Cover Page for Statistical Analysis Plan

Sponsor name:	Novo Nordisk A/S
NCT number	NCT04251156
Sponsor trial ID:	NN9536-4379
Official title of study:	Effect and safety of semaglutide 2.4 mg once-weekly on weight management in subjects with overweight or obesity
Document date*	06 October 2022

*Document date refers to the date on which the document was most recently updated.
Note: The date in the header of Page 2 is the date of compilation of the documents and not of an update to content.

16.1.9 Documentation of statistical methods

List of contents

Statistical Analysis Plan

Effect and safety of semaglutide 2.4 mg once-weekly on weight management in subjects with overweight or obesity

Substance: semaglutide

Table of contents

Page

Table of figures

Page

Table of tables

Page

Version History

This Statistical Analysis Plan (SAP) for study NN9536-4379 is based on the protocol version 6 dated 06 January 2021.

SAP Version	Date	Change	Rationale
1.0	15 November 2019	Not Applicable	Original version
2.0	06 October 2022	Additional analyses and clarifications have been implemented	The rationales for the changes are described in section 3.

List of abbreviations

AD	available but discontinued
AE	adverse event
ANCOVA	analysis of covariance
AT	available on randomised treatment
BMI	body mass index
CI	confidence interval
COA	Clinical Outcome Assessments
dBP	diastolic blood pressure
FAS	full analysis set
FFA	free fatty acids
FPG	fasting plasma glucose
HbA_{1c}	glycated haemoglobin
HDL	high density lipoprotein
IWQoL-lite for CT	Impact of Weight on Quality of Life-lite for Clinical Trials
LAO	last available observation
LDL	low-density lipoprotein
MD	missing and discontinued
MMRM	mixed model for repeated measurements
MT	missing on treatment
OR	odds ratio
PK	pharmacokinetic
PYE	patient years of exposure
PYO	patient years of observation
SAE	serious adverse event
SAP	statistical analysis plan
SAS	safety analysis set
sBP	systolic blood pressure
s.c.	subcutaneous
SD	standard deviation
SF-36	Short Form 36 v2.0 acute
T2D	type 2 diabetes
TEAE	treatment emergent adverse event
VLDL	very low-density lipoprotein
WC	waist circumference

1 Introduction

This SAP includes detailed procedure for executing the statistical analyses of the primary and secondary endpoints. Statistical analyses and a number of clarifications additional to these specified in the trial protocol are pre-specified with this SAP. Changes to the protocol-planned analyses are described in Section 3.

1.1 Objectives, Endpoints, and Estimands

1.1.1 Primary and secondary objectives

1.1.1.1 Primary objective

To compare the effect of semaglutide s.c. 2.4 mg once-weekly versus semaglutide placebo as an adjunct to a reduced-calorie diet and increased physical activity in subjects with overweight or obesity and with or without T2D on body weight

1.1.1.2 Secondary objectives

To compare the effect of semaglutide s.c. 2.4 mg once-weekly versus semaglutide placebo as an adjunct to a reduced-calorie diet and increased physical activity in subjects with overweight or obesity and with or without T2D on:

- Cardiovascular risk factors
- Clinical Outcome Assessments (COAs)
- Glucose metabolism
- Glycaemic status
- Use of medication for hypertension and dyslipidaemia
- Use of oral antidiabetic medication (apply to subjects with T2D at screening)
- Liver indices
- Treatment discontinuation

To compare the safety and tolerability of semaglutide s.c. 2.4 mg once-weekly versus semaglutide placebo as an adjunct to a reduced-calorie diet and increased physical activity in subjects with overweight or obesity and with or without T2D.

Primary estimand

The estimand will quantify the average treatment effect of semaglutide relative to semaglutide placebo after 44 weeks, as an adjunct to a reduced-calorie diet and increased physical activity, in all randomised subjects regardless of adherence to treatment and regardless of starting rescue interventions (weight management drugs or bariatric surgery) (“treatment policy” estimand). The estimand will cover all effect-related objectives.

Secondary estimand

The estimand will quantify the average treatment effect of semaglutide relative to semaglutide placebo after 44 weeks, as an adjunct to a reduced-calorie diet and increased physical activity, in all

randomised subjects had they remained on their randomised treatment for the entire planned duration of the trial and not started any rescue intervention (weight management drugs or bariatric surgery) ("hypothetical" estimand). The estimand will cover all effect-related objectives.

Primary and secondary endpoints

1.1.1.3 Primary endpoint

The primary endpoints addressing the primary objective:

- Change from baseline at week 0 to week 44 in body weight (%)
- Subjects who after 44 weeks achieve (yes/no):
  - Body weight reduction ≥ 5% from baseline at week 0

1.1.1.4 Secondary endpoints

The confirmatory and supportive secondary endpoints addressing the primary and secondary objectives are listed in Section 1.1.2.2.1 and 1.1.2.2.2.

1.1.1.4.1 Confirmatory secondary endpoints

- Subjects who after 44 weeks achieve (yes/no):
  - Body weight reduction ≥ 10% from baseline at week 0
  - Body weight reduction ≥ 15% from baseline at week 0
- Change from baseline at week 0 to week 44 in:
  - Waist circumference (cm)
  - Systolic blood pressure (mmHg)
  - Physical functioning score (SF-36)
  - Physical function domain (5-items) score (IWQoL-Lite for CT)

1.1.1.4.2 Supportive secondary endpoints

Effect endpoints
The supportive secondary endpoints are used to compare effect of semaglutide s.c. 2.4 mg once weekly versus semaglutide placebo.

- Change from baseline at week 0 to week 44 in:
  - Body weight (kg)
  - BMI (kg/m2)
  - HbA_{1c} (%, mmol /mol)
  - Fasting plasma glucose (FPG) (mg/dL)
  - Fasting serum insulin (mIU/L)
  - Diastolic blood pressure (mmHg)
  - Lipids (mg/dL)
    - Total cholesterol
    - High density lipoprotein (HDL) cholesterol
    - Low density lipoprotein (LDL) cholesterol
    - Very low density lipoprotein (VLDL) cholesterol
    - Free fatty acids (FFA)
    - Triglycerides

- Short Form-36 (SF-36)
  - role-physical score
  - bodily pain score
  - general health score
  - vitality score
  - social functioning score
  - role-emotional score
  - mental health score
  - physical component summary
  - mental component summary
- o IWQoL-Lite for CT
  - pain/discomfort domain score
  - psychosocial domain score
  - total score

- Subjects who after 44 weeks achieve (yes/no):
  - Body weight reduction ≥ 20% from baseline at week 0
  - Responder definition value for SF-36 physical functioning score
  - Responder definition value for IWQoL-Lite for CT physical function domain (5 items) score
  -
- Change from baseline at week 0 to week 44 in:
  - Glycaemic category (normo-glycaemia, pre-diabetes, T2D) (only apply to subjects with no T2D at baseline)
  - Antihypertensive medication (decrease, no change, increase)
  - Lipid-lowering medication (decrease, no change, increase)
  - Concomitant oral antidiabetic medication (decrease, no change, increase) (only apply to subjects with T2D at week 0)
  - Fatty liver index (FLI) score category (<30, ≥30 and <60, ≥60)
- Subjects who from randomisation at week 0 to week 44 have permanently discontinued randomised trial product (yes/no)
- Time to permanent discontinuation of randomised trial product (weeks)

The following supportive secondary endpoints are used for subjects with T2D at baseline:

- Subjects who after 44 weeks achieve (yes/no):
  - $HbA_{1c} < 7.0\%$ (53 mmol/mol)
  - $HbA_{1c} \leq 6.5\%$ (48 mmol/mol)

Safety endpoints

- Number of treatment emergent adverse events (TEAEs) from week 0 to week 51
- Number of serious adverse events (SAEs) from week 0 to week 51
- Number of treatment emergent severe or blood glucose (BG) confirmed symptomatic hypoglycaemia episodes (yes/no) from week 0 to week 51 (only applies to subjects with T2D at week 0)
- Change from baseline at week 0 to week 44 in:

- Pulse (bpm)
- Amylase (U/L)
- Lipase (U/L)
- Calcitonin (ng/L)

1.2 Study Design

This is a 44-week, randomised, double-blind, placebo-controlled, two-armed, parallel group, multicentre, multiregional clinical trial (MRCT) comparing semaglutide s.c. 2.4 mg once-weekly with semaglutide placebo in subjects with obesity or with overweight and weight-related comorbidities.

The trial includes a screening visit to assess the subject's eligibility followed by visits/phone contacts every 2nd week during dose escalation. From week 20, visits/phone contacts will take place every 4th week for the remaining maintenance period until end of treatment (week 44). A follow-up visit ('End of trial') for safety assessments is scheduled 7 weeks after end of treatment, to account for the exposure to the long half-life of semaglutide.

The trial design is outlined in Figure 1.

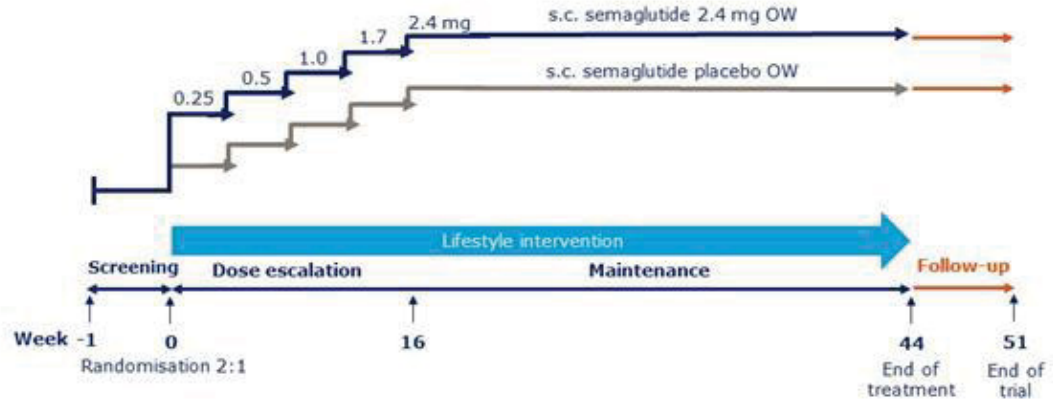

Figure 1 A schematic diagram of the trial design, with the duration of the trial periods including follow-up period

Eligible subjects fulfilling all randomisation criteria at visit 2 will be randomised in a 2:1 manner to receive either semaglutide s.c. 2.4 mg once-weekly or semaglutide placebo as an adjunct to a reduced-calorie diet and increased physical activity.

At the time of randomisation subjects will be stratified according to the following categories:

- Subjects without T2D or

- Subjects with T2D

To mitigate SU or glinide induced hypoglycaemia, subjects treated with SU or glinide (either alone or in combination with other OADs) will be asked to reduce the SU or glinide dose by approximately 50% at the discretion of the investigator, from randomisation.

2 Statistical Analyses

Taxonomy of week 44 assessments

For each subject a given assessment at week 44 may be available or missing and Table 1 describes the taxonomy for this. Note, this is done per assessment and per subject; subjects may be a different type for different assessments (a subject may have "available on randomised treatment (AT)" for body weight but "missing on randomised treatment (MT)" for waist circumference).

Table 1 Taxonomy for subjects based on week 44 assessments

Assessment at week 44	Subjects on randomised treatment at week 44	Type description	Type Abbreviation
Available	Yes	Available on randomised treatment: Subjects who complete the trial on randomised treatment with an assessment at week 44: Includes those that stop and restart trial product.	AT
	No	Available but discontinued Subjects who discontinued randomised treatment prematurely but returned to have an assessment at week 44. These are also called retrieved subjects	AD
Missing	Yes	Missing on randomised treatment: Subjects who complete the trial on randomised treatment without an assessment at week 44: Includes those that stop and restart trial product.	MT
	No	Missing and discontinued: Subjects who discontinued randomised treatment prematurely and did not return to have an assessment at week 44. These are also called non-retrieved subjects	MD

2.1 Sample size determination

The tests of superiority of semaglutide 2.4 mg to semaglutide placebo for the primary and confirmatory secondary endpoints are performed using the fixed-sequence statistical strategy. This strategy tests the endpoints using a predefined hierarchical order, all at the significance level of 5%, moving to test the next endpoint only after a statistically significant superiority result (p-value < 5%) on the previous endpoint. The test hierarchy is given in Table 2 with underlying assumptions, marginal power and effective power. The effective power is calculated under the assumption of independence of endpoints by multiplying the respective marginal powers successively. As the two primary endpoints are included in the statistical testing hierarchy, significant superiority of semaglutide 2.4 mg vs. semaglutide placebo must be demonstrated for each of the primary endpoints.

In the analysis approach addressing the primary estimand, week 44 assessments from retrieved subjects (AD) are used. These data are also used to impute missing measurements at week 44 for non-retrieved subjects (MD). The imputation is done separately within each treatment arm (see description below). However, for the power calculations missing values (MT and MD), regardless of treatment arm, are assumed to be similar to semaglutide placebo subjects. These assumptions are likely conservative with respect to the power, and correspond to the jump to reference sensitivity analysis planned below.

Assumptions

The common assumptions for the power calculations are

- The significance level is 5%
- The randomisation ratio is 2:1
- For continuous endpoints the t-test on the mean difference assuming equal variances is used
- For binary endpoints the Pearson chi-square test for two independent proportions is used
- 70% of subjects are assumed to be treatment completers, i.e. on randomised treatment (AT) at week 44
- Thus 30% of subjects discontinue trial product permanently and
    - ≥ 50% of these are retrieved (AD) at week 44
- 30% of the subjects have T2D
- All subjects in the semaglutide placebo arm are assumed to have same effect as subjects who complete the trial on semaglutide placebo (AT)
- Retrieved subjects (AD) in the semaglutide arm are assumed to have an effect corresponding to half the treatment difference (compared to semaglutide placebo) of subjects who complete the trial on semaglutide (AT)
- Non-retrieved subjects (MD) in the semaglutide arm are assumed to have an effect corresponding to semaglutide placebo arm
- 80% of the total trial population are Chinese subjects

Further assumptions made to calculate the power for each of the primary and confirmatory secondary endpoints are based on findings from other projects conducted by Novo Nordisk (NN8022 (SCALE), NN9535 (SUSTAIN), NN9924 (PIONEER)), and trial NN9536-4153 and are presented in Table 2.Furthermore, the sample size for this trial was determined to reach approximately 95% effective power (corresponding to approximately 90% effective power for the Chinese subpopulation) for the body weight related endpoints in the statistical hierarchy, i.e. body weight change, 5%, 10% and 15% responders and waist circumference change .

Given these assumptions, the sample size of 375 subjects (250 in the s.c. semaglutide 2.4 mg once weekly and 125 in the semaglutide placebo arm), gives an effective power (marginal powers multiplied) of 25% for the last endpoint in the hierarchical testing procedure, i.e. IWQoL-Lite physical function domain score change.

Table 2 Assumptions, marginal power and effective power for each endpoint in the hierarchical testing procedure given an anticipated number of 375 randomised subjects

Order	Endpoint	Assumed mean or proportion for completers [non-T2D][T2D] (±SD)		Expected mean (±SD) or proportion		Expected difference or proportion ratio	Marginal power (%)	Effective power (%)
		Semaglutide 2.4 mg	Semaglutide placebo	Semaglutide 2.4 mg	Semaglutide placebo			
1	% weight change #	[12.6] [10.4] (±10)	[3] [1.7] (±10)	9.8 (±11)	2.6 (±11)	7.2%-points	> 99	99
2	5% responders	[78%] [71%]	[42%] [37%]	68%	41%	1.7	> 99	99
3	10% responders	[60%] [52%]	[24%] [20%]	50%	23%	2.2	> 99	99
4	15% responders	[41%] [32%]	[12%] [9%]	31%	11%	2.8	99	99

5	WC change (cm) #	[9.9] [8.2] (±10)	[4] [2.8] (±10)	8.1 (±11)	3.6 (±11)	4.5 cm	96	95
6	sBP change (mmHg) #	[5.6] [4.6] (±13)	[1.5] [0.4] (±13)	4.4 (±14)	1.2 (±14)	3.2 mmHg	55	52
7	SF-36 PF score change	[5.4] [5.4] (±10)	[2] [2] (±10)	4.6 (±11)	2 (±11)	2.6 score-points	59	31
	IWQoL-Lite PFD score change	[21.6] [21.6] (±20)	[13] [13] (±20)	19.7 (±21)	13 (±21)	6.7 score-points	82	25

SD = standard deviation; WC = waist circumference; sBP = systolic blood pressure; SF-36 = Short Form 36 v2.0 acute; PF = physical functioning; IWQoL-Lite = Impact of Weight on Quality of Life-Lite for Clinical Trials; PFD = physical function domain; # shown as a positive number
All tests in the hierarchy are based on the primary estimand

2.2 Definition of analysis sets

Two analysis sets are defined:

- The *full analysis set* (FAS) includes all randomised subjects according to the intention-to-treat principle.
- The *safety analysis set* (SAS) includes all randomised subjects exposed to at least one dose of randomised treatment.

Any observation excluded from the analysis will be documented before database lock with the reason for exclusion provided.

Two observation periods are defined for each subject:

- In-trial: The *in-trial period* is defined as the uninterrupted time interval from date of randomisation to date of last contact with trial site.
- On-treatment (with trial product): A time-point is considered as 'on-treatment' if any dose of trial product has been administered within the prior 2 weeks (14 days). The *on-treatment period* is defined as all times which are considered on-treatment.
    - In general, the on-treatment period will therefore be from the date of first trial product administration to date of last trial product administration (+14 days) excluding potential off-treatment time intervals triggered by at least two consecutive missed doses.
    - The on-treatment period as described above (i.e. employing a lag time of 2 weeks [14 days]) applies to all effect assessments, safety laboratory assessments, physical examination, pulse and ECG.
    - For the evaluation of adverse events, hypoglycaemic episodes, adjudicated events, eye examination and antibodies the lag time for each on-treatment time interval is 7 weeks (49 days).

The in-trial and on-treatment periods define the patient years of observation (PYO) and patient years of exposure (PYE), respectively, as the total time duration in the periods.

2.3 Statistical analyses

Effect endpoints will be analysed using the FAS; safety endpoints will be analysed using the SAS.

Results from statistical analyses will generally be accompanied by two-sided 95% confidence intervals and corresponding p-values. Superiority will be claimed if p-values are less than 5% and the estimated treatment contrasts favours semaglutide 2.4 mg.

Handling of missing baseline data

The last available and eligible observation at or before randomisation is used as the baseline value. If no assessments are available, the mean value at randomisation across all subjects is used as the baseline value.

2.3.1 Primary Endpoints Analyses

Definition of primary endpoint: % weight change

Change from baseline (week 0) to week 44 in body weight (%) is defined as

$$\%\ \text{weight change} = \frac{(\text{body weight at week 44} - \text{body weight at baseline})}{\text{body weight at baseline}} \times 100.$$

Definition of primary endpoint: 5% responders

A body weight reduction of at least 5% from baseline (week 0) to week 44 is defined as

$$5\%\ \text{responder} = \begin{cases} 1 \text{ if } \%\ \text{weight change} \leq -5\% \\ 0 \text{ if } \%\ \text{weight change} > -5\% \end{cases}.$$

Analyses addressing the primary estimand

The following statistical analyses and imputation methods are designed to address the primary estimand, i.e. to assess the effectiveness of semaglutide 2.4 mg.

The analysis model for % weight change is a linear regression (ANCOVA) of % weight change with randomised treatment and stratification groups as factors and baseline body weight (kg) as covariate. The stratification group is defined by T2D status stratification category. The estimated treatment difference between semaglutide 2.4 mg and semaglutide placebo will be reported together with the associated two-sided 95% confidence interval (CI) and corresponding p-value.

The analysis model for the 5% responder endpoint is a logistic regression using randomised treatment and stratification groups as factors and baseline body weight (kg) as covariate. The stratification group is defined by T2D status stratification category. The estimated odds ratio (OR) between semaglutide 2.4 mg and semaglutide placebo will be reported together with the associated two-sided 95% CI and corresponding p-value.

The superiority tests of semaglutide 2.4 mg vs. semaglutide placebo will be carried out as follows for the two analysis models.

Let $\mu_{\text{semaglutide}}$ and $\mu_{\text{semaglutide placebo}}$ denote the true mean of % weight change for semaglutide 2.4 mg and semaglutide placebo group, respectively. The null and alternative hypotheses tested are

$$H: \mu_{semaglutide} \geq \mu_{semaglutide\ placebo}\ vs$$
$$H_A: \mu_{semaglutide} < \mu_{semaglutide\ placebo}.$$

The hypothesis will be rejected and superiority claimed, if the upper limit of the estimated two-sided 95% CI is below 0.

Let $OR_{semaglutide/semaglutide\ placebo}$ denote the true odds ratio between semaglutide 2.4 mg and semaglutide placebo. The null and alternative hypotheses tested are

$$H: OR_{semaglutide/semaglutide\ placebo} \leq 1\ vs$$
$$H_A: OR_{semaglutide/semaglutide\ placebo} > 1.$$

The hypothesis will be rejected and superiority claimed, if the lower limit of the estimated two-sided 95% CI is above 1.

Handling of missing week 44 values for the primary estimand

All available data at week 44 (AT and AD) are used and missing values (MT and MD) at week 44 will be imputed and the endpoints will be derived from the imputed values. Several approaches for imputation will be applied. First, a description of the primary imputation approach to address the primary estimand for the primary endpoints is given followed by a description of the sensitivity analyses used to assess the robustness of the primary analysis results. The sensitivity analyses investigate how assumptions on body weight development after discontinuation of randomised treatment impact the estimated treatment contrasts between semaglutide 2.4 mg and semaglutide placebo. An illustration of all imputation approaches for the primary estimand is given in Figure 2.

Primary imputation approach for the primary estimand

Multiple imputation approach using retrieved subjects (RD-MI): The primary imputation approach for the primary estimand is a multiple imputation similar to the one described by McEvoy[1]. Missing body weight measurement at week 44 for non-retrieved subjects (MD) are imputed using assessments from retrieved subjects (AD) in each randomised treatment arm. This will be done according to the timing of last available observation on-treatment (LAO-OT) of body weight prior to week 44. Missing body weight measurements at week 44 for subjects on randomised treatment (MT) are imputed in a similar way by sampling from available measurements at week 44 from subjects on randomised treatment (AT) in the relevant randomised treatment arm. The multiple imputation approach is done in three steps:

1. **Imputation:** Defines an imputation model using retrieved subjects (AD) from FAS and done within groups defined by randomised treatment and the timing of the LAO-OT of body weight. The model will be a linear regression of body weight (kg) at week 44 with gender (male/female), baseline BMI (kg/m^2) (in categories <35, 35-<40, ≥40), timing of the LAO-OT of body weight and stratification groups (defined by stratification categories for T2D status) as factors and baseline body weight (kg) and LAO-OT of body weight (kg) as covariates. No interactions will be included. The grouping of timing will be done by quarters (intervals of 11 weeks). If timing by quarters is too restrictive, halves (intervals of 22 weeks) or excluding timing will be used. The timing by quarters or halves is defined as too restrictive if the imputation model cannot be fit due to inadequate numbers of retrieved subjects in one or more groups. If the imputation model still cannot be fit after excluding timing then the model will be

further reduced until the model can be fit. Reduction will be done in a fixed order by first removing gender, then collapsing the two highest baseline BMI groups into one (≥35), then removing stratification factor (T2D status), then removing baseline BMI group and finally removing baseline body weight (kg). If the imputation model with only LAO-OT of body weight (kg) cannot be fit, the imputation will be done regardless of the randomised treatment arm. If no LAO-OT exists post-baseline then LAO-OT will be the baseline body weight and the timing will be the first interval. If any subjects are MT, an imputation model for missing body weight measurements at week 44 for MT subjects will also be defined using AT subjects in a similar way. The estimated posterior distribution for the parameters (regression coefficients and variances) in the imputation models are then used to impute missing week 44 body weight values for each randomised treatment arm. This will be done 1,000 times and results in 1,000 complete data sets.

2. **Analysis:** Analysis of each of the 1,000 complete data sets, using the analysis models (ANCOVA and logistic regression) results in 1,000 times 2 estimations.
3. **Pooling**: Integrates the 1,000 times 2 estimation results into two final results using Rubin's formula.

Based on NN9536-4153 phase 2 results 1,000 copies should be sufficient to establish stable results. If 1,000 copies are insufficient, 10,000 copies will be used. The multiple imputations will be generated using Novo Nordisk trial number 95364379 as seed number.

Sensitivity analyses

Jump to reference multiple imputation approach (J2R-MI): Missing values of body weight at week 44 (MT and MD) for both the semaglutide 2.4 mg and semaglutide placebo group are imputed by sampling among all available assessments at week 44 in the semaglutide placebo group (AT and AD). This approach makes the assumption that subjects instantly after discontinuation lose any effect of randomised treatment beyond what can be expected from semaglutide placebo treatment as adjunct to reduced-calorie diet and increased physical activity[2]. The multiple imputation approach is done as above with the first step replaced by

1. **Imputation**: Defines an imputation model using semaglutide placebo subjects from FAS with a week 44 measurement (AT and AD). The model will be a linear regression of body weight (kg) at week 44 with gender (male/female), BMI (kg/m^2) (in categories <35, 35-<40, ≥40) and stratification groups (defined by stratification categories for T2D status) as factors and baseline body weight (kg) as covariate. No interactions will be included. If the imputation model cannot be fit due to inadequate numbers of retrieved subjects in one or more groups, then the imputation model will be reduced until the model can be fit. Reduction will be done in a fixed order by first removing gender, then collapsing the two highest baseline BMI groups into one (≥35), then removing stratification factor (T2D status) and finally removing baseline BMI group. The estimated posterior distribution for the parameters (regression coefficients and variances) in the imputation models are then used to impute missing week 44 body weight values for each randomised treatment arm. This will be done 1,000 times and results in 1,000 complete data sets.

The jump to reference approach is the basis for the sample size calculations.

A single imputation approach as done by Sacks[3] (S1-SI and S2-SI): Missing weight measurements at week 44 for non-retrieved subjects (MD) are imputed using a weight regain rate of 0.3 kg/month after LAO but truncated at no change from baseline whenever the extrapolation would lead to a positive weight gain relative to baseline. If a subject's weight at drug discontinuation represented a gain in weight relative to baseline, no additional gain will be imputed, and the unfavourable gain is carried forward to week 44. The weight regain imputation will be done for both randomised arms (S1-SI). Additionally, a version where only the semaglutide 2.4 mg arm uses the regain rate while the semaglutide placebo arm uses last available observation (corresponding to a weight regain rate of 0 kg/month) will be performed (S2-SI). For both versions, missing weight measurements at week 44 for subjects on randomised treatment (MT) are imputed by using LAO.

Tipping-point multiple imputation analysis (TP-MI): First, missing data are imputed according to the primary multiple imputation approach. Second, for both treatment arms a penalty will be added to the imputed values at week 44. The approach is to gradually increase this penalty until all confirmed conclusions from the primary analysis are reversed. For each hypothesis tested the specific value of the penalty that reverses the conclusion will be used to evaluate the robustness of the primary analysis results. This sensitivity analysis evaluates the robustness of the superiority conclusions.

Mixed model for repeated measurements (MMRM): This 'MMRM for effectiveness' will use all assessments regardless of adherence to randomised treatment, including assessments at week 44 for retrieved drop-outs (AD). The MMRM for effectiveness will be fitted using the same factors and covariate as for the primary analyses all nested within visit. An unstructured covariance matrix for measurements within the same subject will be employed, assuming that measurements for different subjects are independent. For the 5% responder analysis, the same MMRM will be applied except that body weight (kg) will be used as response variable in the model. Individual missing values for body weight at week 44 will be predicted from the MMRM and used to classify each subject as 5% responder or not. This classification will then be analysed using the same logistic regression model as in the primary analysis of the primary estimand.

Non-retrieved subjects as non-responders: For the 5% responder analysis an analysis using subjects with missing week 44 assessment as non-responders in the logistic regressions will be done.

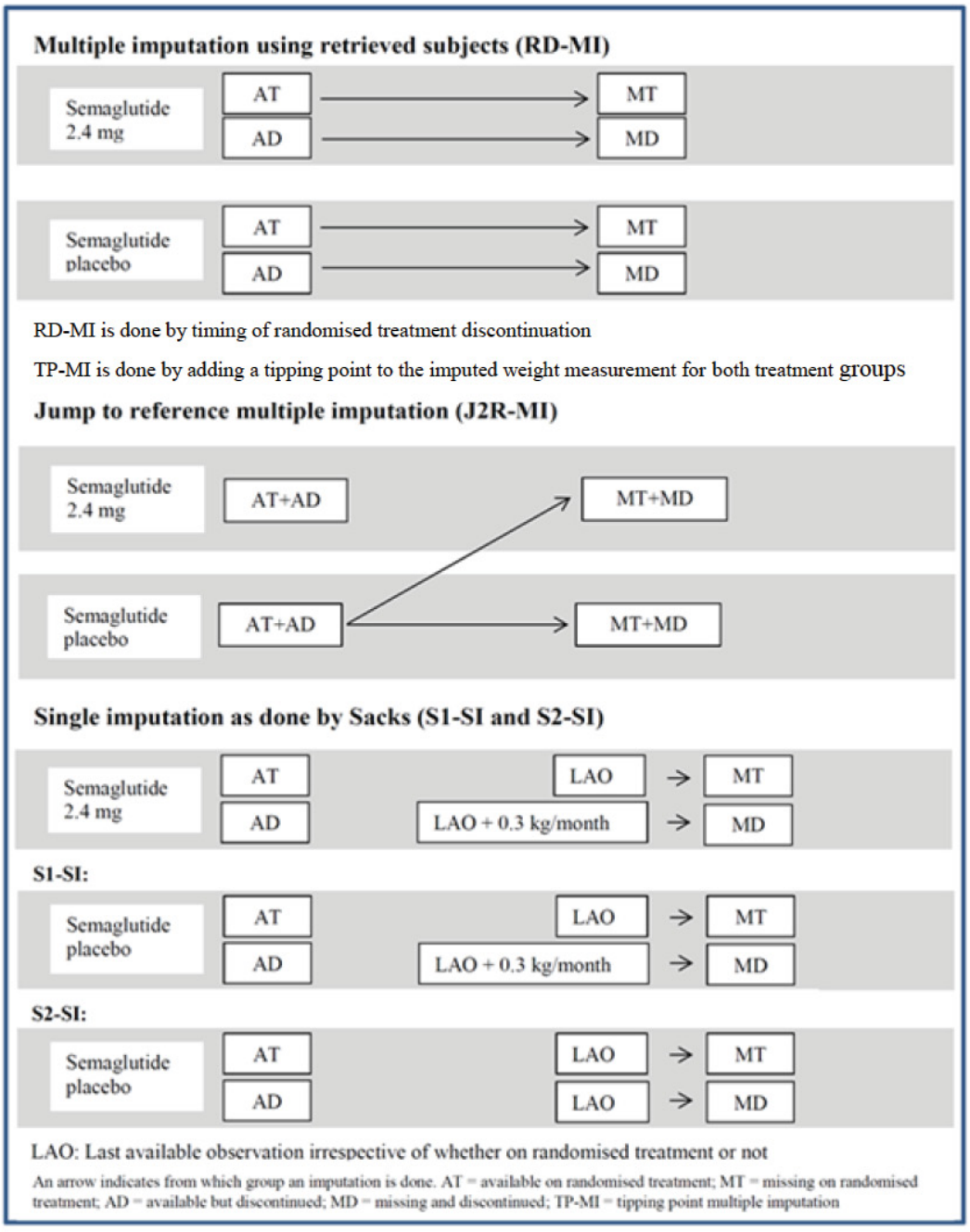

Figure 2 Illustration of imputation approaches for the primary estimand

Analysis addressing the secondary estimand

The secondary estimand for % weight change addresses the efficacy of semaglutide 2.4 mg and will be assessed using a 'MMRM for efficacy'. Week 44 assessments for retrieved drop-outs (AD) are not used in this analysis. The MMRM for efficacy will use assessments only from subjects who are taking the randomised treatment until end of treatment or until first discontinuing of randomised treatment. The derived date of the second consecutive missed dose will be used as the latest date for

using assessments in this MMRM. The assessment closest in time and before the derived date of the second consecutive missed dose will be used as last assessment on randomised treatment. For subjects who initiate anti-obesity therapies before completion or first discontinuing of randomised treatment, the date of starting weight management drugs or undergoing bariatric surgery will be used as latest date for using assessments in this MMRM. Similarly, the assessment closest in time and before the date of starting weight management drugs or undergoing bariatric surgery will be used as last assessment on randomised treatment. The MMRM for efficacy will be fitted using % weight change and the same factors and covariate as for the primary analyses all nested within visit. An unstructured covariance matrix for measurements within the same subject will be employed, assuming that measurements for different subjects are independent.

The secondary estimand for 5% responders will be assessed using the same MMRM for efficacy. For subjects with missing body weight at week 44, individual values for body weight will be predicted from the MMRM and used to classify each subject as 5% responder or not. This classification will then be analysed using the same logistic regression model as in the primary analysis of the primary estimand.

Analyses addressing both estimands

For all analyses of responder endpoints including both primary and secondary endpoints (≥5%, ≥10%, ≥15%, ≥20% body weight reduction, SF-36 physical functioning score and IWQoL-Lite for CT physical function domain (5 items) score responder, and HbA_{1c} responder ($HbA_{1c} < 7.0\%$ and $HbA_{1c} \leq 6.5\%$ for T2D subjects)), in addition to the estimated OR, the estimated treatment differences (ETDs) will be provided by calculating the responder probabilities and treatment differences between responder probabilities based on the logistic regression model, with confidence intervals for treatment differences obtained using the delta method.

An overview of all analysis and imputation methods to address the primary and secondary estimands for the primary endpoints is given in Table 3.

2.3.2 Secondary Endpoints Analyses

2.3.2.1 Confirmatory Secondary Endpoints

Confirmatory secondary endpoints are listed in section 1.1.2.2.1 and are all included in the fixed-sequence statistical strategy, see above. All tests are tests of superiority of semaglutide 2.4 mg to semaglutide placebo.

Analyses addressing the primary estimand

All confirmatory secondary endpoints will be analysed using the same imputation approach as used for the primary endpoints and to address the primary estimand. The imputation model is the same as for the primary endpoints with body weight replaced by assessments of the endpoint to be analysed. The statistical model for continuous endpoints will be ANCOVA with factors and covariate as for the primary endpoint % weight change with baseline body weight replaced by the baseline assessment of the endpoint to be analysed. The statistical model for body weight responder endpoints will be logistic regression with factors and covariate as for the primary endpoint 5% responders.

Analyses addressing the secondary estimand

The confirmatory secondary endpoints which relate to the primary objective will be analysed to address the secondary estimand using the same MMRM for efficacy described for the primary endpoints.

Sensitivity analyses for confirmatory secondary endpoints
For all continuous confirmatory secondary endpoints a sensitivity analysis using jump to reference as imputation approach will be carried out. For all binary confirmatory secondary endpoints a sensitivity analysis using non-retrieved subjects as non-responders will be carried out.

An overview of all analysis and imputation methods to address the primary and secondary estimands for confirmatory secondary endpoints is given in Table 3.

Table 3 Analysis and imputation methods to address the primary and secondary estimands for the primary and confirmatory secondary endpoints in the statistical testing hierarchy

Objective	Endpoint	Test order	Endpoint type	Estimand	Analysis set	Statistical model	Imputation approach	Sensitivity analyses
Primary endpoints								
Primary	% weight change	1	Continuous	Primary	FAS	ANCOVA	RD-MI	J2R-MI S1-SI S2-SI TP-MI MMRM
				Secondary	FAS	MMRM	-	-
Primary	5% responders	2	Binary	Primary	FAS	LR	RD-MI	J2R-MI S1-SI S2-SI TP-MI MMRM Non-responder
				Secondary	FAS	LR	MMRM	-
Confirmatory secondary endpoints								
Primary	10% responders	3	Binary	Primary	FAS	LR	RD-MI	Non-responders
				Secondary	FAS	LR	MMRM	-
Primary	15% responders	4	Binary	Primary	FAS	LR	RD-MI	Non-responders
				Secondary	FAS	LR	MMRM	-
Primary	WC change (cm)	5	Continuous	Primary	FAS	ANCOVA	RD-MI	J2R-MI
				Secondary	FAS	MMRM	-	-
Secondary	sBP change (mmHg)	6	Continuous	Primary	FAS	ANCOVA	RD-MI	J2R-MI
				Secondary	FAS	MMRM	-	-
Secondary	SF-36 PF score change	7	Continuous	Primary	FAS	ANCOVA	RD-MI	J2R-MI
				Secondary	FAS	MMRM	-	-
Secondary	IWQoL-Lite PFD score change	8	Continuous	Primary	FAS	ANCOVA	RD-MI	J2R-MI
				Secondary	FAS	MMRM	-	-

FAS = full analysis set; ANCOVA = analysis of covariance; RD-MI = multiple imputation using retrieved subjects; J2R-MI = jump to reference multiple imputation; S1-SI and S2-SI = single imputation as done by Sacks; TP-MI = tipping point multiple imputation; MMRM = mixed model for repeated measurements; LR = logistic regression; WC = waist circumference; sBP = systolic blood pressure; SF-36 = Short Form 36 v2.0 acute; PF = physical functioning; IWQoL-Lite = Impact of Weight on Quality of Life-Lite for Clinical Trials; PFD = physical function domain

2.3.2.2 Supportive Secondary Endpoints

Supportive secondary endpoints are listed in Section 1.1.2.2.2. All tests are tests of superiority of semaglutide 2.4 mg to semaglutide placebo.

Analyses addressing the primary estimand

The effect-related supportive secondary endpoints will be analysed using the same imputation approach as used for the primary endpoints and to address the primary estimand. The imputation model is the same as for the primary endpoints with body weight replaced by assessments of the endpoint to be analysed. The statistical model for continuous endpoints will be ANCOVA with factors and covariate as for the primary endpoint % weight change with baseline body weight replaced by the baseline assessment of the endpoint to be analysed.

The statistical model for HbA_{1c} responder endpoints for subjects with T2D at baseline will be logistic regression with randomised treatment as factor and the baseline HbA_{1c} as covariate. Missing HbA_{1c} data at week 44 will be imputed using J2R due to the low number of expected retrieved subjects in this subgroup.

The statistical model for responder endpoints relating to COAs will be logistic regression with randomised treatment and stratification groups (defined by stratification categories for T2D status) as factors and the baseline assessment of the endpoint to be analysed as covariate.

For fasting serum insulin, as planned measurement visits differ by design depending on T2D status at baseline, so that for subjects without T2D at baseline, the LAO-OT value always equals to the baseline value. To avoid collinearity, LAO-OT will be removed from the imputation model.

For lipids and fasting serum insulin a multiplicative model will be used, i.e. the ratio between post randomisation measurements and baseline will be calculated instead of differences, and both the dependent variable and covariate will be log-transformed.

For glycaemic status, antihypertensive medication, lipid-lowering medication, concomitant oral antidiabetic medication, fatty liver index score category and treatment discontinuation no analysis will be performed. Observed data will be summarised by descriptive statistics.

Analyses addressing the secondary estimand

The supportive secondary endpoints which relate to the primary objective will be analysed to address the secondary estimand using the same MMRM for efficacy described for the primary endpoints. The effect-related supportive secondary endpoints related to the secondary objective will also be analysed to address the secondary estimand using the same MMRM for efficacy described for the primary endpoints. For change in HbA_{1c}, FPG and fasting serum insulin, whose planned

measurement visits differ depending on T2D status at baseline, only the visits where the measurements are planned for all subjects will be included in the MMRM.

HbA_{1c} responders will be analysed on the subjects with T2D at baseline only using a logistic regression model with randomised treatment as factor and the baseline HbA_{1c} as covariate. For subjects with missing HbA_{1c} at week 44, individual values will be predicted from the MMRM and used to classify each subjects as responder or not. The MMRM will include all scheduled visit data from the T2D subjects as response, randomised treatment as a fixed factor, and baseline HbA_{1c} as covariate.

Sensitivity analyses for supportive secondary endpoints

For supportive secondary endpoints no sensitivity analysis will be carried out.

Analysis of safety endpoints

The safety endpoint pulse will be analysed using an MMRM for efficacy as described in Section 2.3.1. For amylase, lipase and calcitonin descriptive statistics will be provided. The analysis of calcitonin will be stratified by gender.

Adverse events will be defined as "treatment-emergent" (TEAE), if the onset of the event occurs in the on-treatment period (see definition in Section 2.2). TEAEs and SAEs will be summarised by descriptive statistics, such as frequencies and rates. No formal statistical inference will be carried out based on the number of TEAEs and SAEs.

An overview of all analysis and imputation methods to address the primary and secondary estimands for supportive secondary endpoints is given in Table 4.

Table 4 Analysis and imputation methods to address the primary and secondary estimands for supportive secondary endpoints

Objective	Endpoint	Endpoint type	Estimand	Analysis set	Statistical model	Imputation approach	Sensitivity analyses
Supportive secondary endpoints (effect related)							
Primary	Weight change (kg)	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Primary	20% responders	Binary	Primary	FAS	LR	RD-MI	-
			Secondary	FAS	LR	MMRM	
Primary	BMI change (kg/m^2)	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	HbA_{1c} change (%, mmol/mol)	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	FPG change (mg/dL, mmol/L)	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	
Secondary	Fasting serum insulin change (mIU/L)	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	dBP change (mmHg)	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	
Secondary	Total cholesterol change (mg/dL, mmol/L)	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	HDL change (mg/dL, mmol/L)	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-

Secondary	LDL change (mg/dL, mmol/L)	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	
Secondary	VLDL change (mg/dL, mmol/L)	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	
Secondary	FFA change (mg/dL, mmol/L)	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	Triglycerides (mg/dL, mmol/L)	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	SF-36 PF score responders[#]	Binary	Primary	FAS	LR	RD-MI	-
			Secondary	FAS	LR	MMRM	
Secondary	SF-36 RP score change	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	SF-36 BP score change	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	SF-36 GH score change	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	SF-36 VT score change	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	SF-36 SF score change	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	SF-36 RE score change	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	SF-36 MH score change	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	SF-36 PCS score change	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	SF-36 MCS score change	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	IWQoL-Lite PFD score responders[##]	Binary	Primary	FAS	LR	RD-MI	-
			Secondary	FAS	LR	MMRM	-
Secondary	IWQoL-Lite PDD score change	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	IWQoL-Lite PSD score change	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	IWQoL-Lite total score change	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	$HbA_{1c} < 7.0\%$ responders	Binary	Primary	FAS (T2D)*1	LR	J2R-MI	
			Secondary	FAS (T2D)*1	LR	MMRM	
Secondary	$HbA_{1c} \leq 6.5\%$ responders	Binary	Primary	FAS (T2D)*1	LR	J2R-MI	
			Secondary	FAS (T2D)*1	LR	MMRM	
Secondary	Glycaemic category change	Categorical	Primary	FAS	Descriptive statistics	-	-
Secondary	Antihypertensive medication change	Categorical	Primary	FAS	Descriptive statistics	-	-
Secondary	Lipid-lowering medication change	Categorical	Primary	FAS	Descriptive statistics	-	-
Secondary	Concomitant OAD change	Categorical	Primary	FAS	Descriptive statistics	-	-
Secondary	FLI score category change	Categorical	Primary	FAS	Descriptive statistics	-	-

Secondary	Subjects who have permanently discontinued randomised trial product	Binary	Primary	FAS	Descriptive statistics	-	-
Secondary	Time to permanent discontinuation of randomised trial product	Continuous	Primary	FAS	Descriptive statistics	-	-
Supportive secondary endpoints (safety related)							
Secondary	Number of TEAEs	Continuous	-	SAS	-	-	-
Secondary	Number of SAEs	Continuous	-	SAS	-	-	-
Secondary	Number of hypoglycaemia episodes	Continuous	-	SAS	-	-	-
Secondary	Pulse change (bpm)	Continuous	-	SAS	MMRM	-	-
Secondary	Amylase change (U/L)	Continuous	-	SAS	Descriptive statistics	-	-
Secondary	Lipase change (U/L)	Continuous	-	SAS	Descriptive statistics	-	-
Secondary	Calcitonin change (ng/L)	Continuous	-	SAS	Descriptive statistics	-	-

FAS = full analysis set; ANCOVA = analysis of covariance; RD-MI = multiple imputation using retrieved subjects; MMRM = mixed model for repeated measurements; J2R-MI = jump to reference multiple imputation; BMI = body mass index; HbA_{1c} = Glycated Haemoglobin A1c; FPG = fasting plasma glucose; dBP = diastolic blood pressure; HDL = high density lipoprotein; LDL = low density lipoprotein; VLDL = very low density lipoprotein; FFA = free fatty acids; LR = logistic regression; SF-36 = Short Form 36 v2.0 acute; PF= Physical Functioning; RP = Role-Physical; BP = Bodily Pain; GH = General Health; VT = Vitality; SF = Social Functioning; RE = Role-Emotional; MH = Mental Health; PCS = Physical component summary; MCS = Mental component summary; IWQoL-Lite = Impact of Weight on Quality of Life-Lite for Clinical Trials; PFD = physical function domain; PDD = pain/discomfort domain; PSD = psychosocial domain; TEAEs = treatment emergent adverse events; SAEs = serious adverse events; [#] responder value = 3.7; [##] responder value = 14.6; *1 based on subjects with T2D at screening;

2.3.3 Exploratory Endpoints Analyses

Not applicable for this trial.

2.3.4 Other Analyses

All collected data that were not defined as endpoints will be summarised by descriptive statistics.

2.4 Pharmacokinetic and/or pharmacodynamic modelling

Population PK and exposure-response analyses will be used as supportive evidence for the evaluation of efficacy and safety and further to support the recommended dose of semaglutide in subjects with obesity. First, plasma semaglutide concentrations will be analysed using a population pharmacokinetic model, quantifying covariate (such as baseline body weight, age, gender, race, ethnicity, and injection site) effects on semaglutide exposure. Second, model based estimates of steady-state average concentrations will be derived for each subject, in order to facilitate subsequent exposure-response analyses. Relevant efficacy and safety endpoints will be related to steady-state average concentrations and subjected to model based analysis.

The analyses will be conducted separately for each trial and be combined into a meta-analysis, including the phase 2 trial and phase 3a trials with PK sampling. A modelling analysis plan will be prepared before first database lock in the semaglutide phase 3a programme for weight management,

outlining details of the analyses. The modelling will be performed by Quantitative Clinical Pharmacometrics at Novo Nordisk and will be reported separately from the clinical trial reports.

2.5 Subgroup Analysis

All the analyses for the overall population will be repeated for the China subgroup population including Taiwan. The imputation for the missing data of the China subgroup analysis will be based on the overall population.

2.6 Interim Analysis

Not applicable for this trial.

3 Changes to Protocol-planned Analyses

The main analyses were described in the protocol for the trial NN9536-4379. However, clarifications and more detailed descriptions of endpoints and analyses are provided in this SAP. The changes from the protocol of NN9536-4379 are summarised below.

- It has been added that the secondary estimand will cover all effect-related objectives in order to align with the phase 3a programme.

- Body weight responder endpoints of losing ≥20%, HbA_{1c} < 7.0% (53 mmol/mol) and HbA_{1c} ≤ 6.5% (48 mmol/mol) (only apply to subjects with T2D at baseline) have been added to the list of supportive secondary endpoints in order to align with the phase 3a programme.

- In the text describing that "In general, the on-treatment period will therefore be from the date of first trial product administration to date of last trial product administration" the following has been added "(+14 days)" to emphasize that the lag-time after last trial product administration is included in the on-treatment period.

- The definition of the on-treatment period has been clarified generally and for each assessment.

- It is clarified that RD-MI imputation is performed with timing of LAO-OT added in the model as a factor for all endpoints. This is to clarify that the grouping of subjects according to timing is as in the publication by McEvoy.[1] Furthermore, if the imputation model cannot be fit, it has been described how to reduce the imputation model.

- It is clarified that if no post-baseline LAO-OT exist, then the LAO-OT will be the baseline value and the timing of LAO-OT will be the first interval.

- The baseline BMI-grouping "27-<35" in RD-MI and J2R-MI has been changed to "-<35" to accommodate the fact that subjects may lose weight between the screening and the randomisation visit.

- In the J2R-MI imputation model, the model reduction steps have been pre-specified in case the model cannot run.

- The text has been updated so that the supportive secondary endpoint "glycaemic category" will be applied to only subjects with no T2D at baseline.

- It is clarified that TP-MI imputation is performed with penalties applied to both treatment groups and Figure 2 has therefore also been updated.

- It is clarified that, in addition to OR, ETD will be reported for logistic regression analyses.

- A description has been included in the sensitivity analysis of the 5% responder endpoint (primary estimand) using MMRM to clarify precisely how the MMRM will be parameterized, and how missing values will be imputed.

- It has been clarified that the non-responder analysis includes subjects with missing body weight assessment at week 44 as non-responders.

- It has been clarified that the 5% responder analysis using MMRM for the secondary estimand is predicting individual values for % weight change only when % weight change is missing at week 44. Furthermore, it is clarified that the logistic regression includes the same factors and covariate as for the analysis of the primary estimand.

- For analyses of HbA_{1c} responder endpoints, the imputation approach has been changed to J2R because the low number of retrieved subjects is expected in this subpopulation (subjects with T2D at baseline).

- It has been updated that the responder definition value is 3.7 for SF-36 physical functioning score and 14.6 for IWQoLLite for CT physical function domain (5 items) score.

- It has been clarified that for change in HbA_{1c}, FPG and fasting serum insulin, whose planned measurement visits differ depending on T2D status at baseline, only the visits where the measurements are planned for all subjects will be included in the MMRM.

- It has been clarified that for the primary estimand analysis of fasting serum insulin, LAO-OT will be removed from the imputation model to avoid collinearity.

4 References

1. McEvoy BW. Missing data in clinical trials for weight management. J Biopharm Stat. 2016;26(1):30-6.
2. Carpenter JR, Roger JH, Kenward MG. Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. J Biopharm Stat. 2013;23(6):1352-71.
3. Sacks FM, Bray GA, Carey VJ, Smith SR, Ryan DH, Anton SD, et al. Comparison of weight-loss diets with different compositions of fat, protein, and carbohydrates. N Engl J Med. 2009;360(9):859-73.