Cover Page for Statistical Analysis Plan

Sponsor name:	Novo Nordisk A/S
NCT number	NCT03015220
Sponsor trial ID:	NN9924-4282
Official title of study:	PIONEER 10 – Japan OAD combination. Safety and efficacy of oral semaglutide versus dulaglutide both in combination with one OAD in Japanese subjects with type 2 diabetes.
Document date:	02 November 2018

16.1.9 Documentation of statistical methods

List of contents

Redacted statistical analysis plan includes redaction of personal identifiable and company confidential information.

.

Statistical Analysis Plan

Trial ID: NN9924-4282

Safety and efficacy of oral semaglutide versus dulaglutide both in combination with one OAD in Japanese subjects with type 2 diabetes

A 52-week, randomised, open-label, active-controlled trial

Author
[redacted]

Table of contents

Page

List of abbreviations

ADA	American Diabetes Association
AE	adverse event
AACE	American Association of Clinical Endocrinologists
α-GI	alpha-glucosidase inhibitor
ANCOVA	analysis of covariance
ATC	Anatomical Therapeutic Chemical
BG	blood glucose
BMI	body mass index
BP	bodily Pain
CI	confidence interval
CRF	case report form
CTR	clinical trial report
draft AD guideline	Draft Guideline for Clinical Evaluation of Hypoglycaemic Agents
DTR-QOL	Diabetes Therapy-Related Quality of Life
EAC	event adjudication committee
ECG	electrocardiogram
eCRF	electronic case report form
FAS	full analysis set
FPG	fasting plasma glucose
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
MAR	missing at random
MCS	mental component score
MCMC	Markov Chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
MH	mental Health
MI	multiple imputation
MMRM	mixed model for repeated measurements
NBS	norm-based score
OAD	oral antidiabetic drug
PCS	physical component score
PF	physical Functioning
PG	plasma glucose
PRO	patient reported outcome

RE	role Limitations Due to Emotional Problems
REML	restricted maximum likelihood
RP	role Limitations Due to Physical Health
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)
SGLT-2 inhibitor	sodium-glucose cotransporter-2 inhibitor
SNAC	sodium N-[8-(2-hydroxybenzoyl)amino]caprylate
SU	sulfonylurea
SMPG	self-measured plasme glucose
T2DM	Type 2 diabetes mellitus
TEAE	treatment emergent adverse event
TZD	thiazolidinedione
VLDL	very low-density lipoprotein
VT	vitality

1 Introduction

1.1 Trial information

This is a 52-week, randomised, open-label, active-controlled, parallel-group, multi-centre, single country trial with 4 treatment arms, comparing the safety and efficacy of oral semaglutide with dulaglutide in Japanese subjects with T2DM (type 2 diabetes mellitus) inadequately controlled on any of the oral anti-diabetic drugs (OAD) (sulphonylurea (SU), glinide, thiazolidinedione (TZD), alpha-glucosidase inhibitor (α-GI) or sodium-glucose cotransporter-2 inhibitor (SGLT-2)).

Primary objective

To compare the safety and tolerability of once-daily dosing of three dose levels (3, 7 and 14 mg) of oral semaglutide versus once-weekly 0.75 mg dulaglutide subcutaneously both in combination with one OAD (SU, glinide, TZD, α-GI or SGLT-2) in Japanese subjects with T2DM.

Secondary objective

To compare the effect of once-daily dosing of three dose levels (3, 7 and 14 mg) of oral semaglutide versus once-weekly 0.75 mg dulaglutide subcutaneously both in combination with one OAD (SU, glinide, TZD, α-GI or SGLT-2) on glycaemic control and body weight in Japanese subjects with T2DM.

Trial design

Japanese subjects with type 2 diabetes mellitus inadequately controlled on one OAD will be randomised in a 2:2:2:1 manner to receive one of the following treatments as add-on to their pre-trial background OAD medication:

- Oral semaglutide 3 mg once-daily
- Oral semaglutide 7 mg once-daily
- Oral semaglutide 14 mg once-daily
- Dulaglutide 0.75 mg s.c. once-weekly

The background OAD medication must consist of one of the following OADs as monotherapy: SU, glinide, TZD, α-GI or SGLT-2.

Total trial duration for the individual subject will be approximately 59 weeks. The trial includes a 2-week screening period and a 52-week randomised treatment period, followed by a 5-week follow-up period. 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 NN9924-4282 “Safety and efficacy of oral semaglutide versus dulaglutide both in combination with one OAD in Japanese

subjects with type 2 diabetes mellitus", version 3.0 (16 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.

2.1 General considerations

The blinding of the randomised dose level for semaglutide arms 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.

When deemed relevant, additional descriptive tables will be created by background OAD medication.

In statistical analyses where stratification is included, the background OAD medication at screening (SU, glinide, TZD, α-GI or SGLT-2 inhibitor) will be included based on the actual information collected through the eCRF. In case of missing eCRF information the information collected from the IWRS system will be used. In the statistical analyses the stratification factor will refer to background OAD medication at screening.

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 with no multiplicity adjustments:

- oral semaglutide 14 mg vs. dulaglutide 0.75 mg
- oral semaglutide 7 mg vs. dulaglutide 0.75 mg
- oral semaglutide 3 mg vs. dulaglutide 0.75 mg

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.

2.1.1 Primary and secondary estimands

Two estimands addressing different aspects of the trial objective will be defined; a primary 'Treatment-policy' estimand and a secondary 'Hypothetical' estimand:

2.1.1.1 Primary estimand – 'Treatment policy'

- Treatment difference (oral semaglutide versus dulaglutide) 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.

2.1.1.2 Secondary estimand – 'Hypothetical'

- treatment difference (oral semaglutide versus dulaglutide) 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.

2.1.2 Missing data consideration

Based on the results from similar Japanese phase 3 trials investigating liraglutide and s.c. semaglutide in combination with one OAD (NN2211-3924 and NN9535-4091) and the oral semaglutide phase 2 trial (NN9924-3790), it is expected that maximum 20% of the subjects will discontinue trial product or withdrawal from the trial at week 52.

The primary efficacy results will be evaluated at week 26. This approach will result in a lower proportion of missing data, use of rescue medication and premature treatment discontinuation, compared to the expected proportion at week 52. For the primary estimand, data collected regardless of discontinuation of trial product or initiation of rescue medication(s) will be used to draw inference, missing data will therefore mainly be due to withdrawal from trial or lost to

follow-up. When estimating the secondary estimand, the proportion of missing data is expected to be higher compared to the primary estimand, since data collected after discontinuation of trial product or initiation of rescue medication(s) will be set to missing as well.

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

2.1.3 Sample size calculation

The sample size of this trial was determined based on a requirement of the draft AD guideline.[1]

A total of 455 subjects will be randomised in a 2:2:2:1 manner to receive either a dose of 3, 7 or 14 mg of oral semaglutide or of 0.75 mg dulaglutide. Randomisation will be stratified by background OAD medication at screening (SU, glinide, TZD, α-GI or SGLT-2 inhibitor) with the additional requirement of 147 subjects on SU and 77 subjects on each of the other four background OAD medications. It is expected that maximum 20% of the subjects will discontinue trial product or withdrawal from the trial at week 52. Therefore, at least 100 completers in combination with SU and 50 completers in combination with each of other OAD monotherapies (glinide, TZD, α-GI or SGLT-2 inhibitor) on oral semaglutide (3, 7 or 14 mg) are expected.

2.1.4 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".

2.1.5 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 (V14) for subjects on trial product
- the latest occurring visit of the end-of-treatment visit (V13) or the follow-up premature discontinuation visit (V14A), 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 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 (V14)
- the follow-up prematurely discontinuation visit (V14A)
- 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: 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 observation period ends at the first date of:

- 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 primary estimand. The on-treatment without rescue medication observation period will be the primary observation period when estimating the secondary estimand. The on-treatment observation period will be considered supportive for evaluating efficacy. Safety will be evaluated based on the in-trial and the on-treatment observation periods.

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

Before data are locked for statistical analysis 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. 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.

2.2 Primary endpoints

The primary endpoint is number of TEAEs during exposure to trial product, assessed up to approximately 57 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 2.1.5). The evaluation will be based on SAS.

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.

Furthermore summary tables based on system organ class and preferred term are made for:

- Serious TEAEs
- Possibly or probably related TEAEs
- Severe TEAEs
- TEAEs with preferred term that are experienced by at least 5% of the subjects
- TEAEs leading to discontinuation of trial product

2.3 Secondary endpoints

2.3.1 Supportive secondary endpoints

2.3.1.1 Efficacy endpoints

The below supportive secondary efficacy endpoints will be analysed based on FAS using:

- The in-trial observation period
- The on-treatment without rescue medication observation period

Change from baseline to week 26 in HbA_{1c}

The primary estimand will be estimated based on the FAS using week 26 measurements from the in-trial observation period. The primary statistical analysis will be a pattern mixture model using multiple imputation to handle missing data assuming that the missing data mechanism is missing at random (MAR) within the groups used for imputation. Imputation of missing data at week 26 will be done within 5 groups of subjects: one group of subjects who at week 26 will have discontinued treatment or have initiated rescue medication regardless of randomised treatment arm, and 4 groups of subjects defined by randomised treatment arm for subjects who will still be on treatment and not have 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 stratification factor as a categorical fixed effect and baseline HbA_{1c} measurement as a covariate will be fitted to observed values of the change from baseline in HbA_{1c} at week 26. Furthermore, when fitting the ANCOVA model for the one group of subjects who at week 26 will have discontinued treatment or initiated rescue medication, also randomised treatment arm will be included in the model as a categorical fixed effect.
  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:
  - randomised treatment factor (for the group of subjects who at week 26 will have discontinued treatment or initiated rescue medication)
  - stratification factor
- 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 and baseline HbA_{1c}. Thus, 1000 complete data sets will be generated including observed and imputed values. Thus, 1000 complete data sets will be generated including observed and imputed values.

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 and stratification factor as categorical fixed effects

and baseline HbA_{1c} as a covariate. The results obtained from analysing the datasets will be combined using Rubin's rule[2] to draw inference.

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 medication observation period. The secondary analysis 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 and stratification factor as categorical fixed effects and baseline HbA_{1c} as a covariate, all nested within visit. An unstructured covariance matrix for HbA_{1c} measurements within the same subject will be employed, assuming measurements from different subjects are independent. The MMRM is a well-established method that accounts for the uncertainty pertaining to missing data. This analysis assumes that the missing data mechanism is MAR. Under this assumption the statistical behaviour of the missing data (given the observed responses and model fixed effects and covariates) is assumed to be same as the observed data.

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

Continuous efficacy endpoints

Change from baseline to week 52 in:

- HbA_{1c}

Change from baseline to week 26 week and 52 in:

- FPG
- Body weight (kg)
- Body weight (%)
- BMI
- Waist circumference
- Fasting lipid profile (total cholesterol, LDL cholesterol, HDL cholesterol, triglycerides, VLDL)

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

- BMI kg/m^2 = body weight (kg)/(Height (m) x Height (m))

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

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

The above continuous endpoints will be analysed separately using similar model approaches as for change from baseline to week 26 in HbA_{1c} 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 and 52. For the analysis at week 52, the imputation of missing data will be performed similarly to the week 26 evaluation, only using week 52 instead of week 26 as treatment adherence/rescue status. This will lead to 5 groups of subjects: one group of subjects who at week 52 will have discontinued treatment or have initiated rescue medication regardless of randomised treatment arm, and 4 groups of subjects defined by randomised treatment arm for subjects who will still be on treatment and not have initiated rescue medication. Furthermore treatment adherence/rescue status at week 26 will be included as a factor in the imputation model. If the model does not fit due to sparse data the same approach as for the week 26 evaluation will be used with the only exception that the treatment adherence/rescue status at week 26 will be removed from the model before removing the randomised treatment factor.

For evaluation of the secondary estimand the MMRM will include all scheduled post-baseline measurement up to and including week 52. From this model the estimated treatment differences (ratios) will be presented at week 26 and week 52 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 subjects after week 26 and 52 achieves (yes/no):

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

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.

Handling of missing data for binary endpoints

HbA_{1c} and body weight

Missing data for the above five 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 the stratification factor 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 stratification factor 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[2] to draw inference.

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

Time to event endpoint

- 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 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 **at or 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 **at or 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. ≤ 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 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 and stratification factor as categorical fixed effects and baseline HbA_{1c} as a covariate. From this analysis the estimated Hazard ratios between each of the oral semaglutide dose levels and dulaglutide 0.75 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 whether 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.

2.3.1.2 Safety endpoints and assessments

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.

Other safety endpoints

Change from baseline to week 26 and week 52 in:

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

The above safety endpoints will be evaluated using the primary analysis for the primary estimand based on SAS using the in-trial observation period and using the primary analysis for the secondary estimand based on SAS using the on-treatment observation period. Endpoints will be analysed separately as described above for continuous efficacy endpoints. Results will be presented at week 26 and 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 category
- Physical examination

Change from baseline to week 52 in:

- Eye examination category

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 57 weeks

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

Classification of hypoglycaemia:

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

Treatment emergent: hypoglycaemic episodes will be defined as treatment-emergent if the onset of the episode occurs within the on-treatment observation period (see definition of observation periods in Section 2.1.5).

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

Hypoglycaemic episodes are classified according to the Novo Nordisk classification of hypoglycaemia and the ADA classification of hypoglycaemia (see Figure 2-1).

Novo Nordisk classification of hypoglycaemia

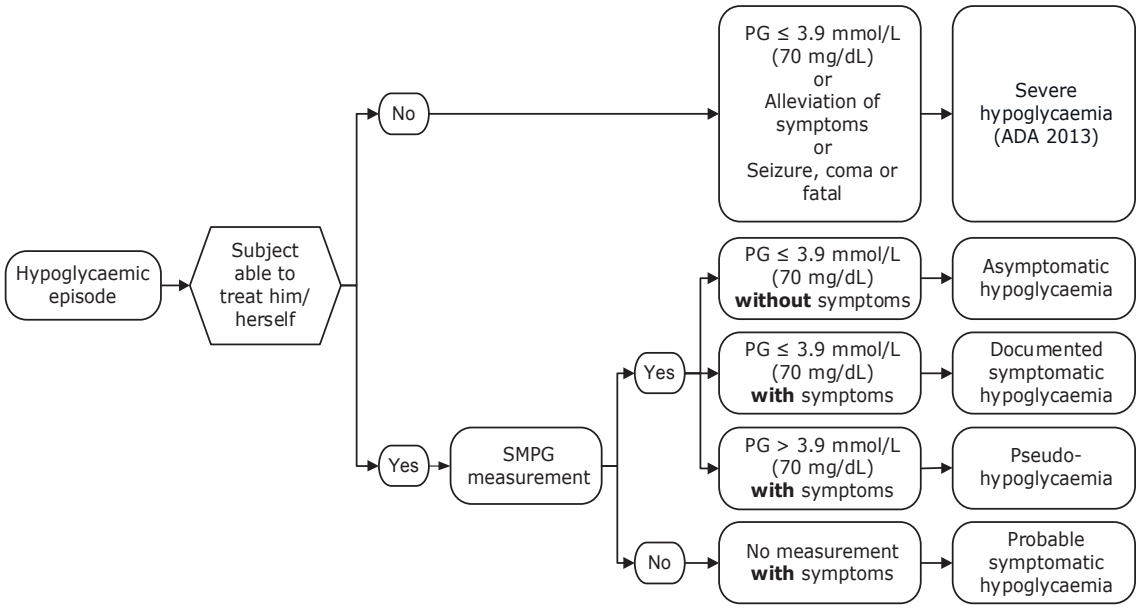
In normal physiology, symptoms of hypoglycaemia occur below a PG level of 3.1 mmol/L (56 mg/dL)[3]. Therefore, Novo Nordisk has included hypoglycaemia with PG levels below this cut-off point in the definition of BG confirmed hypoglycaemia.

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[4] 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 2-1 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

Due to sparse data the protocol pre-specified analyses of severe or BG-confirmed symptomatic hypoglycaemia will not be performed.

2.3.2 Interim analysis

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

2.3.3 Pharmacokinetic and/or pharmacodynamic modelling

Semaglutide plasma concentration or SNAC plasma concentration will not be collected.

2.3.4 Health economics and/or patient reported outcomes

2.3.4.1 PRO endpoints

Change from baseline to week 26 and week 52 in:

- SF-36v2TM (acute version) health survey: Physical component score, mental component score and scores from the 8 domains
- DTR-QOL questionnaire: Total score and scores from the 4 domains

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

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 diabetes[5]. The SF-36v2 (acute version) for adults with a 1-week recall period contains 36 items.

A total of 35 items are used to 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 Software including the 2009 US general population norm. Version 5.0 of the QualityMetric Health Outcomes™ Scoring Software available at time of licensing will be used (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 CRF) is not included in any score.

Table 2-1 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 2-2.

Table 2-2 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 2.3.4.2

Diabetes Therapy-Related Quality of Life (DTR-QoL)

DTR-QoL is a diabetes specific HRQoL measure. DTR-QoL assesses the influence of diabetes treatment on HRQoL on 4 domains: "Burden on social activities and daily activities", "Anxiety and dissatisfaction with treatment", "Hypoglycemia" and "Satisfaction with treatment". DTR -QoL contains 29 items evaluated on a 7-point graded response scale (see Table 2–4). Higher item scores indicate a higher level of HRQoL for items 1-25. For items 26-29 a higher score indicates a lower level of HRQoL.

Domain scores

The response scale used was a 7-point Likert scale (1: completely true – 7: not true at all). The score of each item is reversed so that "7" represents the highest QoL. The domain score is calculated from the mean score of the attribute items, and the scoring range is converted to 0–100. The total score, after simple addition of the item scores, is converted to 0–100 (best-case response=100; worst-case response=0).

Table 2-3 Overview of domains for DTR-QoL

Domain	Items numbers of items included in domain	Comment
Burden on social activities and daily activities	1-13	Formula for domain score derivation[1]: (Sum of item scores -13) * (100/(13*6))
Anxiety and dissatisfaction with treatment	14, 19-25	Formula for domain score derivation[1]: (Sum of item scores -8) * (100/(8*6))
Hypoglycemia	15-18	Formula for domain score derivation[1]: (Sum of item scores -4) * (100/(4*6))
Satisfaction with treatment	26-29	Step 1: Item scores 26-29 to be reversed[2] Step 2: Formula for domain score derivation[1]: (Sum of item scores -4) * (100/(4*6))
Total score	1-29	Step 1: Item scores 26-29 to be reversed[2] Step 2: Formula for domain score derivation[1]: (Sum of item scores -29) * (100/(29*6))

Missing data at instrument level will be handled in the following way. If the number of items with a missing value in a domain is less than 50% of the total items in the domain, the mean value excluding the missing value(s) is calculated and substituted for the missing value(s). If the number of items with a missing value in the domain is 50% or more of the total items in the domain, the domain score is not calculated. The total score is not calculated, if none of the domain scores can be calculated.

Responder threshold values

Half of a standard deviation (SD) of the baseline DTR-QoL total and domain scores were used as distribution-based approach defining the responder thresholds. The thresholds are derived from baseline DTR-QoL data across trial arms. Responder analyses will be based on the responder threshold values and are described in Section 2.3.4.2.

2.3.4.2 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 NN9924-4282. However, clarifications, more detailed descriptions of endpoints and analyses are provided in this SAP. The changes from the protocol of NN9924-4282 are summarised below:

- The names of the primary and secondary estimands have been changed from de-facto and de-jure to treatment policy and hypothetical, respectively.
- In the primary endpoint section it is specified that summary tables will be made for specific groups of TEAEs. “TEAEs leading to discontinuation of trial product” has been added to this list.
- Adjustments to the imputation method for the primary analyses for the primary estimand have been made due to sparse missing data. Instead of using the 8 imputation groups as specified in the protocol only 5 imputation groups will be used: one group of subjects who at week 26 have discontinued treatment or initiated rescue medication regardless of randomised treatment arm, and 4 groups of subjects defined by randomised treatment arm for subjects who are still on treatment and have not initiated rescue medication. Furthermore it has been specified which factors to include in the imputation model for both week 26 and week 52 in case the model is not estimable.

- For the MMRM analyses, it has been specified that for subjects who do not have post-baseline assessments for planned visits available in the on-treatment without rescue medication period, the baseline value will be carried forward to the first planned visit, if the first planned visit do not fall later than 8 weeks after randomisation, to ensure that all randomised subjects will contribute to the statistical analyses.
- The statistical analyses of the two binary efficacy endpoints (HbA1c reduction ≥ 1%-point (10.9 mmol/mol) and body weight loss ≥ 3%) will not be analysed individually but as components of the composite endpoint.
- For the binary efficacy endpoints, imputation of missing data in the analyses for the hypothetical estimand has been specified to use a sequential imputation approach assuming data are MAR.
- The protocol pre-specified statistical analyses of Body weight loss ≥ 10% will be omitted due to low number of such episodes across treatment groups (based on review of blinded data).
- A clarification of the 'no hypoglycaemia' component in the composite binary endpoint has been added
- The definitions of initiation of rescue medication and additional anti-diabetic medication used for the time-to-event endpoints as well as the accompanying statistical analyses have been further clarified.
- It has been specified that all safety laboratory results (except amylase and lipase) are safety assessments and not safety endpoints as written in the trial protocol
- For free fatty acids assessments done by the [redacted] all assessments are considered invalid due to the samples being stored at ambient temperature, hence these results will not be reported
- The protocol pre-specified statistical analyses of severe or BG-confirmed symptomatic hypoglycaemia episodes will be omitted due to low number of such episodes across treatment groups (based on review of blinded data).
- Because PRO endpoints will be further evaluated in a separate PRO report after finalisation of the CTR, it has been specified that SF-36v2 (acute version) will be analysed using the primary analysis of the primary estimand only, and DTR-QoL will be analysed for both the primary and the secondary estimand.
- The responder analyses for SF 36v2 (acute version) and DTR-QoL will be presented in a separate PRO report after finalisation of the CTR.

4 References

1. Ministry of Health, Welfare. La. Draft Guideline for Clinical Evaluation of Hypoglycaemic Agents. . 19 May 2015.
2. Little RJA, Rubin DB. Statistical analysis with missing data: New York: John Wiley & Sons. 1987.
3. 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.
4. 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.
5. Maruish ME, (Ed.). User's manual for the SF-36v2 Health Survey (3rd ed.). Lincoln, RI: QualityMetric Incorporated. 2011.