Cover Page for Statistical analysis plan

Sponsor name:	Novo Nordisk A/S
NCT number	NCT03021187
Sponsor trial ID:	NN9924-4280
Official title of study:	PIONEER 8 – Insulin add-on Efficacy and Safety of Oral Semaglutide versus Placebo in Subjects with Type 2 Diabetes Mellitus treated with insulin A 52-week, randomised, double-blind, placebo-controlled trial
Document date:	08 February 2019

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: NN9924-4280

PIONEER 8 – Insulin add-on

Efficacy and Safety of Oral Semaglutide versus Placebo in Subjects with Type 2 Diabetes Mellitus treated with insulin

A 52-week, randomised, double-blind, placebo-controlled trial

Trial phase: 3a

Author

[redacted]

Table of contents

Page

List of abbreviations

AACE	American Association of Clinical Endocrinologists
ADA	American Diabetes Association
AE	adverse event
ANCOVA	analysis of covariance
BG	blood glucose
BMI	body mass index
BP	bodily pain
CRF	case report form
CTR	clinical trial report
DTSQs	diabetes treatment satisfaction questionnaire – status version
EAC	event adjudication committee
ECG	electrocardiogram
eCRF	electronic case report form
EOT	end-of-treatment
FAS	full analysis set
FPG	fasting plasma glucose
GH	general health
gMCP	graph based multiple comparison procedure
HbA_{1c}	glycosylated haemoglobin
HDL	high-density lipoprotein
HRQoL	health-related quality of life
IWQOL-Lite-CT	impact of weight on quality of life clinical trials version
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 summary
MedDRA	Medical Dictionary for Regulatory Activities
MH	mental health
MI	multiple imputation
MMRM	mixed model for repeated measurements
NBS	norm-based score
PCS	physical component summary
PF	physical functioning
PG	plasma glucose
PIONEER	Peptide InnOvatioN for Early diabEtes tReatment
PK	pharmacokinetics

PRO	patient reported outcomes
RP	role physical
SAP	statistical analysis plan
SAS	safety analysis set
s.c.	subcutaneous
SD	standard deviation
SF	social functioning
SF-36v2 (acute version)	SF-36v2® Health Survey (acute version)
SMPG	self-measured plasma glucose
T2DM	type 2 diabetes mellitus
TE	treatment effect
TEAE	treatment-emergent adverse events
US	United States
VT	vitality

1 Introduction

1.1 Trial information

This is a 52-week, randomised, double-blind, placebo-controlled, four-armed, parallel-group, multicentre, multinational trial. The trial will compare the efficacy and safety of three dose levels of once-daily oral semaglutide versus placebo in subjects with type 2 diabetes mellitus (T2DM) treated with insulin.

Primary objective

To compare the effect of once-daily dosing of three dose levels of oral semaglutide (3, 7 and 14 mg) versus placebo on glycaemic control in subjects with T2DM treated with insulin.

Secondary objectives

To compare the effect of once-daily dosing of three dose levels of oral semaglutide (3, 7 and 14 mg) versus placebo on body weight in subjects with T2DM treated with insulin.

To compare the safety and tolerability of once-daily dosing of three dose levels of oral semaglutide (3, 7 and 14 mg) versus placebo in subjects with T2DM treated with insulin.

Trial design

Subjects will be randomised 1:1:1:1 manner to receive one of the following treatments:

- 3 mg oral semaglutide once-daily
- 7 mg oral semaglutide once-daily
- 14 mg oral semaglutide once-daily
- placebo once-daily

The total trial duration for the individual subject will be approximately 59 weeks. The trial includes a 2-week screening period followed by a 52-week randomised treatment period and a 5-week follow-up period. The 52-week randomised treatment period is split into two treatment periods; an initial 26-week period where the insulin treatment is restricted followed by a 26-week period where the insulin treatment is adjustable without any restrictions. 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 NN9924-4280 “Efficacy and Safety of Oral Semaglutide versus Placebo in Subjects with T2DM treated with insulin”, version 4.0 (15 May 2017) as well as the Original protocol, version 1.0 (19 August 2016), the Protocol amendment no. 1, version 1.0 (06 October 2016), Protocol amendment no. 2, version 2.0 (22 November 2016), Protocol amendment no. 3, version 3.0 (15 May 2017).

This SAP 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 within 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

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

Data from all sites will be analysed and reported together.

In statistical analyses where stratification is included, the background metformin medication at screening (metformin/no metformin) and background insulin medication (basal insulin/basal and bolus in any combination/premixed insulin including combinations of soluble insulin(s)) will be included based on the actual information collected through the electronic case report form (eCRF). In case of missing eCRF information the information collected from the interactive web response system (IWRS) will be used. The information regarding descent (Japanese/non-Japanese) will be included based on country details from the IWRS. In the statistical analyses, stratification factors refer to the background metformin medication at screening and the background insulin medication at screening. Descent (Japanese/non-Japanese) will be included in the statistical analyses as part of region. The regions are Europe, North America, South America and Asia.

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.

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

- Oral semaglutide 14 mg vs. placebo
- Oral semaglutide 7 mg vs. placebo
- Oral semaglutide 3 mg vs. placebo

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

The two different estimands defined below will be used for the evaluation of the efficacy endpoints.

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 (oral semaglutide versus placebo) 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). 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 treatment adherence 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 (oral semaglutide versus placebo) 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. It is considered a clinically relevant estimand as it provides information to treating clinicians about the expected glycaemic efficacy of oral semaglutide 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 are defined for the remaining secondary endpoints.

Missing data considerations at week 26

When estimating the treatment policy 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 is expected to be a maximum 10% based on the oral semaglutide phase 2 trial (NN9924-3790). Thus, missing data will mainly be due to withdrawal from trial or lost to follow-up.

When estimating the hypothetical estimand, the proportion of subjects with 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 proportion of 20% is based on the oral semaglutide phase 2 trial (NN9924-3790), that indicates that a low starting dose with gradual dose escalation

diminishes gastrointestinal adverse events (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 gastrointestinal AEs and eventually initiation of rescue medication. Initiation of rescue medication is expected to be more frequent in the placebo 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

Both the primary endpoint, change from baseline to week 26 in HbA_{1c} and the confirmatory secondary endpoint, change from baseline to week 26 in body weight are planned to be tested for superiority of oral semaglutide vs. placebo at each dose level (3 mg, 7 mg, and 14 mg).

The sample size calculation is made to ensure a power of at least 90% to jointly confirm HbA_{1c} superiority of oral semaglutide vs. placebo at each dose level out of the six pre-specified confirmatory hypotheses shown in Table 2-2. The closed testing procedure described in Bretz et al[1]. is used to control the overall type I error at a nominal two-sided 5% level. The statistical testing strategy is built on the following two principles:

- Within a dose level, glycaemic effect must be established in terms of HbA_{1c} superiority before testing for added benefits in terms of body weight superiority.
- Glycaemic effect in terms of HbA_{1c} superiority 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 graph based multiple comparison procedure (gMCP)[2] using 10000 simulations. All of the six 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 TE and the standard deviations are given in Table 2-1. These are based on the oral semaglutide phase 2 results (NN9924-3790) 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 TE 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 TE used

in the sample size calculation will be adjusted according to a 75% smaller effect in these subjects. The adjusted TE for testing superiority is defined as 0.8×TE + 0.2×TE×0.25.

Table 2-1 Assumptions used in the sample size calculation

Oral semaglutide vs. placebo	HbA_{1c}			Body weight		
Treatment dose	14 mg	7 mg	3 mg	14 mg	7 mg	3 mg
Treatment effect (TE)	-0.8%-points	-0.60%-points	-0.45%-points	-3.0 kg	-2.0 kg	-1.0 kg
Adjusted TE, superiority	-0.68%-points	-0.51%-points	-0.38%-points	-2.55 kg	-1.70 kg	-0.85 kg
Standard deviation	1.1%s	1.1%s	1.1%	4.0 kg	4.0 kg	4.0 kg

With the above assumptions, allocating 180 subjects to each of the oral semaglutide arms and the placebo arm provides 90% power to confirm HbA_{1c} superiority of oral semaglutide vs. placebo at all dose levels. Calculated powers for individual hypotheses are presented in Table 2-2. In total 4×180 = 720 subjects are planned to be randomised.

Table 2-2 Calculated powers for individual hypotheses

Statistical test	**HbA_{1c} superiority**			**Body weight superiority**		
Treatment dose	14 mg	7 mg	3 mg	14 mg	7 mg	3 mg
Power (%)	> 99	99	90	> 99	96	47

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 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".

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 (V19) for subjects on trial product

- The latest occurring visit of the end-of-treatment (EOT) visit (V18) or the follow-up premature discontinuation visit (V19A), 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 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 IWRS) and ends at the date of:

- The last direct subject-site contact, which is scheduled to take place 5 weeks after planned last dose of trial product at a 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 the 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 and AEs including hypoglycaemic episodes, the observation period ends at the first date of any of the following:

- The follow-up visit (V19)
- The follow-up prematurely discontinuation visit (V19A)
- 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. Specifically it starts at date of first dose of trial product and the end date is the first date of any of the following:

- The last dose of trial product + 3 days
- Initiation of rescue medication

The in-trial observation period will be the primary observation period for estimating the treatment policy estimand. The on-treatment without rescue medication observation period will be the primary observation period when estimating the hypothetical 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 event adjudication committee (EAC) adjudicated onset date.

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 should be used restrictively and normally no data should be excluded from the FAS and in-trial period. 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 (CTR).

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 at each dose level of oral semaglutide versus placebo. Let the mean treatment difference be defined as μ = (oral semaglutide minus placebo):

- HbA_{1c} superiority
  H_0: $\mu \geq 0.0$ %-point against H_A: $\mu < 0.0$ %-point
- 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 at the 5% significance level.

Multiplicity and criteria for confirming hypotheses

The type I error for testing the six 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.[1] and outlined in Figure 2-1.

The first hypothesis to be tested is superiority 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, then 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 updated local significance level (α-local). This process will be repeated until no further hypotheses can be confirmed.

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 treatment policy 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.

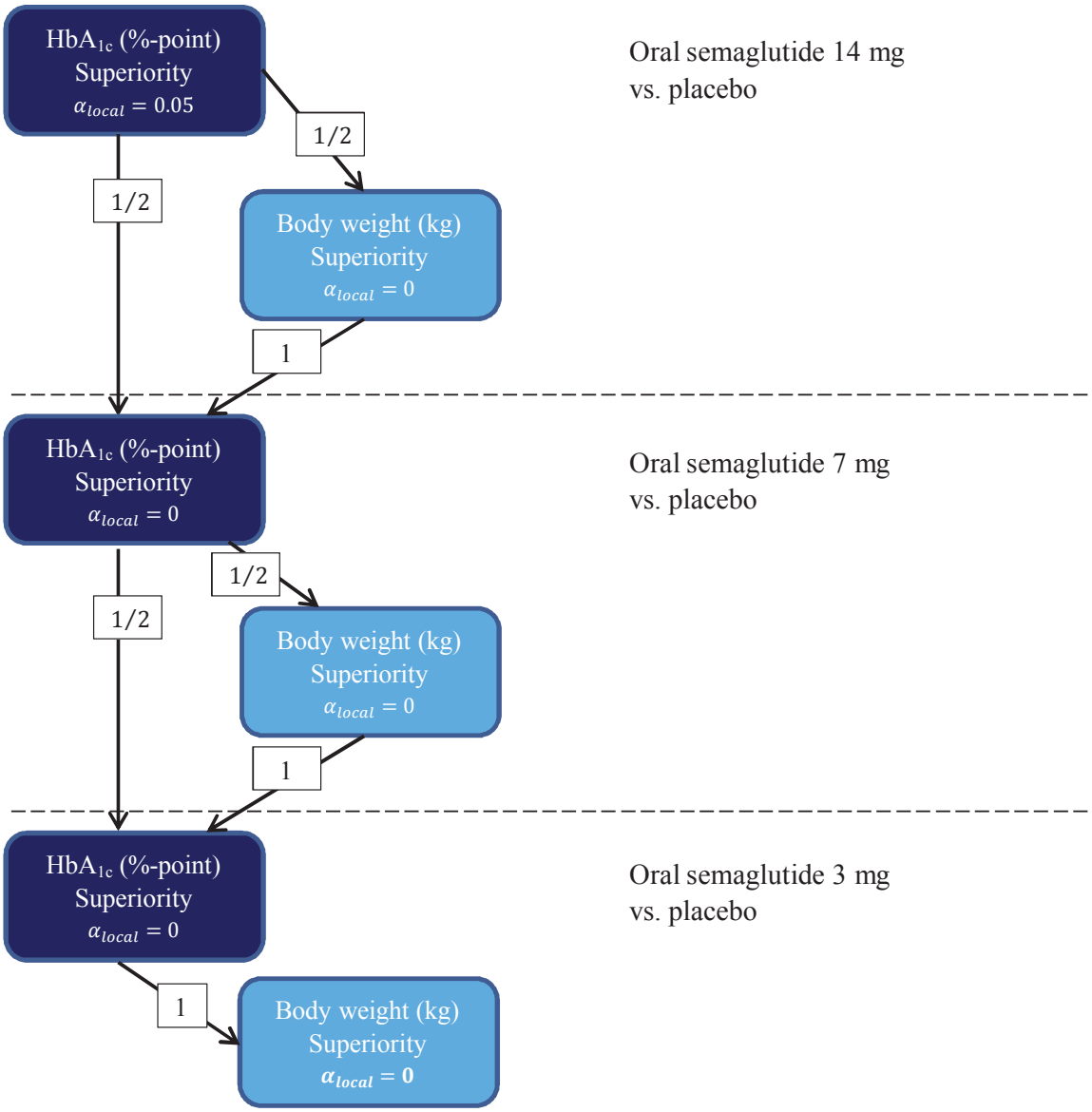


Figure 2-1 Graphical illustration of the closed testing procedure

The overall significance level of α = 0.05 (two-sided) is initially allocated to the HbA_{1c} superiority test on the highest dose level. 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 hypotheses in the dark boxes.

2.3 Primary endpoint

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

2.3.1 Primary analysis for the treatment policy estimand

The treatment policy 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 (MI) 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 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) 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, stratification factors and the interaction between the stratification factors 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. If the model does not fit due to sparse data the following factors will be removed from the imputation model in a step-wise manner, meaning that only baseline HbA_{1c} will be included in the model if using the last approach:
    - region
    - interaction between stratification factors
    - stratification factors
- The estimated parameters for location and dispersion will be used to impute 1000 values for each subject with missing week 26 data based on factors included in the imputation model. Thus, 1000 complete data sets will be generated including observed and imputed values.

Analysis used for confirming superiority versus placebo at week 26:

For each of the 1000 (now complete) imputed data sets the change from baseline to week 26 in HbA_{1c} will be analysed using an ANCOVA with treatment, region, stratification factors and the interaction between the stratification factors as categorical fixed effects and baseline HbA_{1c} as covariate. The results obtained from analysing the datasets will be combined using Rubin's rule[3] to draw inference.

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

2.3.2 Primary analysis for the hypothetical estimand

The hypothetical estimand will be estimated based on the FAS using post-baseline measurements up to and including week 26 from the on-treatment without rescue medication observation period. The primary analysis for the hypothetical 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, region, stratification factors and the interaction between the stratification factors 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 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.

From this model the estimated treatment contrasts (each oral semaglutide dose vs. placebo) will be presented together with 95% CIs and p-values.

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.

2.3.3 Sensitivity analyses

To investigate the robustness of the primary analysis results, complementary and separate analyses will be performed for the treatment policy and hypothetical estimand. In line with EMA recommendations[4] and with a report from the United States (US) National Research Council[5], these analyses will primarily evaluate the sensitivity of the results due to the impact of missing data.

The robustness of the results from the primary analysis is assessed through the following sensitivity analyses.

Sensitivity analyses for the treatment policy estimand

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

- A comparator MI analysis based on FAS using the in-trial observation period
- A comparator MI analysis differentiating between reasons for discontinuing treatment prematurely based on FAS using the in-trial observation period
- A tipping-point MI analysis based on FAS using the in-trial observation period

Sensitivity analysis for the hypothetical estimand

The estimation of the hypothetical estimand will be repeated using the following sensitivity analysis:

A tipping-point MI analysis based on FAS using the on-treatment without rescue medication observation period.

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 of or all of the missing data in the oral semaglutide treatment arms, while maintaining the MAR data assumption for the placebo arm:

- *Comparator MI 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 placebo treatment arm. In effect this imputation approach removes the treatment difference between oral semaglutide and placebo for all subjects randomised to oral semaglutide, given that oral semaglutide is better than placebo.
- *Comparator MI analysis differentiating between reasons for discontinuing treatment prematurely:* In this sensitivity analysis only missing data at week 26 for subjects who discontinue oral semaglutide treatment due to treatment related AEs will be imputed to resemble the distribution of the week 26 values observed in the placebo treatment arm. Treatment related AEs are defined as AEs classified as having a possible or probable relationship to the trial product as reported by the investigator. In effect this imputation approach removes the treatment difference between oral semaglutide and placebo for this selected group of subjects randomised to oral semaglutide.
- *Tipping-point MI analysis:* In this sensitivity analysis, missing data will first 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 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.

2.3.3.2 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 sensitivity analyses 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 treatment policy and hypothetical estimands will be estimated using the same approaches as described for the primary HbA_{1c} endpoint. Baseline body weight will be used as a covariate instead of baseline HbA_{1c} in both the MI and MMRM analysis models.

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 treatment policy estimand is strictly below its local two-sided significance level resulting from the closed testing procedure in Table 2-1. Sensitivity analyses similar to the ones pre-specified 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 treatment policy estimand based on FAS using the in-trial observation period
- The hypothetical estimand based on FAS using the on-treatment without rescue medication observation period

No sensitivity analyses are planned for these supportive secondary efficacy endpoints.

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:

- Fasting plasma glucose (FPG)
- Body weight (%)
- Body mass index (BMI)
- Waist circumference
- Fasting lipid profiles (total cholesterol, low-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL) cholesterol, triglycerides)
- Total daily insulin dose (IU)
- Patient reported outcomes (PROs) (see Section 2.7 for further details)

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

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

Change from baseline to week 26 and week 52 in the below derived endpoints from the self-measured plasma glucose (SMPG) 7-point 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 and PROs, will be log-transformed prior to analysis with the associated log-transformed baseline value as a covariate.

For evaluation of the treatment policy estimand the analyses 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 treatment or initiated rescue medication prior to week 26 or at/after week 26. This will result in imputation of missing data within 12 groups of subjects instead of the 8 groups as described for the week 26 evaluation in Section 2.3.1.If data are too sparse to fit the model in all 12 groups then the imputation will be done 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 hypothetical estimand the MMRM based primary analysis will include all scheduled post-baseline measurement up to and including week 26 (or up to and including week 52 for change from baseline to week 52 endpoints). From this model the estimated treatment contrasts at week 26 (or at week 52 for change from baseline to week 52 endpoints) will be presented with 95% confidence intervals and two-sided p-values for test of no difference. For endpoints where the first planned visit falls later than 8 weeks after randomisation, the baseline will not be carried forward.

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.

Binary efficacy endpoints

If a subject after week 26 and after week 52 achieves (yes/no):

- HbA_{1c} < 7.0 % (53 mmol/mol) (American Diabetes Association (ADA) target)
- HbA_{1c} ≤ 6.5 % (48 mmol/mol) (American Association of Clinical Endocrinologists (AACE) target)
- Body weight loss ≥ 5 %
- Body weight loss ≥ 10 %
- HbA_{1c} < 7.0 % (53 mmol/mol) without hypoglycaemia (severe or blood glucose confirmed (BG-confirmed) symptomatic hypoglycaemic episodes) and no weight gain
- HbA_{1c} reduction ≥ 1 %-point (10.9 mmol/mol) 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 of BG-confirmed symptomatic hypoglycaemia as data collected regardless of discontinuation of trial product or initiation of rescue medication(s) is used.

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 MI techniques. For the treatment policy estimand the binary endpoints will be calculated as dichotomisations of the 1000 MIs 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 are 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 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 EOT 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, stratification factors and the interaction between stratification factors as categorical effects 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 MIs from the sequential imputation.

For both estimands, each of the 1000 data sets will be analysed using a logistic regression model with treatment, region, stratification factors and the interaction between stratification factors 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[3] 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[6] 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 are 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, 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, stratification factors and the interaction between the 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.

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, stratification factors and the interaction between 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.

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 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 insulin and concomitant medication data reported by the investigator, to determine whether or not the recorded anti-diabetic medication is new anti-diabetic medication or intensification of anti-diabetic medication.

1. **New anti-diabetic medication:** Anti-diabetic medication (4th-level ATC code), excluding insulin, 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. **Insulin:** A more than 20% increase in the total dose of insulin after randomisation as compared to the total insulin dose at randomisation and with a dosing duration across 2 consecutive visits or more

    b. **Other 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 (or 2 consecutive visits for insulin) are chosen as a minimum duration for the medication to be considered as 'anti-diabetic medication'. This threshold is set to ensure that short-term durations (i.e., ≤ 21 days (or < 2 visits for insulin)) 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.

As an initial 20% reduction of the total insulin dose is recommended at the randomisation visit for all subjects using insulin as background medication, the background dose is defined as the total insulin dose taken before the 20% reduction.

Treatment policy estimand: Time to additional anti-diabetic medication

The analysis used to estimate 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 whether or not subjects prematurely discontinued treatment. Time from randomisation to additional anti-diabetic medication will be analysed using a Cox proportional hazards model with treatment, region, stratification factors and the interaction between stratification factors as categorical fixed effects and the baseline HbA_{1c} value as a covariate. From this analysis the estimated hazard ratio between oral semaglutide and placebo together with associated two-sided 95% CIs and unadjusted two-sided p-values will be presented.

The endpoint aims to address the need of additional anti-diabetic medication regardless of whether this is due to lack of effect or related to tolerability of the trial product. Switching to a new anti-diabetic treatment (excluding insulin) and/or intensification of anti-diabetic medication are also considered events and withdrawn subjects or subjects lost to follow-up will be considered as having an event (started on additional anti-diabetic medication) on the day of withdrawal. Subjects will be censored on the day before the planned end-of-treatment visit.

Hypothetical estimand: Time to rescue medication

The analysis used to estimate 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 endpoint aims to address a lack of effect of treatment with trial product. Only initiation of rescue medication (new anti-diabetic medication and/or intensification of anti-diabetic medication) as add-on to randomised treatment is considered an event; switching to another anti-diabetic treatment is not considered an event (initiation of rescue medication) 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 on first date of trial product, in order to account for all events of rescue medication initiation.

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 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 MedDRA coding.

A TEAE 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 treatment policy estimand based on SAS using the in-trial observation period and using the primary analysis for the hypothetical 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)

Other safety assessments

Change from baseline to week 26 and week 52 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.

Hypoglycaemia

- Number of treatment-emergent severe or BG confirmed symptomatic hypoglycaemic episodes during exposure to trial product, assessed up to approximately 26 and 57 weeks
- Treatment-emergent severe or BG confirmed symptomatic hypoglycaemic episodes during exposure to trial product, assessed up to approximately 26 and 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 (Figure 2-2).

Novo Nordisk classification of hypoglycaemia

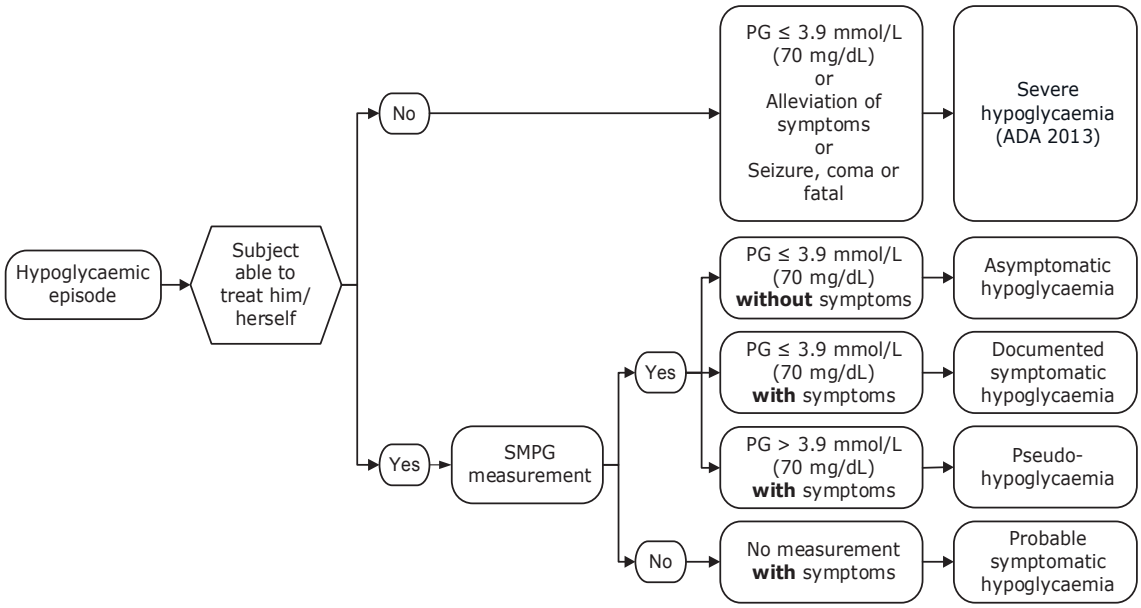
In normal physiology, symptoms of hypoglycaemia occur below a plasma glucose (PG) level of 3.1 mmol/L (56 mg/dL).[7] Therefore, Novo Nordisk has included hypoglycaemia with PG levels below this cut-off point in the definition of BG-confirmed hypoglycaemia.

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

- Severe or BG-confirmed symptomatic hypoglycaemia: An episode that is severe according to the ADA classification[8] or BG confirmed by a PG value < 3.1 mmol/L (56 mg/dL) **with** symptoms consistent with hypoglycaemia.

ADA classification 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

PG: plasma glucose SMPG: Self-measured plasma glucose

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 in the on-treatment period 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 treatment, region, stratification factors and the interaction between the stratification factors as fixed factors and baseline HbA_{1c} as covariate. The estimated rate ratio between each oral semaglutide dose and placebo will be presented together with the 95% CI and p-value.

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, region, stratification factors and the interaction between the stratification factors as fixed factors and baseline HbA_{1c} as covariate. The estimated odds ratio between each oral semaglutide dose and placebo will be presented together with the 95% CI and p-value.

The number of treatment-emergent severe or BG-confirmed symptomatic hypoglycaemic episodes will be presented by time using summary statistics.

2.4.2.3 Pharmacokinetic endpoints

- Semaglutide plasma concentrations for population pharmacokinetic (PK) analysis

The 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 more details in Section 2.6.

2.5 Interim analysis

No interim analyses or other analyses of unblinded data will be performed before the database is locked.

2.6 Pharmacokinetic and/or pharmacodynamic modelling

Data from this trial will be evaluated using population PK analysis and exposure-response for semaglutide. The purpose of the population PK analysis will be:

- To describe the covariate factors (such as weight, age, gender, race and ethnicity) that influence semaglutide exposure
- To estimate a steady-state exposure level for each subject with PK data, in order to facilitate subsequent exposure-response analyses

The purpose of the exposure-response analyses will be to support the recommended dose, by investigating response and potentially side effects across the exposure range.

The population PK and exposure-response analyses will be conducted as a meta-analysis, including all relevant oral semaglutide phase 3a trials with PK assessments. A separate modelling analysis plan will be prepared before first database lock in the oral semaglutide phase 3a programme, 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

PRO endpoints

Change from baseline to week 26 and week 52 in:

- SF-36v2® (acute version) health survey (SF-36v2 (acute version)): Scores from the 8 domains and summary of the physical component summary (PCS) score and the mental component summary (MCS) score

- Impact of weight on quality of life clinical trials version (IWQOL-Lite-CT): Total score and scores from the 4 domains
- Diabetes treatment satisfaction questionnaire – status version (DTSQs): Individual items and treatment satisfaction score (6 of the 8 items summed)

A more detailed description of the handling of the two PRO questionnaires used in this trial is provided in the following sections. No multiplicity adjustments will be done for the PRO questionnaires.

The PRO questionnaire endpoints will be evaluated using the primary analysis for the treatment policy estimand based on FAS using the in-trial observation period and using the primary analysis for the hypothetical estimand based on FAS using the on-treatment without rescue medication period. Scores will be analysed separately as the other continuous efficacy endpoints with the associated baseline value as a covariate.

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

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.[9] 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), MCS, PCS. 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 Software[10] 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 (i.e. Version 5.0).

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 domains and component summary measures (PCS and MCS) indicate better HRQoL/general health status. Item 2 (i.e. Question 2 in case report form (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.[9]

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 2.7.4.

2.7.2 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 22-item version of the IWQOL-Lite-CT was completed by patients.

Domain 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
Psychosocial	6-8, 10-16, 19, 21, 22
Physical	1-5, 17, 18
Physical Function	1-3, 17, 18
Pain/Discomfort	4, 5
IWQOL-Lite-CT Total	1-8, 10-19, 21, 22

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 non-missing, and for the IWQOL-Lite-CT total score if a minimum of 75% of all items are non-missing.

The scoring is done in three steps:

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 [9]

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 2.7.4.

2.7.3 Diabetes Treatment Satisfaction Questionnaire – status version (DTSQs)

The DTSQs questionnaire will be used to assess subject's treatment satisfaction. This questionnaire contains 8 items that measures the treatment satisfaction for subjects' diabetes treatment in terms of convenience, flexibility and general feelings regarding treatment.

Item scores

The DTSQs items are scored on a 7-point graded response scale ranging from 6 to 0. Higher scores indicate higher levels of treatment satisfaction for DTSQs items 1, 4 -8. For items 2 and 3 a higher score indicates a higher patient perceived experience of hyperglycaemia and hypoglycaemia, respectively. Thus, lower scores indicate a perception of BG levels being "none of the time" unacceptably high (item 2) or low (item 3). If data are missing for an item, the item score is treated as missing. No reversal of item scores will be done.

Treatment satisfaction score

The domain score of total treatment satisfaction (total treatment satisfaction score) is computed by adding the six items scores 1, 4-8. The score has a minimum of zero and a maximum of 36. A higher treatment satisfaction score indicates a higher level of treatment satisfaction. No reversals of items are necessary prior to computing the treatment satisfaction score.

Missing data at instrument level will be handled in the following way. For computing the total treatment satisfaction score consisting of six items, missing data from one item is allowed.

Scoring algorithm:

- Step 1: Sum the existing item scores (i.e. either 5 or 6 item scores)
- Step 2: Divide this sum by the number of existing item scores
- Step 3: Multiply by 6 (the number of items in the total treatment satisfaction scale)

Responder threshold values

Half of a standard deviation (SD) of the baseline DTSQs item and domain scores were used as distribution-based approach defining the responder thresholds. The thresholds are derived from baseline DTSQs data across trial arms. Responder analyses will be based on the responder threshold values and are described in Section 2.7.4.

2.7.4 Responder analyses

Responder analyses will be conducted for both estimands, for the same time points that are defined for the analyses of PRO endpoints 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.

3 Changes to the statistical analyses planned in the protocol

The main analyses were described in the protocol for the trial NN9924-4280. However, clarifications, more detailed descriptions of endpoints and analyses are provided in this SAP. The changes from the protocol of NN9924-4280 are summarised below:

- The countries for region in the study have been specified.
- The primary and secondary estimand names 'de-facto' and 'de-jure' have been changed to 'treatment policy' and 'hypothetical' respectively.
- Step-wise factor removal from ANCOVA imputation model for primary analysis.
- The number of sensitivity analyses has been reduced:
    - The MMRM sensitivity analysis of the treatment policy estimand has been omitted. It is considered sufficient to keep just the two current sensitivity analyses to stress test the primary results.
    - Three MI sensitivity analyses of the hypothetical estimand have been omitted. It is considered sufficient to keep the tipping point sensitivity analysis for the secondary analysis of the hypothetical estimand as it can be considered as a progressive stress-testing to assess how severe the departures from MAR must be to reverse the conclusions from the primary MMRM analysis used to address the hypothetical estimand.
    - The last observation carried forward (LOCF) sensitivity analysis specified in the trial protocol 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.
- It has been specified which assessments will be analysed on logarithmic scale.
- The statistical analysis 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 with more details how missing data in the analyses for both the treatment policy estimand and hypothetical estimand will be imputed.
- A clarification of the 'without hypoglycaemia' component in composite binary endpoints has been added. The specification on how to analyses the binary component 'without hypoglycaemia' in the composite endpoint 'HbA_{1c} < 7.0 % (53 mmol/mol) without hypoglycaemia (severe or BG-confirmed symptomatic hypoglycaemia) and no weight gain' has not been described in the statistical section of the protocol. The specification is therefore included in the SAP.
- The definitions of rescue medication and additional anti-diabetic medication used for the time-to-event endpoints were added along with discussion on additional anti-diabetic medication. Furthermore, the accompanying statistical analyses have been further clarified.
- All PROs (SF-36v2 (acute version), IWQOL-Lite-CT, and DTSQs) will be analysed statistically. SF-36v2 (acute version) using the primary analysis for the treatment policy

estimand. IWQOL-Lite-CT and DTSQs using both the primary analysis for the treatment policy estimand and the hypothetical estimand.

- The responder analyses and the primary analysis for the hypothetical estimand of SF 36v2 (acute version) will be presented in a report separate from and after finalisation of the CTR.
- In addition to the total score, four domain scores will be analysed for IWQOL-Lite-CT.

4 References

1. 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.
2. Rohmeyer K, Klinglmueller F. gMCP: Graph Based Multiple Test Procedures. R package version 0.8-8. 3 Oct 2014.
3. Little R, Rubin D. Statistical analysis with missing data. Sons. JW, editor. New York.: John Wiley & Sons. 1987.
4. (CHMP) EMACfMPfHU. Guideline on Missing Data in Confirmatory Clinical Trials (EMA/CPMP/EWP/1776/99 Rev. 1). 7/2/2010 2010.
5. National Academy of Sciences (NAS). The Prevention and Treatment of Missing Data in Clinical Trials. Washington D.C.: The National Academies Press. 2010.
6. Bagger JI, Knop FK, Lund A, Vestergaard H, Holst JJ, Vilsbøll T. Impaired regulation of the incretin effect in patients with type 2 diabetes. J Clin Endocrinol Metab. 2011;96(3):737-45.
7. 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.
8. 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.
9. Maruish ME, (Ed.). User's manual for the SF-36v2 Health Survey (3rd ed.). Lincoln, RI: QualityMetric Incorporated. 2011.
10. 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.

16.1.9.1 Pre-defined MedDRA search – list of preferred terms

Table of contents

Page

Trial: NN9924-4280

Documentation of Statistical methods, Version 1.0 dated 08 February 2019

Overview of deleted pages

Pages	Section	Title
40-249	16.1.9.1	Pre-defined MedDRA search – list of preferred terms