Cover Page for Statistical Analysis Plan

Sponsor name:	Novo Nordisk A/S
NCT number	NCT04102189
Sponsor trial ID:	NN9536-4451
Official title of study:	Effect and safety of semaglutide 2.4 mg once weekly on weight management in adolescents with overweight or obesity
Document date*	16 December 2021

*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

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

Statistical Analysis Plan

Protocol title: Effect and safety of semaglutide 2.4 mg once weekly on weight management in adolescents with overweight or obesity

Substance: semaglutide

Table of contents

List of abbreviations

AD	available but discontinued
ADA	American Diabetes Association
AE	adverse event
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
AT	available on randomised treatment
BMI	body mass index
CDC	Centers for Disease Control and Prevention
COA	clinical outcome assessment
CI	confidence interval
C-SSRS	Columbia Suicidality Severity Rating Scale
ECG	electrocardiogram
ETD	estimated treatment difference
FAS	full analysis set
FPG	fasting plasma glucose
HbA_{1c}	glycated haemoglobin
HDL	high density lipoprotein
HOMA-B	homeostasis model assessment of beta-cell function
HOMA-IR	homeostasis model assessment of insulin resistance
ISPAD	International Society for Paediatric and Adolescent Diabetes
IWQOL	Impact of Weight on Quality of Life
J2R-MI	jump to reference multiple imputation
LAO-OT	last available observation on-treatment
LDL	low-density lipoprotein
LR	logistic regression
MD	missing and discontinued
MMRM	mixed model for repeated measurements
MT	missing on randomised treatment
OR	odds ratio
PHQ-9	Patient Health Questionnaire 9
PK	pharmacokinetic
PYE	patient years of exposure
PYO	patient years of observation
RD-MI	multiple imputation using retrieved subjects
SAE	serious adverse event
SAP	statistical analysis plan
SAS	safety analysis set
s.c.	subcutaneous
SD	standard deviation
T2D	type 2 diabetes
TEAE	treatment-emergent adverse event
TP-MI	tipping-point multiple imputation
VLDL	very low-density lipoprotein
WHO	World Health Organisation

Version history

This statistical analysis plan (SAP) is based on the updated protocol for trial NN9536-4451 "Effect and safety of semaglutide 2.4 mg once weekly on weight management in adolescents with overweight or obesity", version 2.0 (06 January 2021).

SAP Version	Date	Change	Rationale
1.0		Not Applicable	Original version

1 Introduction

Changes to the protocol-planned analyses are described in section 3.

1.1 Trial information

1.1.1 Objectives

1.1.1.1 Primary objective

To compare the effect of semaglutide s.c. once-weekly versus semaglutide placebo as an adjunct to a reduced-calorie diet and increased physical activity on weight management in adolescents (ages 12 to <18 years) with overweight or obesity.

1.1.1.2 Secondary objectives

To compare the effect of semaglutide s.c. once weekly versus semaglutide placebo as an adjunct to a reduced-calorie diet and increased physical activity in adolescents (ages 12 to <18 years) with overweight or obesity on:

- Cardiovascular risk factors
- Glucose metabolism

To compare the safety and tolerability of semaglutide s.c. once weekly versus semaglutide placebo as an adjunct to a reduced-calorie diet and increased physical activity in adolescents with overweight or obesity.

1.1.1.3 Exploratory objectives

To compare the effect of semaglutide s.c. once weekly versus semaglutide placebo as an adjunct to a reduced-calorie diet and increased physical activity in adolescents (ages 12 to <18 years) with overweight or obesity on:

- Clinical Outcome Assessments (COAs)

To compare additional safety and tolerability of semaglutide s.c. once weekly versus semaglutide placebo as an adjunct to a reduced-calorie diet and increased physical activity in adolescents with overweight or obesity.

1.1.2 Estimands

1.1.2.1 Primary estimand

The estimand will quantify the average treatment difference of semaglutide relative to semaglutide placebo after 68 weeks, as an adjunct to a reduced-calorie diet and increased physical activity, in all randomised subjects regardless of adherence to treatment or initiation of rescue interventions (weight management drugs or bariatric surgery) ("effectiveness"/"treatment policy" estimand). The estimand will cover all effect-related objectives.

The following expansion of the primary estimand will cover the exploratory endpoint "Change in body mass index from baseline (week 0) to week 52 (%)". The estimand will quantify the average

treatment effect of semaglutide relative to semaglutide placebo after 52 weeks, as an adjunct to a reduced-calorie diet and increased physical activity, in all randomised subjects regardless of adherence to treatment or initiation of rescue interventions.

The following expansion of the primary estimand will cover the exploratory endpoint "Change in body mass index from baseline (week 0) to week 75 (%)". The estimand will quantify the average treatment effect of semaglutide relative to semaglutide placebo after 75 weeks, as an adjunct to a reduced-calorie diet and increased physical activity, in all randomised subjects regardless of adherence to treatment or initiation of rescue interventions.

1.1.2.2 Secondary estimand

The estimand will quantify the average treatment difference of semaglutide relative to semaglutide placebo after 68 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) ("efficacy"/"hypothetical" estimand). The estimand will cover the primary objective.

1.1.3 Endpoints

1.1.3.1 Primary endpoint

Endpoint title	Time frame	Unit
Change in BMI	From baseline (week 0) to week 68	%

1.1.3.2 Confirmatory secondary endpoints

Endpoint title	Time frame	Unit
Achieving ≥5% reduction of body weight	From baseline (week 0) to week 68	Count of subject

1.1.3.3 Supportive secondary endpoints

Effect endpoints:

Endpoint title	Time frame	Unit
Change in body weight	From baseline (week 0) to week 68	kg
Change in body weight	From baseline (week 0) to week 68	%
Change in BMI percentage of the 95th percentile on gender and age-specific growth charts (CDC.gov)[1]	From baseline (week 0) to week 68	%
Change in BMI (standard deviation score) (WHO.int)[2]	From baseline (week 0) to week 68	Score points
Change in waist circumference	From baseline (week 0) to week 68	cm
Change in systolic blood pressure	From baseline (week 0) to week 68	mmHg
Change in diastolic blood pressure	From baseline (week 0) to week 68	mmHg
Change in HbA_{1c}	From baseline (week 0) to week 68	% point, mmol/mol
Achieving ≥5% reduction of BMI	From baseline (week 0) to week 68	Count of subject

Endpoint title	Time frame	Unit
Change in fasting plasma glucose	From baseline (week 0) to week 68	mmol/L, mg/dL
Change in fasting insulin	From baseline (week 0) to week 68	pmol/L, mU/L
Change in total cholesterol	From baseline (week 0) to week 68	mmol/L, mg/dL
Change in high density lipoprotein (HDL) cholesterol	From baseline (week 0) to week 68	mmol/L, mg/dL
Change in low density lipoprotein (LDL) cholesterol	From baseline (week 0) to week 68	mmol/L, mg/dL
Change in very low density lipoprotein (VLDL) cholesterol	From baseline (week 0) to week 68	mmol/L, mg/dL
Change in triglycerides	From baseline (week 0) to week 68	mmol/L, mg/dL

Safety endpoints:

Endpoint title	Time frame	Unit
Treatment-emergent adverse events (TEAEs)	From baseline (week 0) to week 75	Count of subject
Treatment-emergent serious adverse events (SAEs)	From baseline (week 0) to week 75	Count of subject
Change in pulse	From baseline (week 0) to week 68	bpm
Change in amylase	From baseline (week 0) to week 68	U/L
Change in lipase	From baseline (week 0) to week 68	U/L
Change in calcitonin	From baseline (week 0) to week 68	ng/L
Change in ALT	From baseline (week 0) to week 68	U/L

1.1.3.4 Exploratory endpoints

Effect endpoints:

Endpoint title	Time frame	Unit
Change in BMI	From baseline (week 0) to week 68	kg/m^2
Change in HOMA-B	From baseline (week 0) to week 68	%
Change in HOMA-IR	From baseline (week 0) to week 68	Score
Change in IWQOL-Kids physical comfort	From baseline (week 0) to week 68	Score points
Change in IWQOL-Kids body esteem	From baseline (week 0) to week 68	Score points
Change in IWQOL-Kids social life	From baseline (week 0) to week 68	Score points
Change in IWQOL-Kids family-relations	From baseline (week 0) to week 68	Score points
Change in IWQOL-Kids total	From baseline (week 0) to week 68	Score points
Achieving ≥10% reduction of body weight	From baseline (week 0) to week 68	Count of subject
Change in BMI	From baseline (week 0) to week 52	%
Change in BMI	From baseline (week 0) to week 75	%

Safety endpoints for adolescents with T2D:

Endpoint title	Time frame	Unit
Treatment-emergent severe or blood glucose confirmed symptomatic hypoglycaemia episodes	From baseline (week 0) to week 75	Count of subject
Treatment-emergent hypoglycaemic episodes according to ADA/ISPAD classification (both the 2014[3] and the 2018[4] definition)	From baseline (week 0) to week 75	Count of subject
Treatment-emergent hypoglycaemic episodes according to Novo Nordisk classification	From baseline (week 0) to week 75	Count of subject

Safety assessments:

Endpoint title	Time frame	Unit
Occurrence of anti-semaglutide antibodies	From baseline (week 0) to week 68	Count of subject
Occurrence of anti-semaglutide antibodies	From baseline (week 0) to week 75	Count of subject
Change in bone age assessment, x-ray	From baseline (week 0) to week 68	years
Change in electrocardiogram (ECG)	From baseline (week 0) to week 68	Count of subject
Change in laboratory parameters	From baseline (week 0) to week 68	
Change in pubertal status (Tanner staging)	From baseline (week 0) to week 68	Count of subject
Change in height (standard deviation score) (WHO.int)	From baseline (week 0) to week 68	Score points
Change in mental health assessed by Columbia Suicidality Severity Rating Scale (C-SSRS)	From baseline (week 0) to week 68	Score points
Change in Patient Reported Health Questionnaire 9 (PHQ-9)	From baseline (week 0) to week 68	Score points
Change in ophthalmological evaluation (for subjects with T2D)	From baseline (week 0) to week 68	Count of subject

Pharmacokinetic analysis

The pharmacokinetic analysis will be done as mentioned in the protocol section 10.4.

1.1.4 Type of trial

This is a 68-week, randomised, two-arm, double-blinded, multi-centre clinical trial comparing semaglutide s.c. 2.4 mg once-weekly with semaglutide placebo in adolescents with overweight or obesity.

The trial design is provided in the protocol, section 5.

1.2 Scope of the statistical analysis plan

This statistical analysis plan (SAP) is based on the updated protocol for trial NN9536-4451 "Effect and safety of semaglutide 2.4 mg once weekly on weight management in adolescents with overweight or obesity", version 2.0 (06 January 2021) and includes more detailed procedures for executing the statistical analyses.

2 Statistical considerations

2.1 Statistical hypotheses

The tests of superiority of s.c. semaglutide 2.4 mg to semaglutide placebo for the primary and confirmatory secondary endpoints are performed using a fixed-sequence statistical strategy.

This strategy tests the endpoints using a predefined hierarchical order; first the primary endpoint change in BMI (%) from baseline (week 0) to week 68 is tested at a significance level of 5%. If superiority of s.c. semaglutide 2.4 mg to semaglutide placebo is not confirmed, then the testing will stop. If superiority is confirmed (two-sided p-value <5%), the test of the confirmatory secondary endpoint subjects achieving ≥5% reduction of body weight will be performed.

2.2 Sample size determination

The trial is designed with an effective power of 90% and 72% to detect differences on the primary endpoint and confirmatory secondary endpoint, respectively. The effective power was calculated under the assumption of independence of endpoints by multiplying the respective marginal powers successively which is a conservative approach.

The assumptions for these calculations are conservatively based on findings from the s.c. semaglutide 0.1 mg and 0.2 mg once daily arms in trial NN9536-4153 and are presented in Table 1.

In the analysis approach addressing the primary estimand, week 68 assessments from retrieved subjects are included. These data are also used to impute missing measurements at week 68 for non-retrieved subjects. The imputation is done separately within each treatment arm. However, for the sample size calculations, missing values 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 described below.

Table 1 Assumptions, marginal power and effective power for each endpoint in the hierarchical testing procedure given an anticipated number of 192 randomised subjects

Order	Endpoint	Assumed mean (±SD) or proportion for completers		Expected* mean (±SD) or proportion	Expected* difference or proportion ratio	Marginal power (%)	Effective power (%)
		Semaglutide 2.4 mg	Semaglutide placebo				
1	Change in BMI (%)	10.5 (±10)	3 (±10)	8.5 (±11)	5.5%-points	90	90
2	5% BW responders	71%	42%	63%	1.5	80	72

*Accounting for reduced treatment effect among non-completers

All above outlined sample size and power considerations are for the primary estimand for the primary endpoint or the confirmatory secondary endpoints (treatment policy strategy). It is assumed that up to 35% of subjects discontinue permanently and 50% of these are retrieved at week 68. Any superiority conclusions will be based on the primary estimand.

Further assumptions for the sample size calculations are that the significance level is 5%, and that subjects are randomised in a 2:1 ratio to semaglutide or semaglutide placebo. The power calculations for continuous endpoints are based on a t-test on the mean difference assuming equal variances whereas for the categorical endpoint it is based on the Pearson chi-square test for two independent proportions.

All subjects in the semaglutide placebo arm are assumed to have same effect as subjects who complete the trial on semaglutide placebo. Retrieved subjects 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. In the semaglutide arm, non-retrieved subjects and subjects with data missing on randomised treatment are assumed to have an effect corresponding to the semaglutide placebo arm.

2.3 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 database 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 excluding potential off-treatment time intervals triggered by at least two consecutive missed doses
  - For the evaluation of AEs and hypoglycaemic episodes 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.

2.4 Statistical analyses

2.4.1 General considerations

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.

2.4.2 Primary endpoint

Definition of primary endpoint % change in BMI.

Change from baseline to week 68 in BMI (%) is defined as:

$$\%\ \text{change in BMI} = \frac{(\text{BMI at week 68} - \text{BMI at baseline})}{\text{BMI at baseline}} \times 100.$$

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 % change in BMI is a linear regression (ANCOVA) of % change in BMI on randomised treatment and stratification group (gender *Tanner stage group) as factors, and baseline BMI (kg/m^2) as covariate. The estimated treatment difference between s.c. 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.

Let $\mu_{\text{semaglutide}}$ and $\mu_{\text{semaglutide placebo}}$ denote the true mean of % change in BMI for s.c. 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.

Taxonomy of week 68 assessments being available or missing

For each subject a given assessment at week 68 may be available or missing and Table 2 describes the taxonomy for this. Note, this is done per assessment not per subject; subjects may be a different type for different assessments (a subject may have "available on treatment (AT)" for weight but "missing on treatment (MT)" for waist circumference).

Table 2 Taxonomy for subjects based on week 68 assessments being available or missing

Assessment at week 68	On randomised treatment at week 68	Type description	Type Abbreviation
Available	Yes	**Available on randomised treatment**: Subjects who complete the trial on randomised treatment with an assessment at week 68: 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 68. 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 68: 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 68. These are also called non-retrieved subjects	MD

Handling of missing week 68 values for the primary estimand

All available data at week 68 (AT and AD) are used and missing values (MT and MD) at week 68 will be imputed. 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 BMI development after discontinuation of randomised treatment impact the estimated treatment contrasts between s.c. semaglutide 2.4 mg and semaglutide placebo. An illustration of all imputation approaches for the primary estimand is given in Figure 1.

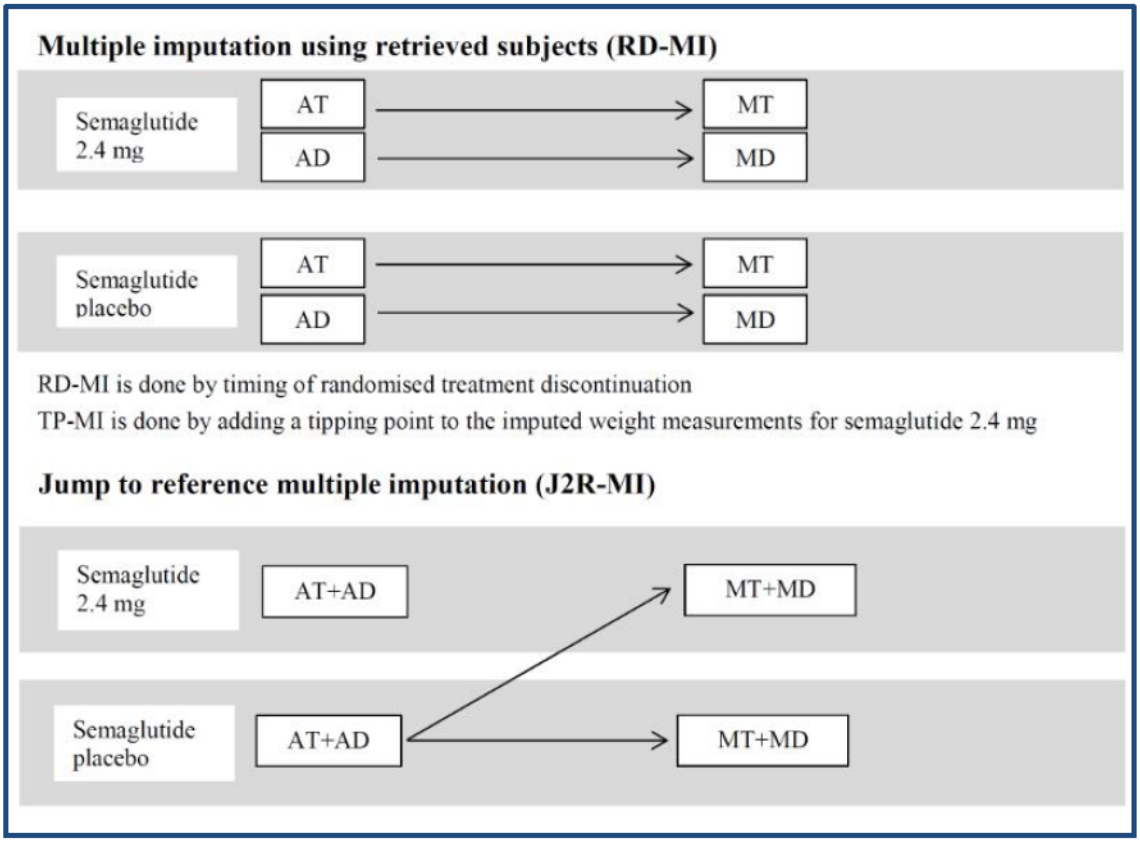

Figure 1 Illustration of imputation approaches for the effectiveness estimand

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[5].

Missing BMI measurements at week 68 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 during the on-treatment period (LAO-OT) of BMI. Missing BMI measurements at week 68 for subjects on randomised treatment (MT) are imputed by sampling from available measurements at week 68 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. The model will be a linear regression of BMI (kg/m^2) at week 68 on stratification group (gender*Tanner stage group) as a factor and baseline BMI (kg/m^2), timing of LAO-OT of BMI (kg/m^2) and LAO-OT of BMI (kg/m^2) as covariates. No interactions will be included. If any subjects are MT, an imputation model for missing BMI measurements at week 68 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 68 BMI values for each randomised treatment arm. This will be done 1000 times and result in 1000 complete data sets

2. **Analysis**: Analysis of each of the 1000 complete data sets, using the analysis model (ANCOVA), resulting in 1000 estimations
3. **Pooling**: Integrate the 1000 estimation results into a final result 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 95364451 as seed number.

Sensitivity analyses

Jump to reference multiple imputation approach (J2R-MI): Missing values of BMI at week 68 (MT and MD) for both the semaglutide 2.4 mg and semaglutide placebo group is by sampling among all available assessments at week 68 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 diet and physical activity[5]. 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 68 measurement (AT and AD). The model will be a linear regression of BMI (kg/m^2) at week 68 on stratification group (gender*Tanner stage group) as a factor and baseline BMI (kg/m^2) as covariate. No interactions will be included. The estimated posterior distribution for the parameters (regression coefficients and variances) in the imputation models are then used to impute missing week 68 BMI values for each randomised treatment arms. This will be done 1000 times and result in 1000 complete data sets.

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

Tipping-point multiple imputation analysis (TP-MI): This analysis will be performed only if superiority is concluded with respect to the primary endpoint. First, missing data are imputed according to the primary multiple imputation approach. Then, a penalty is added to the imputed values at week 68. The approach is to explore a range of penalties for both treatment groups, and the impact these would have on the study conclusions. The 2-dimensional space of penalties covering the range from -30% to 30% will be explored for both treatment groups. This sensitivity analysis evaluates the robustness of the superiority conclusions to departures from the observed change in body weight in both treatment groups.

Excluding subjects from site [redacted]*:* In Belgium site [redacted] where six subjects were randomised it was discovered after all six subjects had been randomised that the stadiometer was not calibrated. A sensitivity analysis of the primary endpoint will be performed in which these six subjects are excluded.

Mixed model for repeated measurements (MMRM): This 'MMRM for effectiveness' will use all assessments regardless of adherence to randomised treatment, including assessments at week 68 for retrieved drop-outs (AD). The MMRM for effectiveness will be fitted using the same factor 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.

ANCOVA assuming unequal variances: An alternative analysis model similar to the primary analysis model (ANCOVA), following the primary imputation approach (RD-MI), but assuming unequal variances instead of equal variances. The analysis model for % change in BMI is a linear regression (ANCOVA) of % change in BMI on randomised treatment and stratification group (gender *Tanner stage group) as factors, and baseline BMI (kg/m^2) as covariate. The estimated treatment difference between s.c. semaglutide 2.4 mg and semaglutide placebo will be reported together with the associated two-sided 95% CI and corresponding p-value.

Analysis addressing the secondary estimand

The secondary estimand for % change in BMI addresses the efficacy of semaglutide 2.4 mg and will be assessed using a 'MMRM for efficacy'. Week 68 assessments for retrieved subjects (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 trial or at first discontinuation of randomised treatment. The derived date of the second consecutive missed dose will be used as the latest date for assessments included 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 rescue interventions (weight management drugs or bariatric surgery) before completion or first discontinuation of randomised treatment, 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 % change in BMI and the same factor and covariate as for the primary analysis 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.

2.4.3 Secondary endpoints

2.4.3.1 Confirmatory secondary endpoint

The confirmatory secondary endpoint is weight loss ≥5% at week 68 (Yes/No) (≥5% body weight responder endpoint), and is included in the fixed-sequence statistical strategy described above.

Analysis addressing the primary estimand

The ≥5% body weight responder endpoint will be analysed using the same imputation approach as used for the primary endpoint and to address the primary estimand. The imputation model is the same as for the primary endpoint, with BMI replaced by body weight, and the resulting imputed values will then be dichotomized to derive the responder endpoint. The statistical model for the ≥5% body weight responder endpoint is a logistic regression using randomised treatment and stratification group (gender *Tanner stage group) as factors, and baseline body weight (kg) as a covariate. 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 $OR_{semaglutide/placebo}$ denote the true odds ratio between semaglutide 2.4 mg and semaglutide placebo. The null and alternative hypotheses tested are

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

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

Sensitivity analysis for the confirmatory secondary endpoint: A sensitivity analysis considering non-retrieved subjects as non-responders will be carried out.

Analyses addressing the secondary estimand

The secondary estimand for the ≥5% body weight responder endpoint will be assessed using the same MMRM as for the primary endpoint. Subjects who are missing week 68 body weight will be replaced by predicted values for % weight change at week 68 from the MMRM, and these predicted values will then be dichotomized to derive the responder endpoint. A logistic regression model with randomised treatment as a factor and baseline body weight (kg) as a covariate will then be used to analyse the 5% responder endpoint.

For all analyses of responder endpoints (≥5% body weight reduction, ≥10% body weight reduction, and ≥5% BMI reduction), 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 effectiveness and efficacy estimands for the primary and confirmatory secondary endpoint is given in Table 3.

Table 3 Analysis and imputation methods to address the effectiveness and efficacy 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 endpoint								
Primary	% change in BMI	1	Continuous	Primary	FAS	ANCOVA	RD-MI	J2R-MI TP-MI Excl. [redacted] MMRM Unequal variance
				Secondary	FAS	MMRM	-	-
Confirmatory secondary endpoint								
Primary	≥5% body weight responders	2	Binary	Primary	FAS	LR	RD-MI	Non-responders
				Secondary	FAS	LR	MMRM	-

FAS = full analysis set; ANCOVA = analysis of covariance; RD-MI = multiple imputation using retrieved subjects; J2R-MI = jump to reference multiple imputation; TP-MI = tipping point multiple imputation; Excl. [redacted] = sensitivity analysis excluding site [redacted]; MMRM = mixed model for repeated measurements; LR = logistic regression.
Test order refers to the order of the endpoint in the statistical test hierarchy outlined in Table 1.

2.4.3.2 Supportive secondary endpoints

Supportive secondary endpoints are listed in section 1.1.3.3.

Analyses addressing the primary estimand

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

The statistical model for the responder endpoint relating to BMI will be logistic regression with randomised treatment as a factor and the baseline assessment of the endpoint to be analysed as covariate.

For fasting insulin, lipids and ALT, 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.

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 endpoint.

Analysis of safety endpoints

The safety endpoints pulse and ALT will be analysed using an MMRM for efficacy as described in section 1.1.3.2. These analyses will be based on the safety analysis set. For amylase, lipase and calcitonin descriptive statistics will be provided. The analysis of calcitonin will be stratified by gender.

AEs will be defined as “treatment-emergent” (TEAE), if the onset of the event occurs in the on-treatment period (see definition in section 2.3). 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. AEs in the run-in period (prior to randomisation) will be presented in listings.

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

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

Objective	Endpoint	Endpoint type	Estimand	Analysis set	Statistical model	Imputation approach
Supportive secondary endpoints (effect related)						
Primary	Weight change (kg)	Continuous	Primary	FAS	ANCOVA	RD-MI
			Secondary	FAS	MMRM	-
Primary	Weight change (%)	Continuous	Primary	FAS	ANCOVA	RD-MI
			Secondary	FAS	MMRM	-
Primary	BMI percentage of the 95th percentile on gender and age-specific growth charts change (CDC.gov)	Continuous	Primary	FAS	ANCOVA	RD-MI
			Secondary	FAS	MMRM	-
Primary	BMI (standard deviation score) (WHO.int) change	Continuous	Primary	FAS	ANCOVA	RD-MI
			Secondary	FAS	MMRM	-
Primary	Waist circumference change (cm)	Continuous	Primary	FAS	ANCOVA	RD-MI
			Secondary	FAS	MMRM	-
Primary	Achieving ≥5% reduction of BMI	Binary	Primary	FAS	LR	RD-MI
			Secondary	FAS	LR	MMRM
Secondary	Systolic blood pressure change (mmHg)	Continuous	Primary	FAS	ANCOVA	RD-MI
Secondary	Diastolic blood pressure change (mmHg)	Continuous	Primary	FAS	ANCOVA	RD-MI
Secondary	HbA_{1c} change (%, mmol/mol)	Continuous	Primary	FAS	ANCOVA	RD-MI
Secondary	FPG change (mmol/L, mg/dL)	Continuous	Primary	FAS	ANCOVA	RD-MI
Secondary	Fasting insulin change (pmol/L, mIU/L)	Continuous	Primary	FAS	ANCOVA	RD-MI
Secondary	Total cholesterol change (mmol/L, mg/dL)	Continuous	Primary	FAS	ANCOVA	RD-MI
Secondary	HDL change (mmol/L, mg/dL)	Continuous	Primary	FAS	ANCOVA	RD-MI
Secondary	LDL change (mmol/L, mg/dL)	Continuous	Primary	FAS	ANCOVA	RD-MI
Secondary	VLDL change (mmol/L, mg/dL)	Continuous	Primary	FAS	ANCOVA	RD-MI
Secondary	Triglycerides change (mmol/L, mg/dL)	Continuous	Primary	FAS	ANCOVA	RD-MI
Supportive secondary endpoints (safety related)						
Secondary	Number of TEAEs	Continuous	-	SAS	-	-
Secondary	Number of SAEs	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	-
Secondary	ALT change (U/L)	Continuous		SAS	MMRM	

FAS = full analysis set; ANCOVA = analysis of covariance; RD-MI = multiple imputation using retrieved subjects; MMRM = mixed model for repeated measurements; BMI = body mass index; HbA1c = Hemoglobin A1c; TEAEs = treatment-emergent adverse events; SAEs = serious adverse events.

2.4.4 Exploratory endpoints

Exploratory endpoints are listed in section 1.1.3.4.

Analyses addressing the primary estimand

The effect-related exploratory endpoints will be analysed using the same imputation approach as used for the primary endpoint and to address the primary estimand. The imputation model is the

same as for the primary endpoint with BMI replaced by assessments of the endpoint to be analysed. The statistical model for continuous endpoints will be ANCOVA with factor and covariate as for the primary endpoint % change in BMI with baseline BMI replaced by the baseline assessment of the endpoint to be analysed.

The statistical model for responder endpoints relating to body weight will be logistic regression with randomised treatment as a factor and the baseline assessment of the endpoint to be analysed as covariate.

For HOMA-B and HOMA-IR, 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.

Analyses addressing the secondary estimand

The exploratory endpoints measured at week 68 which relate to the primary objective will be analysed to address the secondary estimand using the same MMRM for efficacy described for the primary endpoint.

An overview of all analysis and imputation methods to address the effectiveness and efficacy estimands for effect-related exploratory endpoints is given in Table 5.

Analysis of safety endpoints

Observed data for exploratory safety endpoints and other safety assessments will be summarised by descriptive statistics.

Hypoglycaemic episodes will be summarised by descriptive statistics, such as frequencies and rates. No formal statistical inference will be carried out based on the number of hypoglycaemic episodes.

Table 5 Analysis and imputation methods to address the effectiveness and efficacy estimands for exploratory endpoints

Objective	Endpoint	Endpoint type	Estimand	Analysis set	Statistical model	Imputation approach
Exploratory endpoints (effect related)						
Primary	BMI change (kg/m^2)	Continuous	Primary	FAS	ANCOVA	RD-MI
			Secondary	FAS	MMRM	-
Primary	Achieving ≥10% reduction of body weight	Binary	Primary	FAS	LR	RD-MI
			Secondary	FAS	LR	MMRM
Primary	BMI change to week 52 (%)	Continuous	Primary	FAS	ANCOVA	RD-MI
Primary	BMI change to week 75 (%)	Continuous	Primary	FAS	ANCOVA	RD-MI
Secondary	HOMA-B change (%)	Continuous	Primary	FAS	ANCOVA	RD-MI
Secondary	HOMA-IR change (score)	Continuous	Primary	FAS	ANCOVA	RD-MI
Secondary	IWQOL-Kids physical comfort change (score)	Continuous	Primary	FAS	ANCOVA	RD-MI
Secondary	IWQOL-Kids body esteem change (score)	Continuous	Primary	FAS	ANCOVA	RD-MI
Secondary	IWQOL-Kids social life change (score)	Continuous	Primary	FAS	ANCOVA	RD-MI
Secondary	IWQOL-Kids family-relations change (score)	Continuous	Primary	FAS	ANCOVA	RD-MI
Secondary	IWQoL-Kids total change (score)	Continuous	Primary	FAS	ANCOVA	RD-MI

FAS = full analysis set; ANCOVA = analysis of covariance; RD-MI = multiple imputation using retrieved subjects; MMRM = mixed model for repeated measurements; BMI = body mass index; FPG = fasting plasma glucose; HDL = high density lipoprotein; LDL = low density lipoprotein; VLDL = very low density lipoprotein; LR = logistic regression; IWQoL-Kids = Impact of Weight on Quality of Life-Kids for Clinical Trials; PCD = physical comfort domain; BED = body esteem domain; SLD = social life domain; FRD = family relations domain.

2.4.5 Other analyses

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

2.4.5.1 Subgroup analyses

Subgroup analyses of the primary and the confirmatory secondary endpoint will be carried out addressing both the primary and the secondary estimand using the same imputation approaches and statistical models as used in the analyses of the primary and confirmatory secondary endpoints. The subgroups are defined in Table 6. The results of the subgroup analyses will be presented with treatment contrasts and treatment-by-subgroup interaction p-values in table form and forest plots.

If a treatment arm has less than 5 subjects within a subgroup, the subgroup will be merged with another, until the criteria of at least 5 subjects per treatment arm and subgroup is reached. For ordered subgroups (e.g. age, baseline BMI), the merging will be done according to the order. For unordered subgroups (e.g. race, region), it will be done by merging the subgroups with the lowest number of subjects.

The results of the subgroup analyses will be submitted to authorities and will not be reported in the clinical trial report.

Table 6 Definition of subgroups

Subgroup	Levels
Age	12 - <15 years; 15 - <18 years

Subgroup	Levels
Sex	Female; Male
Ethnicity	Hispanic or Latino; Not Hispanic or Latino
Race	White; Other
Tanner stage	2-3; 4-5
BMI	<35 kg/m^2; 35 - <40 kg/m^2; ≥40 kg/m^2
Region	Europe (Austria, Belgium, Croatia, Ireland, Russian Federation, United Kingdom); North America (USA); South America (Mexico)
Region 2	USA; Non-USA

2.5 Endpoint derivation methods

2.5.1 BMI percentage of the 95th percentile (CDC.gov)

The growth charts from the CDC consist of a series of percentile curves that illustrate the distribution of selected body measurements in children from Unites States.

The BMI-for-age reference data applies to children and adolescents (2-20 years).

In practice when using the growth reference data, then the age in months is used to look up the 95th percentile. E.g. if a subject is 180.2 months when being measured, then the 95th percentile to be used is obtained from the 180.5 month reference, since this reference covers the interval from 180 to 180.99 months. According to CDC[1] at a finer age can be obtained using linear interpolation.

Hence, in order to calculate an exact percentage above the 95th percentile for a given age in months, linear interpolation of the 95th percentile is applied between the lower 95th percentile in months and the upper 95th percentile in months.

2.5.2 BMI and height SDS (WHO.int)

The BMI standard deviation scores (SDS) and the height SDS scores will be calculated using growth reference data for children and adolescents (5-19 years) from WHO[2]. The SDS scores are sometimes referred to as z-scores also.

The z-score for a subject with a BMI y at age t is given as:

$$z = \frac{\left(\left(\frac{y}{M(t)}\right)^{L(t)} - 1\right)}{S(t)L(t)} \qquad if\ |z| \leq 3$$

where L, M and S depend on sex and age in months and can be looked up in the reference data. If the z-score is between -3 and 3 then no adjustment is needed. If the z-score is > 3 then the score is adjusted as follows:

$$z = 3 + \left(\frac{y - SD3pos}{SD23\,pos}\right) \qquad if\ z > 3$$

where *SD3pos* is the cut-off 3 SD calculated at *t* by the LMS method[2]:

$$SD3pos = M(t) * \left(1 + L(t) * S(t) * (3)\right)^{1/L(t)}$$

and *SD23pos* is the difference between the cut-offs 3 SD and 2 SD calculated at *t* by the LMS method:

$$SD23pos = M(t) * \left(1 + L(t) * S(t) * (3)\right)^{1/L(t)} - M(t) * \left(1 + L(t) * S(t) * (2)\right)^{1/L(t)}$$

If the z-score is < 3 then the score is adjusted as follows:

$$z = -3 + \left(\frac{y - SD3neg}{SD23\ neg}\right) \qquad if\ z < -3$$

where *SD3neg* is the cut-off -3 SD calculated at *t*, and *SD23neg* is the difference between the cut-offs -2 SD and -3 SD, both calculated by the LMS method:

$$SD3neg = M(t) * \left(1 + L(t) * S(t) * (-3)\right)^{1/L(t)}$$

$$SD23neg = M(t) * \left(1 + L(t) * S(t) * (-2)\right)^{1/L(t)} - M(t) * \left(1 + L(t) * S(t) * (-3)\right)^{1/L(t)}$$

In practice when using the growth reference data, then the age in whole months is used to look up the L, M and S values. E.g. if a subject is 180.2 months when being measured, then the values of L, M and S to use are done for 180 months, i.e. the age is rounded down to the nearest month.

In order use the same derivation with respect to calculating an exact *z*-score for a given age in months as for the CDC derivations in section 2.5.1, linear interpolation of the *z*-score is applied between the lower *z*-score in months and the upper *z*-score in months.

The height SDS scores are calculated in the same manner as for the BMI SDS scores.

Some subjects will turn 19 years old during the trial for which there is no reference data above the age of 19 years (228 months). In these cases, the reference data for being 19 years old will be used.

2.5.3 IWQOL-Kids

The Impact of Weight on Quality of Life (IWQOL)-Kids questionnaire contains 27 questions (also referred to as items) with five response options ranging from never true (1) to always true (5). The 27 items are categorised into four scales: physical comfort (six items); body esteem (nine items); social life (six items); and family relations (six items).

The score on each scale will calculated as an unweighted sum of that scale's constituent items and then transformed to 0 to 100 scoring, where 100 represents the best quality of life, and 0 represents the worst quality of life. At least 50% of the items within a scale should be answered to calculate the scale score. For computation of the total score, the unweighted sum of all of the items will be used, and then scores will similarly transformed to the 0 to 100 scoring. At least 75% of the 27 items should be answered to calculate the total score.

3 Changes to the statistical analyses planned in the protocol

The main analyses were described in the protocol for the trial NN9536-4451. However, clarifications and more detailed descriptions of endpoints and analyses are provided in this SAP. The changes from the protocol of NN9536-4451 are:

- Treatment-emergent hypoglycaemic episodes will also be classified according to ADA/ISPAD 2018[4] classification
- Fasting plasma glucose and lipids have been promoted from exploratory endpoints to supportive secondary endpoints in order to align with the upcoming STEP Young trial
- Based on feedback from authorities ALT has been promoted to a supportive secondary endpoint
- It is clarified that RD-MI imputation is performed with timing of last available observation during the on-treatment period (LAO-OT) added in the model as a covariate. This is true for all endpoints. This is to clarify that the grouping of subjects according to timing is as in McEvoy[5]
- It is clarified that TP-MI imputation is performed with penalties applied to both treatment arms
- Due to the stadiometer not being calibrated at site [redacted], a sensitivity analysis of the primary endpoint has been added
- A sensitivity analysis of the primary endpoint using an ANCOVA assuming unequal variances in the treatment arms has been added
- It is clarified that, in addition to OR, ETD will be reported for logistic regression analyses
- It is clarified that analyses of safety endpoints will be performed based on the safety analysis set
- It is specified that lipids, FPG and fasting insulin will also be analysed in SI-units
- Subgroup analyses of the primary and confirmatory secondary endpoint have been added due to a similar request from the authorities on the SCALE TEENS trial.

4 References

1. Centers for Disease Control and Prevention. Stature-for-age charts, 2 to 20 years, LMS parameters and selected smoothed stature percentiles in centimeters, by sex and age [Internet] 2018 [Cited August 2018].
2. World Health Organization. World Health Organization: Growth reference 5-19 years. Accessed at: http://www.who.int/growthref/who2007_bmi_for_age/en/ [updated 2 March 2018; cited 2018 2 March].
3. Ly TT, Maahs DM, Rewers A, Dunger D, Oduwole A, Jones TW, et al. ISPAD Clinical Practice Consensus Guidelines 2014. Assessment and management of hypoglycemia in children and adolescents with diabetes. Pediatr Diabetes. 2014;15 Suppl 20:180-92.
4. Mayer-Davis EJ, Kahkoska AR, Jefferies C, Dabelea D, Balde N, Gong CX, et al. ISPAD Clinical Practice Consensus Guidelines 2018: Definition, epidemiology, and classification of diabetes in children and adolescents. Pediatr Diabetes. 2018;19 Suppl 27:7-19.
5. McEvoy BW. Missing data in clinical trials for weight management. J Biopharm Stat. 2016;26(1):30-6.