Cover Page for Statistical Analysis Plan

Sponsor name:	Novo Nordisk A/S
NCT number	NCT02970942
Sponsor trial ID:	NN9931-4296
Official title of study:	Investigation of efficacy and safety of three dose levels of subcutaneous semaglutide once daily versus placebo in subjects with non-alcoholic steatohepatitis
Document date:	27 February 2020

16.1.9 Documentation of statistical methods

List of contents

Statistical Analysis Plan

NN9931-4296

Statistical Analysis Plan

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

Name: [redacted]
Department: Biostatistics Obesity & Metabolism

Table of contents

Page

Table of figures

Page

Table of tables

Page

Version history

This Statistical Analysis Plan (SAP) for trial NN9931-4296 is based on the protocol version 5.0 dated 09 March 2018.

Table 1 SAP Version History Summary

SAP Version	Approval Date	Change	Rationale
1		Not Applicable	Original version

1 Introduction

This SAP elaborates on the statistical analyses described in the protocol section 17. Changes to the protocol-planned analyses are documented in section 6.2, appendix 2.

1.1 Objectives and endpoints

1.1.1 Objectives

Primary objective

To compare the effect of semaglutide subcutaneous (s.c.) once daily versus placebo on histological resolution of non-alcoholic steatohepatitis (NASH).

Secondary efficacy objectives

To investigate the dose-response relationship of three dose levels of semaglutide s.c. once daily (0.1 mg/day, 0.2 mg/day and 0.4 mg/day) on histological resolution of NASH.

To compare the effects of semaglutide s.c. once daily to placebo on liver-related histological parameters and biomarkers of NASH disease.

To investigate the effects of semaglutide s.c. once daily versus placebo in subjects with NASH on:
1. Weight-related parameters
2. Glucose metabolism related parameters
3. Cardiovascular risk factors
4. Patient reported outcomes

Secondary safety objectives

To evaluate the safety and tolerability of three dose levels of semaglutide s.c. once daily in subjects with NASH.

1.1.2 Endpoints

Primary endpoint

NASH resolution* without worsening of fibrosis** after 72 weeks (yes/no)

*) Resolution of NASH defined by the NASH Clinical research network (CRN) as “no more than mild residual inflammatory cells (0 – 1) and no ballooning (0)” based on comprehensive interpretation by two independent pathologists (central reading) blinded to treatment allocation.

**) worsening defined by an increase of at least one stage of the Kleiner fibrosis classification

Secondary endpoints

Confirmatory secondary endpoint

At least one stage of liver fibrosis improvement with no worsening of NASH after 72 weeks (yes/no) (worsening defined as an increase of at least one stage of either lobular inflammation or hepatocyte ballooning according to NASH CRN criteria).

Supportive secondary endpoints

See section 5.4.2.

1.2 Trial design

This is a 72-week, randomised, double-blind, placebo-controlled, six-armed, parallel group, multi-centre, multi-national trial comparing once daily administration of semaglutide s.c. in three different doses (0.1 mg, 0.2 mg and 0.4 mg) with placebo in subjects with NASH.

A planned total of 288 subjects will be randomised. Based on an assumption of a 65% screening failure rate, 823 subjects will be screened. Subjects will be randomised in a 3:3:3:1:1:1 manner to receive daily dosing of semaglutide s.c. 0.1 mg, 0.2 mg, 0.4 mg or corresponding injection volumes of placebo once daily (for details see Figure 1–1). To avoid bias in the assessment of the different semaglutide doses, the trial will be double-blinded within dose levels. The dose levels will not be blinded between each other because of different dose escalations and different target doses and volumes required. It is expected that 15% of subjects will withdraw from the trial or prematurely discontinue treatment with trial product.

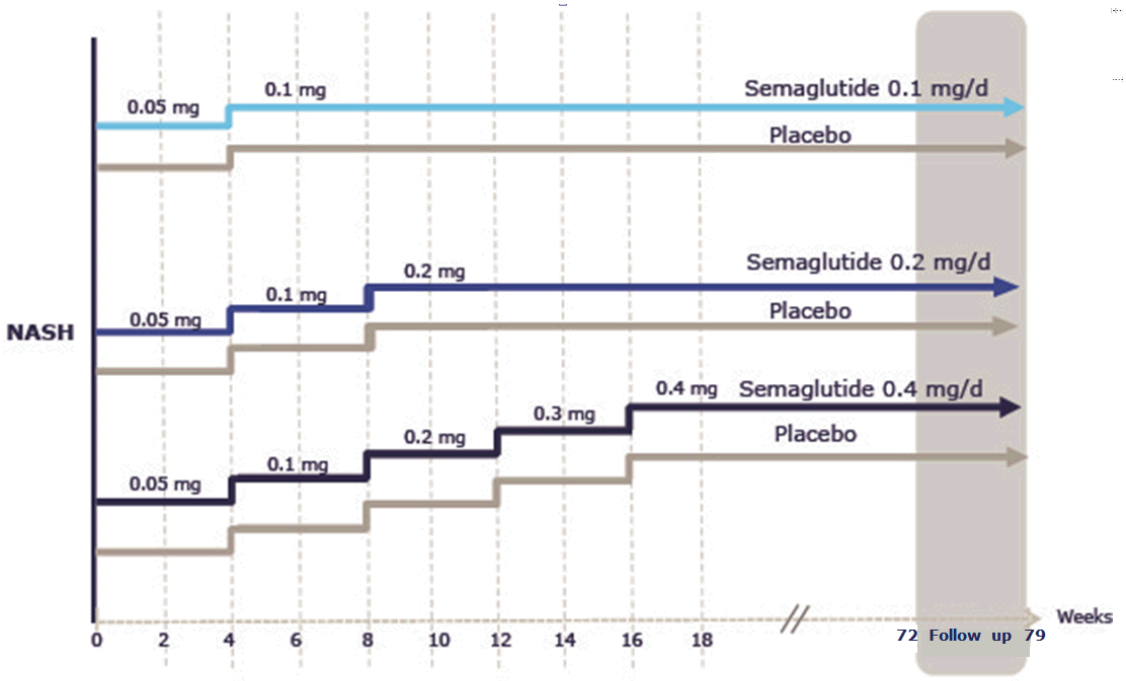

Figure 1–1 Trial design

Subjects in all treatment groups including placebo will receive nutritional and physical activity counselling by a member of the study team according to local site practice.

The randomisation of subjects to treatment will be stratified in five groups as described in protocol section 5.2 and protocol section 11.

The diagnosis of NASH as well as the histology based scores will be subject to central pathologist evaluation.

Subjects who prematurely discontinue the trial product treatment should continue with the originally scheduled site visits/contacts including a liver biopsy and final assessments 72 weeks after randomisation (visit 19A). However the following should not be done after visit 19 for subjects prematurely discontinuing trial product: Semaglutide plasma concentration, anti-semaglutide antibody assessment, hypo reporting and handing out subject diaries.

Subjects who prematurely discontinue trial product treatment will continue to receive nutritional and physical exercise counselling throughout their trial participation.

The total trial duration for the individual subject will be up to 85 weeks (maximum).

2 Statistical hypotheses

For the primary endpoint, NASH resolution without worsening in fibrosis after 72 weeks (yes/no), three confirmatory tests are planned to compare each dose level of semaglutide with placebo. Each comparison will test for superiority of semaglutide versus placebo, i.e. the null hypothesis of equal responder proportions will be tested against the alternative hypothesis of a higher proportion with semaglutide. Likewise, three confirmatory tests (one for each dose level of semaglutide vs. placebo) are planned for the confirmatory secondary histological endpoint, at least one stage of liver fibrosis improvement with no worsening of NASH after 72 weeks (yes/no). All six confirmatory tests will be performed on the subpopulation of subjects with fibrosis stage 2 or 3 at baseline, i.e. excluding subjects with fibrosis stage 1.

In order to control the family-wise type 1 error rate at 2.5% (for one-sided testing), the six confirmatory tests will be performed hierarchically in accordance to the graph in Figure 2–1. The graph defines a closed test procedure as described in Bretz et al 2009[1]. The test procedure will start with the comparison of semaglutide 0.4 mg versus placebo on NASH resolution without worsening in fibrosis. If the test confirms superiority, the comparison of semaglutide 0.4 mg versus placebo will be performed on fibrosis improvement with no worsening of NASH. If this test confirms superiority, the comparison of semaglutide 0.2 mg versus placebo will be performed on NASH resolution without worsening in fibrosis. The procedure will continue by descending dose alternating between the two endpoints. In each case, the null hypothesis will be tested using a local one-sided significance level of 2.5% (equivalent of a two-sided level of 5%). If one of the tests fails to reject the null hypothesis, the test procedure will stop and no further conclusions will be made.

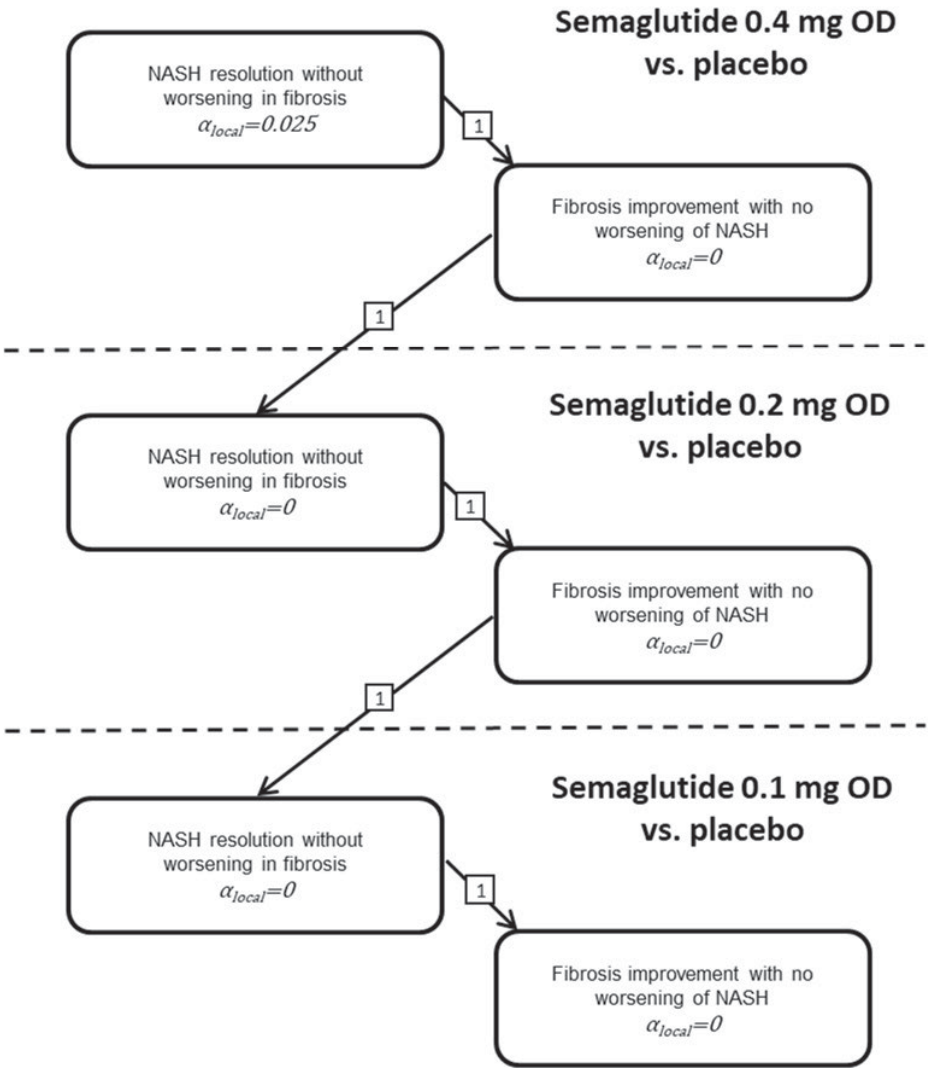

Figure 2–1 Graph of test procedure

3 Sample size determination

The sample size calculation is based on an aim to detect a difference in the primary endpoint between the highest semaglutide dose and the pooled placebo arm, using all subjects irrespective of baseline fibrosis stage (1-3). See protocol section 17.1 for details.

4 Analysis sets

The following analysis sets are defined in accordance with the ICH-E9 guidance[2]:

- The full analysis set (FAS) includes all randomised subjects. Subjects in the FAS will contribute to the evaluation "as randomised".
- The safety analysis set (SAS) includes all subjects receiving at least one dose of randomised treatment. Subjects in the SAS will contribute to the evaluation "as treated".

Observation periods

Data will be evaluated based on different observation periods which will be derived individually for each subject. The following two observation periods are defined:

- In-trial: This period starts on the date of the randomisation visit and ends on the first of the following dates (both inclusive):
  a) follow-up visit (end-of-trial visit)
  b) withdrawal of consent
  c) last contact with subject (for subjects lost to follow-up)
  d) death
- On-treatment: For evaluation of adverse events (AEs) and hypoglycaemic episodes, this period starts on the date of first administration of trial product and ends on the date of whatever comes first of: a) last dose of trial product + 49 days (7 half-lives of semaglutide) or b) end of the in-trial period. For evaluation of all other data, the period ends at the date of the last dose of trial product +7 days.

The statistical analyses of the efficacy endpoints will primarily be based on the in-trial period. The on-treatment period is used for some supportive efficacy analyses and all statistical analyses of safety endpoints. Summary statistics will in general be presented for both observation periods.

Data collected after the observation period in question will be excluded from any summary or analysis based on that observation period.

5 Statistical analyses

5.1 General considerations

The FAS will be used in the analysis of the efficacy endpoints whereas the SAS will be used for the safety endpoints (see the definition of the sets in section 4). If not otherwise specified, analyses for the primary and confirmatory secondary endpoints will be made on the subpopulation of subjects with fibrosis stage 2 or 3 at baseline, whereas analyses for the supportive secondary endpoints will be made on all subjects in the analysis sets.

The estimand addressing the primary objective is defined as the de-facto (effectiveness) treatment effect on NASH resolution without worsening in fibrosis for all randomised subjects after 72 weeks. All post-baseline scheduled visit data will be included in the statistical analysis, including data collected after discontinuation of trial product. The chosen estimand assesses the difference in resolution of NASH in a future population that results from initiating treatment with semaglutide as compared to placebo. Generalisation of this estimand depends among other things on the extent to which treatment adherence in this trial reflects clinical practice.

If not otherwise specified, the three different placebo arms will be pooled into one placebo arm in the statistical analyses. The pooling is based on the assumption that there is no substantial effect of different placebo volumes on the efficacy and safety endpoints. The validity of this assumption will be checked for the primary endpoint and treatment-emergent adverse events by evaluating descriptive statistics where each placebo arm is presented separately. In addition, the assumption will be evaluated based on estimates from a supportive analysis of the primary endpoint (see section 5.3.4).

The statistical analyses will in general consist of the following three pairwise treatment comparisons:

- semaglutide 0.4 mg versus placebo
- semaglutide 0.2 mg versus placebo
- semaglutide 0.1 mg versus placebo

The results of the comparisons will be presented as estimated treatment contrasts together with two-sided 95% confidence intervals and p-values corresponding to two-sided tests of no difference. For confirmatory tests of superiority, one-sided p-values will be used. The issue of multiple testing will be addressed for the primary and confirmatory secondary endpoints (see section 2) but not for any of the supportive secondary endpoints.

The baseline measurement is defined as the latest available measurement at or prior to the randomisation visit. An exception is made for identifying abnormal laboratory findings (including ALT, AST, GGT, bilirubin and INR) in which case the baseline value is defined as the mean of the available measurements at the screening and randomisation visits.

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

5.2 Subject disposition

The number of subjects in the analysis sets will be displayed as well as number of subjects in categories defined by end of treatment status (completed treatment or prematurely discontinued treatment, primary reason for discontinuation) and end of trial status (completed trial or was withdrawn from trial, primary reason for withdrawal, presence of biopsy).

5.3 Primary endpoint analyses

5.3.1 Definition of endpoint

The primary endpoint is the binary outcome NASH resolution without worsening in fibrosis after 72 weeks (yes/no). NASH resolution is derived from the scores given by the pathologists and is defined as lobular inflammation score ≤1 and hepatocyte ballooning score 0. Worsening in fibrosis is defined as an increase of at least one stage of the Kleiner fibrosis classification.

5.3.2 Main analytical approach

The primary analysis will be based on the Cochran-Mantel-Haenszel (CMH) test[3] which will be performed separately for the comparisons between each of the semaglutide arms and placebo. The test will adjust for baseline diabetes status (with or without T2D) and baseline fibrosis stage (F2 or F3). The response data will consist of the outcomes of the week 72 biopsy including assessments taken after premature discontinuation of trial product. Missing response data will be imputed as no resolution of NASH. This approach does not rely on an assumption of missing at random and should be considered conservative for estimating the treatment effect.

The exact one-sided p-value will be calculated for testing superiority. The two-sided p-value will be calculated as two times the one-sided p-value. Besides the p-values, the common odds ratio will be estimated together with an exact 95% confidence interval using the Mantel-Haenszel estimator associated with the CMH test.

5.3.3 Sensitivity analyses

To investigate the sensitivity of the results of the primary analysis with regard to the handling of missing data, the following four sensitivity analyses will be performed:

- Analysis using imputation based on treatment adherence: An analysis based on the same type of non-parametric method as for the primary analysis but with missing data handled by a multiple imputation (MI) method which assumes that the unobserved outcomes are well described by the observed outcomes from subjects who at week 72 are similar in terms of treatment adherence. This will be done as follows:
    1. Missing data are imputed by sampling with replacement from the empirical distribution of observed outcomes separately within the 8 groups of subjects defined

by randomised treatment arm and whether subjects complete the 72-week treatment or not. If a group completely lacks observed outcomes, missing data within this group will be imputed as no resolution of NASH. Multiple (1000) replicates of a complete data set are generated in this way.

2. For each complete data set, the log odds ratios are estimated using the same method as in the primary analysis together with the asymptotic standard errors.
3. The estimates and standard errors for the 1000 complete data sets are pooled using Rubin's rule[4]:

$$m_{MI} = \frac{1}{N}\sum_{i=1}^{N} m_i\,, \qquad SE_{MI} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} SE_i^2 + \left(1+\frac{1}{N}\right)\left(\frac{1}{N-1}\right)\sum_{i=1}^{N}(m_i - m_{MI})^2},$$

where m_i and SE_i are the estimated log odds ratios and standard errors for the N = 1000 data sets, and m_{MI} and SE_{MI} are the pooled estimates. From m_{MI} and SE_{MI}, the 95% confidence intervals for the odds ratios and associated p-values are calculated.

This analysis differs from the primary analysis in that it assumes that subjects with missing week 72 data have the same chances of NASH resolution as subjects with week 72 data in the same treatment group, accounting for whether they completed or discontinued the randomised treatment. This is a less conservative assumption than the one used in the primary analysis since it includes the possibility that subjects with missing week 72 data may have NASH resolution. The analysis intends to address missing data relative to what the measurements would have been had the measurements been taken.

- Analysis using imputation based on unconditional reference: An analysis based on the same type of non-parametric method as for the primary analysis but with missing data handled by an MI method which assumes that the unobserved outcomes are well described by the observed outcomes of the subjects in the placebo arm with similar baseline characteristics. The imputation will be done by random sampling of observed outcomes from subjects with the same baseline diabetes status and baseline fibrosis stage. 1000 replicates of a complete data set will be generated that will then be analysed in the same way as in the MI analysis based on treatment adherence. This analysis differs from the primary analysis in that it assumes that subjects with missing week 72 data have the same chances of NASH resolution as subjects with week 72 data in the placebo group. If there are more missing week 72 data in the semaglutide group than in the placebo group, this analysis is less conservative than the primary analysis and probably gives better estimate of the treatment effect. Compared to the first sensitivity analysis, this analysis is more conservative since it assumes that subjects with missing week 72 data in the semaglutide group have the same chance of NASH resolution as subjects in the placebo group.

- Complete case in-trial analysis: The same as the primary analysis but where subjects with missing week 72 data are excluded from the analysis.

- Complete case on-treatment analysis: The same as the primary analysis but where subjects with missing week 72 data or for whom the data were collected after the on-treatment period are excluded from the analysis.

The two complete case analyses are included as benchmarks. They are not based on the randomisation principle and do not estimate any causal effect of semaglutide treatment. The results are expected to be biased in favour of semaglutide and must be interpreted with extreme caution.

The adjustment for covariates in the primary analysis does not exactly match the stratification used for the randomisation. The rationale is that it will not be possible to adjust for region (Japanese or non-Japanese) and, at the same time, adjust for diabetes status and baseline fibrosis stage within the Japanese group due to small sample sizes. Adjustment for region is not expected to influence the overall results but is included in the stratification to facilitate an evaluation of consistency of treatment effect between the entire population and Japanese subjects. Region has therefore been excluded from the primary analysis. However, a CMH test stratified according to the five strata used for the randomisation will be performed as a sensitivity analysis.

Adjustment for baseline body weight as a continuous covariate may potentially give more precise estimates of the treatment effects. To investigate the influence of such an adjustment, a sensitivity analysis will be performed using a logistic regression model which includes treatment, baseline diabetes status, baseline fibrosis stage and diabetes-by-fibrosis interaction as factors and baseline body weight as a covariate.

5.3.4 Supplementary analyses

Supportive analyses

As a supportive analysis, the main analytical approach will be repeated including all subjects irrespective of baseline fibrosis stage. The test will adjust for baseline diabetes status and baseline fibrosis stage as a three-level factor (F1, F2 or F3).

As part of an evaluation of the appropriateness of pooling the three placebo arms, a supportive analysis will be performed using the same CMH test as in the primary analysis but without pooling the placebo arms. Each semaglutide arm will instead be compared with the placebo arm which received the same injection volume. All subjects irrespective of baseline fibrosis stage will be included in this analysis.

Exploratory analysis

The dose-response relationship with respect to the primary endpoint will be further explored by fitting a modified logistic regression model with baseline diabetes status, baseline fibrosis stage and diabetes-by-fibrosis interaction as factors and the log-transformed dose level and baseline body weight as covariates. The model may be expressed as

$$p_i = c_1 + \frac{(c_2 - c_1)}{1 + e^{-\beta \cdot X_i}}$$

where p_i is the probability of NASH resolution for subject *i*, c_1 and c_2 are the asymptotic probabilities of NASH resolution at zero and infinite dose levels, respectively, and $\beta \cdot X_i$ is the linear prediction function of factors and covariates. The response in the placebo arm will be included in the analysis to help estimate c_1. If the model does not describe the dose-response relationship well or if there are convergence problems, a different approximation may be investigated. All subjects irrespective of baseline fibrosis stage will be included in this analysis.

5.4 Secondary endpoint analysis

5.4.1 Confirmatory secondary endpoint

5.4.1.1 Definition of endpoint

The secondary confirmatory endpoint is a binary outcome defined as at least one stage of liver fibrosis improvement with no worsening of NASH after 72 weeks (yes/no). Worsening of NASH is defined as an increase of at least one point in either lobular inflammation score or hepatocyte ballooning score.

5.4.1.2 Main analytical approach

The main analysis will be the same as for the primary endpoint, i.e. a CMH test adjusted for baseline diabetes status and baseline fibrosis stage. Missing week 72 data will be imputed as no improvement in fibrosis.

5.4.1.3 Sensitivity analysis

The analysis using imputation based on unconditional reference will be performed in the same way as for the primary endpoint.

5.4.1.4 Supplementary analysis

The main analytical approach will be repeated including all subjects irrespective of baseline fibrosis stage.

5.4.2 Supportive secondary endpoints

5.4.2.1 Efficacy endpoints

Analyses will be performed for the change from baseline to week 72 in the following histological feature scores:

- Total NAS (0-8) and each of the components:
  - Steatosis (0-3)
  - Lobular inflammation (0-3)

- Hepatocyte ballooning (0-2)
- Stage of fibrosis according to the Kleiner fibrosis classification (0-4)
- Activity component of SAF score (0-4)

The activity component of the SAF score is defined as the unweighted sum of hepatocyte ballooning and lobular inflammation. The definition of the lobular inflammation score is modified in this calculation so that the scores 2 and 3 on the original scale are merged to a score 2. The possible range of the sum is thus 0 to 4. For all scores, a higher value indicates a more severe state of disease.

The histological feature scores will be analysed by an ordered logistic regression model (also known as a proportional odds model) with the score at week 72 as response; treatment, baseline diabetes status, baseline fibrosis stage, diabetes-by-fibrosis interaction and corresponding baseline score as factors; and baseline body weight as a covariate. The analyses will be based on the in-trial period and missing week 72 data will be imputed as no change from baseline in agreement with the analyses of the binary histological endpoints. The results will be presented as an estimate of the cumulative odds ratio for each treatment comparison together with the associated 95% confidence intervals and p-values.

Biomarkers of NASH disease

Analyses will be performed for the change from baseline to week 72 in the following biomarkers for NASH disease:

- Algorithms
  - Fibrosis-4 (Fib-4) score
  - NAFLD Fibrosis Score (NFS)
- Blood samples
  - Liver enzymes
    - Alanine aminotransferase (ALT)
    - Aspartate aminotransferase (AST)
    - Gamma glutamyltransferase (GGT)
  - Liver synthesis function
    - Albumin
    - International normalized ratio (INR)
  - Exploratory biomarkers
    - Enhanced liver fibrosis (ELF) test
    - Cytokeratin 18 (CK-18) fragments (M30 and M65 assays)
    - microRNA 122 (miR-122)
    - Interleukin-1 receptor (IL-1R) antagonist
    - Monocyte chemoattractant protein 1 (MCP-1)
    - Fibroblast growth factor 21 (FGF-21)

- Imaging
  - Liver stiffness
  - Controlled attenuation parameter (CAP) (liver steatosis)

The Fib-4 score will be derived according to the formula[5]:

$$\text{Fib-4} = \frac{\text{Age (years)} \times \text{AST (U/L)}}{\text{Thrombocyte count } (10^9/\text{L}) \times \sqrt{\text{ALT (U/L)}}}$$

The derivation will be performed at each visit where ALT, AST and thrombocyte count have been assessed. If any of the three laboratory parameters is missing at a specific visit, the Fib-4 score will be considered missing as well.

The NAFLD Fibrosis Score will be derived according to the linear regression formula[6]:

$$\begin{aligned}\text{NFS} = -1.675 + 0.037 \times \text{Age (years)} + 0.094 \times \text{BMI (kg/m}^2) + 1.13 \\ \times \text{Hyperglycaemia (yes/no)} + 0.99 \times \text{AST/ALT} - 0.013 \\ \times \text{Thrombocyte count } (10^9/\text{L}) - 0.66 \times \text{Albumin (g/dL)}\end{aligned}$$

Hyperglycaemia (yes/no) is a binary variable defined as 1 if FPG $\geq$ 6.1 mmol/L (110 mg/dL) at the corresponding visit of assessment or the subject has been diagnosed with T2D at screening; otherwise, the variable is defined as 0. NFS will be derived at each visit where body weight, FPG, ALT, AST, thrombocyte count and albumin have been assessed.

The ELF discriminant score will be derived as a log-linear combination of the markers hyaluronic acid (HA), amino-terminal propeptide of type III collagen (PIIINP) and tissue inhibitor of metalloproteinase 1 (TIMP1). The derivation will be performed at the central laboratory according to the instructions of the equipment manufacturer.

Analyses of liver stiffness and CAP (liver steatosis) will only be applicable to sites where these assessments were possible. Subjects at sites which did not have capability to assess liver stiffness or CAP (liver steatosis) will be excluded from the corresponding analysis.

All biomarkers except for NFS, ELF and CAP (liver steatosis) will be logarithmically transformed before the statistical analysis. The treatment differences will subsequently be back-transformed to the original scale and expressed as treatment ratios.

The main analysis of each of the biomarkers will be analysis of covariance (ANCOVA) with missing data handled by unconditional reference-based imputation. This will be done as follows:

1. An ANCOVA model with baseline diabetes status, baseline fibrosis stage and diabetes-by-fibrosis interaction as factors and baseline body weight and baseline value of the

corresponding biomarker as covariates is fitted to the change from baseline to 72 weeks for the placebo group only.

2. 1000 sets of values of the model parameters are drawn from the posterior distribution. For each replicated set of parameter values, the model is used to generate a complete data set by imputing missing values at 72 weeks for subjects in all treatment groups based on their baseline diabetes status, baseline fibrosis stage, baseline body weight and baseline value of the corresponding biomarker.
3. For each complete data set, the change from baseline to 72 weeks is analysed using an ANCOVA model with treatment, baseline diabetes status, baseline fibrosis stage and diabetes-by-fibrosis interaction as factors and baseline body weight and baseline value of the corresponding biomarker as covariates.
4. The estimated treatment differences and standard errors for the 1000 complete data sets are pooled using Rubin's rule. From the pooled estimates and standard errors, the 95% confidence intervals for the treatment differences and associated p-values are calculated.

A supportive on-treatment analysis using a mixed model for repeated measurements (MMRM) will be performed for each of the biomarkers. In this model, all scheduled post-baseline measurements taken during the individual subject's on-treatment period will enter as response; treatment, baseline diabetes status, baseline fibrosis stage and diabetes-by-fibrosis interaction will be included as factors; and baseline body weight and baseline value of the corresponding biomarker will be included as covariates. All factors and covariates will be nested within visit and an unstructured covariance matrix for measurements within subject will be employed. There will be no explicit imputation of missing values. As for the main analysis, the estimated treatment differences at week 72 with associated 95% confidence intervals and p-values will be presented.

Whereas the main analysis attempts to estimate the de-facto treatment effect (in agreement with the chosen estimand for the primary objective), the supportive analysis aims to estimate the de-jure effect that would have been observed if all subjects had remained on treatment and completed all visits. The latter analysis relies on the assumption that data are missing at random, which means that given the observed data, the events that lead to data being missing are independent of the unobserved data.

Additional exploratory analyses will be conducted to evaluate the performance of the biomarkers as predictors of NASH (yes/no), fibrosis (stage ≥2, ≥3 and 4, respectively) and/or steatosis (score ≥1, ≥2 and 3, respectively) as applicable using the liver biopsy as the gold-standard reference. The results will be reported separately from the clinical trial report.

Weight-related parameters

The secondary endpoints related to weight are defined as change from baseline to 72 weeks in:

- Body weight (% and kg)
- Waist circumference
- Body mass index (BMI)

These endpoints will be analysed separately using the same type of ANCOVA with MI as for the biomarkers with treatment, baseline diabetes status, baseline fibrosis stage and diabetes-by-fibrosis interaction as factors and corresponding baseline value as a single covariate. Supportive on-treatment analyses will be performed based on MMRM in the same way as for the biomarkers with factors and covariates specified as in the ANCOVA model.

In addition to the continuous endpoints, the following binary endpoints related to weight will be analysed separately:

- Weight loss of ≥ 5% of baseline body weight at 72 weeks (yes/no)
- Weight loss of ≥ 10% of baseline body weight at 72 weeks (yes/no)

The binary endpoints will be compared between treatment arms using an MI approach similar to the continuous endpoints but based on logistic regression. The 1000 data sets with imputed values for percent change in body weight will be reused to derive an equal number of complete data sets for the binary outcomes. For each data set, the binary outcomes will be analysed using a logistic regression model which includes treatment, diabetes status, baseline fibrosis stage and diabetes-by-fibrosis interaction as factors and baseline body weight as a covariate. The estimated log odds ratios and standard errors for the 1000 complete data sets are then pooled using Rubin's rule. From the pooled estimates and standard errors, the 95% confidence intervals for the odds ratios and associated p-values are calculated.

Glucose metabolism related parameters

The secondary endpoints related to glucose metabolism are defined as change from baseline to 72 weeks in:

- Glycosylated haemoglobin A1c (HbA_{1c})
- Fasting plasma glucose (FPG)
- Fasting glucagon
- Homeostatic model assessment - insulin resistance (HOMA-IR), derived through the approximation formula[7]:

$$\text{HOMA-IR} = \text{FPG (mmol/L)} \times \text{Fasting insulin (mIU/L)}/22.5$$

These endpoints will be analysed using the same type of ANCOVA with MI as for the biomarkers. For each endpoint, the ANCOVA will be performed separately for subjects with and without type 2 diabetes at screening. The model will include treatment and baseline fibrosis stage as factors and corresponding baseline value as covariate. Supportive on-treatment analyses will be performed based on MMRM in the same way as for the biomarkers with factors and covariates specified as in

the ANCOVA model. Fasting glucagon and HOMA-IR will be logarithmically transformed before analysis.

Cardiovascular risk factors

The secondary endpoints related to cardiovascular risk factors are defined as change from baseline to 72 weeks in:

- Systolic and diastolic blood pressure
- Lipids (total cholesterol, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, very low-density lipoprotein cholesterol, triglycerides, free fatty acids)
- High-sensitivity C-reactive protein (hsCRP)

These endpoints will be analysed separately using the same type of ANCOVA with MI as for the biomarkers with treatment, baseline diabetes status, baseline fibrosis stage and diabetes-by-fibrosis interaction as factors and corresponding baseline value as a single covariate. Supportive on-treatment analyses will be performed based on MMRM in the same way as for the biomarkers with factors and covariates specified as in the ANCOVA model. The lipids and hsCRP will be logarithmically transformed before analysis.

Patient reported outcomes

The results from the short form 36 (SF-36) questionnaire will be analysed as the change from baseline to 72 weeks in overall mental and physical scores, respectively, as well as the change in each of the 8 domains:

- Physical functioning
- Role functioning
- Bodily pain
- General health
- Vitality
- Social functioning
- Role emotional
- Mental health

The change in score will be analysed as a continuous endpoint using the same type of ANCOVA with MI as for the biomarkers with treatment, baseline diabetes status, baseline fibrosis stage and diabetes-by-fibrosis interaction as factors and corresponding baseline score as a single covariate. Supportive on-treatment analyses will be performed based on MMRM in the same way as for the biomarkers with factors and covariates specified as in the ANCOVA model.

5.4.2.2 Safety endpoints

The following secondary endpoints are used to support the safety objectives:

- Number of treatment-emergent adverse events during the trial
- Number of treatment-emergent hypoglycaemic episodes during the trial
- Number of treatment-emergent severe or blood glucose (BG) confirmed symptomatic hypoglycaemic episodes during the trial
- Number of treatment-emergent severe hypoglycaemic episodes during the trial
- Number of subjects discontinuing treatment due to gastrointestinal adverse events
- Change from baseline to 72 weeks in:
    - Pulse
    - Electrocardiogram (ECG)
    - Physical examination
    - Haematology (haemoglobin, haematocrit, thrombocytes, erythrocytes, leucocytes, differential count)
    - Biochemistry (creatinine, eGFR, creatine phosphokinase, urea, bilirubin (total), alkaline phosphatase, ferritin, sodium, potassium, calcium (total), amylase, lipase)
    - Hormones (calcitonin)
- Occurrence of anti-semaglutide antibodies during and after 72 weeks treatment (yes/no):
    - Anti-semaglutide binding antibodies
    - Anti-semaglutide binding antibodies with *in vitro* neutralising effect
    - Anti-semaglutide binding antibodies cross reacting with native GLP-1
    - Cross-reacting anti-semaglutide binding antibodies with *in vitro* neutralising effect to native GLP-1
- Anti-semaglutide antibody binding level during and after 72 weeks treatment

Adverse events

AEs will be coded using version 22.1 of the Medical Dictionary for regulatory Activities (MedDRA) coding. A treatment emergent adverse event (TEAE) is defined as an event that has onset date during the on-treatment period (see section 4).

AE data will be displayed in terms of the number of subjects with at least one event, the percentage of subjects with at least one event, the number of events and the event rate per 100 patient years of exposure. The main AE summaries will only contain TEAEs. Non-treatment emergent AEs will be included in listings and overview summaries.

Additional summaries will be made for select AEs based on predefined MedDRA searches. A list of MedDRA search terms will be specified and documented before database lock.

Hypoglycaemic episodes

Hypoglycaemic episodes will be classified and then summarised descriptively in terms of the number of subjects with at least one event, the percentage of subjects with at least one event, the

number of events and the event rate per 100 patient years of exposure. The summaries will be made separately for subjects with and without type 2 diabetes at randomisation.

For subjects with type 2 diabetes at randomisation, the severe or BG confirmed symptomatic hypoglycaemic episodes will be analysed as a binary outcome using Fisher's exact test, which will be performed separately for the comparisons between each of the semaglutide arms and placebo. The results will be presented as two-sided p-values.

For details on the classification of hypoglycaemia, see protocol section 17.4.1.2.

Pulse

Pulse will be summarised by descriptive statistics and analysed using the same type of MMRM as for the on-treatment analyses of the biomarkers with treatment, baseline diabetes status, baseline fibrosis stage and diabetes-by-fibrosis interaction as factors and baseline pulse as a covariate.

ECG

ECG will be described by summarising the number and percentage of subjects with normal and abnormal readings (separated into clinically significant and not clinically significant). The summaries will be presented by visit and as shift tables from baseline to week 72.

Physical examination

The results of the physical examination will be summarised descriptively in the same way as ECG.

Laboratory assessments

The haematology and biochemistry parameters will be summarised and evaluated by descriptive statistics.

Amylase and lipase will be analysed separately using the same type of MMRM as for the on-treatment analyses of the biomarkers with treatment, baseline diabetes status, baseline fibrosis stage and diabetes-by-fibrosis interaction as factors and the baseline value of the corresponding laboratory parameter as a covariate. A logarithmic transformation will be applied for both amylase and lipase

Calcitonin will be summarised in tables including number and percentage of observations > and ≤ LLOQ, quartiles, minimum and maximum. Abnormal elevations of calcitonin will further be explored by a shift plot showing the maximum post-baseline observation against baseline for each subject.

Antibodies

The occurrence of anti-semaglutide antibodies will be described by summarising the number and percentage of subjects with antibodies in the different treatment arms.

5.4.2.3 Pharmacokinetic endpoints

The plasma concentrations of semaglutide will be summarised by descriptive statistics. In addition, the data will be used for population pharmacokinetic (PK) modelling, see section 17.6 in the protocol.

5.5 Exploratory endpoint analysis

Not applicable for this trial.

5.6 Other safety analyses

Not applicable for this trial.

5.7 Other analyses

Not applicable for this trial.

5.8 Interim analyses

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

6 Supporting documentation

6.1 Appendix 1: List of abbreviations

AE	adverse event
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
AST	aspartate aminotransferase
BG	blood glucose
BMI	body mass index
CAP	controlled attenuation parameter
CK-18	cytokeratin 18
CMH	Cochran-Mantel-Haenszel
ECG	electrocardiogram
eCRF	electronic case report form
ELF	enhanced liver fibrosis
FAS	full analysis set
FGF-21	fibroblast growth factor 21
Fib-4	fibrosis 4
FPG	fasting plasma glucose
GGT	gamma glutamyltransferase
HbA_{1c}	glycosylated haemoglobin A1c
HOMA-IR	homeostatic model assessment - insulin resistance
hsCRP	high-sensitivity C-reactive protein
ICH	International Council on Harmonization

IL-1R	interleukin-1 receptor
INR	international normalized ratio
LLOQ	lower limit of quantification
MCP-1	monocyte chemoattractant protein 1
MedDRA	medical dictionary for regulatory activities
MI	multiple imputation
miR-122	microRNA 122
MMRM	mixed model for repeated measurements
NAFLD	non-alcoholic fatty liver disease
NAS	NAFLD activity score
NASH	non-alcoholic steatohepatitis
NFS	NAFLD fibrosis score
SAP	statistical analysis plan
SAS	safety analysis set
s.c.	subcutaneous
SF-36	short form 36
T2D	type 2 diabetes
TEAE	treatment emergent adverse event

6.2 Appendix 2: Changes to protocol-planned analyses

Statistical hypotheses

The secondary endpoint fibrosis improvement with no worsening of NASH was added to the confirmatory test procedure. In addition, the population for the confirmatory tests was changed to only consist of subjects with fibrosis stage 2 or 3 at baseline based on national scientific advice held with DKMA (DK), MEB (NL), and BfArM (DE) in 2019.

Observation periods

The definition of the in-trial period was clarified. In addition, the end of the on-treatment period was extended from last dose of trial product +1 day to +7 days, in order to match the visit window in the protocol flow chart.

Primary endpoint analyses

As a consequence of the change in the population for the confirmatory tests, the main analysis and the sensitivity analyses will only include subjects with fibrosis stage 2 or 3 at baseline. A supportive analysis was added repeating the main analytical approach including all subjects irrespective of baseline fibrosis stage.

The exact p-value and confidence interval will be used for the confirmatory tests instead of the asymptotic p-value for best accuracy.

The sensitivity analysis including baseline vitamin E use as factor was removed due to an observed low number of subjects using vitamin E.

Confirmatory secondary endpoint

The supportive secondary endpoint fibrosis improvement with no worsening of NASH was changed to a confirmatory endpoint to reflect its inclusion in the confirmatory test procedure. The main analysis was changed from logistic regression to a CMH test and a sensitivity analysis was added with respect to handling of missing data.

As for the primary endpoint, the main analysis and the sensitivity analyses will only include subjects with fibrosis stage 2 or 3 at baseline. A supportive analysis was added repeating the main analytical approach including all subjects irrespective of baseline fibrosis stage.

Supportive secondary endpoints

Biomarkers of NASH disease

For ELF, the value reported by the central laboratory will be used instead of a value derived by Biostatistics in order to use the correct derivation for the specific equipment.

Exploratory analyses to evaluate the performance of the biomarkers as predictors of NASH, fibrosis and/or steatosis have been moved out of the clinical trial report since they do not concern the trial objectives.

HOMA-IR

The unit for fasting insulin was corrected in the derivation formula.

Hypoglycaemic episodes

Due to few observed events, the analysis of severe or BG confirmed symptomatic hypoglycaemic episodes was changed from negative binomial regression to Fisher's exact test.

Calcitonin

Summary tables of abnormal elevations of calcitonin were replaced by a shift plot.

Antibodies

Listings of antibody binding levels and descriptive statistics comparing efficacy and safety between antibody-positive and antibody-negative subjects were removed. An exploratory investigation will instead be data-driven with a consideration of the number of antibody-positive subjects, and will only be presented if informative.

7 References

1. Bretz F, Maurer W, Brannath W, Posch M. A Graphical Approach to Sequentially Rejective Multiple Test Procedures. Statistics in Medicine. 2009;28:586–604.
2. Group IEEW. Statistical principles for Clinical Trials: ICH Harmonised Tripartite Guideline. Statistics in Medicine. 1999. p. 1905-42.
3. Agresti A. Categorical Data Analysis. 2nd ed. New York ed: John Wiley & Sons; 2002.
4. Little RJA, Rubin DB. Statistical analysis with missing data: New York: John Wiley & Sons. 1987.
5. Sterling RK, Lissen E, Clumeck N, Sola R, Correa MC, Montaner J, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. Hepatology. 2006;43(6):1317-25.
6. Angulo P, Hui JM, Marchesini G, Bugianesi E, George J, Farrell GC, et al. The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. Hepatology. 2007;45(4):846-54.
7. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia. 1985;28(7):412-9.