Supplementary Online Content 3

Wadden TA, Bailey TS, Billings LK, et al. Effect of subcutaneous semaglutide vs placebo as an adjunct to intensive behavioral therapy on body weight in adults with overweight or obesity: the STEP 3 randomized clinical trial. *JAMA*. Published online February 24, 2021. doi:10.1001/jama.2021.1831

Statistical Analysis Plan

Note: The secondary estimand (described herein) is referred to as the trial product estimand (also known as the hypothetical estimand) in the manuscript.

Statistical Analysis Plan

Trial ID: NN9536-4375

STEP 3

Effect and safety of semaglutide 2.4 mg once-weekly as adjunct to intensive behavioural therapy in subjects with overweight or obesity

Author
[redacted]
Biostatistics Aalborg 2

This ~~confidential~~ document is the property of Novo Nordisk. ~~No unpublished information contained herein may be disclosed without prior written approval from Novo Nordisk. Access to this document must be restricted to relevant parties.~~

Table of contents

Page

List of abbreviations

AD	*available but discontinued*
AE	*adverse event*
AT	*available on randomised treatment*
ANCOVA	*analysis of covariance*
BMI	*body mass index*
BP	*bodily pain*
CI	*confidence interval*
COA	*clinical outcome assessment*
CRF	*case report form*
CTR	*clinical trial report*
dBP	*diastolic blood pressure*
ECG	*electrocardiogram*
FAS	*full analysis set*
FFA	*free fatty acid*
FPG	*fasting plasma glucose*
GH	*general health*
$HbA1_c$	*glycated haemoglobin*
HDL	*high density lipoprotein*
hsCRP	*high-sensitivity C-Reactive Protein*
LAO-OT	*last available observation during the on-treatment period*
LDL	*low-density lipoprotein*
MCS	*mental component summary*
MD	*missing and discontinued*
MedDRA	*Medical Dictionary for Regulatory Activities*
MH	*mental health*
MMRM	*mixed model for repeated measurements*
MT	*missing on randomised treatment*
OR	*odds ratio*
PAI-1	*plasminogen activator inhibitor-1*
PCS	*physical component summary*
PDD	*pain/discomfort domain*
PF	*physical functioning*
PFD	*physical function domain*
PK	*pharmacokinetics*
PYE	*patient years of exposure*
PYO	*patient years of observation*
RE	*role-emotional*
RP	*role-physical*

SAE	*serious adverse event*
SAP	*statistical analysis plan*
SAS	*safety analysis set*
sBP	*systolic blood pressure*
SD	*standard deviation*
SF	*social functioning*
SF-36	*Short Form-36*
TEAE	*treatment-emergent adverse event*
VLDL	*very low density lipoprotein*
VT	*vitality*
WC	*waist circumference*

1 Introduction

1.1 Trial information

1.1.1 Objective(s)

1.1.1.1 Primary objective

To compare the effect of semaglutide s.c. 2.4 mg once-weekly versus semaglutide placebo as an adjunct to IBT in subjects with overweight or obesity on body weight.

1.1.1.2 Secondary objectives

To compare the effect of semaglutide s.c. 2.4 mg once-weekly versus semaglutide placebo as an adjunct to IBT in subjects with overweight or obesity on:

- Cardiovascular risk factors
- Clinical Outcome Assessments (COAs)
- Glucose metabolism

To compare the safety and tolerability of semaglutide s.c. 2.4 mg once-weekly versus semaglutide placebo as an adjunct to IBT in subjects with overweight or obesity.

1.1.2 Estimands

Primary estimand

The estimand will quantify the average treatment effect of semaglutide relative to semaglutide placebo after 68 weeks, as an adjunct to IBT including an initial 8-week LCD, in all randomised subjects regardless of adherence to treatment and regardless of initiation of other anti-obesity therapies (i.e. weight management drugs or bariatric surgery) ("treatment policy" estimand). The estimand will cover all effect-related objectives.

The following expansion of the primary estimand will cover the supportive secondary endpoint "Change from baseline at week 0 to week 8 in body weight (%)". The estimand will quantify the average treatment effect of semaglutide relative to semaglutide placebo after 8 weeks, as an adjunct to IBT including LCD, in all randomised subjects regardless of adherence to treatment and regardless of initiation of other anti-obesity therapies.

Secondary estimand

The estimand will quantify the average treatment effect of semaglutide relative to semaglutide placebo after 68 weeks, as an adjunct to IBT including an initial 8-week LCD, in all randomised subjects had they remained on their randomised treatment for the entire planned duration of the trial

and not initiated any other anti-obesity therapies (i.e. weight management drugs or bariatric surgery) ("hypothetical" estimand). The estimand will cover all effect-related objectives.

1.1.3 Endpoints

1.1.3.1 Primary endpoints

Primary endpoints addressing the primary objective:

- Change from baseline at week 0 to week 68 in body weight (%)
- Subjects who after 68 weeks achieve (yes/no):
- Body weight reduction ≥5% from baseline at week 0

1.1.3.2 Secondary endpoints

The confirmatory and supportive secondary endpoints addressing the primary and secondary objectives are listed below.

Confirmatory secondary endpoints

Subjects who after 68 weeks achieve (yes/no):

- Body weight reduction ≥10% from baseline at week 0
- Body weight reduction ≥15% from baseline at week 0

Change from baseline at week 0 to week 68 in:

- Waist circumference (cm)
- Systolic blood pressure (mmHg)
- Physical functioning score (SF-36)

Supportive secondary endpoints

Effect endpoints

- Change from baseline at week 0 to week 68 in:
- Body weight (kg)
- BMI (kg/m2)
- Hemoglobin A_{1c} (HbA_{1c}) (%, mmol/mol)
- Fasting plasma glucose (mg/dL)
- Fasting serum insulin (µIU/mL)
- Diastolic blood pressure (mmHg)
- Lipids (mg/mL)
    - Total cholesterol
    - High-density lipoprotein
    - Low-density lipoprotein

  - Very low-density lipoprotein
  - Free fatty acids
  - Triglycerides
- High sensitivity C-reactive protein (hsCRP) (mg/L)
- Plasminogen activator inhibitor-1 (PAI-1) Activity (UA/mL)
- SF-36:
  - Role-Physical score
  - Bodily Pain score
  - General Health score
  - Vitality score
  - Social Functioning score
  - Role-Emotional score
  - Mental Health score
  - Physical component summary
  - Mental component summary

Subjects who after 68 weeks achieve (yes/no):

- Body weight reduction ≥20% from baseline at week 0
- Increase in SF-36 physical function score ≥4.3 from baseline at week 0

Change from baseline at week 0 to week 8 in body weight (%)

Safety endpoints

Number of treatment-emergent adverse events (TEAEs) from baseline at week 0 to week 75
Number of serious adverse events (SAEs) from baseline at week 0 to week 75
Change from baseline at week 0 to week 68 in:

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

1.1.3.3 Exploratory endpoints

Exploratory endpoints addressing the exploratory objectives:
Change from baseline at week 0 to week 68 in:

- Glycaemic category (normo-glycaemia, pre-diabetes, T2D)
- Antihypertensive medication (decrease, no change, increase)
- Lipid lowering medication (decrease, no change, increase)
- Work Productivity and Activity Impairment Questionnaire - Specific Health Problem V2.0 (WPAI-SHP)
  - Work time missed due to weight (%)

- Impairment while working due to weight (%)
- Overall work impairment due to weight (%)
- Activity impairment due to weight (%)

- Total score (WRSSM)

Subjects who from randomisation to week 68 have permanently discontinued randomised trial product (yes/no)
Time to permanent discontinuation of randomised trial product (weeks).

1.1.4 Type of trial

This is a 68-week, randomised, double-blind, placebo-controlled, two-armed, parallel group, multi-centre clinical trial conducted in US, which compare semaglutide s.c. 2.4 mg once-weekly with semaglutide placebo, as an adjunct to IBT, in subjects with overweight or obesity.

The trial includes a screening visit to assess the subject's eligibility followed by weekly visits during the first 12 weeks. From week 12 to 24, visits will take place every 2nd week. From week 24, visits will take place every 4th week for the remaining maintenance period until end of treatment (week 68). A follow-up visit ("End of trial") for safety assessments is scheduled 7 weeks after end of treatment to account for the exposure to the long half-life of semaglutide.

1.2 Scope of the statistical analysis plan

This statistical analysis plan (SAP) is based on the protocol for trial NN9536-4375 "Effect and safety of semaglutide 2.4 mg once-weekly as adjunct to intensive behavioural therapy in subjects with overweight or obesity", version 3.0 (28 May 2019), and includes more detailed procedures for executing the statistical analyses of the primary and secondary endpoints. Statistical analyses and a number of clarifications additional to those specified in the trial protocol are pre-specified with this SAP. All changes to the statistical analyses planned in the trial protocol are documented in section 3.

2 Statistical considerations

Taxonomy of week 68 assessments

For each subject a given assessment at week 68 may be available or missing and **Table 2-1** describes the taxonomy for this. Note, this is done per assessment and per subject; subjects may be a different type for different assessments (a subject may have "available on randomised treatment (AT)" for body weight but "missing on randomised treatment (MT)" for waist circumference).

Table 2-1 Taxonomy for subjects based on week 68 assessments

Assessment at week 68	Subjects on randomised treatment at week 68	Type description	Type Abbreviation
Available	Yes	**Available on randomised treatment**: Subjects who complete the trial on randomised treatment with an assessment at week 68: Includes those that stop and restart trial product.	AT
	No	**Available but discontinued** Subjects who discontinued randomised treatment prematurely but returned to have an assessment at week 68. These are also called retrieved subjects	AD
Missing	Yes	**Missing on randomised treatment**: Subjects who complete the trial on randomised treatment without an assessment at week 68: Includes those that stop and restart trial product.	MT
	No	**Missing and discontinued**: Subjects who discontinued randomised treatment prematurely and did not return to have an assessment at week 68. These are also called non-retrieved subjects	MD

2.1 Sample size determination

The sample size and thereby the power for this trial is primarily defined to support safety. However, no formal statistical inference is planned based on number of adverse events. Given the trial sample size, the power of statistical tests for effect endpoints is described below.

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

In the analysis approach addressing the primary estimand, week 68 assessments from retrieved subjects (AD) is used. These data are also used to impute missing measurements at week 68 for non-retrieved subjects (MD). The imputation is done separately within each treatment arm (see description below). However, for the power calculations missing values (MT and MD), regardless of treatment arm, are assumed to be similar to semaglutide placebo subjects. These assumptions are likely conservative with respect to the power, and correspond to the jump to reference sensitivity analysis planned below.

Assumptions

The common assumptions for the power calculations are

- The significance level is 5%
- The randomisation ratio is 2:1
- For continuous endpoints the t-test on the mean difference assuming equal variances is used
- For binary endpoints the Pearson chi-square test for two independent proportions is used
- Based on data from NN9536-4153
  1. 20% of subjects discontinue permanently and
  2. 60% of these are retrieved (AD) at week 68
- All subjects in the placebo arm are assumed to have same effect as subjects who complete the trial on placebo (AT)
- Retrieved subjects (AD) in the semaglutide 2.4 mg arm are assumed to have an effect corresponding to half the treatment difference (compared to semaglutide placebo) of subjects who complete the trial on semaglutide 2.4 mg (AT)
- Non-retrieved subjects (MD) in the semaglutide 2.4 mg arm are assumed to have an effect corresponding to semaglutide placebo

Further assumptions made to calculate the power for each of the primary and confirmatory secondary endpoints are based on findings from other projects conducted by Novo Nordisk (NN8022 (SCALE), NN9535 (SUSTAIN), NN9924 (PIONEER)), and trial NN9536-4153 and are presented in **Table 2-2**.

Given these assumptions, the sample size of 600 subjects (400 in the semaglutide 2.4 mg once weekly and 200 in the semaglutide placebo arm), gives an effective power (marginal powers multiplied) of 86% for the seven endpoints in the hierarchical testing procedure. As sample size is primarily driven by safety, additional scenarios for assumptions are not included due to the overall high power.

Table 2-2 Assumptions, marginal power and effective power for each endpoint in the hierarchical testing procedure given an anticipated number 600 randomised subjects

Order	Endpoint	Assumed mean (±SD) or proportion for completers		Expected mean (±SD) or proportion Semaglutide 2.4 mg	Expected difference or proportion ratio	Marginal power (%)	Effective power (%)
		Semaglutide 2.4 mg	Semaglutide placebo				
1	% weight loss #	17 (10)	7 (10)	15.6 (11)	8.6 %-points	> 99	> 99
2	5% responders	88%	58%	85%	1.5	> 99	> 99
3	10% responders	76%	38%	71%	1.9	> 99	> 99
4	15% responders	58%	21%	53%	2.5	> 99	> 99
5	WC change (cm) #	17 (11)	8 (11)	15.7 (12)	7.7 cm	> 99	> 99
6	sBP change (mmHg) #	9.1 (13)	4.5 (13)	8.5 (14)	4 mmHg	91	91
7	SF-36 PF score change	6 (±10)	2 (±10)	5.4 (±11)	3.4 score-points	95	86

SD = standard deviation; WC = waist circumference; sBP = systolic blood pressure; SF-36 = Short Form 36 v2.0 acute; PF = physical functioning; # shown as a positive number

All tests in the hierarchy are based on the primary estimand

2.2 Definition of analysis sets

- Two analysis sets are defined:
  - The *full analysis set* (FAS) includes all randomised subjects according to the intention-to-treat principle. Subjects in the FAS will contribute to evaluation "as randomised".
  - The *safety analysis set* (SAS) includes all randomised subjects exposed to at least one dose of randomised treatment. Subjects in the SAS will contribute to evaluation "as treated".

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

Two observation periods are defined for each subject:

- In-trial: The *in-trial period* is defined as the uninterrupted time interval from date of randomisation to date of last contact with trial site.
- On-treatment (with trial product): A time-point is considered as 'on-treatment' if any dose of trial product has been administered within the prior 2 weeks (14 days). The *on-treatment period* is defined as all times which are considered on-treatment.
  - In general, the on-treatment period will therefore be from the date of first trial product administration to date of last trial product administration (+14 days) excluding potential off-treatment time intervals triggered by at least two consecutive missed doses.
  - For the evaluation of adverse events the lag time for each on-treatment time interval is 7 weeks (49 days)

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

2.3 Statistical analyses

If necessary, a statistical analysis plan (SAP) may be written in addition to the protocol, including a more technical and detailed elaboration of the statistical analyses. The SAP will be finalised before database lock.

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

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

Handling of missing baseline data

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

2.3.1 Primary endpoint

- Definition of primary endpoint: % weight change
- Change from baseline (week 0) to week 68 in body weight (%) is defined as
- $$\% \text{ weight change} = \frac{(\text{body weight at week 68} - \text{body weight at baseline})}{\text{body weight at baseline}} \times 100.$$
- Definition of primary endpoint: 5% responders
- A body weight reduction of least 5% from baseline (week 0) to week 68 is defined as
  - $$5\% \text{ responder} = \begin{cases} 1 \text{ if } \% \text{ weight change} \leq -5\% \\ 0 \text{ if } \% \text{ weight change} > -5\% \end{cases}$$

Analyses addressing the primary estimand

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

The analysis model for % weight change is a linear regression (ANCOVA) of % weight change with randomised treatment as a factor and baseline body weight (kg) as covariate. The estimated

treatment difference between semaglutide 2.4 mg and semaglutide placebo will be reported together with the associated two-sided 95% CI and corresponding p-value.

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

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

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

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

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

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

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

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

Handling of missing week 68 values for the primary estimand

All available data at week 68 (AT and AD) are used and missing values (MT and MD) at week 68 will be imputed and the endpoints will be derived from the imputed values. Several approaches for imputation will be applied. First, a description of the primary imputation approach to address the primary estimand for the primary endpoints is given followed by a description of the sensitivity analyses used to assess the robustness of the primary analysis results. The sensitivity analyses investigate how assumptions on body weight development after discontinuation of randomised treatment impact the estimated treatment contrasts between semaglutide 2.4 mg and semaglutide placebo. An illustration of all imputation approaches for the primary estimand is given in **Figure 2-1**.

Primary imputation approach for the primary estimand

Multiple imputation approach using retrieved subjects (RD-MI): The primary imputation approach for the primary estimand is a multiple imputation similar to the one described by McEvoy[1]. Missing body weight measurement at week 68 for non-retrieved subjects (MD) are imputed using assessments from retrieved subjects (AD) in each randomised treatment arm. This will be done according to the timing of last available observation on-treatment (LAO-OT) of body weight prior to week 68. Missing body weight measurements at week 68 for subjects on randomised treatment (MT) are imputed in a similar way by sampling from available measurements at week 68 from subjects on randomised treatment (AT) in the relevant randomised treatment arm. The multiple imputation approach is done in three steps:

3. **Imputation**: Defines an imputation model using retrieved subjects (AD) from FAS and done within groups defined by randomised treatment. . The model will be a linear regression of body weight (kg) at week 68 with gender (male/female), baseline BMI (kg/m^2) (in categories -<35, 35-<40, ≥40) and timing of the LAO-OT of body weight as factors and baseline body weight (kg) and LAO-OT of body weight (kg) as covariates. No interactions will be included. The grouping of timing will be done by quarters (intervals of 17 weeks). If timing by quarters is too restrictive, halves (intervals of 34 weeks) or excluding timing will be used. The timing by quarters or halves is defined as too restrictive if the imputation model cannot be fit due to inadequate numbers of retrieved subjects in 1 or more groups. If the imputation model still cannot be fit after excluding timing then the model will be further reduced until the model can be fit. Reduction will be done in a fixed order by first removing gender, then collapsing the two highest baseline BMI groups into one (≥35) and finally removing baseline BMI group. If no LAO-OT exists post-baseline then the LAO-OT will be the baseline body weight and the timing will be the first interval. If any subjects are MT, an imputation model for missing body weight measurements at week 68 for MT subjects will also be defined using AT subjects in a similar way. The estimated posterior distribution for the parameters (regression coefficients and variances) in the imputation models are then used to impute missing week 68 body weight values for each randomised treatment arm. This will be done 1,000 times and results in 1,000 complete data sets.
4. **Analysis**: Analysis of each of the 1,000 complete data sets, using the analysis models (ANCOVA and logistic regression) results in 1,000 times 2 estimations.
5. **Pooling**: Integrates the 1,000 times 2 estimation results into two final results using Rubin's formula.

Based on NN9536-4153 phase 2 results 1,000 copies should be sufficient to establish stable results. If 1,000 copies are insufficient, 10,000 copies will be used. The multiple imputations will be generated using Novo Nordisk trial number 95364375 as seed number. In addition to the seed number, it is specified that the dataset is sorted by subject ID.

Sensitivity analysis

Jump to reference multiple imputation approach (J2R-MI): Missing values of body weight at week 68 (MT and MD) for both the semaglutide 2.4 mg and semaglutide placebo group are imputed by sampling among all available assessments at week 68 in the semaglutide placebo group (AT and AD). This approach makes the assumption that subjects instantly after discontinuation lose any effect of randomised treatment beyond what can be expected from semaglutide placebo treatment as adjunct to IBT including an initial 8-week LCD[4]. The multiple imputation approach is done as above with the first step replaced by:

1. **Imputation**: Defines an imputation model using semaglutide placebo subjects from FAS with a week 68 measurement (AT and AD). The model will be a linear regression of body weight (kg) at week 68 with gender (male/female), BMI (kg/m^2) (in categories -<35,35-<40, ≥40) as factors and baseline body weight (kg) as covariate. No interactions will be included. If the imputation model cannot be fit due to inadequate numbers of retrieved subjects in one or more groups, then the imputation model will be reduced until the model can be fit. Reduction will be done in a fixed order by first removing gender, then collapsing the two highest baseline BMI groups into one (≥35) and finally removing baseline BMI group. The estimated posterior distribution for the parameters (regression coefficients and variances) in the imputation models are then used to impute missing week 68 body weight values for each randomised treatment arm. This will be done 1,000 times and results in 1,000 complete data sets.

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

A single imputation approach as done by Sacks[5] (S1-SI and S2-SI): Missing weight measurements at week 68 for non-retrieved subjects (MD) are imputed using a weight regain rate of 0.3 kg/month after LAO but truncated at no change from baseline whenever the extrapolation would lead to a positive weight gain relative to baseline. If a subject's weight at drug discontinuation represented a gain in weight relative to baseline, no additional gain will be imputed, and the unfavourable gain is carried forward to week 68. The weight regain imputation will be done for both randomised arms (S1-SI). Additionally, a version where only the semaglutide 2.4 mg arm uses the regain rate while the placebo arm uses last available observation (corresponding to a weight regain rate of 0 kg/month) will be performed (S2-SI). For both versions, missing weight measurements at week 68 for subjects on randomised treatment (MT) are imputed by using LAO.

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

robustness of the superiority conclusions to departures from the observed change in body weight in both treatment groups.

Mixed model for repeated measurements (MMRM): This 'MMRM for effectiveness' will use all assessments regardless of adherence to randomised treatment, including assessments at week 68 for retrieved drop-outs (AD). The MMRM for effectiveness will be fitted using the same factor and covariate as for the primary analyses all nested within visit. An unstructured covariance matrix for measurements within the same subject will be employed, assuming that measurements for different subjects are independent. For the 5% responder analysis, the same MMRM will be applied except that body weight (kg) will be used as response variable in the model. Individual missing values for body weight at week 68 will be predicted from the MMRM and used to classify each subject as 5% responder or not. This classification will then be analysed using the same logistic regression model as in the primary analysis of the primary estimand.

Subjects with missing week 68 assessment as non-responders: For the 5% responder analysis an analysis using subjects with missing week 68 assessment as non-responders in the logistic regressions will be done.

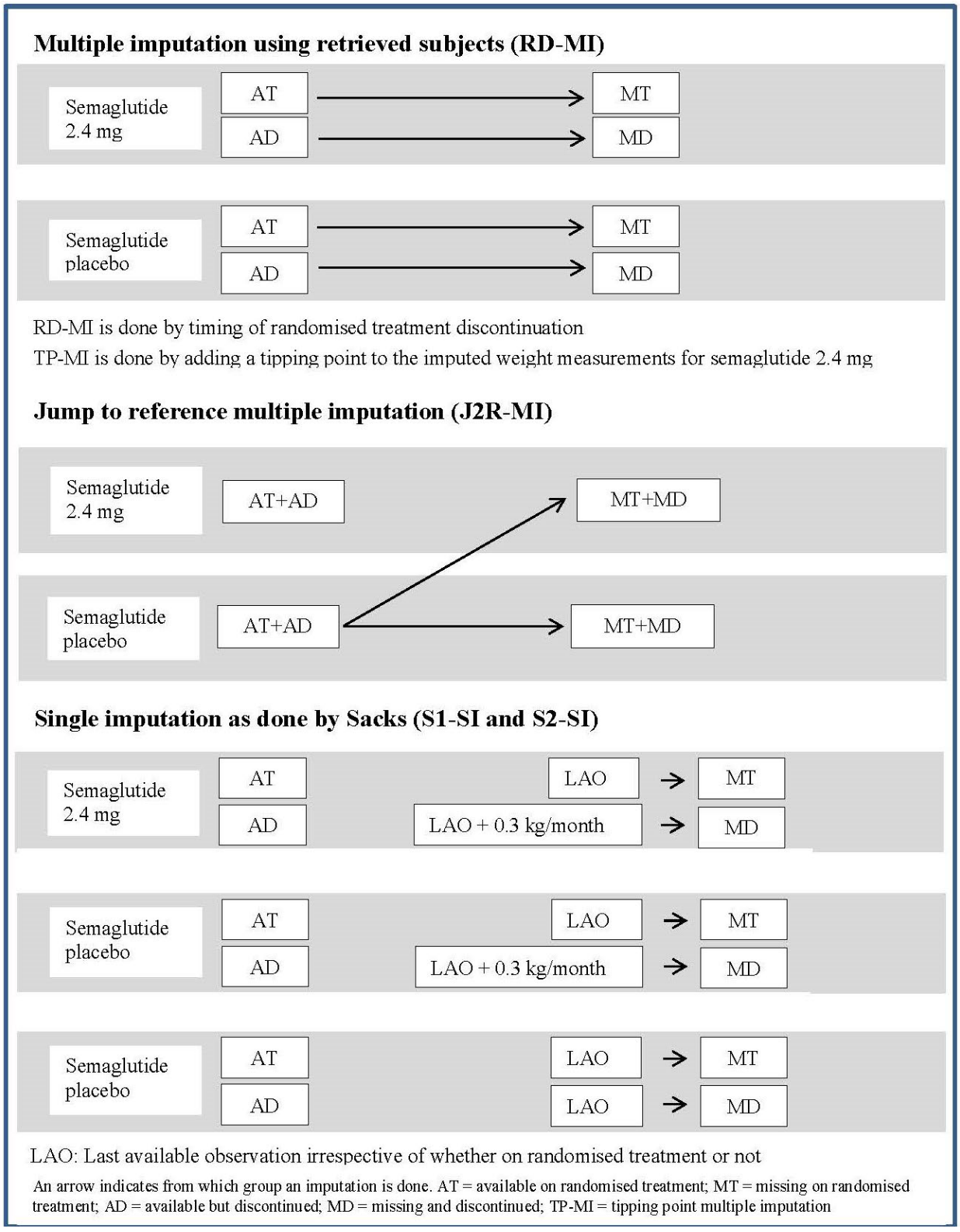

Figure 2-1 Illustration of imputation approaches for the primary estimand

Analysis addressing the secondary estimand

The secondary estimand for % weight change addresses the efficacy of semaglutide 2.4 mg and will be assessed using a 'MMRM for efficacy'. Week 68 assessments for retrieved drop-outs (AD) are not used in this analysis. The MMRM for efficacy will use assessments only from subjects who are taking the randomised treatment until end of treatment or until first discontinuing of randomised treatment. The derived date of the second consecutive missed dose will be used as the latest date for using assessments in this MMRM. The assessment closest in time and before the derived date of the second consecutive missed dose will be used as last assessment on randomised treatment. For subjects who initiate other anti-obesity therapies before completion or first discontinuing of randomised treatment, the date of starting weight management drugs or undergoing bariatric surgery will be used as latest date for using assessments in this MMRM. Similarly, the assessment closest in time and before the date of starting weight management drugs or undergoing bariatric surgery will be used as last assessment on randomised treatment. The MMRM for efficacy will be fitted using % weight change and the same factor and covariate as for the primary analyses all nested within visit. An unstructured covariance matrix for measurements within the same subject will be employed, assuming that measurements for different subjects are independent.

The secondary estimand for 5% responders will be assessed using the same MMRM for efficacy except that body weight (kg) will be used as response variable in the model. For subjects with missing body weight at week 68, individual values for body weight will be predicted from the MMRM and used to classify each subject as 5% responder or not. This classification will then be analysed using a logistic regression model with randomised treatment as a factor and baseline body weight (kg) as covariate.

An overview of all analysis and imputation methods to address the primary and secondary estimands for the primary endpoints is given in **Table 2-3**.

2.3.2 Secondary endpoints

- **Confirmatory secondary endpoints**

Confirmatory secondary endpoints are listed in section 4.2.2.1 of the study protocoland are all included in the fixed-sequence statistical strategy, see above. All tests are tests of superiority of semaglutide 2.4 mg to semaglutide placebo.

Analyses addressing the primary estimand

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

assessment of the endpoint to be analysed. The statistical model for body weight responder endpoints will be logistic regression with factor and covariate as for the primary endpoint 5% responders.

Analyses addressing the secondary estimand

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

Sensitivity analyses for confirmatory secondary endpoints

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

An overview of all analysis and imputation methods to address the primary and secondary estimands for confirmatory secondary endpoints is given **Table 2-3**.

Table 2-3 Analysis and imputation methods to address the primary and secondary estimands for the primary and confirmatory secondary endpoints in the statistical testing hierarchy

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

FAS = full analysis set; ANCOVA = analysis of covariance; RD-MI = multiple imputation using retrieved subjects; J2R-MI = jump to reference multiple imputation; S1-SI and S2-SI = single imputation as done by Sacks; TP-MI = tipping point multiple imputation; MMRM = mixed model for repeated measurements; LR = logistic regression; WC = waist circumference; sBP = systolic blood pressure; SF-36 = Short Form 36 v2.0 acute; PF = physical functioning
Test order refers to the order of the endpoint in the statistical test hierarchy outlined in **Table 2-2**.

- **Supportive secondary endpoints**

Supportive secondary endpoints are listed in section 4.2.2.2 of the study protocol. All tests are tests of superiority of semaglutide 2.4 mg to semaglutide placebo.

Analyses addressing the primary estimand

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

responder endpoints relating to COAs will be logistic regression with randomised treatment as a factor and the baseline assessment of the endpoint to be analysed as covariate.

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

Analyses addressing the secondary estimand

The supportive secondary endpoints evaluated at week 68 will be analysed to address the secondary estimand using the same MMRM for efficacy described for the primary endpoints.

Sensitivity analyses for supportive secondary endpoints

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

Analysis of safety endpoints

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

Adverse events will be defined as “treatment-emergent” (TEAE), if the onset of the event occurs in the on-treatment period (see definition in Section 2.2. TEAEs and SAEs will be summarised by descriptive statistics, such as frequencies and rates. No formal statistical inference will be carried out based on the number of TEAEs and SAEs. All AEs will be coded using the most recent version of the Medical Dictionary for Regulatory Activities (MedDRA).

An overview of all analysis and imputation methods to address the primary and secondary estimands for supportive secondary endpoints is given in **Table 2-4**.

Table 2-4 Analysis and imputation methods to address the primary and secondary estimands for supportive secondary endpoints

Objective	Endpoint	Endpoint type	Estimand	Analysis set	Statistical model	Imputation approach	Sensitivity analyses
Supportive secondary endpoints (effect related)							
Primary	20% responders	Binary	Primary	FAS	LR	RD-MI	-
			Secondary	FAS	LR	MMRM	-
Primary	Weight change (kg)	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Primary	BMI change (kg/m^2)	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Primary	% weight change*	Continuous	Primary	FAS	ANCOVA	J2R-MI	-
Secondary	HbA_{1c} change (%, mmol/mol)	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	FPG change (mg/dL, mmol/mol)	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	Fasting insulin change (mIU/L, pmol/mol)	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	sBP change (mmHg)	Continuous	Secondary	FAS	MMRM	-	-
Secondary	dBP change (mmHg)	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	Total cholesterol change (mg/dL, mmol/mol)	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	HDL change (mg/dL, mmol/mol)	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	LDL change (mg/dL, mmol/mol)	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	VLDL change (mg/dL, mmol/mol)	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	FFA change (mg/dL, mmol/mol)	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	Triglycerides change (mg/dL, mmol/mol)	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	hsCRP change (mg/L)	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	PAI-1 change (AU/mL)	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	SF-36 PF score responders #	Binary	Primary	FAS	LR	RD-MI	-
			Secondary	FAS	LR	MMRM-	-
Secondary	SF-36 PF score change	Continuous	Secondary	FAS	MMRM	-	-
Secondary	SF-36 RP score change	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	SF-36 BP score change	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	SF-36 GH score change	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	SF-36 VT score change	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	SF-36 SF score change	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	SF-36 RE score change	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	SF-36 MH score change	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	SF-36 PCS score change	Continuous	Primary	FAS	ANCOVA	RD-MI	-
			Secondary	FAS	MMRM	-	-
Secondary	SF-36 MCS score change	Continuous	Primary	FAS	ANCOVA	RD-MI	-

			Secondary	FAS	MMRM	-	-
Supportive secondary endpoints (safety related)							
Secondary	Number of TEAEs	Continuous	-	SAS	-	-	-
Secondary	Number of SAEs	Continuous	-	SAS	-	-	-
Secondary	Pulse change (bpm)	Continuous	-	SAS	MMRM	-	-
Secondary	Amylase change (U/L)	Continuous	-	SAS	Descriptive statistics	-	-
Secondary	Lipase change (U/L)	Continuous	-	SAS	Descriptive statistic	-	-
Secondary	Calcitonin change (ng/L)	Continuous	-	SAS	Descriptive statistics	-	-

FAS = full analysis set; ANCOVA = analysis of covariance; RD-MI = multiple imputation using retrieved subjects; MMRM = mixed model for repeated measurements; BMI = body mass index; HbA1c = Hemoglobin A1c; FPG = fasting plasma glucose; dBP = diastolic blood pressure; HDL = high density lipoprotein; LDL = low density lipoprotein; VLDL = very low density lipoprotein; FFA = free fatty acids; hsCRP = high sensitivity C-Reactive Protein; PAI-1 = Plasminogen Activator Inhibitor-1; LR = logistic regression; SF-36 = Short Form 36 v2.0 acute; PF= Physical Functioning; RP = Role-Physical; BP = Bodily Pain; GH = General Health; VT = Vitality; SF = Social Functioning; RE = Role-Emotional; MH = Mental Health; PCS = Physical component summary; MCS = Mental component summary; TEAEs = treatment emergent adverse events; SAEs = serious adverse events; # responder value = 4.3; * from baseline to week 8

2.3.3 Exploratory endpoints

Exploratory endpoints are listed in section 4.2.3 of the study protocol. Observed data for exploratory endpoints will be summarised by descriptive statistics.

2.3.4 Explorative statistical analysis for pharmacogenetics and biomarkers

The statistical analysis of biomarker endpoints is described under section 10.3.2.2 of the study protocol.

2.3.5 Other analyses

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

2.4 Pharmacokinetic and/or pharmacodynamic modelling

Not applicable for this trial.

3 Changes to the statistical analyses planned in the protocol

The main analyses were described in the protocol for the trial NN9536-4375. However, clarifications and more detailed descriptions of endpoints and analyses are provided in this SAP. The changes from the study protocol of NN9536-4375 are summarised below.

3.1 Trial-specific changes

- The following secondary objective has been added (concerning the supportive secondary endpoint "Change from baseline at week 0 to week 8 in body weight (%)). "To compare the effect of semaglutide s.c. 2.4 mg once-weekly versus semaglutide placebo as an adjunct to IBT including LCD in subjects with overweight or obesity on body weight". Furthermore it is specified that the imputation approach for percent change in body weight at week 8 will be a jump to reference multiple imputation, as it is expected that the multiple imputation model using retrieved subjects cannot be fit due to inadequate numbers of retrieved subjects in 1 or more groups.

- For analyses for the supportive secondary endpoints which relate to the primary objective and addressing the secondary estimand, it has been specified that this only includes endpoints evaluated at week 68 analyses (i.e. not week 8). This is in agreement with the secondary estimand and the text has only been updated in the section describing the statistical analyses for the supportive secondary endpoints .

3.2 Changes applied across STEP trials

- The supportive secondary endpoint "Body weight reduction ≥ 20% from baseline at week 0" has been added.
- It has been clarified that subjects in the FAS/SAS will be evaluated "as randomised"/"as treated".
- It has been clarified that the on-treatment period will be from date of first trial product administration to date of last trial product administration '(+14 days)'
- The text explaining how to handle missing baseline values has been changed to make it clear that if no eligible observation at or before randomisation is available then the mean of baseline values across all subjects is used as baseline value.
- Units for PAI-1 corrected to AU/mL (only applies to NN9536-4373, NN9536-4374 and NN9536-4375)
- It is specified that all AEs will be coded using the most recent version of the Medical Dictionary for Regulatory Activities (MedDRA) coding.
- The BMI-grouping "27-<35" has been changed to "-<35", since subjects may loose weight between the screening and the randomisation visit, and therefore have a BMI below 27 kg/m2 at the time of randomisation.
- It is clarified that RD-MI imputation is performed according to the timing of last available observation *during the on-treatment period* (LAO-OT). This is true for all endpoints. This is to clarify that the grouping of subjects according to timing is as in by McEvoy1. Furthermore it is clarified that the LAO-OT must be prior to the landmark visit (week 68).
- In grouping of retrieved subjects by timing of LAO-OT in the RD-MI procedure, it is clarified that timing by quarters or halves is defined as too restrictive if the imputation

model cannot be fit due to inadequate numbers of retrieved subjects in 1 or more groups. Furthermore it is described how a model reduction will be performed if needed.

- It is clarified that if no post-baseline LAO-OT exist, then the LAO-OT will be the baseline value and the timing of LAO-OT will be the first interval.
- In all multiple imputation procedures, in addition to the seed number, it is specified that the dataset is sorted by subject ID.
- The TP-MI procedure has been updated to be a 2-way tipping point analysis in which penalties are applied to both treatment groups (semaglutide 2.4 mg and placebo).
    - First, missing data are imputed according to the primary multiple imputation approach. Then, a penalty is added to the imputed values at week 68. The approach is to explore a range of penalties for both treatment groups, and the impact these would have on the study conclusions. The 2-dimensional space of penalties covering the range from -30% to 30% will be explored for both treatment groups. This sensitivity analysis evaluates the robustness of the superiority conclusions to departures from the observed change in body weight in both treatment groups.
    - The rationale for the change in TP-MI is the following feedback from FDA: “To confirm the robustness of superiority conclusions using a tipping point analysis, we believe that a 2-way tipping point analysis represents the real world situation for missing data from the both treatment arms (semaglutide and placebo). We would like to see departures from the treatment difference by varying both treatment arms rather than only adding a penalty to the active treatment arm (semaglutide). Additionally, please include interpretations for the varying scenarios and how likely they would be seen in a real-world setting.”]
- A description has been included of the sensitivity analysis of the 5% responder endpoint (primary estimand) using MMRM.
- It has been clarified that the non-responder analysis includes subjects with missing body weight assessment at week 68 as non-responders.
- It has been clarified that the 5% responder analysis using MMRM for the secondary estimand will be predicting individual values for body weight only when body weight is missing at week 68. Furthermore, it is clarified that the logistic regression will include both randomised treatment as a factor and baseline body weight as covariate.
- For physical functioning score (SF-36) the range of the scale was described as being 1-100. This has been deleted, since calculations are done on the norm based scores6.
- It is specified that fasting serum insulin will be log-transformed and analysed using a multiplicative model.
- It is specified that lipids, FPG and fasting insulin will also be analysed in SI-units.
- Text and tables have been updated to reflect that analyses for all non-weight related endpoints based on the secondary estimand are to be done.

4 Change log

SAP change log

Version	Reason for change
1.0	New
2.0	The BMI grouping "27-<35" has been changed to "-<35" to accommodate the fact that subjects may lose weight between the screening and the randomisation visit and therefore have a BMI below 27 kg/m^2 at the time of randomisation. **Table 2-3** has been updated to also include the secondary estimand for the sBP change (mmHg) and SF-36 PF score change. This is a consequence of the update in version 1 that stated that the secondary estimand will cover all effect-related objectives.

5 References

1. McEvoy BW. Missing data in clinical trials for weight management. J Biopharm Stat. 2016;26(1):30-6.
2. Carpenter JR, Roger JH, Kenward MG. Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. J Biopharm Stat. 2013;23(6):1352-71.
3. Sacks FM, Bray GA, Carey VJ, Smith SR, Ryan DH, Anton SD, et al. Comparison of weight-loss diets with different compositions of fat, protein, and carbohydrates. N Engl J Med. 2009;360(9):859-73
4. Stevens GA, Singh GM, Lu Y, Danaei G, Lin JK, Finucane MM, et al. National, regional, and global trends in adult overweight and obesity prevalences. Popul Health Metr. 2012;10(1):22.
5. NCD Risk Factor Collaboration. Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19·2 million participants. Lancet. 2016;387(10026):1377-96.
6. Maruish, M. E. (Ed.). User's manual for the SF-36v2 Health Survey (3rd ed.). Lincoln, RI: QualityMetric.