

Cover Page for Statistical analysis plan

Sponsor name:	Novo Nordisk A/S
NCT number	NCT02863419
Sponsor trial ID:	NN9924-4224
Official title of study:	PIONEER 4 – vs. GLP-1 RA Efficacy and safety of oral semaglutide versus liraglutide and versus placebo in subjects with type 2 diabetes mellitus A 52-week randomised, double-blind, active- and placebo-controlled trial
Document date:	29 October 2018

16.1.9 Documentation of statistical methods

List of contents

Statistical analysis plan	Link
16.1.9.1 Pre-defined MedDRA search – list of preferred terms	Link

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

Statistical Analysis Plan

Trial ID: NN9924-4224

PIONEER 4 – vs. GLP-1 RA

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

A 52-week randomised, double-dummy, double-blind, active- and placebo-controlled trial

Author:



Table of contents

Table of contents	2
List of abbreviations	3
1 Introduction	6
1.1 Trial information	6
1.2 Scope of the statistical analysis plan	6
2 Statistical considerations	7
2.1 Sample size calculation	9
2.2 Definition of analysis sets	12
2.3 Primary endpoint	15
2.3.1 Primary analysis for the primary estimand	15
2.3.2 Primary analysis for the secondary estimand	16
2.3.3 Sensitivity analyses	17
2.3.3.1 Pattern mixture models	18
2.3.3.2 Other sensitivity analyses	18
2.3.3.3 Assessment of sensitivity analyses	19
2.4 Secondary endpoints	19
2.4.1 Confirmatory secondary endpoints	19
2.4.2 Supportive secondary endpoints	19
2.4.2.1 Efficacy endpoints	19
2.4.2.2 Safety endpoints	23
2.5 Interim analysis	27
2.6 Patient reported outcomes	27
2.6.1 Diabetes Treatment Satisfaction Questionnaire – status version (DTSQs)	27
2.6.2 Responder analyses	28
3 Changes to the statistical analyses planned in the protocol	28
4 References	30

List of abbreviations

AACE	American Association of Clinical Endocrinologists
ADA	American Diabetes Association
AE	adverse event
ALP	Alkaline phosphatase
ALT	alanine aminotransferase
ANCOVA	Analysis of covariance
AST	aspartate aminotransferase
AUC	Area under the curve
BG	blood glucose
BG meter	blood glucose meter
BMI	body mass index
CLAE	clinical laboratory adverse event
CK	creatinine kinase
CKD-EPI	Chronic Kidney Disease Epidemiology collaboration
CRF	case report form
CRO	contract research organisation
CVD	cardiovascular disease
DFU	direction for use
DPP-4	dipeptidyl peptidase-4
DTSQ	diabetes treatment satisfaction questionnaire
DUN	dispensing unit number
EAC	event adjudication committee
ECG	electrocardiogram
eCRF	electronic case report form
EDC	electronic data capture
eGFR	estimated glomerular filtration rate
EMA	European Medicines Agency
EOT	end of treatment
FAS	full analysis set
FDA	U.S. Food and Drug Administration
FDAAA	U.S. Food and Drug Administration Amendment Act
FPG	fasting plasma glucose
FSFV	first subject first visit
GCP	Good Clinical Practice
GLP-1	glucagon-like peptide-1
GLP-1 RA	glucagon-like peptide-1 receptor agonist
HbA1c	glycosylated haemoglobin
HDL	high density lipoprotein

HOMA-B	homeostatic model assessment index of beta-cell function
HOMA-IR	homeostatic model assessment index of insulin resistance
IB	Investigator's Brochure
ICH	International Conference on Harmonisation
ICMJE	International Committee of Medical Journal Editors
IEC	independent ethics committee
IMP	investigational medicinal product
IRB	institutional review board
IWRS	interactive web response system
LDL	low density lipoprotein
LLoQ	lower limit of quantification
LSFV	last subject first visit
LSLV	last subject last visit
MAR	missing at random
MedDRA	Medical Dictionary for Regulatory Activities
MEN 2	Multiple Endocrine Neoplasia Type 2
MI	Multiple Imputation
MTC	Medullary Thyroid Carcinoma
NIMP	non-investigational medicinal product
NYHA	New York Heart Association
OAD	oral antidiabetic drug
PG	plasma glucose
PK	pharmacokinetics
PP	per protocol
PRO	patient reported outcome
SAE	serious adverse event
SAP	statistical analysis plan
s.c.	subcutaneous(ly)
SGLT-2	sodium-glucose co-transporter-2
SIF	safety information form
SMPG	self-measured plasma glucose
SmPC	summary of product characteristics
SNAC	sodium N-[8-(2-hydroxybenzoyl) amino] caprylate
STEMI	ST-elevation acute myocardial infarction
SUSAR	suspected unexpected serious adverse reaction
T2DM	type 2 diabetes mellitus
TEAE	treatment-emergent adverse events
TIA	transient ischaemic attack
TMM	Trial Materials Manual
UNL	Upper Normal Limit

Statistical Analysis Plan
Trial ID: NN9924-4224
UTN: U1111-1176-6029
EudraCT no.: 2015-005210-30

~~CONFIDENTIAL~~

Date:	25 May 2018	Novo Nordisk
Version:	1.0	
Status:	Final	
Page:	5 of 30	

UTN Universal Trial Number
VLDL very low density lipoprotein

1 Introduction

1.1 Trial information

This is a 52-week, randomised, double-blind, double-dummy, active- and placebo-controlled, parallel-group, multicentre, multinational trial with 3 arms comparing the efficacy and safety of once-daily dosing of oral semaglutide vs. liraglutide and vs. placebo in subjects with T2DM.

Primary objective

To compare the effect of once-daily dosing of 14 mg oral semaglutide versus 1.8 mg liraglutide subcutaneous and versus placebo, all in combination with metformin with or without a SGLT-2 inhibitor, on glycaemic control in subjects with type 2 diabetes mellitus (T2DM).

Secondary objectives

To compare the effect of once-daily dosing of 14 mg oral semaglutide versus 1.8 mg liraglutide subcutaneous and versus placebo, all in combination with metformin with or without a SGLT-2 inhibitor, on body weight in subjects with T2DM.

To compare the safety and tolerability of once-daily dosing of 14 mg oral semaglutide versus 1.8 mg liraglutide subcutaneous and versus placebo, all in combination with metformin with or without a SGLT-2 inhibitor, in subjects with T2DM.

Trial design

Subjects will be randomised in a 2:2:1 manner to receive one of the following treatments:

- 14 mg oral semaglutide once-daily
- 1.8 mg liraglutide subcutaneous (s.c.) injection 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 follow-up period of 5 weeks. For further details, see the trial protocol.

1.2 Scope of the statistical analysis plan

This statistical analysis plan (SAP) is based on the protocol for trial NN9924-4224 “Efficacy and Safety of Oral Semaglutide versus Liraglutide and versus Placebo in Subjects with Type 2 Diabetes Mellitus”, version 2.0 (17 November 2016), and includes more detailed procedures for executing the statistical analyses of the primary and secondary endpoints. Statistical analyses and a number of clarifications additional to those specified in the trial protocol are pre-specified with this SAP. All changes to the statistical analyses planned in the trial protocol are documented in Section [3](#).

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

2 Statistical considerations

General considerations

The blinding of the randomised treatments will be maintained until the database has been released for statistical analysis. Data from all sites will be analysed and reported together.

In statistical analyses where stratification is included, the two combinations of anti-diabetic background medication at screening (metformin or metformin+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. The information regarding descent (Japanese subjects/non-Japanese subjects) will be included based on country details from the eCRF. In the statistical analyses the stratification factor will refer to anti-diabetic background medication at screening. Descent (Japanese subjects/non-Japanese subjects) will be included in the statistical analyses as part of region.

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 $\frac{1}{2}$ LLoQ. Number of values below LLoQ by treatment and visit will be summarised if deemed relevant.

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

Results from a statistical analysis will as a minimum be presented by the estimated treatment contrasts for the below two 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 14 mg vs. liraglutide 1.8 mg

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

Primary and secondary estimands

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

- Primary estimand – ‘Treatment policy’
 - treatment difference (oral semaglutide versus placebo and oral semaglutide versus liraglutide) 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) and the expected glycaemic benefit compared to initiating treatment with liraglutide including potential rescue medication(s). Generalisation of this estimand depends among other things on the extent to which the use of rescue medication in this trial reflects clinical practice and the adherence to trial product administration in this trial reflects the behaviour of the target population. Accordingly, data collected regardless of discontinuation of trial product or initiation of rescue medication(s) will be used to draw inference.

- Secondary estimand – ‘Hypothetical’
 - treatment difference (oral semaglutide versus placebo and oral semaglutide versus liraglutide) 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 and the glycaemic benefit compared to liraglutide. It is considered a clinically relevant estimand as it provides information to treating clinicians about the expected glycaemic efficacy of oral semaglutide and the expected glycaemic efficacy compared to liraglutide for purposes of treating individual subjects. Generalisation of this estimand depends among other things on the extent to which the adherence to trial product administration in this trial reflects the behaviour of the target population. Accordingly, only data collected prior to discontinuation of trial product or initiation of rescue medication will be used to draw inference. This will avoid confounding from rescue medication.

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

Missing data considerations at week 26

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

When estimating the secondary estimand, the proportion of missing data is expected to be higher (20%) since data collected after discontinuation of trial product or initiation of rescue medication(s) will be set to missing. The 20% of missing data is based on the liraglutide phase 3 trial (NN2211-1572), the oral semaglutide phase 2 trial (NN9924-3790) that indicates that a low starting dose with gradual dose escalation diminishes GI AEs compared with more aggressive dosing regimens. Across treatment arms the main reasons for missing data are expected to be early treatment discontinuation due to AEs (GI AEs for oral semaglutide and liraglutide) and initiation of rescue medication. Initiation of rescue medication is expected to be more frequent in the placebo arm than for the oral semaglutide and liraglutide arms. A higher proportion of subjects are expected to discontinue treatment due to AEs in the oral semaglutide and liraglutide arms, compared to the placebo arm. So overall the frequency of missing data is expected to be similar across treatment arms.

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

2.1 Sample size calculation

To assess the effect of liraglutide on glycaemic effect, a similar trial LEAD-2 (NCT00318461) where liraglutide was used as add on to metformin was reviewed. Based on this trials, the chosen margin of 0.4 provides assurance that oral semaglutide has a clinically relevant effect greater than zero. With regards to the constancy assumption, controlled clinical trials have consistently established that liraglutide is an effective anti-diabetic drug. Therefore, lack of trial sensitivity with liraglutide as comparator is not anticipated to be an issue in this trial.

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

The sample size calculation is made to ensure a power of at least 90% for testing the below four out of the five pre-specified confirmatory hypotheses shown in [Figure 2-1](#). The closed testing procedure described in Bretz et al. (2011)⁴ is used to control the overall type I error at a nominal two-sided 5% level.

The four hypotheses are

- HbA_{1c} superiority of oral semaglutide vs. placebo
- HbA_{1c} non-inferiority of oral semaglutide vs. liraglutide (margin 0.4%)
- Body weight superiority of oral semaglutide vs. placebo
- Body weight superiority of oral semaglutide vs. liraglutide

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

- To demonstrate glycaemic effect of oral semaglutide compared to placebo
- Glycaemic effect compared to liraglutide must be established in terms of HbA_{1c} non-inferiority before testing for added benefits in terms of HbA_{1c} superiority and/or body weight superiority.

The sample size is calculated using the calcPower function in the R package, gMCP² using 10000 simulations. All of the five pre-specified confirmatory tests are assumed to be independent. Since some of the tests are positive correlated, the assumption of independence is viewed as conservative.

The sample size assumptions for treatment effects (TE), adjusted treatment effects and the standard deviations (SD) are given in [Table 2-1](#).

These assumptions are primarily based on the oral semaglutide phase 2 results (NN9924-3790), liraglutide phase 3 trial results (Victoza® US prescribing information) and supported by results from the s.c. semaglutide phase 2 trial (NN9535-1821).

Since the equalising effect of rescue medication will be included in the primary analysis as well as a conservative approach for handling of missing data will be performed, an adjustment in treatment effect will be implemented for the 10% of subjects who are expected to discontinue trial product or initiate rescue medication and for the 10% of subjects who are expected to have actual missing data. The treatment effects used in the sample size calculation will be adjusted according to a 75% smaller effect in these subjects. For the 10% of the subjects with missing data, the non-inferiority margin of 0.4% for HbA_{1c} is added to the imputed values, when testing for non-inferiority. The adjusted treatment effects for testing non-inferiority (HbA_{1c} oral semaglutide versus liraglutide only) and superiority are as described below:

- Non-inferiority
 - $0.8 \times TE + 0.2 \times TE \times 0.25 + \text{non-inferiority margin} \times 0.1$
- Superiority
 - $0.8 \times TE + 0.2 \times TE \times 0.25$

Table 2–1 Assumptions used in the sample size calculation

Comparator	Parameter	Treatment effect (TE)	Adjusted TE, non-inferiority	Adjusted TE, superiority	Standard deviation	Non-inferiority margin
Placebo	HbA _{1c}	-1.0%		-0.85%	1.1%	
Liraglutide	HbA _{1c}	0%	+0.04%	0%	1.1%	0.4%
Placebo	Body weight	-3kg		-2.55kg	4kg	
Liraglutide	Body weight	-1.5kg		-1.275kg	4kg	

With the above assumptions using a 2:2:1 randomisation, allocating 276 subjects to each of the oral semaglutide and the liraglutide arms and allocating 138 subjects to the placebo arm, provides 90% power to jointly confirm HbA_{1c} superiority of oral semaglutide vs. placebo, HbA_{1c} non-inferiority of oral semaglutide vs. liraglutide, body weight superiority of oral semaglutide vs. placebo and body weight superiority of oral semaglutide vs. liraglutide. In total 690 subjects are planned to be randomised. Calculated powers for individual hypotheses are presented in [Table 2–2](#).

Table 2–2 Calculated powers for individual hypotheses

Comparator:	placebo		liraglutide		
Statistical test:	HbA _{1c} superiority	Body weight superiority	HbA _{1c} non-inferiority	HbA _{1c} superiority	Body weight superiority
Power:	> 99%	97%	97%	2%	91%

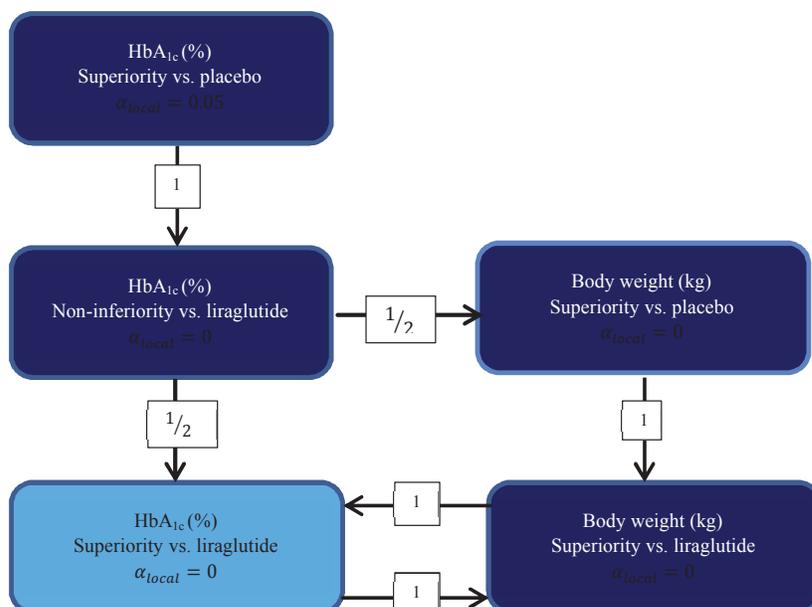


Figure 2–1 Graphical illustration of the closed testing procedure

The overall significance level of $\alpha = 0.05$ (two-sided) is initially allocated to the HbA_{1c} superiority test vs. placebo. 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.

2.2 Definition of analysis sets

The following analysis sets will be defined:

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

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

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

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

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

Data selections and observation periods

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

- the follow-up visit (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 the first step will be selected based on the specified observation period

Definition of the observation periods:

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

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

On-treatment: This observation period represents the time period where subjects are considered treated with trial product. The observation period is a subset of the in-trial observation period. It starts at the date of first dose of trial product. Two slightly different end dates will be needed to cover all assessments appropriately. For adjudicated events, ECGs, eye examination category, anti-semaglutide antibodies, and AEs including hypoglycaemic episodes, the observation period ends at the first date of any of the following:

- the follow-up visit (V14)
- the follow-up prematurely discontinuation visit (V14A)
- the last date on trial product +38 days (5 weeks follow-up period + 3 days visit window)
- 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 has not initiated any rescue medications. The on-treatment without rescue medication observation period starts at first date on trial product and the observation period ends at the first date of any of the following:

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

The in-trial observation period will be the primary observation period when 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.

Confirmatory hypotheses

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

- HbA_{1c} non-inferiority, using a non-inferiority margin of 0.4% (only vs. liraglutide)
 - H₀: $\mu \geq 0.4\%$ against H_a: $\mu < 0.4\%$

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

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

Multiplicity and criteria for confirming hypotheses

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

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

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

2.3 Primary endpoint

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

2.3.1 Primary analysis for the primary estimand

The primary estimand will be estimated based on the FAS using week 26 measurements from the in-trial observation period. The primary statistical analysis will be a pattern mixture model using multiple imputation to handle missing data assuming that the missing data mechanism is missing at random (MAR) within the groups used for imputation. Imputation of missing data at week 26 will be done within 6 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 stratification factor and region as categorical fixed effects and baseline HbA_{1c} measurement as a covariate will be fitted to observed values of the change from baseline in HbA_{1c} at week 26.
- The estimated parameters for location and dispersion will be used to impute 1000 values for each subject with missing week 26 data based on stratification factor and region as categorical and baseline HbA_{1c}. Thus, 1000 complete data sets will be generated including observed and imputed values.

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

Analysis used for confirming superiority versus placebo/liraglutide at week 26:

For each of the 1000 (now complete) imputed data sets, the change from baseline to week 26 will be analysed using an ANCOVA with treatment, stratification factor and region as categorical fixed effects and baseline HbA_{1c} as covariate. The results obtained from analysing the datasets will be combined using Rubin's rule³ to draw inference.

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

Prior to analysing the data using the same model and approach as used for confirming superiority (see above), a value of 0.4% (the non-inferiority margin) will be added to imputed values at week 26 for the oral semaglutide treatment arms only⁴. For evaluating non-inferiority versus liraglutide an unadjusted two sided p-value for testing no difference from the non-inferiority margin will be presented.

2.3.2 Primary analysis for the secondary estimand

The secondary estimand will be estimated based on the FAS using post-baseline measurements up to and including week 26 from the on-treatment without rescue medication observation period. The primary analysis for the secondary estimand will be a Mixed Model for Repeated Measurements

(MMRM). A restricted maximum likelihood (REML) will be used. The model will include all post baseline HbA_{1c} measurements collected at scheduled visits up to and including week 26 as dependent variables. The independent effects included in the model will be treatment, stratification factor and region as categorical fixed effects and baseline HbA_{1c} as a covariate, all nested within visit. An unstructured covariance matrix for HbA_{1c} measurements within the same subject will be employed, assuming measurements from different subjects are independent.

The MMRM is a well-established method that accounts for the uncertainty pertaining to missing data. This analysis assumes that the missing data mechanism is MAR. Under this assumption the statistical behaviour of the missing data (given the observed responses and model fixed effects and covariates) is assumed to be same as the observed data.

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

2.3.3 Sensitivity analyses

To investigate the sensitivity of the primary analysis results, complementary and separate analyses will be performed for the primary and secondary estimand. In line with EMA recommendations⁵ and with a report from the US National Research Council⁶, these analyses will primarily evaluate the sensitivity of the results due to the impact of missing data. Since conservatism, i.e. avoiding bias in favour of oral semaglutide, depends on the context, separate sensitivity analyses will be made for non-inferiority and superiority testing.

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

Sensitivity analyses for the primary estimand

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

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

Sensitivity analyses for the secondary estimand

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

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

2.3.3.1 Pattern mixture models

Common for the three pattern mixture model sensitivity analyses is that they all aim to stress-test the primary HbA_{1c} results by changing the assumptions for part or all missing data in the oral semaglutide treatment arm, while maintaining the missing data assumption for the comparator (placebo or liraglutide depending on the hypothesis).

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

2.3.3.2 Other sensitivity analyses

The following additional sensitivity analysis will be specified

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

2.3.3.3 Assessment of sensitivity analyses

The results from the sensitivity analyses will be collectively used to interpret the robustness of the trial results for HbA_{1c}. Due to the large number of sensitivity analyses and their inherent conservative nature, it will not be a requirement that all confirmatory hypotheses are consistently confirmed across the sensitivity analyses. Thus, no absolute success criteria will be pre-defined for each sensitivity analysis. The sensitivity results in totality will be used to substantiate the credibility of the trial results.

2.4 Secondary endpoints

2.4.1 Confirmatory secondary endpoints

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

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

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

2.4.2 Supportive secondary endpoints

2.4.2.1 Efficacy endpoints

The below supportive secondary efficacy endpoints will be evaluated for

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

No sensitivity analyses are planned for these.

Continuous efficacy endpoints

Change from baseline to week 52 in:

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

Change from baseline to week 26 and week 52 in:

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

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 7-point SMPG profile:

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

The above continuous endpoints will be analysed separately using similar model approaches as for the primary endpoint with the associated baseline response as a covariate. 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 week 52. For the analysis at week 52, the imputation of missing data will be further differentiated by whether subjects have discontinued trial product or initiated rescue medication prior to week 26 or at/after week 26. This will result in imputation of missing data within 9 groups of subjects instead of the 6 groups as described for the week 26 evaluation in Section [2.3.1](#). If less than 5 subjects have available data in one of the 9 groups, the imputation will be made within the 6 groups specified for the primary evaluation. The frequency of missing data is expected to be slightly larger at week 52 compared to week 26. The rate of missing data is expected to decline over time.

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

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

Binary efficacy endpoints

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

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

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

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

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

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

For both estimands, each of the 1000 data sets will be analysed using a logistic regression model with treatment, stratification factor and region as fixed effects and baseline value as covariate (i.e. baseline HbA_{1c} for binary HbA_{1c} endpoints, baseline body weight for body weight endpoints and both HbA_{1c} and baseline body weight for the composite endpoints that comprises both parameters). The results will be combined using Rubin's rule³ to draw inference.

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

Time to event endpoints

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

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

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

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

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

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

Treatment policy estimand: Time to additional anti-diabetic medication

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

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

Hypothetical estimand: Time to rescue medication

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

2.4.2.2 Safety endpoints

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

Adverse events

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

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

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

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

Other safety endpoints

Change from baseline to week 26 and week 52 in:

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

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

Change from baseline to week 26 and week 52 in:

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

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

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

Anti-semaglutide binding antibodies up to approximately 57 weeks:

- Anti-semaglutide binding antibody levels

Other safety assessments

Change from baseline to week 26 and week 52 in:

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

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

Hypoglycaemia

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

Classification of hypoglycaemia:

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

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

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

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

Novo Nordisk classification of hypoglycaemia

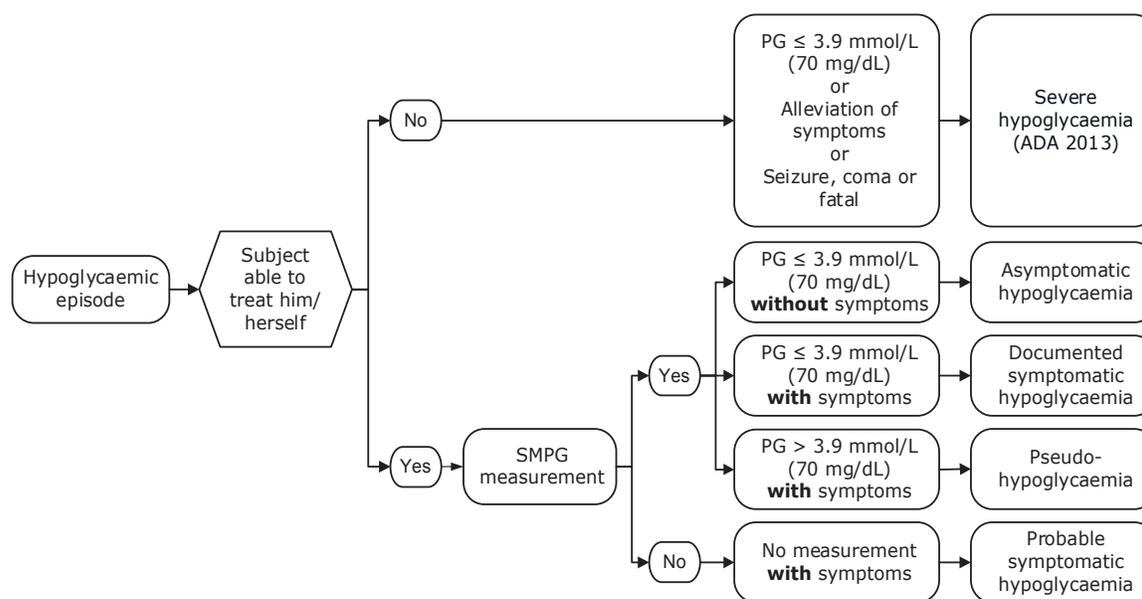
In normal physiology, symptoms of hypoglycaemia occur below a PG level of 3.1 mmol/L (56 mg/dL)⁷. Therefore, Novo Nordisk has included hypoglycaemia with PG levels below this cut-off point in the definition of BG-confirmed hypoglycaemia.

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

Severe or BG-confirmed symptomatic hypoglycaemia: An episode that is severe according to the ADA classification⁸ or BG-confirmed by a PG value < 3.1 mmol/L (56 mg/dL) with symptoms consistent with hypoglycaemia.

ADA classification of hypoglycaemia⁸

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



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

Figure 2–2 ADA classification of hypoglycaemia

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

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

Analysis of severe or BG-confirmed symptomatic hypoglycaemic endpoints

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

2.5 Interim analysis

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

2.6 Patient reported outcomes

Change from baseline to week 26 and week 52 in:

- DTSQs: individual items and treatment satisfaction score (6 of the 8 items summed)

The PRO endpoint will be evaluated using the primary analysis for the primary estimand based on FAS using the in-trial observation period and using the primary analysis for the secondary estimand based on FAS using the on-treatment without rescue medication period. The individual items and the treatment satisfaction score will be analysed separately as the other continuous efficacy endpoints with the associated baseline response as a covariate.

2.6.1 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.

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 blood glucose 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.

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)

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 [2.6.2](#).

2.6.2 Responder analyses

Responder analyses will be conducted for both estimands, for the same time points that are defined for the analyses of PRO endpoint (see protocol) 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 \geq 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 \leq 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 \geq 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 NN9924-4224. However, clarifications, more detailed descriptions of endpoints and analyses are provided in this SAP. The changes from the protocol of NN9924-4224 are summarised below:

- ‘On-treatment without rescue medication’ added to the criteria ‘*is exposed to trial product and have at least one HbA1c measurement at or after week 14*’ for Per protocol analysis set.

- The primary and secondary estimands have changed names from de-facto and de-jure to treatment policy and hypothetical, respectively.
- The MMRM sensitivity analysis of the primary estimand has been omitted in section [2.3.3](#). It is considered sufficient to keep the three current sensitivity analyses to stress test the primary results.
- For the MMRM analyses, it is specified that for subjects who do not have post-baseline assessments for planned visits available in the on-treatment without rescue medication period, the baseline value will be carried forward to the first planned visit, if the first planned visit do not fall later than 8 weeks after randomisation, to ensure that all randomised subjects will contribute to the statistical analyses.
- The three MI sensitivity analyses of the secondary estimand have been omitted in section [2.3.3](#). It is considered sufficient to keep the tipping point sensitivity analysis for the secondary hypothetical estimand as it can be considered as a progressive stress-testing to assess how severe departures from MAR must be in order to reverse the conclusions from the primary MMRM analysis used to address the hypothetical estimand.
- The LOCF sensitivity analysis specified in the trial protocol (section 17.3.3.2) has been omitted, as it is not realistic that subjects with missing data would have had stable results from the point of drop out to trial completion.
- The complete case analysis is based on the assumption that missing data are MCAR, this assumption seem questionable and since a per-protocol analysis is performed, this extra sensitivity analysis is excluded.
- The statistical analyses of the two binary effect endpoints (HbA_{1c} reduction \geq 1%-point (10.9 mmol/mol) and body weight loss \geq 3%) have been omitted, because they are being analysed as a part of the two composite binary effect endpoints.
- For the binary efficacy endpoints, it has been specified how missing data in the analyses for the hypothetical estimand will be imputed using a sequential imputation approach assuming MAR.
- A clarification of the ‘without hypoglycaemia’ component in composite binary endpoints has been added.
- The definitions of initiation of rescue medication and additional anti -diabetic medication used for the time-to-event endpoints as well as the accompanying statistical analyses have been further clarified.
- Standard laboratory parameters have been downgraded from ‘endpoints’ to ‘assessments’ because these parameters are included to ensure that the investigator can monitor the

individual subject and to allow for detection of potential safety signal on a trial-level that is not reflected in adverse event reporting.

- In section [2.6](#) details on the analyses of DTSQs endpoints have been included.

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 RJA, Rubin DB. *Statistical analysis with missing data*: New York: John Wiley & Sons. 1987.
4. Koch GG. Comments on 'Current issues in non-inferiority trials' by Thomas R. Fleming, *Statistics in Medicine*, DOI: 10.1002/sim.2855. *Stat Med*. 2008;27(3):333-42.
5. Committee for Medicinal Products for Human Use (CHMP). *Guideline on Missing Data in Confirmatory Clinical Trials*. EMA/CPMP/EWP/1776/99 Rev. 1. 2 Jul 2010.
6. National Academy of Sciences (NAS). *The Prevention and Treatment of Missing Data in Clinical Trials*. Washington D.C.: The National Academies Press. 2010.
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. Bradley C. Diabetes Treatment Satisfaction Questionnaire. In: Bradley C, editor. *Handbook of Psychology and Diabetes*. Chur, Switzerland: Harwood Academic Publishers; 1994. p. 111-32