Cover Page for Statistical Analysis Plan

Sponsor name:	Novo Nordisk A/S
NCT number	NCT03136484
Sponsor trial ID:	NN9535-4270
Official title of study:	Efficacy and Safety of Semaglutide Versus Canagliflozin as add-on to Metformin in Subjects With Type 2 Diabetes
Document date:	03 June 2019

16.1.9 Documentation of statistical methods

List of contents

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

.

Statistical Analysis Plan

Trial ID: NN9535-4270

SUSTAIN 8 – semaglutide versus canagliflozin

Efficacy and safety of semaglutide versus canagliflozin as add-on to metformin in subjects with type 2 diabetes

[redacted]

Table of contents Page

List of abbreviations

AACE	American Association of Clinical Endocrinologists
ADA	American Diabetes Association
AE	adverse event
ANCOVA	analysis of covariance
BG	blood glucose
BMI	body mass index
CoEQ	Control of Eating Questionnaire
DTSQ	Diabetes Treatment Satisfaction Questionnaire
DXA	DXA analysis set / dual X-ray absorptiometry
ECG	electrocardiogram
FAS	full analysis set
FPG	fasting plasma glucose
HbA_{1c}	glycosylated haemoglobin
HDL	high-density lipoprotein
LDL	low-density lipoprotein
LLOQ	lower limit of quantification
MAR	missing at random
MCMC	Markov Chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
OW	once weekly
PP	per protocol
PRO	patient reported outcome
PT	preferred term
SAS	safety analysis set
SAP	statistical analysis plan
SD	standard deviation
SE	standard errors
SF-36v2™	Short form healthy survey
SMPG	self-measured plasma glucose
SUSTAIN	Semaglutide Unabated Sustainability in Treatment of Type 2 Diabetes
T2D	type 2 diabetes
TEAE	treatment emergent adverse events

1 Introduction

1.1 Trial information

This is a 52-week, confirmatory, randomised, double-blind, double dummy, active-controlled, multicentre, multinational, two-arm, parallel-group trial.

Primary objective

To compare the effect of once-weekly (OW) dosing of subcutaneous semaglutide (1.0 mg) versus once-daily dosing of oral canagliflozin (300 mg) on glycaemic control in subjects with type 2 diabetes (T2D) on a background treatment of metformin.

Secondary objectives

To compare the effects of semaglutide s.c. 1.0 mg once-weekly versus canagliflozin 300 mg once daily after 52 weeks of treatment in subjects with T2D with regards to:

- Weight management
- Other parameters of effect, safety and Patient Reported Outcomes

See the protocol for trial NN9535-4270 for further details.

1.2 Scope of the statistical analysis plan

This SAP is based on the protocol "SUSTAIN 8 – semaglutide versus canagliflozin, Efficacy and safety of semaglutide versus canagliflozin as add-on to metformin in subjects with type 2 diabetes", version 3.0.

2 Statistical considerations

2.1 General considerations

No interim analyses or other analyses of un-blinded or between group data will be performed before the database is locked.

Laboratory values below the lower limit of quantification (LLOQ) will be set to ½LLOQ.

Results from a statistical analysis will be presented by the estimated treatment contrasts at week 52 with associated two-sided 95% confidence intervals and p-values corresponding to two-sided tests of no difference if not otherwise specified.

The comparison presented from a statistical analysis will be semaglutide 1.0 mg versus canagliflozin 300 mg

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

Data from all trial sites will be analysed and reported together.

The regions used in the statistical analyses are defined as:

- North America (United States and Canada)
- Region Europe (United Kingdom, Ireland and Sweden)
- International Operations (Lebanon, Malaysia, Argentina, Mexico, Brazil, India)

2.1.1 Data transformations

A number of the continuous parameters will be log-transformed prior to statistical analysis. The output tables and figures will show the results of the analysis back-transformed to the original scale, implying that log-treatment-differences are reported as treatment ratios. Confidence intervals for the treatment ratios will be calculated as exponentiated upper and lower limits for log-treatment difference confidence intervals. The standard errors (SE) of the back-transformed mean and ratio to baseline estimates are also provided; these SEs are calculated using the delta-method (first order Taylor approximation), whereby the SE on the original scale is calculated as the product of the SE on log-scale and the exponentiated estimate of the mean (geometric mean).

2.1.2 Definition of baseline

For each assessment, the baseline assessment is defined as the latest available measurement at or prior to the randomisation visit. This specifically implies that if a visit 2 assessment is missing (whether it was planned or not planned) then the screening assessment (from visit 1), if available, will be used as the baseline assessment.

2.1.3 Primary estimand

To further detail the trial objective an estimand is defined which is a *de-jure* (efficacy) estimand:

Primary estimand

- The treatment difference between semaglutide and canagliflozin at week 52 for all randomised subjects if all subjects completed treatment and did not initiate rescue medication

This primary *de-jure* estimand is considered clinically relevant as it assesses the glycaemic benefit a person with T2D is expected to achieve if initiating and continuing treatment with semaglutide compared to canagliflozin. Accordingly, only data collected prior to discontinuation of trial product or initiation of rescue medication will be used to draw inference. This will avoid confounding from rescue medication.

2.1.4 Trial completion

Unless subjects withdraw their informed consent, data collection will continue for the full duration of the trial. The full duration of the trial is defined as up to and including the follow-up visit (P11). Subjects completing the follow-up visit (P11) will be considered trial completers.

2.1.5 Missing data considerations at week 52

The actual rate of missing data at week 52 is expected to be maximum 10% based on the rate of trial completers from the subcutaneous semaglutide phase 3a clinical development programme. The frequency of missing data is expected to be similar in the semaglutide and the canagliflozin groups.

When estimating the primary estimand, the combined rate of missing data, subjects discontinuing treatment prematurely or initiating rescue medication on top of trial product, is expected to be maximum 30%. This is based on the results from the subcutaneous semaglutide phase 3a clinical development program. Based on these data, premature treatment discontinuation due to gastrointestinal adverse events (AEs) is expected to be low but more frequent in semaglutide compared to canagliflozin. Other reasons for discontinuing treatment are assumed to be unrelated to treatment and therefore occur with similar rates, so overall the frequency of missing data or data not used at week 52 in the primary analysis is expected to be slightly larger in semaglutide as compared to canagliflozin.

To document the extent and reason(s) for missing data, descriptive summaries and graphical representation of extent, reason(s) for and pattern of missing data will be presented by treatment group.

2.2 Sample size calculation

The primary endpoint, change from baseline to week 52 in HbA_{1c} (%-point) will be tested for non-inferiority and superiority of semaglutide vs. canagliflozin. The confirmatory secondary endpoints, change from baseline to week 52 in body weight (kg) and change from baseline to week 52 in total fat mass (kg) are planned to be tested for superiority of semaglutide vs. canagliflozin.

The sample size calculation is made to ensure a power of at least 90% for meeting HbA_{1c} superiority of semaglutide vs. canagliflozin out of the four pre-specified confirmatory hypotheses shown in Table 2-2. The closed testing procedure described in Bretz et.al. 2011[1] combined with a hierarchical approach is used to control the overall type-1 error at a two-sided 5% level. The statistical testing strategy is built on the following principle:

- Glycaemic efficacy must be established by HbA_{1c} non-inferiority before testing for added benefits in terms of superiority in terms of HbA_{1c} or body weight.
- HbA_{1c} and body weight superiority must be established before testing for added benefits in terms of superiority in terms of total fat mass.

The sample size is calculated using the calcPower function in the R package, gMCP[2] using 10,000 simulations. All of the four pre-specified confirmatory tests are assumed to be independent. Since

some of these tests are positively correlated, the assumption of independence is viewed as conservative. The four hypotheses are:

- HbA_{1c} non-inferiority of semaglutide 1.0 mg vs. canagliflozin 300 mg with a non-inferiority margin of 0.3
- HbA_{1c} superiority of semaglutide 1.0 mg vs. canagliflozin 300 mg
- Body weight superiority of semaglutide 1.0 mg vs. canagliflozin 300 mg
- Total fat mass superiority of semaglutide 1.0 mg vs. canagliflozin 300 mg

The sample size assumptions for efficacy based on on-treatment data without rescue medication, a treatment effect based on in-trial data (see Section 2.3.1) and the standard deviations (SD) are given in Table 2-1. The HbA_{1c} and body weight assumptions are based on the efficacy results and an observed reduction of approximately 20% and 15% respectively in in-trial treatment effect compare to efficacy in the subcutaneous semaglutide phase 3a clinical development programme.[4-7]

A similar reduction in the in-trial treatment effect compared to efficacy is assumed with canagliflozin as comparator. The total fat mass assumption is based on the relevant literature focusing on fat mass[8-10], which indicates a smaller SD for total fat mass as compared to body weight.

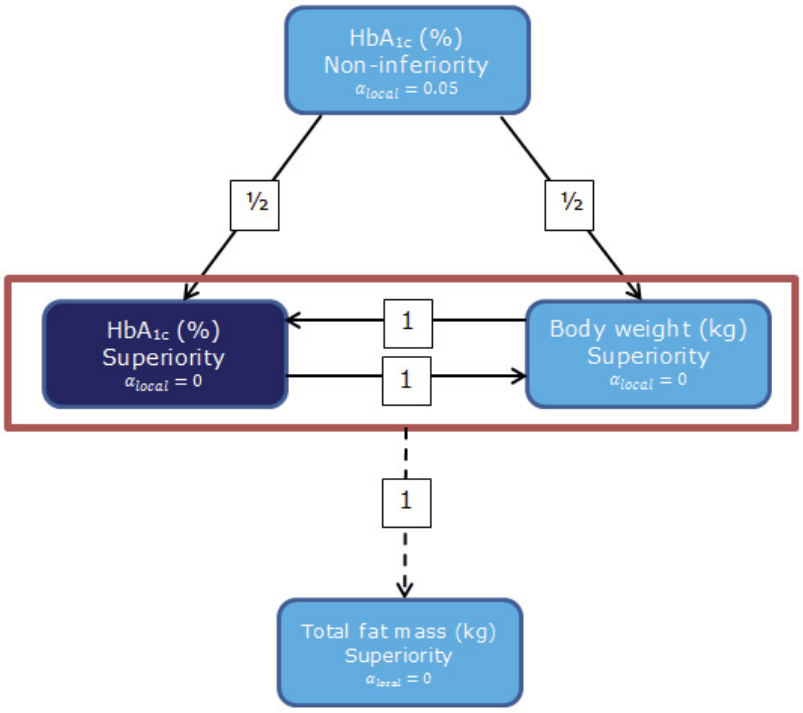
Table 2-1 Assumptions used in the sample size calculation

Semaglutide vs. canagliflozin	HbA_{1c} (%-points)	Body weight (kg)	Total fat mass (kg)
Efficacy	-0.32	-2.4	-1.8
In-trial treatment effect	-0.256	-2.04	-1.53
Standard deviation	1.1	4.0	3.5

With the above assumptions, allocating 392 subjects to the semaglutide arm and the canagliflozin arm provides 90% power to confirm HbA_{1c} superiority of semaglutide vs. canagliflozin across plausible assumptions.

Table 2-2 Calculated powers for meeting individual hypotheses

Statistical test	HbA_{1c} non-inferiority	HbA_{1c} superiority	Body weight superiority	Total fat mass superiority
Efficacy Power (%)	>99%	90%	>99%	91%
In-trial effect power (%)	>99%	90%	>99%	74%



The overall significance level of α = 0.05 (two-sided) is initially allocated to the HbA_{1c} non-inferiority test. The local significance level (α-local) will be reallocated if a hypothesis is confirmed according to the weight given by the directed edges between nodes (hypotheses). The total fat mass superiority test will receive the overall significance of α = 0.05 (two-sided) if and only if both HbA1c and body weight superiority are confirmed at their respective local significance levels.

Figure 2-1 Graphical illustration of the closed testing procedure

The overall significance level of α = 0.05 (two-sided) is initially allocated to the HbA_{1c} non-inferiority test. The local significance level (α-local) will be reallocated if a hypothesis is confirmed according to the weight given by the directed edges between nodes (hypotheses). The total fat mass superiority test will receive the overall significance of α = 0.05 (two-sided) if and only if both HbA_{1c} and body weight superiority are confirmed at their respective local significance levels.

2.2.1 Sample size for the sub-study (dual X-ray absorptiometry)

For the sub-study on body composition assuming an efficacy treatment difference of 1.8 kg and a SD of 3.5 kg, 174 subjects (87 subjects in each arm) will provide 92% power to establish a statistical significant difference resulting in 91% power for confirming superiority in the testing strategy in terms of fat mass loss (kg) at week 52 using a two-sided significance level of 5%.

2.3 Definition of analysis sets

The following analysis sets will be defined:

Full analysis set (FAS): includes all randomised subjects. Subjects in the FAS will contribute to evaluation "as randomised".

DXA analysis set: includes all subjects in FAS who are included in the DXA sub-study. Subjects in the DXA analysis set will contribute to the evaluation "as randomised".

Safety analysis set (SAS): includes all subjects exposed to at least one dose of trial product. Subjects in the SAS will contribute to the evaluation based on the trial product received for the majority of the period they were on treatment. This will be referred to as contributing to the evaluation "as treated".

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

- Have not violated any inclusion criteria
- Have not fulfilled any exclusion criteria
- Have a non-missing HbA_{1c} measurement at screening and/or randomisation
- Is on trial product at visit 8 and have at least one non-missing HbA_{1c} measurement at or after visit 8.

Subjects in the PP analysis set will contribute to the analysis "as treated" as defined for the SAS.

2.3.1 Data selections and observation periods

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

- Firstly, subjects will be selected based on the specified analysis set
- Secondly, data points on the selected subjects from first step will be selected based on the specified observation period

Definition of the observation periods

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

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

For subjects not randomised but exposed to trial product the in-trial period starts at the date of first dose of trial product

For DXA assessments the last direct subject-site contact is defined as the date of the last collected data for the subject.

On-treatment: This observation period represents the time period where subjects are considered treated with trial product. The observation period is a sub-set of the in-trial observation period. It starts at the date of first dose of trial product. Two slightly different end dates will be needed to cover all assessments appropriately according to the flow chart. For adjudicated events, electrocardiograms (ECGs) and AEs including hypoglycaemic episodes, the observation period ends at the first date of any of the following:

- the follow-up visit (P11)
- the follow-up prematurely discontinuation visit (P11A)
- the last date on trial product + 42 days
- the end-date for the in-trial observation period

The follow-up visit is scheduled to take place 5 weeks after the last date on trial product corresponding to approximately five half-lives of subcutaneous semaglutide. The visit window for the follow-up visit is + 7 days, which is the reason for the 42 days specified in the bullet above. Hence, for those assessments this period reflects the period in which subjects are exposed.

For efficacy and other safety assessments (laboratory assessments, physical examination and vital signs) the observation period ends at the last date on trial product + 7 days. This ascertainment window corresponds to the dosing interval and will be used to avoid attenuation of a potential treatment effect on endpoints for which the effect is reversible shortly after treatment discontinuation. Hence, for those assessments this period reflects the period in which subjects are treated.

On-treatment without rescue medication: This observation period is a sub-set of the on-treatment observation period, where subjects are considered treated with trial product, but have not initiated any rescue medications. Specifically it starts at date of first dose of trial product and the observation period ends at the first date of any of the following:

- the last dose of trial product +7 days
- initiation of rescue medication

The 'on-treatment without rescue medication' observation period will be the primary observation period for efficacy evaluations. The in-trial observation period will be considered supportive for efficacy evaluation. Safety will be evaluated based on the in-trial and the on-treatment observation periods unless otherwise specified.

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

Before data are locked for statistical analysis, a review of all data will take place. Any decision to exclude either a subject or single observations from the statistical analysis is the joint responsibility of the members of the Novo Nordisk study group.

Exclusion of data from analyses will be used restrictively and normally no data should be excluded from the FAS. The subjects or observations to be excluded, and the reasons for their exclusion will be documented and signed by those responsible before database lock. The subjects and observations excluded from analysis sets, and the reason for this, will be described in the clinical trial report.

2.4 Primary endpoint

The primary endpoint is change from baseline to week 52 in HbA_{1c}.

2.4.1 Primary analysis for the primary estimand

The primary estimand will be estimated based on the FAS using post-baseline measurements up to and including week 52 from the 'on-treatment without rescue medication' observation period. Imputation of missing data will be handled using multiple imputation assuming that missing data is missing at random (MAR). Missing data will be imputed using observed data within the same group defined by the randomised treatment (semaglutide/canagliflozin). It is hereby assumed that the likely values of what the missing data would have been if available are best described by information from subjects who receive the same treatment.

Technically missing values will be imputed as follows:

- Intermittent missing values are imputed using a Markov Chain Monte Carlo (MCMC) method, in order to obtain a monotone missing data pattern. This imputation is done for each of the treatment groups separately and 500 copies of the dataset will be generated
- A sequential regression approach for imputing monotone missing values at planned visits will be implemented starting with the first visit after baseline and sequentially continuing to the last planned visit at week 52. A model used to impute missing values at each planned visit will be fitted for each of the treatment groups using observed data. The model will include stratification factor (sub-study, non-sub-study) and region as categorical effects and baseline and post-baseline HbA_{1c} values observed or imputed prior to the visit in question as covariates.
- An ANCOVA with treatment, stratification factor (sub-study, non-sub-study) and region as categorical effects and baseline HbA_{1c} as a covariate will be used to analyse HbA_{1c} values at week 52 for each of the 500 complete data sets generated as part of the imputation of missing values. Rubin's rule will be used to combine the analysis results in order to draw inference.

From this analysis, the estimated treatment difference between semaglutide and canagliflozin at week 52 will be presented together with the associated two-sided 95% confidence interval and unadjusted two sided p-values for testing non-inferiority and superiority.

2.4.2 Primary hypotheses

For the primary HbA_{1c} endpoint the following confirmatory one-sided hypotheses are planned to be tested for semaglutide versus canagliflozin. Let the mean treatment difference be defined as μ = (semaglutide minus canagliflozin):

- Non-inferiority, using a non-inferiority margin of 0.3%-point
  - H0: $\mu \geq 0.3$%-point against Ha: $\mu < 0.3$%-point
- Superiority
  - H0: $\mu \geq 0.0$%-point against Ha: $\mu < 0.0$%-point

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

The non-inferiority margin of 0.3 is chosen based on the diabetes guideline[11, 12] and the effect of canagliflozin on glycaemic effect seen in a similar trial (DIA3006[3]) where canagliflozin was used as add on to metformin. In this trial canagliflozin showed an HbA_{1c} treatment difference to placebo of -0.77%-point Hence, based on this trial, the chosen margin of 0.3 provides assurance that semaglutide has an effect compared to placebo greater than 0 with a clinically relevant size. With regards to the constancy assumption, controlled clinical trials have consistently established that canagliflozin is an effective anti-diabetic drug. Therefore, lack of trial sensitivity with canagliflozin as comparator is not anticipated to be an issue in this trial.

2.4.3 Multiplicity and criteria for confirming hypotheses

The Type-I error for testing the four confirmatory hypotheses related to the HbA_{1c}, body weight, and fat mass endpoints will be preserved in the strong sense at 5% (two-sided) using the weighted Bonferroni-based closed testing procedure described in Bretz et. al.[1] and outlined in Figure 2-1. The first hypothesis to be tested is non-inferiority of HbA_{1c}. It will be tested at the overall significance level (5%) while allocating 0% local significance level to the remaining three hypotheses. For this hypothesis, and in general, if a hypothesis is confirmed the significance level will be reallocated according to the weight and the direction of the edges going from the confirmed hypothesis to the next hypotheses as specified in Figure 2-1. Total fat mass will be tested at the overall significance level if each of the other 3 hypotheses is confirmed, otherwise its local significance level will remain 0%. Each of the following hypotheses will be tested at their local significance level (α-local). This process will be repeated until no further hypotheses can be confirmed.

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

2.4.4 Statistical subgroup analyses of HbA_{1c}

Five subgroups based on baseline HbA_{1c} values are defined as follows:

1. $\leq 7.5\%$
2. $> 7.5\%$ to 8.0% (inclusive)
3. $> 8.0\%$ to 8.5% (inclusive)
4. $> 8.5\%$ to 9.0% (inclusive)
5. $> 9.0\%$

Change from baseline in HbA_{1c} at week 52 for subgroups based on baseline HbA_{1c} values will be analysed for the primary estimand using a similar multiple imputation approach as described in section 2.4.1. The complete data sets from the primary analysis will be reused. However the ANCOVA model used to analyse the 500 complete data sets will additionally include the interaction effect of subgroup and treatment as a categorical effect. Rubin's rule will then be used to combine the results and the p-value for the interaction effect and estimated treatment differences at 52 weeks with corresponding two-sided 95% confidence intervals for each subgroup will be presented.

2.4.5 Sensitivity analyses

In order to investigate the robustness of the conclusions from the primary analysis and to stress test the MAR assumption for missing data tipping point sensitivity analyses will be performed for the primary estimand both for the sensitivity of the non-inferiority and the superiority HbA_{1c} hypotheses.

2.4.5.1 Sensitivity analyses for the primary estimand

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

- Tipping-point analysis (pattern mixture model based) based on the FAS using the 'on-treatment without rescue medication' observation period. In this analysis, subjects from the semaglutide group with missing observations will be given a penalty, i.e., it is assumed that subjects with missing observations who are randomised to semaglutide will receive a treatment that is worse than subjects with observed values who are randomised to semaglutide. The idea is to gradually increase the penalty to evaluate at which level the superiority conclusion of the analyses in terms of statistical significance is changed. The tipping point is the penalty level, at which the magnitude of efficacy reduction in subjects with missing data creates a shift in the treatment effect of semaglutide from being statistically significantly better than canagliflozin to being non-statistically significantly better for the superiority test and similarly for the non-inferiority test. Technically, this analysis will be implemented by replicating the primary analysis including the assumption of MAR but subsequently adding increasing penalty values at week 52 to imputed observations in the semaglutide group before applying ANCOVA on the 500 complete data sets.

2.4.5.2 Other sensitivity analyses

The following additional sensitivity analyses are specified:

- Retrieved drop-out analysis based on the FAS using post-baseline measurements up to and including week 52 from the in-trial observation period. Missing data will be imputed using the same approach as described for the primary analysis of the primary estimand. However the imputation will be done within the same group defined not only by the randomised treatment (semaglutide/canagliflozin) but also by the status of treatment completion (still on randomised treatment at week 52 yes/no) (4 groups in total). It is hereby assumed that the likely values of what the missing data would have been if available are best described by information from subjects who at week 52 are similar in terms of randomised treatment and treatment completion status. In addition in the imputation step stratification factor and region is not included in the model in order to avoid potential issues with sparse data. This analysis could be considered addressing an effectiveness estimand. The retrieved drop-out is carried out for the superiority testing only.
- PP analysis based on the PP data set using the 'on-treatment without rescue' observation period. This analysis will be carried out for non-inferiority testing only. The statistical analysis will be the same as the primary analysis for the primary estimand.

2.5 Secondary endpoints

2.5.1 Confirmatory secondary endpoints

Change from baseline to week 52 in body weight (kg) and change from baseline to week 52 in total fat mass (kg) will be confirmatory secondary endpoints.

The primary estimand will be estimated using the same approach as described for the primary HbA_{1c} endpoint. Body weight and total fat mass will be tested for superiority. Baseline and post-baseline body weight will be used as covariates instead of HbA_{1c} for the analysis of body weight. The analysis of total fat mass will be based on the DXA analysis set, stratification factor will not be included in the model and baseline fat mass will be used as covariate instead of baseline HbA_{1c}. Since only baseline and end-of-treatment DXA scans are performed, the missing data pattern will be monotone by default. As a consequence MCMC-imputation is not needed and no post-baseline data will be included as covariates in the imputation model.Superiority will be considered confirmed if the mean treatment difference is supporting the corresponding hypothesis and the two-sided p-value from the primary analysis of the primary estimand is strictly below its local two-sided significance level resulting from the closed testing procedure in Figure 2-1.

The tipping point sensitivity analysis pre-specified to evaluate the robustness of the conclusions from the primary analysis of HbA_{1c} will also be performed to evaluate the robustness of the conclusions from the body weight and total fat mass superiority tests. The analyses will be based on FAS and the DXA analysis set respectively. In addition, the retrieved drop-out sensitivity analysis will also be performed for body weight. .For total fat mass, the data collection does not support a retrieved drop-out analysis as there are no systematic data collection at visit 10 for subjects

discontinuing treatment prematurely. Therefore, a supplementary in-trial analysis will be performed in which the imputation is done within the same group defined by randomised treatment only. The observation period for this analysis is the in-trial period. Besides this, the imputation procedure follows that of the confirmatory analysis for total fat mass, i.e. region is included in the imputation model and no MCMC imputation is performed.

2.5.2 Supportive secondary endpoints

No sensitivity analyses are planned for the supportive secondary endpoints.

2.5.2.1 Efficacy endpoints

Continuous endpoints

The continuous endpoints are change from baseline to week 52 in**:**

- Fasting plasma glucose (FPG)
- Self-measured plasma glucose (SMPG), 7-point profile:
  - Mean 7-point profile
  - Mean post prandial increment (over all meals)
- Fasting blood lipids (total cholesterol, low-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL) cholesterol, triglycerides)
- Body mass index (BMI) and waist circumference
- Systolic and diastolic blood pressure
- Body weight (%)
- Total fat mass (%-point)
- Total lean mass (kg)
- Total lean mass (%-point)
- Visceral fat mass (kg)
- Visceral fat mass (%-point)
- Ratio between total fat mass and total lean mass

The above continuous endpoints will be analysed for the primary estimand separately using a similar model approach as for the primary endpoint with the associated baseline and post-baseline responses as covariates instead of HbA_{1c} for their respective analyses. The DXA endpoints (total fat mass, total lean mass, visceral fat mass and ratio between total fat mass and total lean mass) will be analysed using a similar approach as for the confirmatory secondary endpoint, total fat mass (kg), with the associated baseline values as covariate instead of total fat mass (kg).

Fasting lipid profile endpoints will be log-transformed prior to analysis with the associated log-transformed baseline value as a covariate.

Mean 7-point profile self-measured plasma glucose definition

Subjects will be asked to perform SMPG measurements before and 90 minutes after breakfast, lunch, dinner, and at bedtime.

Mean of the 7-point profile is defined as the area under the profile, calculated using the trapezoidal method, divided by the measurement time.

Binary endpoints

The binary endpoints are subjects who after 52 weeks treatment achieve (yes/no)**:**

- HbA_{1c} <7.0% (53 mmol/mol), American Diabetes Association (ADA) target
- HbA_{1c} ≤6.5% (48 mmol/mol), American Association of Clinical Endocrinologists (AACE) target
- Weight loss ≥3%
- Weight loss ≥5%
- Weight loss ≥10%
- HbA_{1c} <7.0% (53 mmol/mol) without severe or blood glucose confirmed symptomatic hypoglycaemia episodes and no weight gain
- HbA_{1c} reduction ≥1%-point
- HbA_{1c} reduction ≥1%-point and weight loss ≥3%
- HbA_{1c} reduction ≥1%-point and weight loss ≥5%
- HbA_{1c} reduction ≥1%-point and weight loss ≥10%

The above 10 endpoints will be analysed for the primary estimand. The analyses for the primary estimand for all 10 endpoints will be based on the 'on-treatment without rescue medication' observation period. They will be analysed separately using the same type of logistic regression model with treatment, stratification factor (sub-study, non- sub-study), region and associated baseline and post-baseline response(s) (i.e. HbA_{1c} responses for HbA_{1c} endpoints, body weight responses for weight endpoints and both HbA_{1c} and body weight responses for the binary endpoints that combine both parameters) as covariates. To account for missing data, the analysis will be made using a sequential multiple imputation approach as described below:

- The binary endpoint will be derived based on the 500 complete data sets from the primary analysis of HbA1c and the confirmatory analysis of body weight.
- Each of the created complete data sets will be analysed with the logistic regression model. Estimated odds ratios will be log transformed and inference will be drawn using Rubin's rule.[13]

The results after applying Rubin's rule will be back-transformed and described by the odds ratio between treatments and the associated 95% confidence interval and p-value for no treatment difference.

2.5.2.2 Safety endpoints

The safety endpoints will be evaluated based on SAS using the on-treatment observation period and the in-trial observation period unless otherwise stated.

Adverse Events

The following endpoint related to AEs is used to support the safety objective;

- Number of treatment emergent adverse events (TEAEs)

A treatment-emergent AE is an event that has onset date (or increase in severity) during the on-treatment observation period. These will therefore be referred to as 'on-treatment AEs' hereafter. On-treatment AEs are summarised descriptively in terms of the number of subjects with at least one event (N), the percentage of subjects with at least one event (%), the number of events (E) and the event rate per 100 years (R). These summaries are replicated by outputs including all 'in-trial' AEs (i.e., AEs with onset date [or increase in severity] during the 'in-trial' observation period). AEs with onset after the end of the 'in-trial' observation period will be reported in a listing. The development over time in gastrointestinal AEs will be presented graphically.

The most frequent AEs will be defined as preferred terms (PTs) that are experienced by at least 5% of the subjects in any of the treatment arms.

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

Hypoglycaemic episodes

The following two endpoints related to hypoglycaemic episodes are used to support the safety objective:

- Number of treatment-emergent severe or blood glucose (BG)-confirmed symptomatic hypoglycaemic episodes
- Treatment-emergent severe or BG-confirmed symptomatic hypoglycaemic episodes (yes/no)

Data on treatment-emergent hypoglycaemic episodes are presented in terms of the number of subjects with at least one episode, the percentage of subjects with at least one episode (%), the total number of episodes and the episode rate per 100 years of exposure. Summaries of treatment-emergent hypoglycaemic episodes will be presented as an overview including all episodes and episodes by severity.

Classification of Hypoglycaemia:

Treatment emergent: hypoglycaemic episodes will be defined as treatment emergent if the onset is in the on-treatment period (see definition of observation period in section 2.3.1)

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

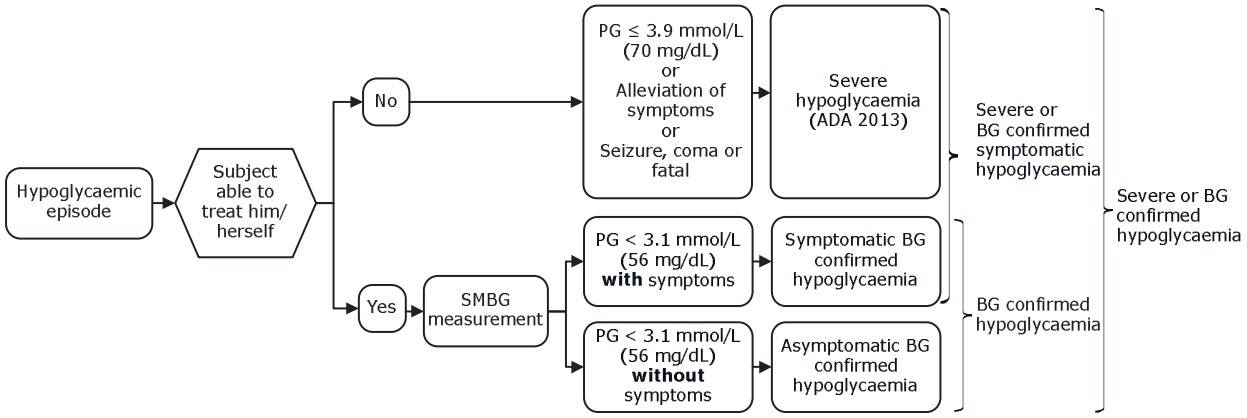
Hypoglycaemic episodes are classified according to the Novo Nordisk classification of hypoglycaemia (see Figure 2-2) and the ADA classification of hypoglycaemia (see Figure 2-3).

Novo Nordisk classification of hypoglycaemia

In normal physiology, symptoms of hypoglycaemia occur below a plasma glucose level of 3.1 mmol/L (56 mg/dL).[14] Therefore, Novo Nordisk has included hypoglycaemia with plasma glucose levels below this cut-off point in the definition of BG confirmed hypoglycaemia.

Novo Nordisk uses the following classification (see Figure 2-2) in addition to the ADA classification:

- Severe or BG confirmed symptomatic hypoglycaemia: An episode that is severe according to the ADA classification[15] or BG confirmed by a plasma glucose value <3.1 mmol/L (56 mg/dL) **with** symptoms consistent with hypoglycaemia.



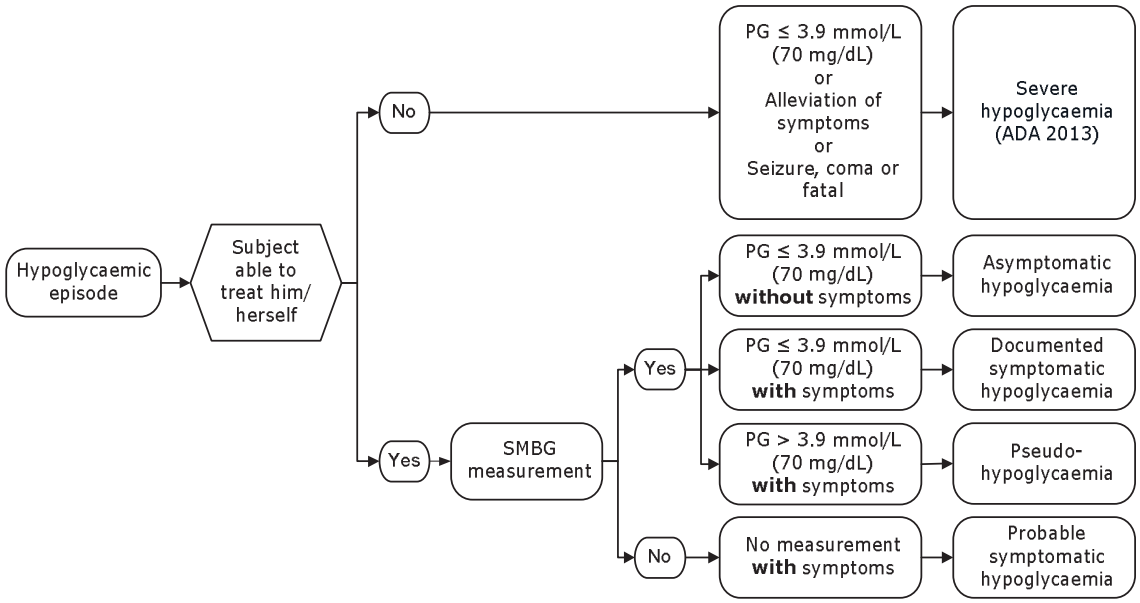
Note: Glucose measurements are performed with capillary blood calibrated to plasma equivalent glucose values

Figure 2-2 Novo Nordisk classification of hypoglycaemia

American Diabetes Association classification[15] of hypoglycaemia

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

- Probable symptomatic hypoglycaemia: An episode during which symptoms of hypoglycaemia are not accompanied by a plasma glucose determination but that was presumably caused by a plasma glucose concentration ≤3.9 mmol/L (70 mg/dL).



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

Figure 2-3 American Diabetes Association classification of hypoglycaemia

Number of treatment emergent severe or blood glucose confirmed symptomatic hypoglycaemic episodes

Number of treatment emergent severe or BG confirmed symptomatic hypoglycaemic episodes during 56 weeks will be analysed using a negative binomial regression model with a log -link function and the logarithm of the time period, from the randomisation and up to the time point in which an occurrence of a hypoglycaemic episode is considered treatment emergent as offset assuming MAR. The model will include factors for treatment and stratification factor (sub-study, non- sub-study) as categorical factors and baseline HbA_{1c} as covariate. The SAS will be used for the analysis.

The results will be described by the rate ratio between treatments and the associated 95% confidence interval and p-value for no treatment difference.

Treatment emergent severe or blood glucose confirmed symptomatic hypoglycaemia episodes (yes/no)

The binary endpoint indicating whether a subject has no treatment-emergent severe or BG confirmed symptomatic hypoglycaemic episodes or at least one will be analysed using a logistic regression model. The model will include factors for treatment and stratification factor (sub-study, non- sub-study) as categorical factors and baseline HbA_{1c} as covariate. The SAS will be used for the analysis.

The results will be described by the odds ratio between treatments and the associated 95% confidence interval and p-value for no treatment difference.

Laboratory assessments

The laboratory assessments supporting the safety objective are change from baseline to week 52 in:

- Haematology
- Biochemistry
- Calcitonin

The above continuous laboratory assessments will be summarised and evaluated by descriptive statistics.

In addition amylase and lipase will be analysed separately using an analysis similar to the primary analysis of the primary endpoint. However this analysis will be based on SAS using the on-treatment observation period.

Both analyses will use the associated log-transformed baseline and post-baseline responses as covariates instead of HbA_{1c}. Lipase and amylase values will be log-transformed prior to the analysis.

Pulse

Change from baseline to week 52 in pulse will be analysed separately with the same model approach as for amylase and lipase but with the pulse value (not log-transformed) at baseline and post-baseline as covariates instead of HbA_{1c}.

Categorical safety assessments

The categorical assessments supporting the safety objective are change from baseline to week 52 in:

- ECG category
- Physical examination
- Eye examination category

The above assessments will be summarised descriptively

2.6 Health economics and/or patient reported outcomes

Change from baseline to week 52 in scores for selected PROs:

- SF-36v2™ Short Form health survey: Total scores (physical component and mental component) and scores from the 8 domains
- Diabetes Treatment Satisfaction Questionnaire (DTSQ): Treatment satisfaction score (sum of 6 of 8 items) and the 8 items separately
- Control of Eating Questionnaire (CoEQ): Scores from the 4 domains and scores from 19 individual items

The PRO questionnaires, SF-36v2™, DTSQ and CoEQ will be used to evaluate the objective regarding Quality of Life. Each of the PRO endpoints will be analysed separately as the other continuous efficacy endpoints for the primary estimand using a similar model approach as for the primary endpoint with the associated baseline and post-baseline responses as covariates.

3 Changes to the statistical analyses planned in the protocol

The changes to the statistical analyses planned in the protocol are described in the table below.

Change to planned statistical analysis	Rationale for change
Italy removed from pre-defined region Europe to be used in the statistical analysis (Section 2.1).	Italy was not included in the study.
References updated in Section 2.2and 2.4.2.	Updated for correctness.
The word nominal removed from the sentence "The closed testing procedure described in Bretz et.al. 20111 combined with a hierarchical approach is used to control the overall type-1 error at a nominal two-sided 5% level. "	Updated for clarification and to clearly distinguish the level used for testing and the level at which the overall type I-error is controlled.
Definition of DXA analysis set added (Section 2.3)	Was not specified in the protocol.

Specification of the PP-analysis set criterion on including 'subjects on trial product at week 28 and having at least one non-missing HbA_{1c} measurement at or after week 28'. This was revised to 'subjects on trial product at visit 8 and having at least one non-missing HbA_{1c} measurement at or after visit 8 (Section 2.3) .	Revision was done to ease programming. Visit 8 corresponded to week 28 ±7 days.
Wording on the multiple imputation model for the primary analysis updated to specify that observed or imputed values will be used as covariates (Section 2.4.1)	Updated for clarification.
The 'in-trial treatment policy' sensitivity analysis is renamed to 'retrieved drop-out' analysis and it is clarified that the model will only be conducted to test the robustness of the superiority hypotheses (Section 2.4.5.2).	Per new preferred terminology, this type of analysis is no longer called an 'in-trial' analysis, but rather a retrieved drop-out analysis.
It was specified that the tipping point analyses for the confirmatory secondary endpoints are carried out on FAS and the DXA analysis set respectively.	Updated for clarification.
The following clarifications for the analyses of the confirmatory secondary endpoint, total fat mass (kg), in the DXA sub-study was added (Section 2.5.1): • For all analyses it is clarified that the analyses are based on the DXA analysis set and that stratification factor will not be included in the models. • Clarification that no MCMC-imputation will be performed • For the in-trial sensitivity analysis, imputation will be done within the same group of randomised treatment irrespective of status of treatment completion and region will be included in the imputation model. The analysis was re-categorised to a supplementary analysis. • Clarification of the in-trial period for DXA assessments	 • The DXA analysis set is the relevant population and stratum DXA/non-DXA is redundant in the analysis of DXA endpoints. • With only 1 post-baseline measurement, non-monotone missingness is not possible and MCMC-imputation is redundant. • No systematic collection of off-treatment DXA scans are done according to protocol (only premature treatment discontinuers not completing the premature end of treatment DXA scans are planned to have an off-treatment scan at the last visit). The data therefore does not support imputation by status of treatment completion. The coarser imputation approach is not expected to lead to sparse data issues, so there is no reason not to include region in the imputation model. • Re-scans for DXA can occur after the P11 follow-up visit.

It was clarified for statistical analyses of the supportive secondary body composition endpoints in the DXA sub-study that the analyses were to follow the same approach as for the confirmatory secondary endpoint, total fat mass (kg), (Section 2.5.2.1):	Data is collected in the same way and similar analysis considerations as for total fat mass (kg) apply.
Wording updated for the description of multiple imputation for binary endpoints (Section 2.5.2.1)	Updated to clarify that no new imputations are done.
Wording updated on analyses of amylase and lipase (Section 2.5.2.2)	Updated to clarify that baseline values of amylase and lipase should be log-transformed before being used as covariates in the analyses.
Wording updated on analysis on pulse rate (Section 2.5.2.2)	Updated to clarify that pulse rate should not be log-transformed.
For HbA_{1c}, total fat mass, total lean mass and visceral fat mass the unit is corrected to'%-point'(multiple places).	Updated for correctness.
Number of imputations revised from 200 to 500 (multiple places).	Revised to align with the other NN9535 phase 3b trials.

4 References

1. Bretz F, Posch M, Glimm E, Klinglmueller F, Maurer W, Rohmeyer K. Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. Biometrical Journal. 2011;53(6):894-913.
2. Rohmeyer K, Klinglmueller F. gMCP: Graph Based Multiple Test Procedures. R package version 0.8-8. 3 Oct 2014.
3. Lavalle-González FJ, Januszewicz A, Davidson J, Tong C, Qiu R, Canovatchel W, et al. Efficacy and safety of canagliflozin compared with placebo and sitagliptin in patients with type 2 diabetes on background metformin monotherapy: a randomised trial. Diabetologia. 2013;56(12):2582-92.
4. Ahrén B, Masmiquel L, Kumar H, Sargin M, Karsbøl JD, Jacobsen SH, et al. Efficacy and safety of once-weekly semaglutide versus once-daily sitagliptin as an add-on to metformin, thiazolidinediones, or both, in patients with type 2 diabetes (SUSTAIN 2): a 56-week, double-blind, phase 3a, randomised trial. Lancet Diabetes Endocrinol. 2017;5(5):341-54.
5. Sorli C, Harashima SI, Tsoukas GM, Unger J, Karsbøl JD, Hansen T, et al. Efficacy and safety of once-weekly semaglutide monotherapy versus placebo in patients with type 2 diabetes (SUSTAIN 1): a double-blind, randomised, placebo-controlled, parallel-group, multinational, multicentre phase 3a trial. Lancet Diabetes Endocrinol. 2017;5(4):251-60.
6. Ahmann AJ, Capehorn M, Charpentier G, Dotta F, Henkel E, Lingvay I, et al. Efficacy and Safety of Once-Weekly Semaglutide Versus Exenatide ER in Subjects With Type 2 Diabetes (SUSTAIN 3): A 56-Week, Open-Label, Randomized Clinical Trial. Diabetes Care. 2018;41(2):258-66.

7. Aroda VR, Bain SC, Cariou B, Piletič M, Rose L, Axelsen M, et al. Efficacy and safety of once-weekly semaglutide versus once-daily insulin glargine as add-on to metformin (with or without sulfonylureas) in insulin-naive patients with type 2 diabetes (SUSTAIN 4): a randomised, open-label, parallel-group, multicentre, multinational, phase 3a trial. Lancet Diabetes Endocrinol. 2017;5(5):355-66.
8. Bolinder J, Ljunggren O, Kullberg J, Johansson L, Wilding J, Langkilde AM, et al. Effects of dapagliflozin on body weight, total fat mass, and regional adipose tissue distribution in patients with type 2 diabetes mellitus with inadequate glycemic control on metformin. J Clin Endocrinol Metab. 2012;97(3):1020-31.
9. Bolinder J, Ljunggren O, Johansson L, Wilding J, Langkilde AM, Sjostrom CD, et al. Dapagliflozin maintains glycaemic control while reducing weight and body fat mass over 2 years in patients with type 2 diabetes mellitus inadequately controlled on metformin. Diabetes Obes Metab. 2014;16(2):159-69.
10. Cefalu WT, Leiter LA, Yoon KH, Arias P, Niskanen L, Xie J, et al. Efficacy and safety of canagliflozin versus glimepiride in patients with type 2 diabetes inadequately controlled with metformin (CANTATA-SU): 52 week results from a randomised, double-blind, phase 3 non-inferiority trial. Lancet. 2013;382(9896):941-50.
11. Administration FaD. Guidence for Industry. Diabetes Mellitus - Evaluating Cardiovascular Risk in New Antidiabetic Therapies to Treat Type 2 Diabetes. 12/1/2008 2008.
12. U.S. Department of Health and Human Services, Food and Drug Administration. Guidance for Industry. Diabetes Mellitus: Developing Drugs and Therapeutic Biologics for Treatment and Prevention, Draft Guidance. 2008.
13. Little RJA, Rubin DB. Statistical analysis with missing data: New York: John Wiley & Sons. 1987.
14. Schwartz NS, Clutter WE, Shah SD, Cryer PE. Glycemic thresholds for activation of glucose counterregulatory systems are higher than the threshold for symptoms. J Clin Invest. 1987;79(3):777-81.
15. Seaquist ER, Anderson J, Childs B, Cryer P, Dagogo-Jack S, Fish L, et al. Hypoglycemia and diabetes: a report of a workgroup of the American Diabetes Association and the Endocrine Society. Diabetes Care. 2013;36(5):1384-95.