Cover Page for Statistical Analysis Plan

Sponsor name:	Novo Nordisk A/S
NCT number	NCT02963935
Sponsor trial ID:	NN8022-4274
Official title of study:	Effect and safety of liraglutide 3.0 mg as an adjunct to intensive behaviour therapy for obesity in a non-specialist setting (SCALE™ IBT)
Document date:	19 December 2018

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: NN8022-4274

SCALE™ IBT

Effect and safety of liraglutide 3.0 mg as an adjunct to intensive behaviour therapy for obesity in a non-specialist setting

Author:
Name: [redacted]
Department: [redacted] *Biostatistics Obesity*

Table of contents

Page

List of abbreviations

AD	available drop-out
AE	adverse event
ANCOVA	analysis of covariance
AT	available on treatment
BMI	body mass index
CMS	Centers for Medicare & Medicaid Services
DBL	Database Lock
ECG	Electrocardiogram
FAS	full analysis set
FFA	free fatty acids
HbA_{1c}	glycosylated haemoglobin
HDL	high-density lipoprotein
IBT	Intensive behaviour therapy
ITT	intention- to-treat
IWQoL-Lite for CT	Impact of Weight on Quality of Life-Lite for Clinical Trials
LAO	last available observation
LAO-OT	last available observation (of body weight) on randomised treatment
LDL	low-density lipoprotein
LLOQ	lower limit of quantification
MCS	mental component summary
MD	missing drop-out
MedDRA	Medical Dictionary for Drug Regulatory Activities
MI	Multiple Imputation
MMRM	Mixed model for repeated measurements
MT	missing on treatment
OR	odds ratio
PCS	physical component summary
PYE	patient years of exposure
PYO	patient years of observation
SAE	serious adverse event

SAP	statistical analysis plan
SD	standard deviations
SF-36	short form-36 version 2.0 acute
TC	total cholesterol
TG	Triglycerides
UNL	upper normal limit
UTN	Universal Trial Number
VLDL	very low density lipoprotein
wANCOVA	weighted analysis of covariance
WC	waist circumference
WRSS	Weight related sign and symptom measure

1 Introduction

1.1 Trial information

Trial design

This is a 56-week, randomised, double-blind, placebo-controlled, two-armed multi-centre trial in subjects with obesity, who have not been diagnosed with type 1 or 2 diabetes mellitus.

A total of 282 subjects will be randomised in a 1:1 manner to receive either liraglutide 3.0 mg or placebo, as an adjunct to CMS-IBT.

Throughout the 56-week treatment period, subjects will attend site visits to monitor response to treatment and 23 brief (approximately 15-minute) CMS-IBT counselling visits.

A 30-day follow-up period is included after the 56-week treatment period in accordance with FDA guidance.

Primary objective

To confirm superiority of liraglutide 3.0 mg vs. placebo, as an adjunct to CMS-IBT*, on weight loss effectiveness in subjects with obesity

*Intensive Behaviour Therapy for obesity in a primary care setting according to Centers for Medicare & Medicaid Services (CMS) visit schedule

Secondary objectives

- To establish the effects of liraglutide 3.0 mg vs. placebo, as an adjunct to CMS-IBT, on cardiometabolic and relevant efficacy endpoints in subjects with obesity
- To establish the degree of adherence to randomised trial product, caloric diet and physical activity, and to explore the effect thereof on weight loss in subjects with obesity
- To establish the safety and tolerability of liraglutide 3.0 mg vs. placebo, as an adjunct to CMS-IBT, in subjects with obesity

1.2 Scope of the statistical analysis plan

This SAP is based on the protocol "SCALE™ IBT. Effect and safety of liraglutide 3.0 mg as an adjunct to intensive behaviour therapy for obesity in a non-specialist setting", version 5, and amendment 1.

2 Statistical considerations

Results from the statistical analysis will generally be accompanied by two-sided 95% confidence intervals and two-sided p-values.

Superiority will be claimed if the two-sided p-value is less than 5% and the estimated treatment difference/ratio favours liraglutide 3.0 mg.

Primary endpoints

The two primary endpoints are defined as:

- Change in body weight (%) from baseline to week 56
- Proportion of subjects losing at least 5% of baseline body weight at week 56

The primary objective is met if superiority of liraglutide 3.0 mg vs. placebo is demonstrated for each of the primary endpoints.

Definition of primary endpoint: % weight change

Change in body weight (%) from baseline to week 56, denoted % weight change, is calculated as body weight at week 56 minus body weight at baseline divided by body weight at baseline and multiplied by 100; therefore

$$\%\ \text{weight change} = \frac{(\text{body weight at week 56} - \text{body weight at baseline})}{\text{body weight at baseline}} \times 100.$$

A negative value of % weight change indicates a body weight loss from baseline to week 56.

Definition of primary endpoint: 5% responders

The proportion of subjects having lost at least 5% of baseline body weight at week 56, denoted 5% responders, is defined as the proportion having

$$\%\ \text{weight change} \leq -5\%.$$

A 5% responder is defined as a subject fulfilling this, therefore

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

Definition of estimands

Effectiveness estimand

The primary estimand is an effectiveness estimand (*de facto*) quantifying the average treatment effect of liraglutide 3.0 mg relative to placebo 56 weeks after randomisation, as adjunct to reduced caloric diet, increased physical activity and CMS-IBT, in all randomised subjects regardless of adherence to treatment.

Efficacy estimand

A secondary efficacy estimand (*de jure*) is quantifying the average treatment effect of liraglutide 3.0 mg relative to placebo 56 weeks after randomisation, as adjunct to reduced caloric diet, increased physical activity and CMS-IBT, if all randomised subjects had adhered to assigned treatment regimen for the entire duration of the trial.

Effectiveness estimand at week 16

In addition, an effectiveness estimand (*de facto*) is quantifying the average treatment effect of liraglutide 3.0 mg relative to placebo 16 weeks after randomisation, as adjunct to reduced caloric diet, increased physical activity and CMS-IBT, in all randomised subjects regardless of adherence to treatment.

Taxonomy of week 56 assessments being available or missing

A given assessment at week 56 may be available or missing and Table 2–1 defines the taxonomy for this. Note this is done per assessment not per subject, as subjects may belong to different types for different assessments (a subject may have “available on treatment (AT)” for weight but “missing on treatment (MT)” for waist circumference).

Table 2–1 Taxonomy of week 56 assessments being available or missing

Assessment at week 56	On drug at week 56	Type description	Type Abbreviation
Available	Yes	**Available on randomised treatment**: Subjects who did not discontinue randomised treatment prematurely. Include those that stop and restart trial product.	AT
	No	**Available drop-outs**: Subjects who discontinued randomised treatment prematurely but returned to have an assessment at week 56; so-called retrieved drop-outs.	AD
Missing	Yes	**Missing on randomised treatment**: Subjects who did not discontinue randomised treatment prematurely. Include those that stop and restart trial product.	MT
	No	**Missing drop-outs**: Subjects who discontinued randomised treatment prematurely and did not return to have an assessment at week 56; so-called non-retrieved drop-outs.	MD

2.1 Sample size calculation

The sample size calculation is based on the analysis approach addressing the primary estimand for the primary endpoints change in body weight (%) from baseline to week 56 (% weight change) and proportion of subjects losing ≥ 5% of baseline body weight at week 56 (5% responders).

The two primary endpoints will be tested in a hierarchical order.

The study design is a 1:1 randomisation with 282 subjects in total (141 subjects in each arm). Please see Table 2–2 for general specifications on sample size and power calculation.

Table 2–2 General specifications for sample size and power calculation

Endpoint	Statistical test	Minimum required power (%)
% weight change	Two-group Satterthwaite unpooled t test on the mean difference with α=0.05 assuming unequal variances	90.0 (marginal)
5% responders	Pearson chi-square test for two independent proportions with α=0.05	90.0 (marginal)

α: two-sided significance level

In the following paragraphs treatment differences and standard deviations (SD) are given without units.

Body weight measurements from returning discontinuing subjects will be used in the primary analysis. But for the sample size calculations, it is assumed that all subjects who discontinue will not return at week 56 and that they have a % weight change like placebo subjects who completed the trial. It is expected that % weight change for liraglutide 3.0 mg subjects who discontinue might be somewhere in between completing placebo subjects and completing liraglutide 3.0 mg subjects. The assumption is thus expected to lead to a conservative effect estimate. In trial NN8022-1839 in non-diabetic subjects and with obesity, around 31% treatment discontinuations were observed in the placebo group, around 25% in the liraglutide 3.0 mg group. The calculated sample size ensures adequate power if the proportion of discontinuing subjects should be as high as 30%.

Based on findings from NN8022-1839, a difference in % weight change between liraglutide 3.0 mg and placebo of -5 (-9% vs. -4%, SD=7) is assumed among completing subjects. For sample size calculation, a common SD of 7 was assumed for completers of liraglutide 3.0 mg or placebo. In NN8022-1839, a SD of 6.65 was seen for % weight change among subjects completing liraglutide 3.0 mg at week 56; 5.88 was observed among subjects completing placebo. Adjusting for 30% discontinuation and using a mixture distribution, a difference of -3.5 is expected in the primary analysis. A sample size of 282 subjects (141 in each arm), gives a marginal power of 98.3%.

For the primary endpoint 5% responders, the expected proportions were calculated based on the same assumptions as above. Adjusting for 30% discontinuation gives an expected 5% responders proportion of 63% in the liraglutide 3.0 mg arm and 44% in the placebo arm. With 282 subjects, this results in a marginal power of 90%.

Given these assumptions, the sample size of 282 subjects (141 in each arm), results in a combined power of 88.5%.

The tests of superiority of liraglutide 3.0 mg to placebo for each of the confirmatory endpoints are performed hierarchically in the order in which the endpoints are presented (see Sections 4.2.1 and 4.2.2.1 in the protocol). The two primary endpoints are included in the statistical testing hierarchy, and the primary objective will be met if superiority of liraglutide 3.0 mg vs. placebo is demonstrated for each of the primary endpoints. The test hierarchy is given in Table 2–3 with underlying assumptions, marginal power and effective power. The effective was calculated using a naïve and conservative approach, which assumes no correlation of endpoints by multiplying the respective marginal powers. Assumptions made to calculate the power for the two primary endpoints and all confirmatory secondary endpoints are also presented in Table 2–3.

Table 2–3 Assumptions, marginal power and effective power for each confirmatory endpoint in the hierarchical testing procedure given an anticipated number of 282 randomised subjects

	Assumed mean (±SD) / proportion for completers liraglutide 3.0 mg	placebo	Expected mean (±SD) / proportion for liraglutide 3.0 mg #	Expected difference #	Marginal power (%)	Effective power (%)
% weight change	-9.0 (±7.0)	-4.0 (±7.0)	-7.5 (±7.4)	-3.5	98.3	98.3
5% responders	72%	44%	63%	19%	90.0	88.5
10% responders	44%	20%	37%	17%	90.3	79.9
15% responders	20%	6%	15%	9%	75.0	59.9
4% responders at week 16	70%	45%	63%	18%	84.2	50.4
WC change	-9.0 (±7.0)	-5.0 (±7.0)	-7.8 (±7.2)	-2.8	90.8	45.8
SF-36 PF	3.7 (±8.0)	2.4 (±8.0)	3.3 (±8.0)	0.9	15.8	7.2
IWQoL-Lite for CT PF	16.0 (±18.0)	11.0 (±18.0)	14.5 (±18.1)	3.5	36.7	2.7
6MWT change	8.2 (±12.0)	5.8 (±12.0)	7.5 (±12.1)	1.7	21.5	0.6

SD standard deviation, WC waist circumference, SF-36 PF Short Form-36 v2.0 acute physical functioning score, IWQoL-Lite for CT PF Impact of Weight on Quality of Life-Lite for Clinical Trial Version Physical function domain (5- items) score, 6MWT six minute walking distance test;
Adjusted for 30% discontinuation;
Assumptions are based on findings from NN8022-1839

2.2 Definition of analysis sets

Two analysis sets are defined:

- The *full analysis set* (FAS) includes all randomised subjects according to the intention-to-treat (ITT) principle. Subjects in the FAS will be analysed as randomised.
- The *safety analysis set* includes all randomised subjects exposed to at least one dose of trial drug. Subjects will be analysed as treated.

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

2.3 Primary endpoints

The two primary endpoints are as mentioned previously (see start of Section 2):

- % weight change
- 5% responders

2.3.1 Analytical methods addressing the effectiveness estimand for the primary endpoints

These analyses will use the FAS. The analyses of the primary endpoints use baseline body weight and body weight at week 56, and assessments at week 56 may be missing ("missing on treatment [MT]" and "missing drop-outs [MD]"). First, a description of the statistical models for the primary endpoints is given under the assumption of no missing values (i.e. all subjects have body weight measurements at baseline and week 56). Subsequently, handling of missing data is described.

2.3.1.1 Statistical model for the primary endpoint % weight change

The primary endpoint % weight change will be analysed using analysis of covariance (ANCOVA) including the factors and covariates listed in Table 2–4.

Table 2–4 Factors and covariates for the analysis of the primary endpoints

Factors and covariates at baseline	Type	Categories
Randomised treatment	Factor	Liraglutide 3.0 mg, placebo
Gender	Factor	Male, female
BMI (kg/m^2)	Factor	[30,35[, [35,40[, ≥40
Body weight (kg)	Covariate	*Not applicable*

The factors and covariates will be included in the model as main effects in an additive structure. The estimated treatment difference between liraglutide 3.0 mg and placebo will be reported together with the associated two-sided 95% CI and corresponding p-value.

The superiority test of liraglutide 3.0 mg vs. placebo will be carried out as follows:

Let $\mu_{liraglutide}$ and $\mu_{placebo}$ denote the true mean of % weight change for liraglutide 3.0 mg and placebo group, respectively. The hypothesis and the alternative are:

H: $\mu_{liraglutide} \geq \mu_{placebo}$ against the alternative H_A: $\mu_{liraglutide} < \mu_{placebo}$.

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

2.3.1.2 Statistical model for the primary endpoint 5% responders

This binary endpoint will be analysed using a logistic regression model. Factors and covariates will be those listed in Table 2–4. The estimated odds ratio (OR) between liraglutide 3.0 mg and placebo will be reported together with the associated two-sided 95% CI and corresponding p-value.

Let $OR_{liraglutide/placebo}$ denote the true odds ratio between liraglutide 3.0 mg and placebo. The hypothesis and the alternative are:

H: $OR_{liraglutide/placebo} \leq 1$ against the alternative H_A: $OR_{liraglutide/placebo} > 1$

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

2.3.2 Handling of missing values for the effectiveness estimand

2.3.2.1 Handling of missing values at baseline

If an assessment has been made both at screening and randomisation, the value from the randomisation visit will be used as the baseline value. If the value measured at the randomisation visit is missing and the assessment has been made at screening, then the screening value will be used as the baseline value.

2.3.2.2 Handling of missing values at week 56

Missing values at week 56 will be imputed and the relevant endpoints will be derived from the imputed values. Several approaches for imputation of missing values at week 56 will be applied. First, a description of the primary imputation approach used to address the effectiveness estimand for the primary endpoints is given. This is followed by a description of a number of sensitivity analyses.

Primary approach for handling of missing values

The primary approach for multiple imputations of missing values of body weight at week 56 (type MT+MD) for both the liraglutide 3.0 mg and placebo group is by sampling among all available assessments at week 56 in the placebo group (type AT+AD). This approach is also known as jump to reference and makes the assumption that subjects instantly after discontinuation lose any effect of randomised treatment beyond what can be expected from placebo treatment as adjunct to diet and exercise[1]. Body weight measurements at visits between baseline and week 56 are not used for this imputation approach. The multiple imputation approach is done in three steps.

- **Imputation**: Step 1 defines an imputation model based on placebo subjects, which is used to impute missing body weight values at week 56 in both arms. This will be done 100 times and results in 100 complete data sets. A more detailed explanation of this step is given below.
- **Analysis**: Step 2 analyses each of the 100 complete data sets, using the statistical model defined in Section 2.3.1.1 and 2.3.1.2 and saves the 100 estimation results.
- **Pooling**: Step 3 integrates the 100 estimation results into a final result using Rubin's formula.

The imputation model in step 1, uses placebo subjects from FAS with non-missing body weight measurements at baseline and week 56. The imputation model is a linear regression of body weight

(kg) at week 56 on the factors and covariates listed in Table 2–4 (except randomised treatment) with no interactions. The estimated posterior distribution for the parameters (regression coefficients and variances) in the imputation model is then used to impute missing week 56 body weight values for both randomised treatment arms.

If 100 copies are not sufficient to establish stable results, a higher number will be used. The multiple imputations will be generated using Novo Nordisk trial number 80224274 as seed number.

Sensitivity analyses of the primary endpoints for the effectiveness estimand

The sensitivity analyses will address the robustness of the primary approach for handling of missing values. In particular the sensitivity analyses address how assumptions on how body weight progresses after discontinuation of randomised treatment impact the estimated treatment difference between liraglutide 3.0 mg and placebo. The sensitivity analyses include the following.

- A multiple imputation approach as described by McEvoy[2] where missing body weight measurement at week 56 for non-retrieved drop-outs (type MD) are imputed by sampling from values obtained from retrieved drop-outs (type AD) in each randomised treatment arm and according to the timing (monthly) of last available observation (of body weight) on randomised treatment (LAO-OT). Missing body weight measurements at week 56 for subjects on drug treatment (type MT) are imputed by sampling from type AT in the relevant randomised treatment arm. Thus, the imputation model for each randomised treatment arm and timing of LAO-OT is a linear regression of body weight (kg) at week 56 on the factors and covariates listed in Table 2–4 (except randomised treatment arm) with no interactions and LAO-OT of body weight. If timing by month is too restrictive, quarters, half-years, or excluding timing will be used.
- A weighted ANCOVA (wANCOVA) where returning drug discontinuing subjects (AD) are up-weighted relative to their proportion of all drug discontinuing subjects (AD+MD) to account for the subjects not returning for assessments at week 56[2]. Similarly, AT subjects are up-weighted relative to their proportion of all drug continuing subjects (AT+MT). The up-weighing is done by randomised treatment arm and the timing of LAO-OT. Subjects who are missing the weight measurement at week 56 (MD+MT) are assigned a weight of 0 (zero).
- A single imputation approach as done by Sacks[3]. Missing weight measurement at week 56 for subjects who drop-out (type MD) are imputed using a weight regain rate of 0.3 kg/month after last available observation (LAO) of body weight. Change from baseline is truncated whenever the extrapolation would lead to a positive weight gain relative to baseline. LAO does not need to be the same as LAO-OT as subjects are allowed to come to scheduled visits after discontinuing randomised treatment. When a subject's LAO represents a gain in weight relative to baseline, no additional gain will be imputed, and the unfavourable gain is carried forward to week 56. Missing weight measurement at week 56 for subjects on treatment (type MT) will be imputed using the LAO. The weight regain imputation will be done for both randomised treatment arms.

Additionally, a version where only the liraglutide 3.0 mg arm uses the regain rate while the placebo arm uses LAO (corresponding to a weight regain rate of 0 kg/month) will be performed.

- Tipping point analysis. In a similar manner as above for a range of weight regain rates (starting from 0.1 kg/month and in intervals of 0.1 kg/month) for MD in liraglutide 3.0 mg arm will be used to define a tipping point in which superiority of liraglutide 3.0 mg disappears. In this analysis, the placebo arm will be imputed by LAO.

Figure 2–1 and Table 2–5 give an overview of the handling of missing values at week 56 assessments.

Additional sensitivity analyses

Additional analyses will investigate how sensitive the primary statistical models (for primary endpoints) are to the choice of factors and covariates in the model as follows: A model resembling the primary analysis models using the same imputation method (jump to reference) adjusting only for baseline body weight and randomised treatment arm. The imputation model (step 1) will be same as for the primary approach.

For the binary primary endpoint, the risk difference (unadjusted for any factors and covariates) and corresponding 95% CI will also be calculated.

2.3.3 Analysis method addressing the efficacy estimand for the primary endpoints

The efficacy estimand for % weight change will be assessed using a mixed model for repeated measurements (MMRM). The MMRM will use assessments only from subjects who are taking the randomised treatment until end of trial or at first discontinuing of trial drug (either temporarily or permanently). A pause of less than eight consecutive days is not regarded as discontinuation for this analysis. This means that assessments at week 56 for retrieved drop-outs (type AD) will be discarded. The MMRM will be fitted using % weight change and the same factors and covariates (Table 2–4) as for the primary analyses all nested within visit. An unstructured covariance matrix for measurements within the same subject will be employed, assuming that measurements for different subjects are independent.

The efficacy estimand for 5% responders will be assessed using the same MMRM. From the MMRM individually predicted values for % weight change at week 56 will be used to classify each subject as 5% responder or not. This classification will then be analysed using a logistic regression model with treatment as the only factor.

2.3.4 Subgroup analyses for primary endpoints

For the primary estimand, subgroup analyses will be done for the factors listed and categorised as in Table 2–4. For baseline weight, categories based on quartiles will be used. The subgroup analyses will be done separately for each of these by including interaction term(s) with randomised treatment arm in the respective models (ANCOVA or logistic regression).

Figure 2–1 Illustration of imputation approaches for the effectiveness estimand

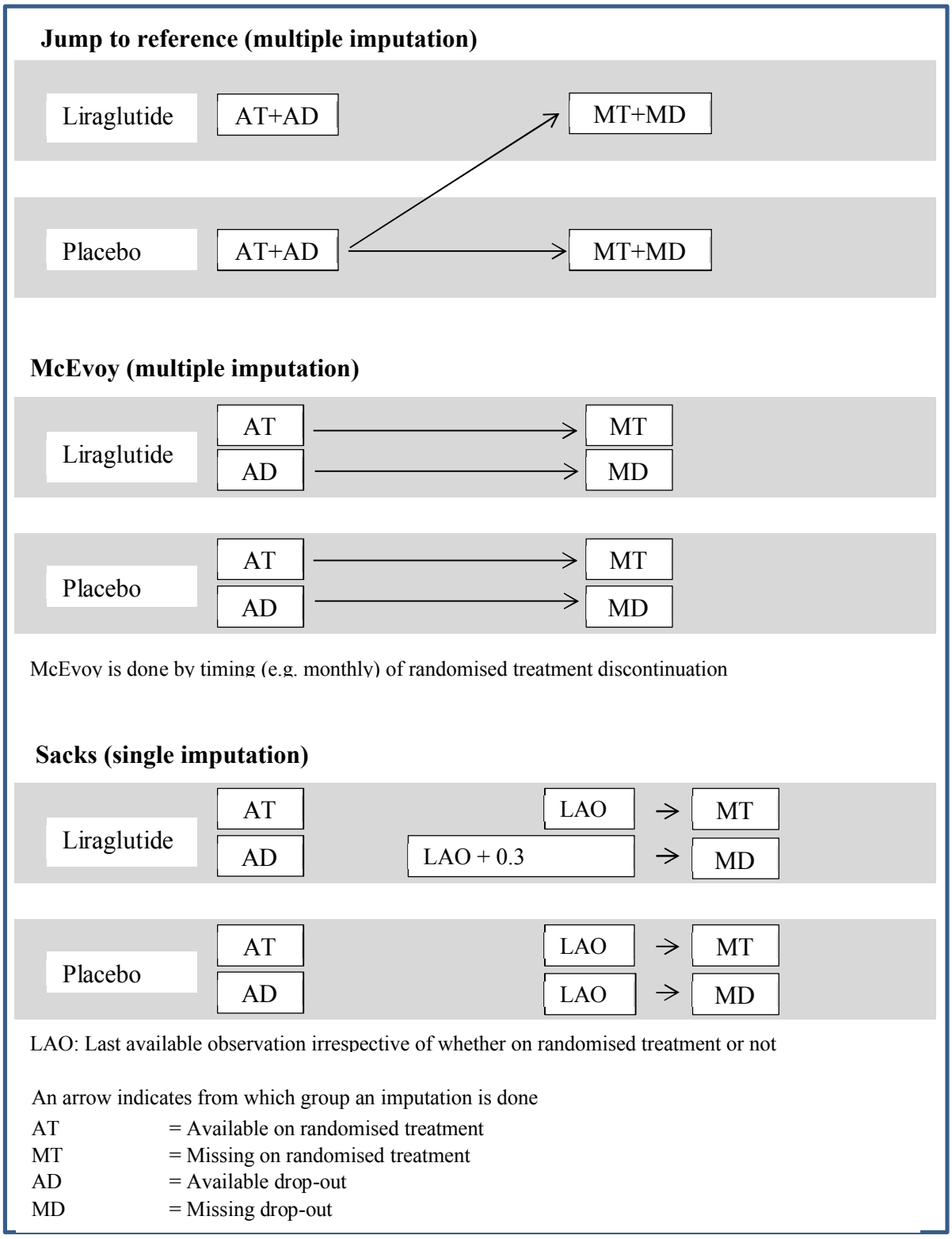

Table 2–5 Overview of handling of missing values

Approach	Description	Assumptions
Jump to reference (multiple imputation)	Liraglutide and placebo MT+MD imputed from placebo AT+AD (i.e., all available placebo assessments at week 56)	- Liraglutide MD lose any treatment effect instantly after drop-out - Placebo AT and placebo AD are assumed to have the same weight loss
McEvoy (multiple imputation)	Separately for the two randomised treatment arms: MD imputed from AD by matching on time (month) of drop-out MT imputed from AT	- AD is representative of the MD for each randomised treatment arm - AT is representative of the MT for each randomised treatment arm
Weighted ANCOVA (wANCOVA)	No imputation. Separately for the two randomised treatment arms: AD subjects are up-weighted relative to their proportion of AD+MD and timing of drop-out AT subjects are up-weighted relative to their proportion of AT+MT	- AD is representative of the MD for each randomised treatment arm - AT is representative of the MT for each randomised treatment arm
Sacks (single imputation)	Liraglutide MD imputed by minimum of LAO+0.3 kg and baseline value for each month from drop-out to week 56 and either a) Placebo MD imputed by minimum of LAO+0.3 kg and baseline value for each month from drop-out to week 56 or b) MT (both randomised treatment arms) and placebo MD are imputed by LAO	a) Liraglutide and placebo MD lose any treatment effect linearly after drop-out. b) Only liraglutide MD lose any treatment effect linearly after drop-out. No change in treatment effect since LAO for others.
Tipping point (single imputation)	Liraglutide MD imputed by minimum of LAO+X kg (X in steps of 0.1) and baseline value for each month from drop-out to week 56. MT (both randomised treatment arms) and placebo MD are imputed by LAO	Only liraglutide MD lose any treatment effect linearly after drop-out. No change in treatment effect since LAO for others

AT available on randomised treatment, MT missing on randomised treatment, AD available drop-out, MD missing drop-out, LAO last available observation, LAO-OT last available observation on randomised treatment.

2.4 Secondary endpoints

2.4.1 Confirmatory secondary endpoints

Confirmatory secondary endpoints are listed in Section 4.2.2.1 in the protocol.

All confirmatory secondary endpoints, except proportion of subjects losing at least 4% of baseline body weight at week 16, are planned to be assessed at week 56 and will be analysed using the same primary MI approach as used for the primary endpoints and to address the effectiveness estimand. The statistical model for continuous confirmatory secondary endpoints will be ANCOVA with covariates listed in Table 2–4 with baseline body weight replaced by baseline measurement of endpoint to be analysed. Binary confirmatory secondary endpoints will be analysed using logistic regression in the same way as for 5% responders.

The effectiveness estimand at week 16 will be assessed using the confirmatory secondary endpoint of proportion of subjects losing at least 4% of baseline body weight at week 16. Subjects with missing values for weight at week 16 will be considered non-responders.

Sensitivity analyses will be carried out for all confirmatory secondary endpoints, except for proportion of subjects losing at least 4% of baseline body weight at week 16. See Table 2–6 for details on planned analysis methods, multiple imputation approach and sensitivity analyses.

The efficacy estimand will also be assessed for confirmatory secondary endpoints, except for proportion of subjects losing at least 4% of baseline body weight at week 16, using MMRM as described for the primary endpoints.

2.4.2 Description of the hierarchical testing procedure to address the effectiveness estimand for primary and confirmatory secondary endpoints

The tests of superiority of liraglutide 3.0 mg to placebo for each of the endpoints are tested in a hierarchical manner in the order in which the endpoints are presented. The two primary endpoints are included in the statistical testing hierarchy below, even though the primary objective will only be met if superiority of liraglutide 3.0 mg vs. placebo is demonstrated for each of the primary endpoints. The test hierarchy is given in Table 2–6. The first endpoints to be tested are all assessing aspects of weight loss (e.g., relative change in body weight, achieving a certain magnitude of weight loss, and change in waist circumference); these endpoints are followed by endpoints assessing change in weight-related comorbidities and/or consequences of excess body weight and ‘feeling and function’ indicators.

Table 2–6 Hierarchical order and type of statistical method to address the effectiveness estimands for primary and confirmatory secondary endpoints

Test order	Endpoint	Endpoint Type	Statistical model	MI approach	Sensitivity analyses #
	Primary endpoints				
1	Change in body weight (%) from baseline to week 56	Continuous	ANCOVA	Jump to reference	McEvoy wANCOVA Sacks Tipping point Unadjusted ANCOVA
2	Proportion of subjects losing at least 5% of baseline body weight at week 56	Binary	Logistic Regression	Jump to reference	McEvoy Sacks Tipping point Risk difference
	Confirmatory secondary endpoints				
3	Proportion of subjects losing more than 10% of baseline body weight at week 56	Binary	Logistic Regression	Jump to reference	McEvoy Tipping point
4	Proportion of subjects losing more than 15% of baseline body weight at week 56	Binary	Logistic Regression	Jump to reference	McEvoy Tipping point
5	Proportion of subjects losing at least 4% of baseline body weight at week 16 (addresses the effectiveness estimand for week 16)	Binary	Logistic Regression	Not needed, as subjects with missing body weight values at week 16 will be considered non-responders.	
6	Change from baseline to week 56 in waist circumference (cm)	Continuous	ANCOVA	Jump to reference	McEvoy wANCOVA Tipping point
7	Change from baseline to week 56 in Short Form-36 (SF-36) v2.0 acute physical functioning score	Continuous	ANCOVA	Jump to reference	McEvoy wANCOVA Tipping point
8	Change from baseline to week 56 in impact of Weight on Quality of Life-Lite (IWQoL-Lite for CT), physical function domain (5-items) score	Continuous	ANCOVA	Jump to reference	McEvoy wANCOVA Tipping point
9	Change in six minute walking distance test (m) from baseline to week 56	Continuous	ANCOVA	Jump to reference	McEvoy wANCOVA Tipping point

MI Multiple Imputation;

See Table 2–5.

2.4.3 Supportive secondary endpoints

2.4.3.1 Efficacy endpoints

Supportive secondary efficacy endpoints addressing the secondary efficacy objectives are listed in Section 4.2.2.2 in the protocol. The questionnaire for Weight related sign and symptom (WRSS)

measure will not be validated until after DBL. Therefore the total score cannot be calculated and the supportive secondary endpoint "Weight related sign and symptom (WRSS) measure, total score" will not be analysed. Instead frequencies for categorical responses will be presented for each item and by visit.

Secondary endpoints addressing the first of the secondary efficacy objectives

Supportive secondary endpoints will be evaluated to address the effectiveness estimand. These endpoints will be analysed using the same MI approach as used for the primary and confirmatory endpoints and to address the effectiveness estimand.

The statistical model for continuous endpoints will be ANCOVA with covariates listed in

Table 2–4 with baseline body weight replaced by baseline measurement of endpoint to be analysed. The statistical model for proportions will be analysed using logistic regression.

In addition to the supportive secondary efficacy endpoints regarding IWQoL-Lite for CT listed in the protocol, analyses and outputs for the IWQoL-Lite for CT physical domain (not to be confused with IWQoL-Lite for CT physical *function* domain) will be made.

Secondary endpoints addressing the second of the secondary efficacy objectives

Adherence, used both as an endpoint and as a factor for weight loss will be investigated in an exploratory manner. In addition to the adherence definitions, analyses and displays presented below, alternative definitions, analyses and displays may be made if deemed relevant. These will be described as post-hoc analyses when reporting the trial.

- Adherence to trial product is defined as at least one administration of trial product per week
- Adherence to caloric diet is defined as at least one food diary entry for five days per week
- Adherence to physical activity is defined as at least 50% of the target minutes per week
- Adherence to caloric diet and physical activity is defined as being adherent to both caloric diet and physical activity
- Adherence to trial product, caloric diet and physical activity is defined as being adherent to all three randomised trial product, caloric diet and physical activity

Number of weeks adherent to treatment will be analysed by fitting a regression model for count data (e.g. negative binomial) with factors and covariates listed in Table 2–4.

Further, the association of adherence to trial product, caloric diet and/or physical activity with relative change of body weight will be explored across the total planned treatment duration. For that purpose, the following variables will be derived: First, the total treatment period (from randomisation to end-of-treatment visit) will be divided into four periods (each e.g. 12 (visit 2 to

10), 12 (visit 10 to 16), 16 (visit 16 to 20) and 16 (visit 20 to 24) weeks). Second, number of weeks adherent to trial product, caloric diet or physical activity during each of the four periods will be divided by total number of weeks of the respective period. Then, plots of relative change in body weight vs. adherence to trial product, caloric diet and/or physical activity will be created.

In addition, relative change of body weight will be summarized descriptively stratified by visits and different adherence categories, which will be defined based on the distribution of observed data.

Exploratory analysis of week 16 responders

In this exploratory analysis, subjects who lost ≥4% of their baseline body weight at week 16 after randomisation are considered as 'week 16 responders'. Subjects who lost <4% of their baseline body weight at week 16 after randomisation are considered as 'week 16 non-responders'. Hence, four 'week 16 response' groups can be defined:

1. ≥4% weight loss at 16 weeks (liraglutide 3.0 mg)
2. <4% weight loss at 16 weeks (liraglutide 3.0 mg)
3. ≥4% weight loss at 16 weeks (placebo)
4. <4% weight loss at 16 weeks (placebo)

The change in body weight (%) from baseline to week 56 will be explored among the four 'week 16 response' groups based on descriptive statistics only. Since the comparisons of the four 'week 16 response' groups are influenced by post-randomisation factors, no formal testing will be done.

2.4.3.2 Safety endpoints

Descriptive statistics for all safety endpoints will be provided with the aim to compare liraglutide 3.0 mg and placebo. All analyses and tabulations will be done using the safety analysis set. Unless otherwise stated, no formal statistical analyses are planned for the safety endpoints.

Adverse events

AEs will be coded using the current version of Medical Dictionary for Regulatory Activities (MedDRA). All AEs will be presented based on system organ class, high level group term and preferred terms.

Adverse events will be classified and analysed as 'in trial' and 'on drug' defined in terms of patient years of observation (PYO) and patient years of exposure (PYE), respectively. For each subject, PYO is defined as number of days from date of randomisation until and including date of follow-up visit or date of last contact. For each subject PYE is defined as time intervals of exposure to trial drug including an ascertainment window of 14 days for each exposure interval.

AEs, which occurred while the subject was in trial or on drug will be summarised descriptively, whereas AEs, which occurred before first exposure to trial drug or after the follow-up visit will only

be presented in listings. AEs will be displayed 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). Furthermore, AEs will be summarised by seriousness, severity, relation to trial drug, premature treatment discontinuation due to AE and outcome.

Summary tables by system organ class, high level group term and preferred term will be made for all AEs, SAEs, AEs possibly or probably related to trial drug, severe AEs, AEs occurring in at least 5% of the subjects in any arm, and AEs occurring in at least 1% of the subjects in any treatment arm.

AEs requiring completion of specific event forms will be presented in tables and listings. In addition, time to occurrence of these AEs will be presented in cumulative incidence plots.

Physical examination

Outcome of the physical examination at screening and change in the physical examination category at week 56 will be summarised in tables.

Pulse

Pulse (beats/min) at baseline and change of pulse at week 56, will be summarised in tables by week and treatment. Additionally, categories based on the maximum change from baseline until week 56 (>0, >5, >10, >20 beats/min) and categories based on the maximum value until week 56 (>80, >90, >100 beats/min) will be included in summary tables. Additionally, change in pulse from baseline to week 56 will be evaluated in the trial observation period similar to the primary analysis for the primary estimand but using the safety analysis set.

ECG

Shifts in the ECG category from screening to week 56 will be summarised in tables.

Laboratory measurements

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

Laboratory safety parameters are measured throughout the trial and comprise haematology and biochemistry as defined in the flow chart.

The distribution of each continuous laboratory parameter will be presented using box plots by treatment and week. Continuous laboratory parameters will be compared to the relevant reference ranges and results will be presented.

Amylase and lipase

Shifts from baseline to highest value in trial period to UNL, 2xUNL and 3xUNL will be summarised in tables.

Mean plots, geometric mean plots and box plots by gender, visit and treatment will be prepared. Change from week 0 to week 56 will also be presented by empirical distribution plots.

Number and percentage of subjects with amylase and lipase levels ≥UNL, ≥2xUNL or ≥3xUNL will be tabulated by week and treatment. Subjects with values ≥2xUNL will additionally be presented with spaghetti plots.

Additionally, change in amylase and lipase from baseline to week 56 will be evaluated in the in trial observation period similar to the primary analysis for the primary estimand but using the safety analysis set.

Calcitonin

Number, percentage and incidence of subjects with persistent (all post-baseline measurements) and incidental (at least one post-baseline measurement) increases in calcitonin for the criteria below will be tabulated for all subjects, males and females.

- From baseline <UNL to ≥UNL
- From baseline <UNL to ≥20 ng/L
- From baseline <UNL to ≥50 ng/L
- From baseline <20 ng/L to ≥20 ng/L
- From baseline <50 ng/L to ≥50 ng/L

Number and percentage of subjects with calcitonin levels ≥UNL, ≥1.5xUNL or ≥20 ng/L and ≥50 ng/L will be tabulated by visit and treatment.

A summary table showing number and percentage of observations < and ≥ LLOQ, minimum, Q25, median, Q75, maximum and geometric mean will be prepared by gender, visit and treatment.

The distribution of the calcitonin values across treatment group and time will be shown in plots by gender and total actual levels. Geometric means will be plotted by visit and treatment in order to assess the pattern of the longitudinal changes.

In addition, a scatter plot of baseline vs. maximum post-baseline calcitonin value will be prepared.

Longitudinal changes with calcitonin ≥20ng/L will be evaluated with spaghetti plots.

Subjects with at least one post-baseline calcitonin value above 20 ng/L will be listed.

The listings will include treatment, age, gender, smoking habits at baseline, risk factor information (use of relevant concomitant medication at time of assessment [proton pump inhibitors and H2 blockers] and medical history of thyroid disorder) and calcitonin values.

3 Changes to the statistical analyses planned in the protocol

3.1 Deviation from analyses as described in the protocol

- A MMRM is used for the efficacy estimand and only assessments from subjects who are taking the randomised treatment until end of trial or until first discontinuing of trial drug (either temporarily or permanent) are included. In the protocol, a trial drug pause of more than two days is defined as a treatment discontinuation, but a treatment discontinuation should have been defined as a trial drug pause of more than 7 days, which is also the information collected in the eCRF.

The following text:

”A pause of less than three consecutive days is not regarded as discontinuation for this analysis.”

has been replaced by:

”A pause of less than eight consecutive days is not regarded as discontinuation for this analysis.”

- The WRSS questionnaire will not be validated until after DBL. Therefore the total score cannot be calculated and the supportive secondary endpoint “Weight related sign and symptom (WRSS) measure, total score” cannot be analysed.

The following text has been added:

“The questionnaire for Weight related sign and symptom (WRSS) measure will not be validated until after DBL. Therefore the total score cannot be calculated and the supportive secondary endpoint “Weight related sign and symptom (WRSS) measure, total score” will not be analysed. Instead frequencies for categorical responses will be presented for each item and by visit.”

- The MMRM for relative change of body weight as function of adherence to trial product, caloric diet and/or physical activity across the total planned treatment duration will not be performed. The analysis was planned as exploratory and additional discussion/clarification is needed to perform the analysis, which will not be in scope for the CTR. Instead plots will be created for investigation of the association between the relative change of body weight and the adherence categories.

The following text:

"Then, MMRM will be performed including randomised treatment, BMI, gender, baseline body weight, period, normalised adherence to trial product, caloric diet and physical activity as factors and covariates as well as the interaction terms between randomised treatment, period and normalised adherence to trial product, caloric diet and physical activity."

has been replaced by:

"Then, plots of relative change in body weight vs. adherence to trial product, caloric diet and/or physical activity will be created."

3.2 Additional analyses not described in the protocol

The following text has been added:

"In addition to the supportive secondary efficacy endpoints regarding IWQoL-Lite for CT listed in the protocol, analyses and outputs for the IWQoL-Lite for CT physical domain (not to be confused with IWQoL-Lite for CT physical function domain) will be made."

4 References

1. Carpenter JR, Roger JH, Kenward MG. Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. J Biopharm Stat. 2013;23(6):1352-71.

2. McEvoy BW. Missing data in clinical trials for weight management. J Biopharm Stat. 2016;26(1):30-6.

3. Sacks FM, Bray GA, Carey VJ, Smith SR, Ryan DH, Anton SD, et al. Comparison of weight-loss diets with different compositions of fat, protein, and carbohydrates. N Engl J Med. 2009;360(9):859-73.