL-Lite Clinical Trial Version: Total score and scores from the 4 domains
- CoEQ: Scores from the 4 domains and scores from 19 individual items

The PRO endpoints will be analysed separately as the other continuous efficacy endpoints using a similar model approaches as for the primary endpoint with the associated baseline response as a covariate.

2.6.1 SF-36v2® (acute version) health survey

The SF-36v2® Health Survey (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 diabetes10. The SF-36v2 (acute version) for adults with a 1-week recall period contains 36 items.

A total of 35 items measure eight domains of functional health and well-being as well as two component summary scores: 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 summary (MCS) score, physical component summary (PCS) score. There is an additional single item giving information on health change over the past week.

Domain scores

Norm-based scores (NBS) will be derived using the QualityMetric Health Outcomes™ Scoring Software1 including the 2009 US general population norm. The most recent version of the QualityMetric Health Outcomes™ Scoring Software available at time of licensing was used for the specific trial (version 4.5 for PIONEER 3). Table 2–3 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 do mains and component summary measures (PCS and MCS) indicate better HRQoL/general health status. Item 2 (i.e. Question 2 in CRF) is not included in any score.

Table 2–3 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 PF domain must be one of the seven domains having valid data. Also, to calculate MCS, the MH 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 2–4.[11]

Table 2–4 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

Responder analyses will be based on the responder threshold values and are described in Section 2.6.4.

2.6.2 Impact of Weight on Quality of Life Clinical Trials Version (IWQOL-Lite-CT)

The IWQOL-Lite-CT is designed to assess the impact of changes in weight on patients' quality of life within the context of clinical trials. The items of the IWQOL-Lite-CT pertain to physical functioning and psychosocial domains and all items employ a 5-point graded response scale (never, rarely, sometimes, usually, always; or not at all true, a little true, moderately true, mostly true, completely true). The 23-item version of the IWQOL-Lite-CT was completed by patients.
Composite scores

All IWQOL-Lite-CT composite scores range from 0 to 100, with higher scores reflecting better levels of functioning. Table 2–5 provides an overview of domains.

Table 2–5 Overview of domains for IWQOL-Lite-CT

Domain	Items numbers of items included in domain 23 items version
Psychosocial	6-8, 10, 12-17, 20, 22, 23
Physical	1-5, 18, 19
Physical Function	1-3, 18, 19
Pain/Discomfort	4, 5
IWQOL-Lite-CT Total	1-8, 10, 12-20, 22, 23

Missing data at instrument level will be handled in the following way. Raw scores for each subscale are computed if a minimum of 50% of the items for that subscale are nonmissing, and for the IWQOL-Lite-CT total score if a minimum of 75% of all items are nonmissing.

The scoring is done in three steps12.

1. If the minimum number of items are answered for a composite, compute the average of the valid item responses for that composite where 1 = ”never” or “not at all true” and 5 = ”always” or “completely true.” This average must be a number between 1 and 5, inclusive.
2. Multiply that average by the total number of items in that composite. The total number of items in each IWQOL-Lite-CT composite is as follows: Psychosocial = 13, Physical = 7, Physical Function = 5, Pain/Discomfort = 2, IWQOL-Lite-CT Total = 20. Round to the nearest integer.
3. To convert the IWQOL-Lite-CT raw score (as calculated above) to the 0 (worst) to 100 (best) metric:

(a) Subtract the raw score from the maximum raw score for each composite. The maximum raw scores are as follows: Psychosocial = 13x5 = 65, Physical = 7x5 = 35, Physical Function = 5x5 = 25, Pain/Discomfort = 2x5 = 10, IWQOL-Lite-CT Total = 20x5 = 100.

(b) Divide the difference in (a) by the raw score range for each composite. The ranges of the raw scores are as follows: Psychosocial = 65-13 = 52, Physical = 35-7 = 28, Physical Function = 25-5 = 20, Pain/Discomfort = 10-2 = 8, IWQOL-Lite-CT Total = 100-20 = 80.

(c) Multiply the total in (b) by 100.

Responder threshold values

Data from the oral semaglutide phase 3a monotherapy trial (PIONEER 1; NN9924-4233) data were used as part of the confirmation of the psychometric properties of IWQOL-Lite-CT in patients with T2DM . To estimate potential responder threshold values for the composite scores in IWQOL-Lite-CT anchor-based and distribution-based methods were applied to PIONEER 1 data .

The resulting responder threshold values, in terms of score points for change from baseline are defined in Table 2–6.[11]

Table 2–6 Responder thresholds for IWQOL-Lite-CT

Domain	Responder threshold
Psychosocial	10
Physical	8
Physical Function	10
Pain/Discomfort	5
IWQOL-Lite-CT Total	8

Responder analyses will be based on the responder threshold values and are described in Section 2.6.4.

2.6.3 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. 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.

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 per trial are used as distribution-based approach defining the responder thresholds. The thresholds are derived from baseline CoEQ data across Oral semaglutide arms (3 mg, 7 mg, 14 mg) and Sitagliptin 100 mg arm. Responder analyses will be based on the responder threshold .

2.6.4 Responder analyses

Responder analyses will be conducted for both estimands, for the same time points that are defined for the analyses of PRO endpoints (see protocols) and separately for each score.

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

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. Estimated proportions and differences in proportions will be reported in addition to odds and odds ratios.

The responder analyses will not be included in the CTR, but in a separate PRO report.

3 Changes to the statistical analyses planned in the protocol

The main analyses were described in the protocol for the trial NN9924-4222. However, clarifications, more detailed descriptions of endpoints and analyses are provided in this SAP. The changes from the protocol of NN9924-4222 are summarised below:

- It has been specified which regions will be used in the statistical analysisThe primary and secondary estimands have changed names from de-facto and de-jure to treatment policy and hypothetical, respectively.
- The MMRM sensitivity analysis and the comparator multiple imputation analysis based on FAS using the on-treatment observation period of the primary estimand have 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, the baseline value will only be carried forward to the first planned visit, if the first planned visit do not fall later than 8 weeks after randomisation. The 8 weeks are chosen as a balance to keep as many subjects as possible in the analysis and not carrying baseline assessments too long forward in time.
- The two MI sensitivity analyses of the secondary estimand have been omitted in Section 2.3.3. 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 2.3.3.2) 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 complete case analysis is based on the assumption that missing data are MCAR, this assumption seem questionable and since a per-protocol analysis is performed, this extra sensitivity analysis is excluded.
- The statistical analyses of the two binary effect endpoints (HbA_{1c} reduction $\geq$ 1%-point (10.9 mmol/mol) and body weight loss $\geq$ 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 'without hypoglycaemia' component in composite binary endpoints has been added. The specification on how to analyse the binary component 'without hypoglycaemia' in the composite endpoint 'HbA1c <7% (53 mmol/mol) without hypoglycaemia (severe or BG-confirmed symptomatic hypoglycaemia) and no weight gain' has not been described in the statistical section in the protocol. The specification is therefore included below.
- 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.

- Standard laboratory parameters have been downgraded from 'endpoints' to 'assessments' because these parameters are included to ensure that the investigator can monitor the individual subject and to allow for detection of potential safety signal on a trial-level that is not reflected in adverse event reporting.
- All PROs, SF-36v2 (acute version), IWQOL-Lite, and CoEQ, will be analysed statistically. All using the primary analysis of the primary estimand and in addtion, IWQOL-Lite will be analysed using secondary estimand.
- The responder analyses and the primary analysis for the secondary estimand of all PROs, SF-36v2 (acute version), IWQOL-Lite, and CoEQ, will be presented in a report separate from and after finalisation of the CTR.

4 References

1. Merck Sharp & Dohme Limited. JANUVIA - European Summary of Product Characteristics Annex 1. 26 May 2015.
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. (CHMP) EMACfMPfHU. Guideline on Missing Data in Confirmatory Clinical Trials (EMA/CPMP/EWP/1776/99 Rev. 1). 7/2/2010 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. Keene ON, Roger JH, Hartley BF, Kenward MG. Missing data sensitivity analysis for recurrent event data using controlled imputation. Pharmaceutical Statistics. 2014;13(4):258-64.
9. 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.
10. 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.
11. Fehnel S, Qin S, Williams V, RTI Health Solutions. Psychometric Evaluation of the IWQOL-Lite-CT in Overweight and Obese Patients With Type 2 Diabetes (RTI-HS Project No. 0303824.003). IWQOL-Lite-CT Scoring and Responder definitions (v. 1.0 Final). In: Meincke H, Novo Nordisk A/S, editors. 17 May 2018.