Cover Page for Statistical Analysis Plan

Sponsor name:	Novo Nordisk A/S
NCT number	NCT05579249
Sponsor trial ID:	NN9536-4741
Official title of study:	Effectiveness of semaglutide 2.4 mg vs. commercially available medications for chronic weight management in participants with obesity in a multi-employer setting in the US – a pragmatic clinical study
Document date*:	17 March 2025

*Document date refers to the date on which the document was most recently updated.

Statistical Analysis Plan

Effectiveness of semaglutide 2.4 mg vs. commercially available medications for chronic weight management in participants with obesity in a multi-employer setting in the US – a pragmatic clinical study

Substances: semaglutide 2.4 mg, orlistat, phentermine/topiramate, naltrexone/bupropion, liraglutide 3.0 mg

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

.

[redacted], Biostatistics

Table of contents

Page

Table of contents 2
Table of figures 3
Table of tables 4
Version History 5
List of abbreviations 6
1 Introduction 8
1.1 Objectives, Endpoints, and Estimands 8
1.2 Study Design 10
2 Statistical Hypotheses 11
2.1 Multiplicity Adjustment 11
3 Analysis Sets 12
4 Statistical Analyses 13
4.1 General Considerations 13
4.2 Primary Estimand Analysis 13
4.2.1 Definition of Endpoints 13
4.2.2 Main Analytical Approach 13
4.2.3 Sensitivity Analysis 14
4.2.4 Supplementary Analysis 15
4.3 Secondary Estimands Analysis 15
4.3.1 Confirmatory Secondary Estimands 15
4.3.1.1 Definition of Endpoints 15
4.3.1.2 Main Analytical Approach 15
4.3.1.3 Sensitivity Analysis 16
4.3.1.4 Supplementary Analysis 16
4.3.2 Supportive Secondary Endpoints 16
4.3.2.1 Definition of Endpoints 16
4.3.2.2 Analytical Approach 18
4.4 Exploratory Endpoints Analysis 19
4.4.1 Definition of Endpoints 19
4.4.2 Analytical approach 20
4.5 Safety Analysis 20
4.5.1 Extent of Exposure 20
4.5.2 Adverse Events 20
4.6 Other Analysis 21
4.6.1 Other Variables and/or Parameters 21
4.6.2 Subgroup Analysis 21
4.7 Interim Analysis 21
4.8 Changes to Protocol-planned Analysis 21
5 Sample size determination 23
6 Supporting Documentation 24
7 References 25

Table of figures

Page

No table of figures entries found.

Table of tables

Page

Table 1-1 Objectives and endpoints 8
Table 3-1 Analysis sets 12
Table 3-2 Defined data point sets 12

Version History

This Statistical Analysis Plan (SAP) for study NN9536-4741 is based on the protocol version 2.0 dated 15Sep2022.

SAP Version	Date	Change	Rationale
1.0	17 Jan 2023	Not Applicable	Original version
2.0	02 Feb 2023	Added analyses of subgroup that by the investigator is categorised as having “Poor” physical functioning at baseline. Vault ID number for e-mail communication with the author of WLQ-25, Dr. Lerner, changed from VV-TMF-5555841 to VV-TMF-5883040.	It is relevant to know the level of improvement in physical functioning for the subgroup that has poor physical functioning at baseline. It is expected that this subgroup will benefit even more from loosing weight than the overall study population. Since this update is done before any randomisation visits have been performed, the specified subgroup analyses are considered pre-specified. Correction of error.
3.0	21 Mar 2025	Removed analyses of subgroup that by the investigator is categorised as having “Poor” physical functioning at baseline. Added LR-NR method as sensitivity analysis for primary endpoint.	The investigator found it challenging to gather the necessary data, resulting in the exclusion of the subgroup classified as having "Poor" physical functioning at baseline from the analyses.

List of abbreviations

AE	adverse event
ANCOVA	analysis of covariance
AOM	anti-obesity medication
BMI	body mass index
CI	confidence interval
COVID-19	coronavirus disease 2019
DPS	data point set
FAS	full analysis set
GS	global satisfaction
GSTS	global Satisfaction, total score
ICH	International council for harmonisation
IWQOL-Lite-CT	Impact of Weight on Quality of Life – Lite – Clinical Trials
LAO-OT	last available observation on treatment
LR	logistic regression
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	mixed model for repeated measurements
OR	odds ratio
PDC	proportion of days covered by study product
PDMW	proportion of days missed from work due to participant illness or not feeling well
PF	physical function domain
PGI-S	Patient Global Impression of Status
SAE	serious adverse event
SAP	statistical analysis plan
SAS	safety analysis set

TFL	tables, figures and listings
TP-MI	tipping point multiple imputation
TS	total score
TSQM	Abbreviated Treatment Satisfaction Questionnaire for Medication
US	United states
WLQ	Work Limitations Questionnaire

1 Introduction

The following is a brief introduction to this pragmatic study. See section 2 in the study protocol for more details including references.

Changes to the analyses described in the protocol are described in Section 4.8.

1.1 Objectives, Endpoints, and Estimands

Objectives and Endpoints

Table 1-1 Objectives and endpoints

Objectives	**Endpoints**		
Primary	Title	Time frame	Unit
To demonstrate the superiority of semaglutide 2.4 mg s.c once weekly versus Other AOMs[1], as an adjunct to lifestyle management, with respect to achieving ≥10.0% body weight reduction from baseline, in adults with obesity.	***Primary**:*		
	Body weight reduction ≥10.0% (yes/no)	At week 52	Count of participants
Secondary Confirmatory	Title	Time frame	Unit
To demonstrate the superiority of semaglutide 2.4 mg s.c once weekly versus Other AOMs, as an adjunct to lifestyle management, with respect to the relative change in body weight from baseline, in adults with obesity	***Confirmatory***		
	Change in body weight	Week 0 to week 52	Percent
To demonstrate the superiority of semaglutide 2.4 mg s.c once weekly versus Other AOMs, as an adjunct to lifestyle management, with respect to the change in physical functioning, in adults with obesity	Change in IWQOL-Lite-CT physical function domain[2]	Week 0 to week 52	Score
Secondary Supportive	Title	Time frame	Unit
To compare the effect of semaglutide 2.4 mg s.c once weekly versus Other AOMs, as an adjunct to lifestyle management, with respect to additional body weight parameters, in adults with obesity	***Supportive***		
	Body weight reduction ≥15.0% (yes/no)	At week 52	Count of participants
	Body weight reduction ≥20.0% (yes/no)	At week 52	Count of participants
To compare the effect of semaglutide 2.4 mg s.c once weekly versus Other AOMs, as an adjunct to lifestyle management, with respect to achieving responder threshold for physical functioning, in adults with obesity	Change in IWQOL-Lite for CT physical function domain score ≥ 14.6(yes/no)[2,3]	At week 52	Count of participants
To compare the effect of semaglutide 2.4 mg s.c once weekly versus Other AOMs, as an adjunct to lifestyle management, with respect to adherence to medication for chronic weight management and work limitations, in adults with obesity	Proportion of days covered (PDC) by study product	Week 0 to week 52	Percent
	Covered by study product ≥80% of days (yes/no)	At week 52	Count of participants

Objectives	**Endpoints**		
	Change in Work Limitations Questionnaire 25-item version (WLQ-25), total score[2]	Week 0 to week 52	Score
Exploratory	Title	Time frame	Unit
To compare the effect of semaglutide 2.4 mg s.c once weekly versus Other AOMs, as an adjunct to lifestyle management, with respect to absence from work, in adults with obesity	***Exploratory:***		
	Proportion of days missed from work due to participant illness or not feeling well	Week 0 to week 52	Percent
To compare the effect of semaglutide 2.4 mg s.c once weekly versus Other AOMs, as an adjunct to lifestyle management, with respect to treatment satisfaction, in adults with obesity	Treatment satisfaction assessed by Abbreviated Treatment Satisfaction Questionnaire for Medication (TSQM-9), Global Satisfaction, total score[2]	End of treatment at week 52	Score points

[1] Other AOMs are defined as one of the following medications approved for chronic weight management: Xenical®, Qsymia®, Contrave®, and Saxenda®.

[2]Please refer to Section 8.1 in the study protocol for detailed information on scales and scoring of patient reported outcomes forms

[3]Responder threshold is the threshold for meaningful within participant change. The threshold is anchor based using the Patient Global Impression of Status, PGI-S.

Primary estimand

The primary clinical question of interest is: What is the effectiveness of semaglutide 2.4 mg versus Other AOMs, as an adjunct to lifestyle management, in adults employed in the US living with obesity. This will be measured by the number of participants achieving ≥10% body weight loss at week 52, regardless of adherence to randomised treatment, and regardless of initiating other anti-obesity therapies (weight management drugs not included in the randomized treatment arm, see Table 6-1 in the study protocol, or bariatric surgery).

The estimand is described by the following attributes (according to International Council for Harmonisation (ICH) E9(R1)[1]):

- **Population**: Adults employed in the US living with obesity
- **Endpoint**: Body weight reduction ≥10.0% (yes/no) from baseline to week 52
- **Treatment condition**: The randomised treatment regardless of discontinuation or dose reduction of investigational medicinal product, and regardless of initiating other anti-obesity therapies (as defined above).
- **Remaining intercurrent events**: None, all intercurrent events (discontinuation or dose reduction of investigational medicinal product and initiation of other anti-obesity therapies) are captured by the treatment condition attribute and handled by the treatment policy strategy.
- **Population-level summary**: The treatment effect will be quantified by the difference in proportions (calculated from odds) between treatment conditions.

Rationale for estimand: The primary estimand takes into account both safety and efficacy and reflect clinical practice in an employer setting. The primary estimand is thus relevant when evaluating treatment effect in an employer setting.

Secondary estimand

For secondary objectives assessed by binary endpoints, the estimands are similar to the primary estimand except for the endpoint attribute. For secondary objectives assessed by continuous endpoints, the estimands are similar to the primary estimand except for the endpoint and the population level summary, which is the difference between treatment conditions for means in change from baseline to week 52.

Additional estimand

An additional clinical question of interest for the primary objective is: What is the effectiveness of semaglutide 2.4 mg, as an adjunct to usual employer-offered lifestyle management, in adults employed in the US living with obesity had they remained on their randomised treatment (possibly on reduced dose) for the entire planned duration of the study, and not initiated other anti-obesity therapies (weight management drugs not included in the randomized treatment arm, see Table 6-1 in the study protocol, or bariatric surgery). This will be measured by the number of participants achieving ≥10% body weight loss at week 52

The estimand is described by the following attributes (according to ICH E9(R1)[1]):

- **Population**: Adults employed in the US living with obesity
- **Endpoint**: Body weight reduction ≥10.0% (yes/no) from baseline to week 52
- **Treatment condition**: The randomised treatment if participants had remained on randomized treatment (regardless of dose reductions) for the entire duration of the study and not initiated any other anti-obesity therapies (as defined above).
- **Remaining intercurrent events**: None, all intercurrent events (discontinuation or initiation of other anti-obesity therapies and dose reduction) are captured by the treatment condition attribute. Discontinuation or initiation of other anti-obesity therapies are handled by the hypothetical strategy. Dose reduction of investigational medicinal product is captured by the treatment policy strategy.
- **Population-level summary**: The treatment effect will be quantified by the difference in proportions (calculated from odds) between treatment conditions.

Rationale for estimand: The additional estimand aims at reflecting the treatment effect without the confounding effects of other anti-obesity therapies or treatment discontinuation.

Similar additional estimands apply to secondary objectives assessed by binary endpoints, except for the endpoint attribute. For secondary objectives assessed by continuous endpoints, the estimands are similar except for the endpoint attribute and the population level summary, which is the difference between treatment conditions for means in change from baseline to week 52.

1.2 Study Design

See Section 4 of the study protocol.

2 Statistical Hypotheses

The tests of superiority of semaglutide 2.4 mg to the group of other commercially available medication for chronic weight management (Other AOMs) for the primary and confirmatory secondary endpoints will be based only on analyses addressing the primary and secondary estimands.

For the primary estimand, the following confirmatory 1-sided hypothesis is planned to be tested for semaglutide 2.4 mg versus Other AOMs. Let the odds ratio be defined as OR = (odds[semaglutide 2.4 mg] divided by odds[Other AOMs]):

$$\underline{\text{Superiority:}}\ H0_1 : OR \leq 1 \text{ against } Ha_1 : OR > 1$$

For the two confirmatory secondary estimands, the following confirmatory 1-sided hypotheses are planned to be tested for semaglutide 2.4 mg versus Other AOMs. Let the mean difference be defined as μ = ([semaglutide 2.4 mg] minus [Other AOMs]):

$$\underline{\text{Superiority for change in body weight:}}\ H0_2 : \mu \geq 0.0 \text{ against } Ha_2 : \mu < 0.0$$

$$\underline{\text{Superiority for change in IWQOL-Lite-CT phys. func. domain:}}\ H0_2 : \mu \leq 0.0 \text{ against } Ha_2 : \mu > 0.0$$

Operationally, the hypotheses will be evaluated by 2-sided tests.

2.1 Multiplicity Adjustment

The type I error will be controlled in the strong sense using a hierarchical (fixed sequence) testing procedure. This is based on priority ordering of the null hypotheses and testing them in this order using the 2-sided 95% confidence interval approach until an insignificant result appears. For example, the second null hypothesis will only be tested if the first null hypothesis has been rejected in favour of semaglutide 2.4 mg.

The steps in the hierarchical testing procedure are as follows:

- Step 1: ≥10% body weight reduction (yes/no) from baseline (week 0) to end of treatment (week 52) superiority of semaglutide 2.4 mg versus Other AOMs.
- Step 2: Change in body weight (%) from baseline (week 0) to end of treatment (week 52) superiority of semaglutide 2.4 mg versus Other AOMs.
- Step 3: Change in IWQOL-Lite-CT physical function domain from baseline (week 0) to end of treatment (week 52) superiority of semaglutide 2.4 mg versus Other AOMs

3 Analysis Sets

The following analysis sets and data point sets are defined:

Table 3-1 Analysis sets

Participant Analysis Set	Description
Full analysis set (FAS)	Consists of all randomized participants. Participants will be included in the analyses according to the planned intervention.
Safety analysis set (SAS)	All participants who are exposed to study intervention. Participants will be included in the analyses according to the intervention they actually received.

Table 3-2 Defined data point sets

Defined data point sets (DPS)	Description
In-study (DPS1)	The time period where the participant is assessed in the study. The in-study observation period for a participant begins on the date of randomization and ends at the first of the following dates (both inclusive): • End of study visit • Withdrawal of consent • Last contact with participant (for participants lost to follow-up) • Death
Adherent (DPS2)	The consecutive time period[1] where the participant is adherent to treatment (possibly on a lower than intended dose) and has not initiated other anti-obesity therapies (weight management drugs not included in randomised treatment arm (see Table 6-1 in the study protocol) or bariatric surgery). End of the adherent period is defined as: • End of study visit • Treatment discontinuation[2] • Initiation of other anti-obesity therapies[3]

[1]The consecutive time period refers to the duration between the dispensing of the drug from the pharmacy as per the prescription, until the participant returns with a new prescription.

[2] Treatment discontinuation is defined as when the period with no coverage of prescription claims exceeds 20% of the expected treatment duration.

[3] Other anti-obesity therapies are defined as weight management drugs not included in the randomized treatment arm, see Table 6-1 in the study protocol, or bariatric surgery.

FAS and DPS1 are used to estimate the primary estimand and the secondary estimands.

FAS and DPS2 are used to estimate the additional estimands.

SAS and DPS1 are used to present safety data.

4 Statistical Analyses

4.1 General Considerations

This section is a summary of the planned statistical analyses.

The last available observation at or before randomization is used as the baseline value. If no assessments are available, the mean value at randomization across all participants is used as the baseline value.

All tests are tests of superiority of semaglutide s.c. 2.4 mg once weekly to Other AOMs. All estimated treatment contrasts between semaglutide 2.4 mg and Other AOMs will be reported together with the associated two-sided 95% confidence interval (CI) and corresponding p-value.

4.2 Primary Estimand Analysis

4.2.1 Definition of Endpoints

Body weight reduction ≥10.0% (yes/no)

A body weight reduction of at least 10% from baseline (week 0) to week 52 is defined as:

$$10\%\text{ weight responder} = \begin{cases}\text{yes if } \%\text{ weight change} \leq -10.0\% \\ \text{no if } \%\text{ weight change} > -10.0\%\end{cases}$$

where change from baseline (week 0) to week 52 in body weight (%) is defined as

$$\text{Change in body weight (\%)} = \frac{(\text{body weight at week 52} - \text{body weight at baseline})}{\text{body weight at baseline}} \times 100$$

4.2.2 Main Analytical Approach

The primary estimand will be estimated by logistic regression (LR) with randomised treatment as factor and baseline body weight (kg) as covariate.

Multiple imputation approach using retrieved subjects (RD-MI):

All available weight measurements (kg) at week 52 are used, and missing values at week 52 will be imputed. The imputation method for the primary analysis is a multiple imputation approach similar to the one described by McEvoy et al[4]. Imputation is done separately for each treatment group and also separately for participants being on and off treatment at week 52.

Average probabilities and treatment differences in probability with corresponding CIs will also be reported alongside the odds and odds ratios.

The multiple imputation approach is done in 3 steps:

1. **Imputation:** Defines imputation models using all available weight measurements at week 52. A separate imputation model will be fitted for each combination of treatment arm and treatment

status (on-/off-drug) at week 52, i.e. for four separate groups of participants, but using identical model parameterisations. The imputation model will be a linear regression of body weight (kg) at week 52 with gender (male/female) and baseline BMI (kg/m^2) (in categories -<35, 35-<40, ≥40) as factors and baseline body weight (kg), last available observation on treatment (LAO-OT) of body weight (kg) and timing of LAO-OT of body weight as covariates. If no LAO-OT exists post-baseline then the LAO-OT will be the baseline body weight and the timing will be 0. No interactions will be included. If the imputation model cannot be fit, 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. If the imputation model with only LAO-OT of body weight (kg) cannot be fit, the imputation will be done regardless of the randomised treatment arm. The reduced model will be used for four groups of participants. The estimated posterior distribution for the parameters (regression coefficients and variances) in the imputation models are then used to impute missing week 52 body weight values for each randomised treatment arm. This will be done 1,000 times and results in 1,000 complete data sets.

2. **Analysis:** Analysis of each of the 1,000 complete data sets, using the analysis model (logistic regression) results in 1,000 estimations. and the results from this logistic regression model will be used to predict the probability of achieving the responder endpoint for all participants had they (counterfactually) been assigned to each specific treatment. This is based on Steingrimsson et al. (2017)[2] and the estimation of the variance covariance matrix done by robust sandwich type estimator described by Stefanski & Boos (2002)[3]. Average probabilities and treatment difference in probability will be estimated by subtracting the average predicted probabilities and the corresponding standard errors will be calculated using the delta-method.
3. **Pooling:** Integrates the 1,000 estimation results into the final result using Rubin's formula.

The multiple imputations will be generated using Novo Nordisk trial number 95364741 as seed number. In addition to the seed number, it is specified that the dataset is sorted by subject ID prior to imputation.

For the estimands with binary endpoints, in any cases where response rates close to 0% or 100% in any treatment group lead to non-convergence, the Firth's maximum-likelihood estimation will be used when performing the logistic regression.

4.2.3 Sensitivity Analysis

A tipping point multiple imputation analysis (TP-MI) will be performed as sensitivity analysis. First, missing weight measurements (kg) at week 52 will be imputed as described for the primary estimand. Second, for the treatment a penalty (a weight change) is added to the imputed values at week 52 and then the logistic regression described for the primary estimand will be performed. The penalty will be increased until the conclusion from the primary estimand analysis is reversed. The 2-dimentional space of penalties covering the range from -30% to 30% will be explored for both treatment groups. This maximum is imposed since it may not be possible to reverse the conclusion for a binary endpoint and since systematic weight gains greater than 30% on top of the imputed values are highly unrealistic.

Non-retrieved participants as non-responders: For the analysis of body weight reduction ≥ 10% an analysis using non-retrieved participants as non-responders in the logistic regressions will be done. This analysis also targets the MAR assumption.

4.2.4 Supplementary Analysis

The statistical model used to estimate the additional estimand is the same as the one used for the primary estimand.

All available weight measurements (kg) at week 52 for participants that are adherent (see Section 3) until and including the week 52 assessment will be used. Measurements for participants that are not adherent at week 52 will be predicted. Prediction of missing values will be done by modelling all available body weight (kg) data from the adherent period for all participants using a mixed model for repeated measurements (MMRM) approach. The MMRM will be fitted using the same factors and covariate as for the primary analysis all nested within timing of visit (monthly, 30.5 days, intervals of time since randomisation). If the model does not converge, results from non-planned intermediate visits will be excluded from the analysis and nesting will then be within visit. An unstructured covariance matrix for measurements within the same participant will be employed, assuming that measurements for different participants are independent. If the analysis fails to converge with the unstructured covariance matrix, a suitable covariance structure will be used. The predicted values will be used to classify each participant as 10% responder or not. The logistic regression model will include randomised treatment as factor and baseline body weight (kg) as covariate. The results from the logistic regression model will be used to predict probabilities of achieving the response condition for all participants had they (counterfactually) been assigned to each specific treatment. Odds and treatment odds ratios will be estimated from the averaged predicted probabilities. Confidence intervals will be calculated using sandwich estimator. Average probabilities and treatment differences in probability with corresponding CIs will also be estimated.

4.3 Secondary Estimands Analysis

4.3.1 Confirmatory Secondary Estimands

4.3.1.1 Definition of Endpoints

Change in body weight (%) is defined in Section 4.2.1.

Change in IWQOL-Lite-CT physical function domain (PF) score is defined as:

$$\text{Change in PF score} = \text{PF score at week 52} - \text{PF score at baseline},$$

where scores are on the 0 - 100 scale.

4.3.1.2 Main Analytical Approach

The secondary estimand for change in body weight (%) will be analysed by a linear regression model (ANCOVA) with randomised treatment as factor and baseline body weight (kg) as covariate.

All available weight measurements (kg) at week 52 are used and missing values at week 52 will be imputed. The imputation method is identical to the imputation method for the primary estimand.

The secondary estimand for change in physical function assessed using IWQOL-Lite-CT will be analysed in a linear regression model identical to the one for estimation of the secondary estimand related to change in body weight (%) except that baseline physical function score is included as covariate instead of baseline body weight. Imputation of missing values will be done using the same approach as for the primary estimand.

4.3.1.3 Sensitivity Analysis

A tipping point multiple imputation analysis (TP-MI) will be performed as sensitivity analysis for each of the two endpoints. This will be done similar to the sensitivity analysis for the primary estimand, except that the penalty is added to the imputed values at week 52. The approach is to explore a range of penalties for both treatment groups, and the impact these would have on the study conclusions. The range of penalties is increased and a 2-dimensional space of penalties covering the range from -50 to 50 for IWQOL-Lite-CT score will be explored for both treatment groups.

4.3.1.4 Supplementary Analysis

The additional estimand for change in body weight (%) will be estimated using an MMRM approach. The MMRM will use the adherent data point set for all participants. The MMRM will be fitted using change in body weight (%) and the same factor and covariate as for the primary analysis all nested within timing of visit (monthly, 30.5 days, intervals of time since randomisation). If the model does not converge, results from non-planned intermediate visits will be excluded from the analysis and nesting will then be within visit. An unstructured covariance matrix for measurements within the same participant will be employed, assuming that measurements for different participants are independent. If the analysis fails to converge with the unstructured covariance matrix, a suitable covariance structure will be used.

The additional estimand for change in physical function assessed using IWQOL-Lite-CT is estimated as for change in body weight (%) except that baseline physical function score is included as covariate instead of baseline body weight.

4.3.2 Supportive Secondary Endpoints

4.3.2.1 Definition of Endpoints

Body weight reduction ≥15.0% (yes/no)

A body weight reduction of at least 15% from baseline (week 0) to week 52 is defined as:

$$15\%\ \text{weight responder} = \begin{cases} \text{yes if } \%\ \text{weight change} \leq -15.0\% \\ \text{no if } \%\ \text{weight change} > -15.0\% \end{cases}$$

Body weight reduction ≥20.0% (yes/no)

A body weight reduction of at least 20% from baseline (week 0) to week 52 is defined as:

$$\text{20\% weight responder} = \begin{cases} \text{yes if } \%\text{ weight change} \leq -20.0\% \\ \text{no if } \%\text{ weight change} > -20.0\% \end{cases}$$

Change in IWQOL-Lite for CT physical function domain score ≥ 14.6 (yes/no)

An improvement of at least 14.6 score points in IWQOL-Lite for CT physical function domain score is defined as:

$$\text{physical function responder} = \begin{cases} \text{yes if PF score change} \geq 14.6 \\ \text{no if PF score change} < 14.6 \end{cases}$$

where scores are on the 0 - 100 scale and calculation of IWQOL-Lite-CT PF is done according to the scoring manual.

Proportion of days covered (PDC) by study product.

The proportion of days covered by study product will be calculated based on the time on study product and length of planned study duration as follows:

$$\text{PDC (\%)} = \frac{\text{duration on study product}}{\text{planned study duration}} \times 100\%.$$

The duration on study product will be based on pharmacy card information, where the duration on study product will start on the date when the first study product is dispensed according to the pharmacy card (information is obtained from the pharmacy dispensing database). The duration on study product will then be calculated as "days supply", where "days supply" will be added up for each dispensing of study product based on the following rules:

- For consecutive dispensing of the same brand of study product, it is assumed that the study product dispensed previously will be used as planned and that the newer dispensed study product will be used subsequently.
- If a participant discontinues using a study product and does not initiate another, then it is assumed that all of the "days supply" of the previous study product will be used
- If a participant in the Other AOMs arm switches from one brand of study product to another then it is assumed that the participant discontinues on the previous study product on the date of dispensing of the new study product.
- If – based on the above rules – the amount of dispensed study product covers a period exceeding the date of the week 52 visit, then only the number of days until and including the date of the week 52 visit will be included in the duration on study product. If the week 52 visit is not attended, then the date of the planned week 52 visit will be used as the last possible date on study product.

Planned study duration is from the date of randomisation when the prescription card is handed out and until the week 52 visit or until the planned date of the week 52 visit if the week 52 visit is not performed.

Covered by study product ≥80% of days (yes/no).

Coverage by study product ≥80% of days is defined as:

$$80\%\ \text{PDC} = \begin{cases} \text{yes if } \text{PDC} \geq 80\% \\ \text{no if } \text{PDC} < 80\%' \end{cases}$$

where PDC is calculated as defined above.

Change in Work Limitations Questionnaire 25-item version (WLQ-25), total score.

Change in WLQ-25 total score (TS) is defined as:

$$\text{Change in WLQ-25 TS} = \text{WLQ-25 TS at week 52} - \text{WLQ-25 TS at baseline},$$

where calculation of WLQ-25 TS is done according to the WLQ-25 scoring manual (version 1). If WLQ-25 TS is multiplied by 100% then it according to the scoring manual is to be interpreted as the percentage of at-work productivity loss in the past two weeks relative to a healthy benchmark sample.

The analysis of WLQ-25 will only be performed for the subjects employed at baseline and only those assessments taken until the subject was employed will be included. In addition, subjects for whom a baseline score cannot be calculated according to the scaling and scoring manual due to too many items being rated as "does not apply to my job" are excluded from the analysis. If the participants become unemployed at post-baseline, then their data will not be included in subsequent analyses.

WLQ-25 physical demands score, WLQ-25 time management score, WLQ-25 mental-interpersonal score, WLQ-25 output score, WLQ-25 at-productivity loss score and WLQ-25 total score will be analysed as mentioned above.

In the scoring manual, it is specified that "If the answers to the physical scale items contradict the answers in the other 3 scales, we can reasonably infer that the respondent didn't notice the change in the response options." In such cases, the scoring manual describes a reversal of the physical scoring scale, but such a step would be a modification of the respondent's responses based on interpretation of the responses, and this contradicts with the fundamental definition of Patient Reported Outcomes. In agreement with the author of WLQ-25, Dr. Lerner, it has been decided to omit this step involving potential modification of responses (e-mail communication with the author of WLQ-25, Dr. Lerner, Vault eTMF: VV-TMF-5883040).

4.3.2.2 Analytical Approach

The secondary estimands for body weight reductions of at least 15% and 20%, respectively, will be analysed like the primary estimand for a body weight reduction of at least 10%.

All available weight measurements (kg) at week 52 are used and missing values at week 52 will be imputed. The imputation method is identical to the imputation method for the primary estimand.

The secondary estimand for change in IWQOL-Lite for CT physical function domain score ≥ 14.6 (yes/no) will be analysed like the primary estimand for a body weight reduction of at least 10% except that baseline physical function score is included as covariate instead of baseline body weight. Imputation of missing values will be done using the same approach as for the primary estimand.

PDC by study product will be reported descriptively by treatment arm based on observed data. Mean, standard deviation, median, 5% and 95% quantiles, minimum and maximum values will be presented.

Covered by study product ≥80% of days (yes/no) will be reported based on observed data as proportion of participants and number of participants by treatment arm.

The secondary estimand for change in WLQ-25 TS will be analysed like the secondary estimand for body weight (kg) except that baseline WLQ-25 TS is included as covariate instead of baseline body weight. Imputation of missing WLQ-25 TS will be done using same approach as for the primary estimand.

Analyses addressing the additional estimand

The additional estimand for body weight reductions of at least 15% and 20%, as well as the IWQOL-Lite CT physical function domain score ≥ 14.6, will be estimated using an MMRM approach, similar to the primary endpoint analysis.

The additional estimand for changes in WLQ-25 TS will be estimated using an MMRM approach, similar to the confirmatory secondary endpoint analysis.

4.4 Exploratory Endpoints Analysis

4.4.1 Definition of Endpoints

Proportion of days missed from work due to participant illness or not feeling well (PDMW)

PDMW is defined as:

$$\text{PDMW (\%)} = \frac{\text{number of days missed during the last month due to feeling unwell}}{30.5} \times 100\%.$$

assuming average number of working days in a month as 30.5, i.e. disregarding potential vacation periods, national holidays etc. Data for participants being unemployed at the time of reporting will be excluded.

Treatment satisfaction assessed by Abbreviated Treatment Satisfaction Questionnaire for Medication (TSQM-9), Global Satisfaction, total score (GS)

TSQM-9 GS at week 52 is simply defined as the score at week 52 of the TSQM-9 global satisfaction domain, where TSQM-9 GS is calculated according to the TSQM-9 user manual (version 2.3).

4.4.2 Analytical approach

PDMW will be reported descriptively by treatment arm based on observed data. Mean, standard deviation, median, 5% and 95% quantiles, minimum and maximum values will be presented. Results will be presented by time of reporting (by month, 30.5 days, since randomisation) and total of reported periods.

TSQM-9 GSTS will be reported descriptively by treatment arm based on observed data. Mean, standard deviation, median, 5% and 95% quantiles, and minimum and maximum values will be presented.

4.5 Safety Analysis

No formal safety analyses are planned.

4.5.1 Extent of Exposure

Extent of exposure is covered by reporting of the PDC endpoint.

4.5.2 Adverse Events

In general, results for AEs, medication errors (including misuse and abuse) and pregnancies will be reported by counts (proportions) of participants and counts of events.

An overview table including, but not limited to the following will be created:

- Serious adverse events (SAEs). Also split into following subcategories:
  - Severity
  - Relationship to study product
  - Outcome (recovered/resolved, not recovered/not resolved, …)
  - Action taken with study product as a result of the SAE
- Adverse events (AEs) leading to discontinuation of randomized treatment). Also split into following subcategories:
  - Serious (Y/N)
  - Severity
  - Relationship to study product
  - Outcome (recovered/resolved, not recovered/not resolved, …)
- AEs of COVID-19, irrespective of seriousness. Also split into following subcategories:
  - Serious (Y/N)
  - Severity
  - Relationship to study product
  - Outcome (recovered/resolved, not recovered/not resolved, …)
  - Action taken with study product as a result of the SAE
- Medication errors, misuse, and abuse. Also split into following subcategories:
  - Serious (Y/N)
  - Severity
  - Relationship to study product
  - Outcome (recovered/resolved, not recovered/not resolved, …)

- Action taken with study product as a result of the medication error, misuse, and abuse

The reported safety assessments (AEs) will be coded using the Medical Dictionary for Regulatory Activities (MedDRA Version 27.1) and descriptively summarised by System Organ Class (SOC) and preferred term. No AE-rates will be calculated.

Furthermore, adverse events (SAEs, AEs leading to treatment discontinuation and AEs for COVID-19) will be reported by system organ class, high level group term and preferred term for the in-study data point set (DPS1) for safety analysis set.

Pregnancies in female participants and pregnancy outcome [until age 1 month] and AEs in the fetus or newborn infant will be listed.

4.6 Other Analysis

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

4.6.1 Other Variables and/or Parameters

Not applicable.

4.6.2 Subgroup Analysis

There is no subgroup analysis planned for this study.

4.7 Interim Analysis

There is no interim analysis planned for this study.

4.8 Changes to Protocol-planned Analysis

- In the protocol, the responder threshold for IWQOL-Lite-CT is described as ">14.6". This is updated to "≥14.6" in this SAP.
- For the MMRM analyses, it is specified in the protocol that factors and covariate are "all nested within visit". This has been changed to: "The MMRM will be fitted using the same factors and covariate as for the primary analysis all nested within timing of visit (monthly, 30.5 days, intervals of time since randomisation). If the model does not converge, results from non-planned intermediate visits will be excluded from the analysis and nesting will then be within visit." The background for this change is that the timing of visits may vary substantially between participants and therefore "timing" and not "visit" has to be used. Nevertheless, since differences in the timing of visits may result in a structure with few observations for some time intervals, it may happen that the MMRM does not converge, and in such a case only planned visits are included in the model.

Changes of a typographical nature:

- In the estimand descriptions, "Rationale for estimand" has been changed from being in the list of estimand attributes to being body text, since the rationales are not attributes of the estimand.

- In the protocol in section 9.3.3.1 “Confirmatory Secondary Estimands” the first line says: “The primary estimand will…”. This has been corrected to read: “The secondary estimand will…”. Furthermore, for the IWQOL-Lite-CT PF endpoint analysis, it has now been clarified that it addresses the secondary estimand for IWQOL-Lite-CT PF.

5 Sample size determination

See Section 9.5 in the study protocol.

6 Supporting Documentation

See Section 10 in the study protocol.

7 References

1. European Medicines Agency. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials, Step 2b (EMA/CHMP/ICH/436221/2017). . 30 Aug 2017.
2. Steingrimsson JA, Hanley DF, Rosenblum M. Improving precision by adjusting for prognostic baseline variables in randomized trials with binary outcomes, without regression model assumptions. Contemp Clin Trials. 2017;54:18-24.
3. Stefanski LA, Boos DD. The Calculus of M-Estimation. *The American Statistician*. 56(1), 29–38. 2002.
4. McEvoy BW. Missing data in clinical trials for weight management. J Biopharm Stat. 2016;26(1):30-6.