Cover Page for Statistical Analysis Plan

Sponsor name:	Novo Nordisk A/S
NCT number	NCT02918279
Sponsor trial ID:	NN8022-4180
Official title of study:	Effect of liraglutide for weight management in pubertal adolescent subjects with obesity
Document date:	11-December-2019

16.1.9 Documentation of statistical methods

List of contents

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

.

Statistical Analysis Plan

Trial ID: *NN8022-4180*

Effect of liraglutide for weight management in pubertal adolescent subjects with obesity

56-week, double-blind, randomised, parallel-group, placebo-controlled multi-national trial followed by a 26-week period off study-drug

Trial Phase: 3a

Author
Name: [redacted]
Department: [redacted]

Table of contents

Page

List of abbreviations

AE	*adverse event*
ADA	*American Diabetes Association*
AD	*available droup-out*
ANOVA	*analysis of variance*
AT	*available on treatment*
BMI	*body mass index*
BMI SDS	*body mass index standard deviation score*
BOCF	*baseline observation carried forward*
CI	*confidence interval*
CRF	*case report form*
CTR	*clinical trial report*
CV	*coefficient of variance*
EoT	*end-of-text*
FAS	*full analysis set*
HbA1c	*glycosylated haemoglobin*
ISPAD	*international society for pediatric and adolescent*
LOCF	*last observation carried forward*
MD	*missing drop-out*
MedDRA	*Medical Dictionary for Regulatory Activities*
MT	*missing on treatment*
PD	*pharmacodynamics*
PK	*pharmacokinetics*
PP	*per protocol*
SAE	*serious adverse event*
SAP	*statistical analysis plan*
SAS	*safety analysis set*
SD	*standard deviation*
SE	*standard error*
LAR	*legally acceptable representative*

1 Introduction

1.1 Trial information

Trial design

This is a 56-week double-blind, randomised, parallel-group, placebo-controlled, multi-national trial followed by a 26-week period off study-drug. The trial will be conducted in pubertal adolescents with obesity aged 12 to less than 18 years. Subjects will be randomised 1:1 to receive liraglutide or placebo. The randomisation will be stratified according to pubertal and glycaemic status (see protocol Sections 5.2 and 11).

At least 30% of subjects will be from areas with lifestyle and nutrition comparable to those in the European Union. All subjects will undergo counselling in healthy nutrition and physical activity for weight loss and must be prescribed a structured programme from the beginning of the 12-week run-in period and continuing through the 26-week follow-up period off trial drug.

Primary objective

- To compare the efficacy of liraglutide versus placebo on weight loss in adolescent subjects with obesity after 56 weeks of treatment

Secondary objectives

- To compare the efficacy of liraglutide versus placebo on glycaemic control, cardiovascular risk factors and Impact of Weight on Quality of Life-Kids (IWQOL-Kids) in adolescent subjects with obesity after 30 and 56 weeks of treatment
- To compare the safety of liraglutide versus placebo in adolescent subjects with obesity after 30 and 56 weeks of treatment
- To examine the potential rebound effect from end of treatment at week 56 to week 82

Further details are described in the protocol section 4.2.

1.2 Scope of the statistical analysis plan

This SAP is based on the protocol "Effect of liraglutide for weight management in pubertal adolescent subjects with obesity", version [final protocol version 1.0].

2 Statistical considerations

Results from the statistical analysis will generally be presented by two-sided confidence intervals (CIs) with a confidence level of 95%. Superiority will be claimed if the two-sided p-value is less than 5% and the treatment estimate favours liraglutide. If the upper limit is below 0, superiority of liraglutide against placebo can be concluded.

The full analysis set (FAS) will be used in the analysis of the efficacy endpoints. For the safety endpoints, the safety analysis set (SAS) will be used. The definition of the analysis sets is given section 2.1.

Laboratory values below the lower limit of quantification (LLOQ) will be set to ½LLOQ.
The baseline value will be defined as the last measured and available value from randomisation (V9) and end of run-in (V8), if not otherwise specified.

2.1 Definitions of analysis sets

The following analysis sets are defined in accordance with the ICH E9[1, 2].

- FAS: includes all randomised subjects who have received at least one dose of trial product and have any post-randomisation data. The statistical evaluation of the FAS will follow the intention-to-treat (ITT) principle and subjects will contribute to the evaluation "as randomised".
- SAS: includes all subjects exposed to at least one dose of trial product. Subjects in the SAS will contribute to the evaluation "as treated".

Before data are locked for statistical analysis, a blinded review of all data will take place. Any decision to exclude a subject or single observation from the statistical analysis is the joint responsibility of the Novo Nordisk trial statistician, the international trial manager and the international medical specialist. Exclusion of data from analyses will be used restrictively and normally no data should be excluded from the FAS. The subjects or observations to be excluded, and the reasons for their exclusion must be documented and signed by those responsible before database lock. The subjects and observations excluded from analysis sets, and the reason for this, will be described in the clinical trial report.

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

2.2 Primary endpoint

The primary endpoint is:

- Change from baseline in BMI SDS after 56 weeks of treatment.

The objective is to show that liraglutide is superior to placebo in obtaining weight loss. Let $\mu_{liraglutide}$ and $\mu_{placebo}$ denote the mean change in BMI SDS for liraglutide and placebo, respectively. The null hypothesis and the alternative hypothesis are:

H0: $\mu_{liraglutide} = \mu_{placebo}$ against the alternative HA: $\mu_{liraglutide} \neq \mu_{placebo}$

The null-hypothesis will be rejected on a 5% level if the two-sided 95% CI of the treatment FAS difference $\mu_{liraglutide} - \mu_{placebo}$ excludes 0. If the upper limit is below 0, superiority of liraglutide against placebo can be concluded.

The hypothesis will be tested using an analysis of covariance (ANCOVA) model using including the factors, covariates and interaction term listed in below table.

Table 2-1: Factors and covariates for the analysis of the primary endpoints

Factors and covariates at baseline	Type	Categories
Randomised treatment	Factors	Liraglutide 3.0 mg, Placebo
Sex	Factors	Female, Male
Region	Factors	Europe, North America
Glycaemic category	Factors	Yes, No*
Tanner stage, Glycaemic category	Interaction factor	Not applicable
Tanner stage	Factors	Stage 2 and 3 together, Stage 4 and 5 together
Baseline BMI SDS	Covariate	Not applicable
Age	Covariate	Not applicable

*Yes: dysglycaemic, No: non-dysglycaemic.

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.

Missing data in the main analysis will be handled by the following multiple imputation (MI) method. A pattern mixture model approach is applied where withdrawn subjects or treatment discontinued subjects without a follow-up visit are assumed to respond as if treated with placebo for the entire trial. Multiple copies (100 copies) of the full dataset will be generated by imputing missing values based on estimated parameters for the placebo group. These missing data will be handled as mentioned in protocol section 17.3.

Table 2-2: Taxonomy of week 30, 56 and 82 assessments being available or missing

Assessment at week 30, 56 and 82	On randomised treatment at week 30, 56 and 82	Type description	Type abbreviation
Available	Yes	Available on randomised treatment: Subjects who did not discontinue randomised treatment prematurely.	AT
	No	Available drop-outs: Subjects who discontinued randomised treatment prematurely but returned to have an assessment at week 30, 56 and 82; so-called retrieved drop-outs.	AD
Missing	Yes	Missing on randomised treatment: Subjects who did not discontinue randomised treatment prematurely.	MT
	No	Missing drop-outs: Subjects who discontinued randomised treatment prematurely and did not return to have an assessment at week 30, 56 and 82; so-called non-retrieved drop-outs.	MD

2.2.1 Sensitivity analysis

To investigate the sensitivity of the results of the main analysis of the primary endpoint about the handling of missing data, a sensitivity analysis will be performed as mentioned below (further details refer protocol section 17.3).

- An ANCOVA will be performed with imputation of missing values according to the last observation carried forward (LOCF) method. The model will include terms for treatment, sex, region, baseline glycaemic category, stratification factor for Tanner stage and interaction between baseline glycaemic category and stratification factor for Tanner stage will be included as fixed effect and baseline BMI SDS and baseline age as covariates. The response variable will be the last available measurement of BMI SDS obtained within the 56-week double-blind period of the trial.
- The same type of ANCOVA as above will be performed but using an imputation of missing values according to the baseline observation carried forward (BOCF) method. Missing measurements of BMI SDS at 56 weeks will be imputed by the corresponding baseline values with this method.

- The same type of ANCOVA as above will be performed without imputation by only including subjects who completed the 56 weeks double-blind period.
- A mixed model for repeated measurements (MMRM) will be applied where all post baseline BMI SDS measurements obtained at planned visits during the 56-week double-blind period will enter as the dependent variables, and visit, treatment, sex, region, baseline glycaemic category, stratification factor for Tanner stage and interaction between baseline glycaemic category and stratification factor for Tanner stage will be included as fixed effects, and baseline BMI SDS and baseline age as covariates. All these factors and covariates will be nested under visit, which is technically the same as introducing the corresponding interaction terms in the model. An unstructured covariance matrix for the BMI SDS measurements within subject will be employed.

  For further details refer protocol section 17.3.

2.3 Handling of missing values

2.3.1 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 is missing at the randomisation visit and an assessment was available at screening, then the screening value will be used as the baseline value.

2.3.2 Handling of missing values at weeks 30, 56 and 82

Missing values at weeks 30, 56 and 82 will be imputed and the relevant endpoints will be analysed from the imputed values. Several approaches for imputation of missing values at weeks 30, 56 and 82 will be applied. First, a description of the primary imputation approach used to address the effectiveness for the primary endpoint is given. This is followed by a description of several sensitivity analyses.

Primary approach for handling missing values
The primary approach for multiple imputations of missing values of BMI SDS at week 30, 56 and 82 (type MT+MD Table 2-2) for both the liraglutide 3.0 mg and placebo group is by sampling all available assessments at respective landmark visits in the placebo group (type AT+AD). This approach is also known as jump to reference and makes the assumptions 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[3]. The multiple imputation approach is done in four steps.

Imputation: Step 1: 100 copies of the dataset will be generated.

Analysis: Step 2: Impute the missing value from placebo completers by fitting enriched regression model. Model will be fitted with the factors, covariates and interaction term in the same order as mentioned in Table 2-3. The estimated parameters and their variances from this model were used to impute missing values at 56 weeks for subjects in both treatment arms, based on their factor levels and the values of the covariates.

Analysis: Step 3: For each of the 100 complete datasets, the change from baseline in BMI SDS at 56 weeks will be analysed using the main ANCOVA model with factors and covariates as mentioned in Table 2-1.

Pooling: Step 4: Pool the 100 estimation results into a final result using Rubin's formula.

The imputation model in step 2 uses placebo subjects from FAS with non-missing BMI SDS at baseline and week 56. The imputation model is a linear regression of BMI SDS at week 56 on the factors and covariates listed in Table 2-3 (except randomised treatment arm) with interaction term included in the model. The order of the factors and covariates has been retained as mentioned in Table 2-3 while fitting imputation model. The estimated posterior distribution for the parameters (regression coefficients and variances) in the imputation model is then used to impute missing week 56 BMI SDS values for both randomised treatment arms.

Table 2-3: Factors and Covariate for imputation model

Factors and covariates at baseline	Type	Categories	Order
Sex	Factors	Female, Male	1
Region	Factors	Europe, North America	2
Baseline Glycaemic category	Factors	Yes, No	3
Tanner stage	Factors	Stage 2 and 3 together, Stage 4 and 5 together	4
Tanner stage, Glycaemic category	Interaction factor	Not applicable	5
Baseline BMI SDS	Covariate	Not applicable	6
Baseline Waist circumference	Covariate	Not applicable	7
Baseline Age	Covariate	Not applicable	8
Baseline HbA1c	Covariate	Not applicable	9

The multiple imputations will be generated using Novo Nordisk trial number 80224180 as seed number.

A similar procedure of imputation will be followed for all the secondary endpoints as mentioned in Table 2-4 except while imputing missing values for waist circumference assessment, baseline waist circumference which will not be included in the imputation model. Similarly, for HbA_{1c} assessments, baseline HbA_{1c} will not be included in the imputation model.

For missing glycaemic category, the imputation will be done for missing FPG and HbA1c assessments using the imputation model mentioned in Table 2-4 except for HbA1c, baseline HbA1c which will not be used while imputing the missing values.

Table 2-4: Statistical analysis to address primary and secondary objectives

SI No.	Endpoints	Landmark visits	Endpoint type	Imputation approach and Statistical model	Analysis set	Sensitivity analyses
	Primary Endpoint					
1	Change from baseline in BMI SDS	Week 56	Continuous	J2R-MI ANCOVA	FAS	• ANCOVA (LOCF) • ANCOVA (BOCF) • ANCOVA (no imputation, with 56 week completers) • MMRM
	Secondary Endpoint					
2	Change from baseline in BMI SDS	Week 30, Week 82	Continuous	J2R-MI ANCOVA	FAS	NA
3	Change from baseline in BMI SDS (%)	Week 30, Week 56	Continuous	J2R-MI ANCOVA	FAS	NA
4	Change from baseline in BMI SDS	Week 56 to 82 (Week 56 as baseline)	Continuous	J2R-MI ANCOVA	FAS	NA
5	Change from baseline BMI	Week 30, Week 56	Continuous	J2R-MI ANCOVA	FAS	NA

	(Kg/m^2)					
6	Percentage of Subjects achieving >=5% reduction in baseline BMI	Week 30, Week 56, Week 82	Categorical	J2R-MI Logistic Regression	FAS	NA
7	Percentage of Subjects achieving >=10% reduction in baseline BMI	Week 30, Week 56, Week 82	Categorical	J2R-MI Logistic Regression	FAS	NA
8	Change from Baseline in PRO score (IWQoL)	Week 30, Week 56	Continuous	J2R-MI ANCOVA	FAS	NA
9	Change from baseline in Glycaemic category	Week 30, Week 56	Categorical	J2R-MI Logistic Regression	FAS	NA
10	Change from baseline in Body weight (kg and %)	Week 30, Week 56	Continuous	J2R-MI ANCOVA	FAS	NA
11	Change from baseline in Waist circumference (cm)	Week 30, Week 56	Continuous	J2R-MI ANCOVA	FAS	NA
12	Change from baseline in Waist-to-hip circumference ratio (ratio)	Week 30, Week 56	Continuous	J2R-MI ANCOVA	FAS	NA
13	Change from baseline Systolic Blood Pressure (mmHg)	Week 30, Week 56	Continuous	J2R-MI ANCOVA	FAS	NA
14	Change from	Week 30,	Continuous	J2R-MI	FAS	NA

	baseline Diastolic Blood Pressure (mmHg)	Week 56		ANCOVA		
15	Change from baseline Pulse (beats/min)	Week 30, Week 56	Continuous	J2R-MI ANCOVA	SAS	NA
16	Change from baseline in Cardiovascular biomarker (hsCRP) (mg/L)	Week 30, Week 56	Continuous	J2R-MI ANCOVA (with log transformed response and baseline values)	FAS	NA
17	Change from baseline in TC (mmol/L)	Week 30, Week 56	Continuous	J2R-MI ANCOVA (with log transformed response and baseline values)	FAS	NA
18	Change from baseline in LDL-cholesterol (mmol/L)	Week 30, Week 56	Continuous	J2R-MI ANCOVA (with log transformed response and baseline values)	FAS	NA
19	Change from baseline in HDL-cholesterol (mmol/L)	Week 30, Week 56	Continuous	J2R-MI ANCOVA (with log transformed response and baseline values)	FAS	NA
20	Change from baseline in non-HDL cholesterol (mmol/L)	Week 30, Week 56	Continuous	J2R-MI ANCOVA (with log transformed response and baseline values)	FAS	NA
21	Change from baseline in VLDL (mmol/L)	Week 30, Week 56	Continuous	J2R-MI ANCOVA (with log transformed response and baseline values)	FAS	NA
23	Change from baseline in TG (mmol/L)	Week 30, Week 56	Continuous	J2R-MI ANCOVA (with log transformed response and baseline values)	FAS	NA

24	Change from baseline in FFA (mmol/L)	Week 30, Week 56	Continuous	J2R-MI ANCOVA (with log transformed response and baseline values)	FAS	NA
25	Change from baseline in HbA1c (%)	Week 30, Week 56	Continuous	J2R-MI ANCOVA	FAS	NA
26	Change from baseline in FPG (mmol/L)	Week 30, Week 56	Continuous	J2R-MI ANCOVA	FAS	NA
27	Change from baseline in fasting insulin (pmol/L)	Week 30, Week 56	Continuous	J2R-MI ANCOVA (with log transformed response and baseline values)	FAS	NA
28	Change from baseline in fasting C-peptide (nmol/L)	Week 30, Week 56	Continuous	J2R-MI ANCOVA (with log transformed response and baseline values)	FAS	NA
29	Change from baseline in HOMA-B	Week 30, Week 56	Continuous	J2R-MI ANCOVA (with log transformed response and baseline values)	FAS	NA
30	Change from baseline in HOMA-IR	Week 30, Week 56	Continuous	J2R-MI ANCOVA (with log transformed response and baseline values)	FAS	NA

2.4 Secondary endpoints

All the secondary supportive efficacy endpoints listed below will be analysed as per protocol section 17.4.

2.4.1 Supportive secondary endpoints

2.4.1.1 Efficacy endpoints

- From Baseline to week 82:
  - Percent of subjects achieving ≥5% reduction in baseline BMI at weeks 30, 56 and 82
  - Percent of subjects achieving ≥10% reduction in baseline BMI at weeks 30, 56 and 82

- Change in BMI SDS from baseline to 30 and 82 weeks and change from 56 weeks to 82 weeks
- Change from baseline to 30 and 56 weeks in:
    - BMI
    - Body weight (kilogram [kg], pounds [lb] and percent [%])
    - Waist circumference
    - Waist-to-hip circumference ratio
    - Cardiovascular risk factors: high sensitivity C-reactive protein (hsCRP) and fasting lipids: total cholesterol (TC), low density lipoprotein cholesterol (LDL-cholesterol), high density lipoprotein cholesterol (HDL-cholesterol), non-HDL cholesterol, very low density lipoprotein cholesterol (VLDL-cholesterol), triglycerides (TG) and free fatty acids (FFA)
    - Systolic and diastolic blood pressure
    - Glucose metabolism: glycosylated haemoglobin (HbA1c), fasting plasma glucose (FPG), fasting insulin, fasting C-peptide, glycaemic category and homeostasis model assessment of beta-cell function and insulin resistance parameters (HOMA-B and HOMA-IR)
    - Patient reported outcome (PRO) assessed by Impact of Weight on Quality of Life-Kids (IWQOL-Kids)

- Change from 56 weeks to 82 weeks in:
    - BMI
    - Body weight (kilogram [kg], pounds [lb] and percent [%])
    - Waist circumference
    - Waist-to-hip circumference ratio
    - Cardiovascular risk factors: high sensitivity C-reactive protein (hsCRP) and fasting lipids: total cholesterol (TC), low density lipoprotein cholesterol (LDL-cholesterol), high density lipoprotein cholesterol (HDL-cholesterol), non-HDL cholesterol, very low density lipoprotein cholesterol (VLDL-cholesterol), triglycerides (TG) and free fatty acids (FFA)
    - Systolic and diastolic blood pressure
    - Glucose metabolism: glycosylated haemoglobin (HbA1c), fasting plasma glucose (FPG), fasting insulin, fasting C-peptide, glycaemic category and homeostasis model assessment of beta-cell function and insulin resistance parameters (HOMA-B and HOMA-IR)

In addition to all the supportive secondary efficacy analysis following added endpoints will be analysed similarly as per primary endpoint statistical model mentioned in protocol section 17.3.

- Change from baseline to 30 and 56 weeks:
    - BMI SDS (%)

- Glycaemic category will be summarised by frequency count for each treatment group
- Nutritional compliance will be summarised using descriptive statistics for each treatment group.

2.4.1.2 Safety endpoints

Descriptive statistics for all safety endpoints mentioned below 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.

- Number of treatment emergent adverse events
- Number of treatment emergent hypoglycaemic episodes:
    - According to ADA/ISPAD classification
    - According to Novo Nordisk/ISPAD classification
- Occurrence of anti-liraglutide antibodies
- Change from baseline to 56 weeks in bone age assessment
- Change from baseline to 30 and 56 weeks in:
    - Pulse
    - Electrocardiogram (ECG)
    - Laboratory parameters:
        - Haematology: haemoglobin, haematocrit, thrombocytes, erythrocytes, leucocytes, differential count (eosinophils, neutrophils, basophils, lymphocytes, monocytes)
        - Biochemistry: creatinine, creatinine kinase, urea (BUN), albumin, bilirubin (total), alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (ALP), sodium, potassium, calcium total, calcium, albumin- corrected, amylase, lipase, and carcinoembryonic antigen (CEA)
        - Hormone levels: calcitonin, insulin-like growth factor-1 (IGF-1), thyroid stimulating hormone (TSH), free thyroxine (free T4), dehydroepiandrosterone sulfate (DHEAS), luteinising hormone (LH), follicle stimulating hormone (FSH), estradiol (females), testosterone (males), prolactin, adrenocorticotropic hormone (ACTH), cortisol
        - Bone metabolism markers: Type 1 collagen N-telopeptide (NTX1), type 1 C-telopeptide (CTX1), procollagen 1 N-terminal propeptide (P1NP), alkaline phosphatase (bone)
    - Pubertal status
    - Physical examination
    - Height standard deviation score (SDS)
    - Mental health assessed by Columbia Suicidality Severity Rating Scale (C-SSRS) and Patient Reported Health Questionnaire 9 (PHQ-9)

- Change from 56 weeks to 82 weeks in:
  - Pulse
  - Laboratory parameters:
    - Haematology: haemoglobin, haematocrit, thrombocytes, erythrocytes, leucocytes, differential count (eosinophils, neutrophils, basophils, lymphocytes, monocytes)
    - Biochemistry: creatinine, creatinine kinase, urea (BUN), albumin, bilirubin (total), alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (ALP), sodium, potassium, calcium total, calcium, albumin- corrected, amylase, lipase, and carcinoembryonic antigen (CEA)
    - Hormone levels: calcitonin, insulin-like growth factor-1 (IGF-1), thyroid stimulating hormone (TSH), free thyroxine (free T4), dehydroepiandrosterone sulfate (DHEAS), luteinising hormone (LH), follicle stimulating hormone (FSH), estradiol (females), testosterone (males), prolactin, adrenocorticotropic hormone (ACTH), cortisol
    - Bone metabolism markers: Type 1 collagen N-telopeptide (NTX1), type 1 C-telopeptide (CTX1), procollagen 1 N-terminal propeptide (P1NP), alkaline phosphatase (bone)
  - Pubertal status
  - Physical examination
  - Height standard deviation score (SDS)
  - Patient Reported Health Questionnaire 9 (PHQ-9)

2.5 Endpoint derivation methods

2.5.1 BMI Standard deviation score (BMI SDS) and Height Standard deviation score (Height SDS)

BMI SDS and height SDS score will be calculated using external reference data on BMI and height from WHO[4]

The following procedure is recommended to calculate a z-score for an individual child with measurement y at age t:

1: Calculate

$$z_{ind} = \frac{[y/M(t)]L(t) - 1}{S(t)\,L(t)}$$

2 : Compute the final z-score (Z^*_{ind}) of the child for the indicator as:

$$z^*_{ind} = \begin{cases} z_{ind} & if \quad |z_{ind}| \leq 3 \\ 3 + \left(\frac{y - SD3pos}{SD23pos} \right) & if \quad z_{ind} > 3 \\ -3 + \left(\frac{y - SD3neg}{SD23neg} \right) & if \quad z_{ind} < -3 \end{cases}$$

Where:

- L(t), M(t) and S(t): Box-cox power, median and CV respectively
- y: Individual BMI or height value
- SD3pos: is the cut-off 3 SD calculated at t (age) by LMS method:
  $SD3pos=M(t)[1+L(t)*S(t)*(3)]^{1/L(t)}$

- SD3neg : is the cut-off -3 SD calculated at t by the LMS method:
  $SD3neg=M(t)[1+L(t)*S(t)*(-3)]^{1/L(t)}$

- SD23pos : is the difference between the cut-offs 3 SD and 2 SD calculated at t by LMS method:
  $SD23pos=M(t)[1+L(t)*S(t)*(3)]^{1/L(t)} - M(t)[1+L(t)*S(t)*(2)]^{1/L(t)}$
- SD23neg : is the difference between the cut-offs -2 SD and -3 SD calculated at t by LMS method:
  $SD23neg = M(t)[1+L(t)*S(t)*(-2)]^{1/L(t)} - M(t)[1+L(t)*S(t)*(-3)]^{1/L(t)}$

To illustrate the procedure, an example with BMI-for-age for boys is provided below and displayed in Figure 1.

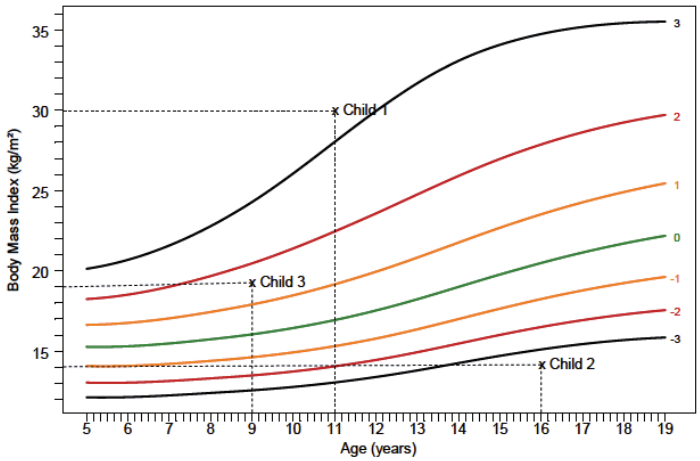

Figure 1 **Examples of children/adolescents ranked according to the 2007 WHO BMI-for-age reference.**

Child 1: 11 year-old boy with BMI=30

L=1.7862 M=16.9392 S=0.11070

$$Z=\frac{[30.0/16.9392]^{((-1.7862))}-1}{0.11070*(-1.7862)} = 3.24 > 3$$

$$SD3=16.9392*[1+(-1.7862)*0.11070*(3)]^{1/(-1.7862)} = 28.03$$
$$SD2=16.9392*[1+(-1.7862)*0.11071*(2)]^{1/(-1.7862)} = 22.45$$

SD23 = 28.03 – 22.45 = 5.58

$$Z^{*}_{ind}= 3+(30-28.03/5.58) = 3.35 \text{ (BMI SDS)}$$

Similar calculation will be done for height SDS.

2.5.2 IWQoL-Kids

The impact of weight on Quality of Life-Kids for Clinical Trials Version contains 27 questions with categorical answers (never, rarely, sometimes, usually, always). All these 27 questions are categorised into 4 domains (Physical Comfort, Body Esteem, Social Life, Family Life) scores. The Total score is the sum of all the 4 domain scores.

Table 2-5: Composite scores

Composite score	Items	Total number of items
Physical Comfort	1 - 6	6
Body Esteem	7 - 15	9
Social Life	16 - 21	6
Family Life	22 - 27	6
Total	1 - 27	27

The derivation of the composite scores within the ADQSIWQL dataset can be done in several ways, but below procedure was followed on this trial.

The categorical values of the questionnaire are converted into numbers (in ADQSIWQL) as below:

Table 2-6: Conversion of categorical values to numeric values

	Categorical values	Numeric values
Items 1- 27		
	Never True	1
	Rarely True	2
	Sometimes True	3
	Usually True	4
	Always True	5

The maximum and minimum possible score for each item will be 5 and 1, respectively. Composite score will be calculated for all four domains, including a total composite score will be calculated including all the domains.

For all four composite scores, at least 50% of the questions within a composite should be answered to calculate the score. For the total score, at least 75% of the 27 items should be answered.

The algorithm used for the derivation of a composite score (per subject and per visit):

((T*S_max)-(T*C_avg))/((T*S_max)-(T*S_min))*100%

where,

T is the total number of items in the composite score (see Table 2-5).
C_avg is the average score of the composite (for subject X at visit X).
S_max is the maximum possible score value (5).
S_min is the minimum possible score value (1).

2.5.3 C-SSRS questionnaire

The Columbia-Suicide Severity Rating Scale questionnaire is used to assess the subject's mental state. It contains two sets of questions; one meant for the screening/randomisation visit (C-SSRS-01; baseline) and the other for the rest of the visits (C-SSRS-02; since last visit). At screening/randomisation, the questionnaire is part of the exclusion criteria (suicidal behaviour, previous suicide attempts). During the trial, the questionnaire is used to evaluate the subject's mental status between visits. The first part of the questionnaire is to build up around 'yes/no' questions, which depending on the answers trigger other parts of the questionnaire

2.5.4 PHQ-9 questionnaire

The Patient Health Questionnaire 9 is used to assess the mental health of the subject. PHQ-9 is done throughout the trial period. It contains 9 questions which are answered with a categorical rating (not at all, several days, more than half the days, nearly every day) translated directly into a number (0-3). The numbers are summed at the bottom of the questionnaire by the investigator. Similar to C-SSRS-01, this questionnaire is used at screening/randomisation as part of the exclusion criteria, i.e., if the total score is ≥15 then the subject should be excluded from the trial.

2.6 Pharmacokinetic and Pharmacodynamic modelling

The Pharmacokinetic and pharmacodynamic analysis will be done as mentioned in the protocol section 17.5.

2.7 Reporting Endpoints

All the continuous endpoints mentioned in section 2.2 and 2.4 will be summarised by treatment group for in-trial period using descriptive statistics such as mean, SD, 5^{th} percentile and 95^{th} percentile, minimum and maximum.

Lipids will be summarised by treatment group using descriptive statistics such as mean, geometric mean, SD, CV, 5th percentile and 95th percentile, minimum and maximum accept HbA1c and FPG which will be summarised as rest of the continuous endpoints.

For the categorical endpoints mentioned in section 2.2 and 2.4 will summarised by treatment group using frequency counts and percentage by treatment groups.

3 Changes to the statistical analyses planned in the protocol

In this SAP the following changes is been made to the statistical consideration in the protocol are as elaborated below for:

3.1 Primary endpoint

- Primary analysis will be performed on in-trial period.
- Tanner stage and glycaemic category collected at randomisation visit will be considered as fixed effect while performing ANCOVA for change from baseline to week 56.
- Tanner stage for each subject at baseline is been derived and grouped using the information received from eCRF.

3.2 Supportive secondary efficacy endpoints

- All the secondary supportive analysis will be performed on in-trial period.
- To analyse the change in BMI SDS from week 56 to week 82 subjects who have completed the treatment (treatment completers), only their assessment at week 56 will be considered as baseline. No change from week 56 to 82 will be calculated for treatment discontinued subjects and for withdrawn subjects unless they have week 56 visit assessment collected.
- Multiple imputation will not to be considered for change in BMI SDS from week 56 to 82 parameter if there is no missing assessment.
- Tanner stage and glycaemic category collected at randomisation visit will be considered as fixed effect while performing ANCOVA for change from baseline to week 30, 56 and 82 and change from week 56 to 82.
- For "percentage of subjects achieving ≥5% reduction" and "Percentage of subjects achieving ≥10% reduction" in baseline BMI at week 30, 56 and 82. Missing imputation will be handled as mentioned in section 2.3 with the baseline BMI as covariate instead of baseline BMI SDS.
- Body weight with standard units (Kg) will only be used in all summary and analysis outputs.

Below additional supportive secondary endpoint is been included in section 2.4

- Change from baseline to 30 and 56 weeks in:
    - BMI SDS (%)

3.3 Supportive secondary safety endpoints

- For overall tanner staging, the maximum scale irrespective of the categorical question will be considered.
- For adverse events in run-in period all the subjects who has attended at least one run-in visit will be included in the run-in outputs.
- C-SSRS will only be summarised using summary statistics (frequency count, percentage) for screening, baseline, week 30, week 56 and week 82 and the shift from baseline to 30 and 56 week and from week 56 to 82 will not be produced.
- Only standard units will be reported for all laboratory parameters.

- Boxplots will not be produced for laboratory parameters.
- ECG and bone age assessment will be summarised using summary statistics (frequency count, percentage) by treatment group for each visit will be produced. Shift table will be produced for ECG at week 30 and week 56.
- SAS will be used to analyse change from baseline pulse values at week 30 and 56.

The following text in protocol:

"Bone age assessment at baseline and change in bone age assessment will be summarised using descriptive statistics"

is replaced to:

"Bone age assessment is summarised using frequencies count and percentages. A shift table will be provided for change in baseline to week 56."

3.4 Analysis Rules

3.4.1 Visit Windowing:

Primary endpoints analysis is performed at Visit 25 (week 56) and secondary endpoints are performed at Visit 19 (week 30), Visit 25 (week 56) and Visit 30 (week 82).
For subjects who prematurely discontinues trial product or subject and/or the subject's LAR(s) withdraw consent, the investigator must aim to undertake procedures like those for the end of treatment (V25) as soon as possible, and a follow-up visit (V26) two weeks later.

The subjects were asked to attend additional visits depending on when the trial product discontinuation takes place during the trial:

- If the trial product is discontinued before V19 subjects were asked to attend three additional visits (V19x, V25x and V30x) taking place weeks 30, 56 and 82 after randomisation, respectively.

- If the trial product is discontinued after V19 and before V25 subjects were asked to attend two additional visits (V25x and V30x) taking place weeks 56 and 82 after randomisation, respectively.

The end of treatment (V25), and a follow-up visit (V26) collected for prematurely discontinued subject and withdrawn subjects did not exactly taken place in respective endpoint weeks (i.e., at week 30, week 56 and week 82). Hence these records will be reallocated to the respective weeks based on analysis day on when the assessments are done by visit windowing concept.

3.4.1.1 Visits windows:

Generally, visit-windows will be used for the data that is summarized by visit; they are based on the study evaluation schedule and comprise a set of days around the nominal visit day. For any assessment, there are the protocol defined scheduled visits around which visit windows were created to cover the complete range of days within the study.

In 4180 the visit windowing will be done only for V25 and V26 visits for prematurely discontinued/withdrawn subjects.

The visit windows are shown in Table 3-1.

In this table, the days are counted since the date of randomization for both safety and efficacy assessments. These visit windows apply to measurements taken at every visit. For assessments collected less often different visit windows will be applied as detailed below.

When visit windows are used, all visits will be re-aligned, i.e., they will be mapped into one of the visit windows. For example, if the Week 4 visit of a subject is delayed and occurs on Day 46 instead of on Day 29, it will be re-aligned to visit window Week 8. In the case of major deviations from the visit schedule, or due to unscheduled visits, several assessments of a subject may fall in a particular visit window (either scheduled or unscheduled). Statistical approaches to handle multiple assessments in a given visit window are specified below.

In general, if two consecutive visits Vt and Vs are x days apart, the upper limit of the visit window for Vt will be Vt+x/2 and the lower limit for the visit Vs will be Vs-x/2 (if x is even, the lower limit for Vs will be Vs-x/2+1, and the upper limit for Vt will be Vt+x/2). The algorithm needs to ensure that visit windows are not overlapping and that there are no gaps, such that each assessment can be uniquely allocated to one visit window, e.g., if Week 8 visit is scheduled for day 57, Week 12 is scheduled at Day 85 and Week 16 is scheduled at Day 113, then the visit window for week 12 extends from Day 72 to Day 99.

Note: The lower limit of first visit will be assigned with -999 which accommodate all assessments occurred before the scheduled day.

The upper limit of last visit will be assigned with 999 which accommodate all assessments occurred after the scheduled day.

Table 3-1: Assessment windows for scheduled visits

Analysis Visit	Week	Scheduled Day	Visit Window
Visit 2 (screening)	week -14	-98	Day -999 to -91
Visit 3 (week -12)	week -12	-84	Day -90 to -70
Visit 5 (week -8)	week -8	-56	Day -69 to -42
Visit 7 (week -4)	week -4	-28	Day -41 to -21
Visit 8 (week -2)	week -2	-14	Day -20 to -7
Visit 9 (week 0)	Randomisation	1	Day -6 to 4
Visit 10 (week 1)	week 1	8	Day 5 to 11
Visit 11 (week 2)	week 2	15	Day 12 to 18
Visit 12 (week 3)	week 3	22	Day 19 to 25
Visit 13 (week 4)	week 4	29	Day 26 to 43
Visit 14 (week 8)	week 8	57	Day 44 to 71
Visit 15 (week 12)	week 12	85	Day 72 to 99
Visit 16 (week 16)	week 16	113	Day 100 to 127
Visit 17 (week 20)	week 20	141	Day 128 to 158
Visit 18 (week 25)	week 25	176	Day 159 to 193
Visit 19 (week 30)	week 30	211	Day 194 to 225
Visit 19x (week 30 follow-up)	week 30	211	Day 194 to 225
Visit 20 (week 34)	week 34	239	Day 226 to 253
Visit 21 (week 38)	week 38	267	Day 254 to 281
Visit 22 (week 42)	week 42	295	Day 282 to 309
Visit 23 (week 46)	week 46	323	Day 310 to 340

Analysis Visit	Week	Scheduled Day	Visit Window
Visit 24 (week 51)	week 51	358	Day 341 to 375
Visit 25 (week 56)	week 56	393	Day 376 to 400
Visit 25x (week 56 follow-up)	week 56	393	Day 376 to 400
Visit 26 (week 58)	week 58	407	Day 401 to 428
Visit 27 (week 64)	week 64	449	Day 429 to 470
Visit 28 (week 70)	week 70	491	Day 471 to 512
Visit 29 (week 76)	week 76	533	Day 513 to 554
Visit 30 (week 82)	week 82	575	Day 555 to 999
Visit 30x (week 82 follow-up)	week 82	575	Day 555 to 999

3.4.1.2 For parameters which are not collected at every visit (e.g. HEIGHT):

Visit windows defined in Table 3-1will be combined. For example, Table 3-2show visit windows for HEIGHT.

Table 3-2: Assessment windows for scheduled visits

Analysis Visit	Week	Scheduled Day	Visit Window
Visit 2 (screening)	week -14	-98	Day -999 to -56
Visit 8 (week -2)	week -2	-14	Day -55 to -7
Visit 9 (week 0)	week 0	1	Day -6 to 43
Visit 15 (week 12)	week 12	85	Day 44 to 148
Visit 19 (week 30)	week 30	211	Day 149 to 253
Visit 19x (week 30 follow-up)	week 30	211	Day 149 to 253
Visit 22 (week 42)	week 42	295	Day 254 to 344

Analysis Visit	Week	Scheduled Day	Visit Window
Visit 25 (week 56)	week 56	393	Day 345 to 484
Visit 25x (week 56 follow-up)	week 56	393	Day 345 to 484
Visit 30 (week 82)	week 82	575	Day 485 to 999
Visit 30x (week 82 follow-up)	week 82	575	Day 485 to 999

3.4.1.3 For parameters which are collected at 'x' visits:

For premature discontinued subjects a parameter such as 'BODY_WEIGHT', 'HEIGHT' is collected at additional visits like 19x, 25x and 30x, hence if the visit reallocation for these parameters and related parameters such as 'BMI', 'BMI_SDS' and 'HEIGHT_SDS' are reallocated to x visit instead of normal visit (i.e., visits 19, 25 and 30). Other parameters are reallocated to normal visits.
Note: Withdrawal subjects are reallocated to normal visits.

3.4.2 Analysis eligible records:

If the records are reallocated to the visit which is already collected, then the collected visit record is used for analysis instead of record which is reallocated to that visit.
if the subjects have records for both normal visit (Visit 19 (week 30)) and x visit (Visit 19x (week 30)) then the records which are reallocated are not eligible for analysis.

4 References

1. International Conference on Harmonisation. ICH Harmonised Tripartite Guideline. Guideline for Good Clinical Practice E6 (R1), Step 4. 10 June 1996.
2. International Conference on Harmonisation. ICH Harmonised Tripartite Guideline. Guideline for Statistical principle for clinical trials Step 4. 5 February 1998.
3. 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.
4. Computation of centiles and z-scores for height-for-age, weight-for-age and BMI-for-age. (http://www.who.int/growthref/computation.pdf?ua=1)

Statistical Documentation

Trial ID: NN8022-4180

Program code and results for the sample size calculation in trial NN8022-4180

Author

[redacted]

Trial ID: NN8022-4180

Program code and results for the sample size calculation in trial NN8022-4180, Version 1.0, dated 20 May 2016

Overview of deleted pages

Pages	Title
2-11	Program code and results for the sample size calculation in trial NN8022-4180