Cover Page for Statistical Analysis Plan

Sponsor name:	Novo Nordisk A/S
NCT number	NCT03018028
Sponsor trial ID:	NN9924-4281
Official title of study:	PIONEER 9 – Japan Monotherapy. Dose-response, Safety and Efficacy of Oral Semaglutide versus Placebo and versus Liraglutide, All as Monotherapy in Japanese Subjects with Type 2 Diabetes
Document date:	28 February 2019

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: NN9924-4281

PIONEER 9 – Japan Monotherapy

Dose-response, Safety and Efficacy of Oral Semaglutide versus Placebo and versus Liraglutide, All as Monotherapy in Japanese Subjects with Type 2 Diabetes

Author:

[redacted]

Table of contents

Page

List of abbreviations

ADA	American Diabetes Association
AE	adverse event
AACE	American Association of Clinical Endocrinologists
ANOVA	analysis of variance
ATC	Anatomical Therapeutic Chemical
BG	blood glucose
BMI	body mass index
CI	confidence interval
CTR	clinical trial report
DTR-QOL	Diabetes Therapy-Related Quality of Life
EAC	event adjudication committee
ECG	electrocardiogram
eCRF	electronic case report form
FAS	full analysis set
HbA_{1c}	glycosylated haemoglobin
HDL	high density lipoprotein
HRQoL	health-related quality of life
HOMA-B	homeostatic model assessment index of beta-cell function
HOMA-IR	homeostatic model assessment index of insulin resistance
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
PK	pharmacokinetics
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)
SMPG	self-measured plasma glucose
T2DM	type 2 diabetes mellitus
TEAE	treatment-emergent adverse events
VLDL	very low density lipoprotein
VT	vitality

1 Introduction

1.1 Trial information

This is a combined phase 2 dose-response and phase 3 safety and efficacy trial. It is a 52-week, randomised double-blind placebo-controlled and open-label active-controlled, parallel-group, multicentre, single country trial with 5 treatment arms in Japanese subjects with T2DM to:

- assess dose-response relationship of once-daily dosing of 3 dose levels of oral semaglutide (3, 7 and 14 mg) versus placebo on glycaemic control at week 26
- compare the safety and efficacy of once-daily dosing of 3 dose levels of oral semaglutide (3, 7 or 14 mg) versus placebo and versus once-daily 0.9 mg liraglutide subcutaneously at week 26 and week 52

Primary objective

To assess the dose-response relationship of once-daily dosing of three dose levels (3, 7 and 14 mg) of oral semaglutide versus placebo as monotherapy on glycaemic control in Japanese subjects with type 2 diabetes mellitus.

Secondary objective

To compare the safety and tolerability of once-daily dosing of three dose levels (3, 7 and 14 mg) of oral semaglutide versus placebo and versus once-daily 0.9 mg liraglutide subcutaneously, all as monotherapy in Japanese subjects with type 2 diabetes mellitus.

To compare the effect of once-daily dosing of three dose levels (3, 7 and 14 mg) of oral semaglutide versus placebo and versus once-daily 0.9 mg liraglutide subcutaneously, all as monotherapy on glycaemic control and body weight in Japanese subjects with type 2 diabetes mellitus.

Trial design

Japanese subjects with T2DM treated with diet and exercise therapy alone or OAD as monotherapy (half maximum approved dose or below) in addition to diet and exercise therapy will be included in this trial. For subjects treated with OAD as monotherapy, pre-trial OAD will be discontinued at screening and washed-out before randomisation. Subjects treated with diet and exercise therapy alone will be randomised without a wash-out period after their eligibility has been confirmed.

Subjects will be randomised in a 1:1:1:1:1 manner to receive either a oral semaglutide 3, 7 or 14 mg once-daily, oral placebo once-daily or liraglutide 0.9 mg subcutaneous (s.c.) injection once-daily. Randomisation will be stratified based on pre-trial OAD treatment at screening (yes/no).

The trial is double-blinded with regard to dose levels of oral semaglutide and placebo. In order to maintain the blinding of the trial for oral semaglutide and placebo all tablets containing oral

semaglutide or placebo are identical with regards to visual appearance. The treatment administration of liraglutide will be open-label.

The total trial duration for the individual subject will be 59 to 65 weeks depending on pre-trial treatment. 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 subjects treated with OAD as monotherapy prior to the screening an 8-week wash-out period including the screening period will replace the 2-week screening period.

1.2 Scope of the statistical analysis plan

This statistical analysis plan (SAP) is based on the protocol for trial NN9924-4281 "Dose-response, Safety and Efficacy of Oral Semaglutide versus Placebo and versus Liraglutide, all as Monotherapy in Japanese Subjects with Type 2 Diabetes", version 3.0 (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

2.1 General considerations

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

Data from all sites will be analysed and reported together.

In statistical analyses where stratification is included, the pre-trial OAD treatment at screening (yes/no) 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 pre-trial OAD treatment 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.

The primary efficacy endpoints 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 of missing data at week 52 and therefore considered a meaningful representation and confirmation of the dose-response relationship of oral semaglutide.

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

- oral semaglutide 14 mg vs. placebo
- oral semaglutide 14 mg vs. liraglutide 0.9 mg
- oral semaglutide 7 mg vs. placebo
- oral semaglutide 7 mg vs. liraglutide 0.9 mg
- oral semaglutide 3 mg vs. placebo
- oral semaglutide 3 mg vs. liraglutide 0.9 mg

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

2.1.1 Primary and secondary estimands

Two estimands addressing different aspect of the trial objective will be defined; a primary "Hypothetical" estimand and a secondary "Treatment policy" estimand.

2.1.1.1 Primary estimand - 'Hypothetical'

- Treatment difference (oral semaglutide versus placebo and oral semaglutide versus liraglutide) at week 26 in all randomised subjects if all subjects adhered to treatment and did not initiate rescue medication

The hypothetical estimand assesses the expected glycaemic benefit under adherence to trial product and is considered the primary interest for evaluating the dose-response relationship of oral semaglutide. 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.

2.1.1.2 Secondary estimand - 'Treatment policy'

- Treatment difference (oral semaglutide versus placebo and oral semaglutide versus liraglutide) at week 26 in all randomised subjects regardless of discontinuation of trial product or initiation of rescue medication(s)

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 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.

2.1.2 Missing data considerations

Based on the results from similar Japanese phase 3 trials investigating liraglutide and s.c. semaglutide as monotherapy (NN2211-1700, NN9535-4091 and NN9535-4092) and the oral semaglutide phase 2 trial (NN9924-3790), it is expected that maximum 30% of the subjects will discontinue trial product or withdrawal from the trial at week 52.

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

When estimating the primary estimand, the proportion of subject with missing data is expected to be higher (20%) than for the secondary estimand because data collected after discontinuation of trial product or initiation of rescue medication(s) will be set to missing. The proportion of subject with missing data at week 26 is expected to be maximum 20% based on the liraglutide phase 3 trial (NN2211-1700), s.c. semaglutide phase 3 trials (NN9535-4091 and NN9535-4092) and the oral semaglutide phase 2 trial (NN9924-3790).

Across treatment arms, the main reasons for missing data are expected to be early treatment discontinuation due to AEs (gastrointestinal 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.

2.2 Sample size calculation

The total sample size of 240 subjects and duration of this trial have been determined to allow for adequate (i) evaluation of the dose-response relationship and (ii) safety evaluation in a monotherapy setting in line with the Japanese draft AD guideline[1]. With a maximum expected premature

treatment discontinuation rate of 30% after week 52, this trial ensures that at least 100 subjects are exposed to oral semaglutide for one year.

Based on 48 subjects per treatment arm and a standard deviation of 1.1% in change from baseline to week 26 in HbA1c, the width of the 95% confidence interval (precision) for the mean difference in HbA1c will at most be +/- 0.5 %-point with 90% probability.

Furthermore, with 48 subjects per treatment arm the trial will be able to show a significant HbA_{1c} treatment difference of 0.74 %-point with 90% power and a treatment difference of 0.64 %-point with 80% power. As a reference, the assumed treatment effects (TE) and adjusted treatment effects are presented in Table 2–1. These are based on the oral semaglutide phase 2 results (NN9924-3790) and supported by results from the s.c. semaglutide phase 2 trial (NN9535-1821).

The adjustment of the treatment effects are done differently when estimating the primary and the secondary estimand, due to the different analyses and the inclusion of data collected after treatment discontinuation and initiation of rescue medication for the secondary estimand. The treatment effects will be adjusted according to a 50% smaller effect in subjects with missing data when estimating the primary estimand. Whereas a 75% smaller effect will be used for the 10% of the subjects who are expected to discontinue trial product or initiate rescue medication and for the 10% of subjects who are expected to have missing data when estimating the secondary estimand.

Primary estimand
- Adjusted TE = 0.8×TE + 0.2×TE×0.50

Secondary estimand
- Adjusted TE = 0.8×TE + 0.1×TE×0.25 + 0.1×TE×0.25

Table 2–1 Treatment effects in HbA1c (%-point)

Oral semaglutide	3 mg	7 mg	14 mg
Treatment effects (TE)	0.45	0.75	1.0
Primary estimand: Adjusted TE	0.405	0.675	0.9
Secondary estimand: Adjusted TE	0.3825	0.6375	0.85

2.3 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 treated. This will be referred to as contributing to the evaluation "as treated".

Data selections and observation periods

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

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

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

- Firstly, subjects will be selected based on the specified analysis set
- Secondly, data points on the selected subjects from 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, anti-semaglutide antibodies and AEs including hypoglycaemic episodes, the observation period ends at the first date of any of the following:

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

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

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

On-treatment without rescue medication: This observation period is a subset of the on-treatment observation period, where subjects are considered treated with trial product, but have not initiated any rescue medications. 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 when estimating the secondary estimand. The on-treatment without rescue medication observation period will be the primary observation period when estimating the primary estimand. Safety will be evaluated based on the in-trial and the on-treatment observation periods.

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

Before data are locked for statistical analysis 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.

2.4 Primary endpoint

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

2.4.1 Analysis for the primary estimand

The primary 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 dose-response will be evaluated by a mixed model for repeated measurements (MMRM). A restricted maximum likelihood will be used. The model will include all post-baseline HbA_{1c} measurements collected at scheduled visits up to and including week 26 as dependent variables. The independent effects included in the model will be treatment and 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. From this analysis estimated differences with corresponding two sided p-values and 95% confidence intervals after 26 weeks of treatment will be presented for all pairwise comparison of oral semaglutide versus placebo. In addition to the dose-response evaluation, oral semaglutide will be compared to liraglutide 0.9 mg and estimated differences with corresponding two sided p-values and 95% confidence intervals after 26 weeks of treatment will be presented for all pairwise comparison of oral semaglutide versus liraglutide 0.9 mg.

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 missing at random (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.

The dose-response relationship of oral semaglutide will be supported by an exploratory analysis. In this model data from all oral semaglutide treatment arms will be included and data from the placebo arm (corresponding to a dose of zero mg) could be included if appropriate. The model will be developed ad hoc after unblinding of the trial to aim for a suitable fit of the dose-response relationship using e.g. a linear or a four-parameter logistic relation of HbA_{1c} vs. the logarithm of the dose.

2.4.2 Analysis for the secondary estimand

The secondary estimand will be estimated based on the FAS using week 26 measurements from the in-trial observation period. The statistical analysis for the secondary estimand will be a pattern mixture model using multiple imputation (MI) to handle missing data assuming that the missing data mechanism is MAR within the groups used for imputation. Imputation of missing data at week 26 will be done within 6 groups of subjects: one group of subjects regardless of randomised treatment arm who at week 26 will have discontinued treatment or have initiated rescue medication, and 5 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 medication 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.

Analysis used for testing superiority versus placebo at week 26:

For each of the 1000 (now complete) imputed data sets the change from baseline to week 26 in HbA_{1c} will be analysed using an analysis of covariance (ANCOVA) with treatment and stratification factor as categorical fixed effects and baseline HbA_{1c} as covariate. The results obtained from analysing the datasets will be combined using Rubin's rule[2] to draw inference.

From this analysis the estimated treatment differences between each of the oral semaglutide dose levels and placebo together with associated two-sided 95% CIs and unadjusted two-sided p-values will be presented for all pairwise comparison of oral semaglutide versus placebo. In addition to this analysis, oral semaglutide will be compared to liraglutide 0.9 mg and estimated differences with corresponding two sided p-values and 95% CIs after 26 weeks of treatment will be presented for all pairwise comparison of oral semaglutide versus liraglutide 0.9 mg.

2.5 Secondary endpoint

2.5.1 Supportive secondary endpoint

2.5.1.1 Efficacy endpoints

The below supportive secondary efficacy endpoints will be evaluated for:

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

Continuous efficacy endpoints

Change from baseline to week 52 in:

- HbA_{1c}

Change from baseline to week 26 and week 52 in:

- FPG
- Body weight (kg)
- Body weight (%)
- BMI
- Waist circumference
- Fasting lipid profile (total cholesterol, LDL cholesterol, HDL cholesterol and triglycerides, VLDL)
- Fasting insulin, C-peptide, glucagon, pro-insulin
- Pro-insulin/insulin ratio
- Insulin resistance (HOMA-IR)
- Beta-cell function (HOMA-B)

BMI will be calculated based on body weight and height based on the formulae:
BMI kg/m^2 = body weight (kg)/(height (m) × height (m)) or (kg/m^2 = [lb/in^2 × 703])

Change from baseline to week 26 and 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 the primary endpoint with the associated baseline response as a covariate. The following endpoints will be log-transformed prior to analysis with the associated log-transformed baseline value as a covariate: fasting insulin, C-peptide, glucagon, pro-insulin, pro-insulin/insulin ratio, HOMA-IR, HOMA-B and fasting lipid profile. No model development of dose-response relationship is planned for these.

For evaluation of the primary estimand the MMRM based primary analysis will include all scheduled post-baseline measurement up to and including week 52. From this analysis the estimated treatment differences (ratios) between each of the oral semaglutide dose levels and placebo at week 26 and 52 and between each of the oral semaglutide dose levels and liraglutide 0.9 mg at week 26 and week 52 together with associated two-sided 95% CIs and unadjusted two-sided p-values will be presented for all pairwise comparison of oral semaglutide versus placebo and versus liraglutide 0.9 mg. The baseline will not be carried forward to first planned visit if the first planned visit falls later than 8 weeks after randomisation.

For evaluation of the secondary 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 6 groups of subjects: one group of subjects regardless of randomised treatment arm who at week 52 will have discontinued treatment or have initiated rescue medication, and 5 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.

The frequency of missing data is expected to be slightly larger at week 52 compared to week 26. The rate of missing data is expected to decline over time.

Binary efficacy endpoints

If a subject after week 26 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 ≥ 1 %-point (10.9 mmol/mol) and body weight loss $\geq 3\%$

The above five (5) endpoints will also be evaluated after week 52.

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 (5) 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 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 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 ANCOVA model will be used to impute missing values at each planned visit. The model will include 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 HbA1c and body weight the imputed data sets will be combined by imputation number.

Only observed data within the corresponding observation period will be included for the 'without hypoglycaemia' component of composite endpoint. Because the number of hypoglycaemic episodes is expected to be very low in this trial, the observed data is considered sufficient when addressing both estimands.

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 initiated at or after randomisation and before (planned) end-of-treatment.

Definition of rescue medication:
New 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 can be considered a ***new anti-diabetic medication***:
- Anti-diabetic medication (4th-level ATC code) that is initiated at or after randomisation and with a dosing duration of more than 21 days

More than 21 days are 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 likelihood ratio tests obtained from 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 placebo and between each of the oral semaglutide dose levels and liraglutide 0.9 mg together with associated two-sided 95% CIs and unadjusted two-sided p-values will be presented for all pairwise comparison of oral semaglutide versus placebo and versus liraglutide 0.9 mg.

The analysis aims to address the need of additional anti-diabetic medication regardless of whether this is due to lack of effect or tolerability of the trial product. Switch to another 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 (that is, initiated additional anti-diabetic medication) 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 of the treatment with trial product. Only initiation of rescue medication as add-on to randomised treatment is considered an event. Switch to another anti-diabetic treatment is not considered an event (that is, initiation of rescue medication) and as a consequence subjects will be censored on the day before date of last trial product. Potential events occurring between randomisation and first date on trial product will be included in the analysis as events at day 0, in order to count all events of rescue medication.

2.5.1.2 Safety endpoints

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

Adverse events

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

All AEs will be coded using version 20.1 of the 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 2.3).

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 52 in:

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

The above safety endpoints will be evaluated using the primary analysis for the secondary estimand based on SAS using the in-trial observation period and using the analysis for the primary 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 52 in:

- ECG category

- Physical examination

Change from baseline to week 52 in:

- Eye examination category

Occurrence of anti-semaglutide antibodies (yes/no):

- 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 antibodies:

- 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 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 2.3).

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

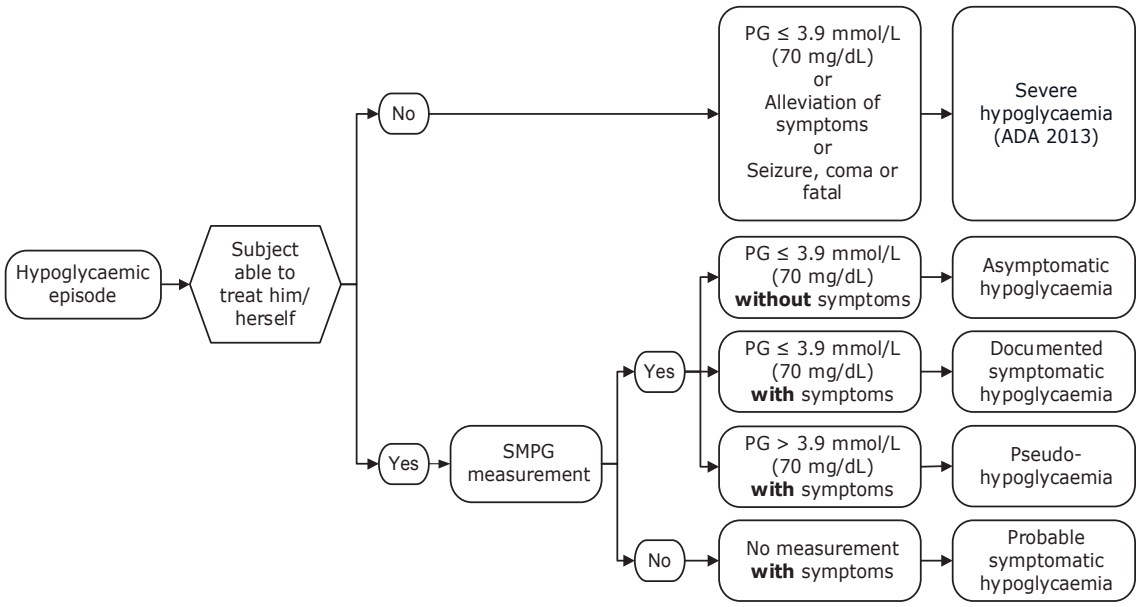
In normal physiology, symptoms of hypoglycaemia occur below a PG level of 3.1 mmol/L (56 mg/dL)[3]. Therefore, Novo Nordisk has included hypoglycaemia with PG levels below this cut-off point in the definition of BG confirmed hypoglycaemia.

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[4] 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.

2.5.1.3 Pharmacokinetic endpoints

- Semaglutide plasma concentrations for population PK analysis

The semaglutide plasma concentrations collected in this trial will be evaluated using relevant summary statistics. In addition, the semaglutide plasma concentration will be part of a meta-analysis across the oral semaglutide phase 3a trials, see more details in Section 2.7.

2.6 Interim analysis

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

2.7 Pharmacokinetic and/or pharmacodynamic modelling

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

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

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

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

2.8 Health economics and/or patient reported outcomes

Change from baseline to week 26 and week 52 in:

- SF-36v2® Health Survey (SF-36v2) (acute version): Physical component score, mental component score and scores from the 8 domains
- DTR-QOL questionnaire: Total score and scores from 4 domains

The PRO 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) for the treatment policy estimand will be reported in the CTR. SF-36v2 (acute version) for the hypothetical estimand will not be included in CTR but will be included in a separate PRO report. DTR-QOL questionnaire for both estimand will be reported in the CTR.

2.8.1 SF-36v2® Health Survey (acute version)

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 2–2 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 2–2 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 2–3.

Table 2–3 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 2.8.3.

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

DTR-QoL is a diabetes specific HRQoL measure that 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 tr eatment”. 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 2–4 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]

Total score	1-29	Step 2: Formula for domain score derivation[1]: (Sum of item scores -4) * (100/(4*6)) 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))

1 General formula: (Sum of item scores - Number of items included in domain) * (100/ (Number of items included in domain * Scaling point range))

2 Item scores are reversed in the following way: 7 to 1, 6 to 2, …, 1 to 7

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 per trial were used as distribution-based approach defining the responder thresholds. The thresholds are derived from baseline DTR-QoL data across trial arms per trial. Responder analyses will be based on the responder threshold values and are described in Section 2.8.3.

2.8.3 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 categorisation is applied for all relevant time points and domain:

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

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

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

The binary responder endpoints will be analysed as the other supportive secondary binary efficacy endpoints (see section 2.5.1.1). 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 NN9924-4281. However, clarifications, more detailed descriptions of endpoints and analyses are provided in this SAP. The changes from the protocol of NN9924-4281 are summarised below:

- The primary and secondary estimands have changed names from de-jure and de-facto to hypothetical and treatment policy, respectively.
- Adjustments to the imputation method for the analyses for the secondary estimand have been made due to sparse missing data. Instead of using the 10 imputation groups as specified in the protocol only 6 imputation groups will be used: one group of subjects regardless of randomised treatment arm who at week 26 will have discontinued treatment or have initiated rescue medication, and 5 groups of subjects defined by randomised treatment arm for subjects who will still be 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 effect endpoints (HbA1c 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, 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 $\geq$ 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.
- It has been specified which assessments will be analysed on logarithmic scale.
- The definitions of initiation of rescue medication and additional anti-diabetic medication used for the time-to-event endpoints as well as the accompanying statistical analyses have been further clarified.
- It has been specified that all safety laboratory results (except amylase and lipase) are safety assessments and not safety endpoints as written in the trial protocol.

- 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, it has been specified that SF-36v2 (acute version) will be analysed using the primary analysis of the treatment policy estimand only, and DTR-QoL will be analysed for both the treatment policy estimand and the hypothetical estimand for the CTR.
- The responder analyses for SF-36v2 (acute version) and DTR-QoL will not be included in the CTR, but in a separate PRO report.

4 References

1 Ministry of Health, Labour and Welfare. 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.

16.1.9.1 Pre-defined MedDRA search – list of preferred terms

Table of contents

Page

Trial ID: NN9924-4281
16.1.9 Documentation of statistical methods, Version 1.0, dated 19-June-2018
Overview of deleted pages

Pages	Section	Title
2-211	16.1.9.1	Pre-defined MedDRA search – list of preferred terms