Cover Page for Statistical Analysis Plan

Sponsor name:	Novo Nordisk A/S
NCT number	NCT04017832
Sponsor trial ID:	NN9924-4309
Official title of study:	PIONEER 12 China multi-regional clinical trial: Efficacy and safety of oral semaglutide versus sitagliptin in subjects with type 2 diabetes mellitus treated with metformin
Document date*:	18 November 2021

*Document date refers to the date on which the document was most recently updated.
Note: The date in the header from Page 2 is the date of compilation of the documents and not of an update to content.

16.1.9 Documentation of statistical methods

List of contents

Statistical Analysis Plan

Trial ID: NN9924-4309

PIONEER 12 China multi-regional clinical trial: Efficacy and safety of oral semaglutide versus sitagliptin in subjects with type 2 diabetes mellitus treated with metformin

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

.

Author:

[redacted]

Clinical Medical Regulatory and Quality-[redacted]

Table of contents

Version History

This Statistical Analysis Plan (SAP) for trial NN9924-4309 is based on the protocol version 3.0 dated 28JAN2019.

SAP Version	Date	Change	Rationale
1.0	16Feb2017	Not Applicable	Original version
2.0	09Nov2018	Detailed description	Update to align with protocol version 2.0
3.0	18Nov2021	Removal of Anti-semaglutide antibodies analyses according to protocol amendment 2.	Anti-semaglutide antibodies data was not collected.
		Removed IWQOL in Section 2.6.2 in SAP.	IWQOL data is not collected.
		Insert description on handling sparse data in Section 2.3.2 Primary analysis for the secondary estimand.	Propose alternative statistical model when sparse data encountered.
		Negative binomial regression and logistic regression on severe or BG confirmed symptomatic hypoglycaemic endpoints that described in protocol Section 2.4.2.2 Safety endpoints are replaced by descriptive statistics.	Too few hypoglycaemic episodes were observed in the study.
		Remove statistical modelling on SF-36 responder binary endpoint analyses in Section 2.6.1, keep descriptive analyses instead.	Statistical testing result from SF36 is not informative in other Pioneer trials
		Chinese subgroup analyses has changed to Chinese subpopulation analyses in Section 2.3.3.1.	More detailed analyses on Chinese subpopulation are performed to assess the efficacy and safety endpoints on subpopulation and document in separate CTR. And region factor is removed from covariate.

List of abbreviations

AACE	American Association of Clinical Endocrinologists
ADA	American Diabetes Association
AE	adverse event
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
AST	aspartate aminotransferase
BG	blood glucose
BMI	body mass index
EAC	event adjudication committee
ECG	electrocardiogram
EOT	end-of-treatment
FAS	full analysis set
FPG	fasting plasma glucose
GI	gastrointestinal
GLP-1	glucagon-like peptide-1
HbA_{1c}	glycosylated haemoglobin
HDL	high-density lipoprotein
IMP	investigational medicinal product
LDL	low-density lipoprotein
LLoQ	lower limit of quantification
MAR	missing at random
MI	myocardial infarction
MMRM	mixed model for repeated measurements
PG	plasma glucose
PIONEER	Peptide InnOvatioN for Early diabEtes tReatment
PK	pharmacokinetics
PRO	patient-reported outcome
SAE	serious adverse event

SAS	safety analysis set
s.c.	subcutaneous(ly)
SMPG	self-measured plasma glucose
T2D	type 2 diabetes mellitus
TE	treatment effect
TEAE	treatment-emergent adverse events
UTN	Universal Trial Number

1 Introduction

1.1 Trial information

This is a 26-week, randomised, double-blinded, double-dummy, active-controlled, four-armed, parallel-group, multicentre, multinational trial comparing the efficacy and safety of three once-daily dose levels of oral semaglutide versus sitagliptin once-daily in subjects with T2D inadequately controlled on metformin.

Subjects will be randomised in a 1:1:1:1 ratio to receive one of the following treatments:

- oral semaglutide 3 mg and sitagliptin placebo
- oral semaglutide 7 mg and sitagliptin placebo
- oral semaglutide 14 mg and sitagliptin placebo
- sitagliptin 100 mg and oral semaglutide placebo

The randomisation will be stratified according to whether the subject is from the China region (including Taiwan and Hong Kong) or not.

For further details see the trial protocol.

1.2 Scope of the statistical analysis plan

This SAP is based on the statistical analyses in the protocol version 3.0 dated 28JAN2019.

2 Statistical considerations

The blinding of the randomised treatments will be maintained until the database has been released for statistical analysis. No interim analyses or other analyses of unblinded data will be performed before the database is locked.

Data from all sites will be analysed and reported together.

China including Taiwan and Hong Kong will be considered as a separate region unless otherwise specified. The information regarding region (China/non-China) will be included based on country details from the IWRS.

The latest available measurement, at or prior to the randomisation visit, will be used as the baseline measurement. If no measurement(s) have been obtained, at or prior to randomisation, the baseline value will be left missing.

Laboratory values below the lower limit of quantification (LLoQ) will be set to ½LLoQ. Number of values below LLoQ by treatment and visit will be summarised if deemed relevant.

Results from a statistical analysis will as a minimum be presented by the estimated treatment contrasts for the below three comparisons with associated two-sided 95% confidence intervals and p-values corresponding to two-sided tests of no difference:

- oral semaglutide 14 mg vs. sitagliptin 100 mg
- oral semaglutide 7 mg vs. sitagliptin 100 mg
- oral semaglutide 3 mg vs. sitagliptin 100 mg

If no statistical analysis is specified, data will be presented using relevant summary statistics.

Primary and secondary estimands

Two estimands addressing different aspects of the trial objective will be defined; a primary 'Hypothetical' estimand and a secondary 'Treatment-policy estimand:

Primary estimand – 'Hypothetical'

- Treatment difference at week 26 for all randomised subjects if all subjects adhered to treatment and did not initiate rescue medication

The hypothetical estimand assesses the glycaemic benefit a future subject is expected to achieve if initiating and continuing treatment with oral semaglutide as compared to sitagliptin. It is considered a clinically relevant estimand as it provides information to treating clinicians about the expected glycaemic efficacy of oral semaglutide compared to sitagliptin for purposes of treating individual subjects. Generalisation of this estimand depends among other things on the extent to which the adherence to trial product administration in this trial reflects the behaviour of the target population. Accordingly, only data collected prior to discontinuation of trial product or initiation of rescue medication will be used to draw inference. This will avoid confounding from rescue medication.

Secondary estimand 'Treatment-policy'

- Treatment difference at week 26 for all randomised subjects regardless of adherence to randomised treatment and initiation of rescue medication.

The treatment policy estimand assesses the expected glycaemic benefit in a future population that results from subjects initiating treatment with oral semaglutide including potential rescue medication(s) as compared to initiating treatment with sitagliptin including potential rescue medication(s). Generalisation of this estimand depends among other things on the extent to which the use of rescue medication in this trial reflects clinical practice and the adherence to trial product administration in this trial reflects the behaviour of the target population. Accordingly, data collected regardless of discontinuation of trial product or initiation of rescue medication(s) will be used to draw inference.

The confirmation of the confirmatory hypotheses will be based on the primary analysis results for the primary estimand (hypothetical). However, to address a different aspect of the trial objective and to be able to compare the trial results with the global trial (PIONEER 3), the secondary estimand (treatment-policy) has been included as well.

Missing data considerations at week 26

When estimating the primary estimand, data collected after discontinuation of trial product or initiation of rescue medication(s) will be set to missing. Across treatment arms the main reasons for missing data are expected to be early treatment discontinuation due to GI AEs and initiation of rescue medication. The proportion of missing data is expected to be 20% based on the sitagliptin phase 3 trials[1], the oral semaglutide phase 2 trial (NN9924-3790) that indicated that a low starting dose with gradual dose escalation diminishes GI AEs compared with more aggressive dosing regimens. Initiation of rescue medication is expected to be more frequent in the sitagliptin arm and in the oral semaglutide 3 mg arm than for the two highest dose levels of oral semaglutide. A higher proportion of subjects are expected to discontinue treatment due to AEs in the oral semaglutide 14 mg arm compared to the other treatment arms. So, overall, the frequency of missing data is expected to be similar across treatment arms.

When estimating the secondary estimand, the proportion of missing data, i.e., data that do not exist even though subjects are intended to stay in the trial regardless of treatment status and initiation of rescue medication(s) is expected to be maximum 10% based on the oral semaglutide phase 2 trial (NN9924-3790). Thus, missing data will mainly be due to withdrawal from trial or lost to follow-up.

Descriptive summaries and graphical representation of extent, reason(s) for and pattern of missing data will be presented by treatment arm.

2.1 Sample size calculation

The primary endpoint is change from baseline to week 26 in HbA_{1c}. For HbA_{1c}, both non-inferiority and superiority versus sitagliptin are planned to be tested at each dose level. The confirmatory secondary endpoint is change from baseline to week 26 in body weight (kg). For body weight, superiority versus sitagliptin is planned to be tested at each dose level. The sample size calculation is made for the primary estimand and ensuresa power of at least 85% for testing the below three out of the nine pre-specified confirmatory hypotheses shown in Figure 2–1. The closed testing procedure described in Bretz et al 2011[2] is used to control the overall type I error at a nominal two-sided 5% level. The three hypotheses are:

- HbA_{1c} superiority of semaglutide 14 mg vs. sitagliptin 100 mg
- HbA_{1c} superiority of semaglutide 7 mg vs. sitagliptin 100 mg
- HbA_{1c} non-inferiority of semaglutide 3 mg vs. sitagliptin 100 mg (margin of 0.3%-point)

The statistical testing strategy is based on the following two principles:

- Within a dose level, glycaemic effect must be established in terms of HbA_{1c} non-inferiority before testing for added benefits in terms of superiority for HbA_{1c} and/or superiority of body weight.
- Glycaemic effect in terms of HbA_{1c} non-inferiority must be established on all higher dose levels before continuing testing hypotheses on lower dose levels.

The sample size is calculated using the calcPower function in the R package, gMCP[3] using 10000 simulations. All of the nine pre-specified confirmatory tests are assumed to be independent. Since positive correlations among the tests are expected, the assumption of independence is viewed as conservative.

The sample size assumptions for treatment effects (TE), adjusted TE and the common standard deviations (SD) used across dose levels are given in Table 2–1.These are based on the oral semaglutide phase 2 results (NN9924-3790), sitagliptin phase 3a trial results[1] and supported by results from the s.c. semaglutide phase 2 trial (NN9535-1821).

An adjustment in TE will be implemented for the 20% of subjects who are expected to have missing data. The TE used in the sample size calculation will be adjusted according to a 50% smaller effect in these subjects. The adjusted treatment effect is defined as $0.8 \times TE + 0.2 \times TE \times 0.50$.

Table 2–1 Assumptions for sample size calculation

Oral semaglutide vs. sitagliptin	HbA_{1c} (%-point)			Body weight (kg)		
Treatment dose	14 mg	7 mg	3 mg	14 mg	7 mg	3 mg
Treatment effect (TE)	-0.5	-0.3	-0.1	-3.0	-2.0	-1.0
Adjusted TE[1]	-0.45	-0.27	-0.09	-2.7	-1.8	-0.9
Standard deviation	1.1	1.1	1.1	4.0	4.0	4.0

[1]The adjusted TE is according to a 50% smaller effect in 20% subjects expected to discontinue treatment

With the above assumptions, allocating 361 subjects to each of the oral semaglutide treatment arms and sitagliptin 100 mg provides 85% power to jointly confirm HbA_{1c} superiority of oral semaglutide 14 mg vs. sitagliptin 100 mg, HbA_{1c} superiority of semaglutide 7 mg vs. sitagliptin 100 mg and HbA_{1c} non-inferiority of semaglutide 3 mg vs. sitagliptin 100 mg for the primary estimand. Calculated powers for selected individual hypotheses are presented in Table 2–2. In total 4×361 = 1444 subjects are planned to be randomised. The planned number of subjects from China region (including Hong Kong and Taiwan) is about ¾ of the total sample size equal to 1084.

Table 2–2 Calculated powers for individual hypotheses

Statistical test	HbA_{1c} superiority			Body weight superiority			HbA_{1c} non-inferiority (margin = 0.3%)
Treatment dose	14 mg	7 mg	3 mg	14 mg	7 mg	3 mg	3 mg
Power (%)	> 99	85	16	> 99	> 99	84	> 99

For the secondary estimand the TE in the 10% of subjects who are expected to discontinue trial product or initiate rescue medication and for the 10% of subjects who are expected to have missing data is expected to be less compared to the primary estimand. In addition, when testing for non-inferiority, the non-inferiority margin of 0.3% for HbA_{1c} is added to the imputed values for the 10% of the subjects with missing data. Assuming a 75% adjustment , the same TE and standard deviations as presented in Table 2–1, the power to jointly meet HbA_{1c} superiority of oral semaglutide 14 mg vs. sitagliptin 100 mg, HbA_{1c} superiority of semaglutide 7 mg vs. sitagliptin 100

mg and HbA_{1c} non-inferiority of semaglutide 3 mg vs. sitagliptin 100 mg is 80% for the secondary estimand.

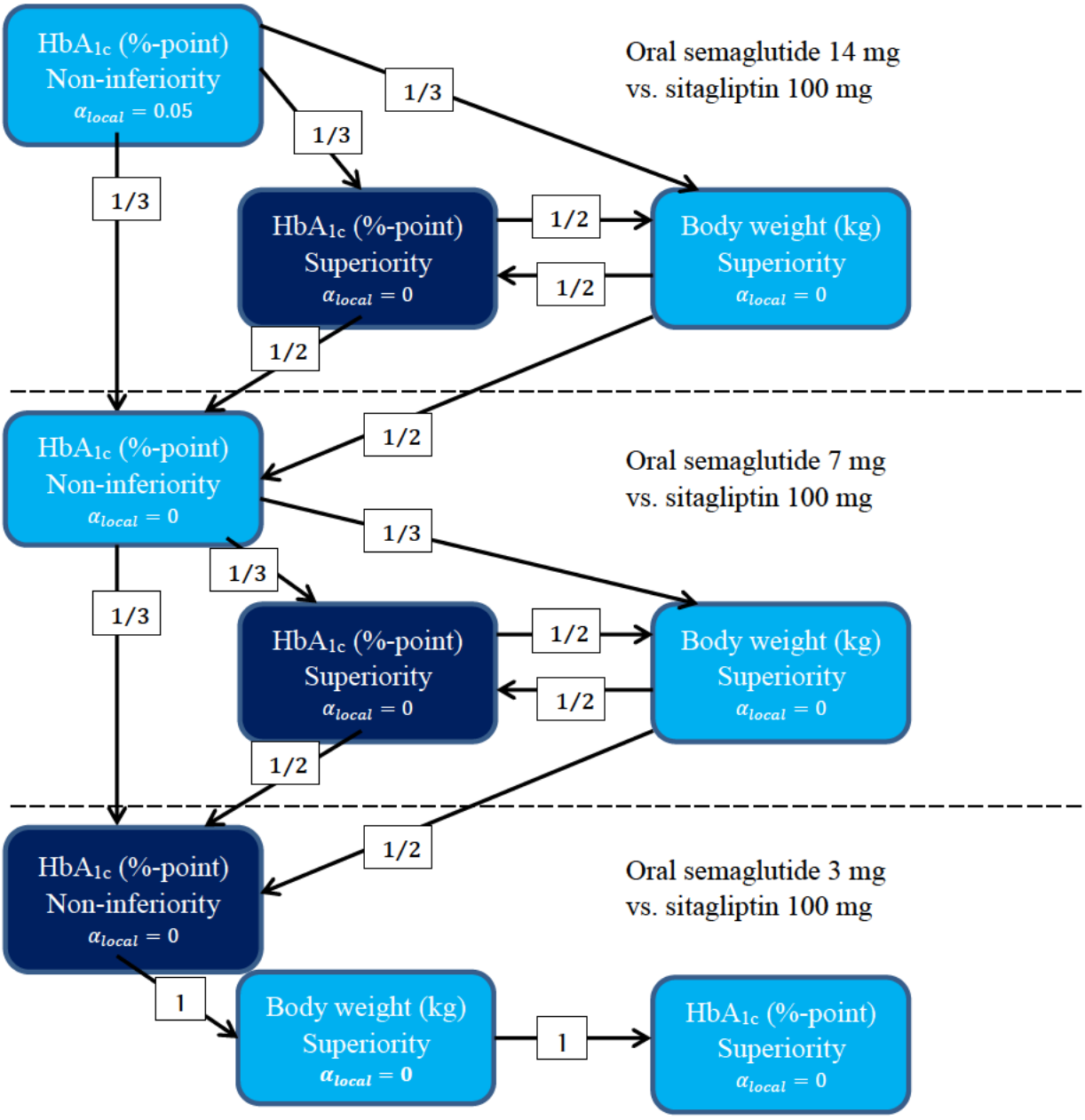


Figure 2–1 Graphical illustration of the closed testing procedure

The overall significance level of α=0.05 (two-sided) is initially allocated to the HbA_{1c} non-inferiority test of semaglutide 14 mg vs. sitagliptin 100 mg. The local significance level (α-local) will be reallocated if a hypothesis is confirmed according to the weight given by the directed edges between nodes (hypotheses). The sample size is based on the hypotheses in the dark boxes.

2.2 Definition of analysis sets

The following analysis sets will be defined:

Full analysis set (FAS): Includes all randomised subjects. Subjects in the FAS will contribute to the evaluation "as randomised".

Safety analysis set (SAS): Includes all subjects exposed to at least one dose of trial product. Subjects in the SAS will contribute to the evaluation based on the trial product received for the majority of the period where they were on treatment. This will be referred to as contributing to the evaluation "as treated".

Per protocol (PP) analysis set: Includes all subjects in the FAS who fulfil the following criteria:
- have not violated any inclusion criteria
- have not fulfilled any exclusion criteria
- have a baseline HbA_{1c} measurement
- is exposed to trial product and have at least one valid HbA_{1c} measurement while on treatment without rescue medication at or after week 14

Subjects in the PP analysis set will, as in the SAS, contribute to the analysis "as treated".

Data selections and observation periods

Unless subjects withdraw their informed consent, data collection will continue for the full duration of the trial. The full duration of the trial is defined as up to and including:
- the follow-up visit (visit 9) for subjects on trial product
- the latest occurring visit of the EOT visit (visit 8) or the follow-up premature discontinuation visit (visit 9A), for subjects who have discontinued trial product prematurely

Subjects and data to be used in an analysis will be selected in a two-step manner:
- Firstly, subjects will be selected based on the specified analysis set
- Secondly, data points on the selected subjects from first step will be selected based on the specified observation period

Definition of the observation periods:

In-trial: This observation period represents the time period where subjects are considered to be in the trial, regardless of discontinuation of trial product or initiation of rescue medication. The

in-trial observation period starts at randomisation (as registered in the IWRS) and ends at the date of:

- The last direct subject-site contact, which is scheduled to take place 5 weeks after planned last dose of trial product at the follow-up visit
- Withdrawal for subjects who withdraw their informed consent
- The last subject-investigator contact as defined by the investigator for subjects who are lost to follow-up
- Death for subjects who dies before any of the above

On-treatment: This observation period represents the time period where subjects are considered treated with the trial product. The observation period is a subset of the in-trial observation period. It starts at the date of first dose of trial product. Two slightly different end dates will be needed to cover all assessments appropriately. For adjudicated events, ECGs, eye examination category, and AEs including hypoglycaemic episodes, the observation period ends at the first date of:

- The follow-up visit (visit 9)
- The follow-up prematurely discontinuation visit (visit 9A)
- The last date on trial product + 38 days
- The end-date for the in-trial observation period

The follow-up visit is scheduled to take place 5 weeks after the last date on trial product corresponding to approximately five half-lives of oral semaglutide. The visit window for the follow-up visit is + 3 days.

For efficacy and other safety assessments (laboratory assessments, physical examination and vital signs) the observation period ends at the last date on trial product + 3 days. This will be used in order to ensure specificity to reversible effects of treatment.

On-treatment without rescue medication: This observation period is a subset of the on-treatment observation period, where subjects are considered treated with trial product, but have not initiated any rescue medications. Specifically, it starts at date of first dose of trial product and the observation period ends at the first date of:

- the last dose of trial product +3 days
- initiation of rescue medication

The on-treatment without rescue medication observation period will be the primary observation period when estimating the primary estimand (hypothetical). The in-trial observation period will be the primary observation period when estimating the secondary estimand (treatment-policy). Safety will be evaluated based on the in-trial and the on-treatment observation periods.

Data points collected outside an observation period will be treated as missing in the analysis. Baseline data will always be included in an observation period. For adjudicated events, the onset date will be the EAC adjudicated onset date.

Before data are locked for statistical analysis and the randomisation code is broken, a review of all data will take place. Any decision to exclude either a subject or single observations from the statistical analysis is the joint responsibility of the members of the Novo Nordisk study group. Exclusion of data from analyses should 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 CTR.

Confirmatory hypotheses

For the primary HbA_{1c} endpoint and the secondary confirmatory body weight endpoint, the following three confirmatory one-sided hypotheses are planned to be tested at each dose level of oral semaglutide versus sitagliptin. Let the mean treatment difference at week 26 be defined as μ = (oral semaglutide minus sitagliptin):

- HbA_{1c}, non-inferiority, using a non-inferiority margin of 0.3%-point
  - H_0: $\mu \geq 0.3$%-point against Ha: $\mu < 0.3$-point%
- HbA1c superiority
  - H_0: $\mu \geq 0.0$%-point against Ha: $\mu < 0.0$%-point
- Body weight superiority
  - H_0: $\mu \geq 0.0$ kg against Ha: $\mu < 0.0$ kg

Operationally the hypotheses will be evaluated by two-sided tests at the 5% significance level.

Multiplicity and criteria for confirming hypotheses

The type I error for testing the nine confirmatory hypotheses related to the HbA_{1c} and body weight endpoints will be preserved in the strong sense at 5% (two-sided) using the weighted Bonferroni-based closed testing procedure described in Bretz et al 2011 and outlined in Figure 2–1.

The first hypothesis to be tested is non-inferiority of HbA_{1c} at the highest dose level. It will be tested at the overall significance level (5%) while allocating 0% local significance level to the remaining of the hypotheses. For this hypothesis, and in general, if a hypothesis is confirmed, then the significance level will be reallocated according to the weight and the direction of the edges going from the confirmed hypothesis to the next hypotheses as specified in Figure 2–1. Each of the following hypotheses will be tested at their updated local significance level (α-local). This process will be repeated until no further hypotheses can be confirmed.

Non-inferiority and/or superiority will be considered confirmed if the mean treatment difference is supporting the corresponding alternative hypothesis and the two-sided p-value from the primary analysis of the primary estimand is strictly below its local two-sided significance level as defined by the closed testing procedure in Figure 2–1. This is equivalent to using a one-sided p-value (nominal α= 0.025) and a one-sided 2.5% overall significance level in the closed testing procedure.

2.3 Primary endpoint

The primary endpoint is change from baseline to week 26 in HbA_{1c}.

2.3.1 Primary analysis for the primary estimand

The primary estimand will be estimated based on the FAS using post-baseline measurements up to and including week 26 from the on-treatment without rescue medication observation period. The primary analysis for the primary estimand will be a Mixed Model for Repeated Measurements (MMRM). A restricted maximum likelihood (REML) will be used. The model will include all post-baseline HbA_{1c} measurements collected at scheduled visits up to and including week 26 as dependent variables. The independent effects included in the model will be treatment and region as categorical fixed effects and baseline HbA_{1c} as a covariate, all nested within visit. An unstructured covariance matrix for HbA_{1c} measurements within the same subject will be employed, assuming measurements from different subjects are independent. For subjects who do not have post-baseline assessments for planned visits available in the on-treatment without rescue medication period, the baseline value will be carried forward to the first planned visit to ensure that all randomised subjects will contribute to the statistical analysis.

The MMRM is a well-established method that accounts for the uncertainty pertaining to missing data. This analysis assumes that the missing data mechanism is missing at random (MAR). Under this assumption, the statistical behaviour of the missing data (given the observed responses and model fixed effects and covariates) is assumed to be same as for the observed data.

2.3.2 Primary analysis for the secondary estimand

The secondary estimand will be estimated based on the FAS using week-26 measurements from the in-trial observation period. The primary statistical analysis will be a pattern mixture model using multiple imputation to handle missing data assuming that the missing data mechanism is MAR within the groups used for imputation. Imputation of missing data at week 26 will be done within 8 groups of subjects defined by randomised treatment arm, and whether subjects at week 26; (i) have discontinued treatment or initiated rescue medication or (ii) are still on treatment and have not initiated rescue medication. It is hereby assumed that the likely values of what the missing data would have been if available are best described by information from subjects who at week 26 are similar in terms of randomised treatment arm and treatment adherence/rescue status.

Missing values for each group will be imputed as follows:

- An analysis of covariance (ANCOVA) with region as categorical fixed effects and baseline HbA_{1c} measurement as a covariate will be fitted to observed values of the change from baseline in HbA_{1c} at week 26.
- The estimated parameters for location and dispersion will be used to impute 1000 values for each subject with missing week-26 data based on region as categorical and baseline HbA_{1c}. Thus, 1000 complete data sets will be generated including observed and imputed values.

- If data is too sparse to fit the above defined imputation model for subjects who have discontinued trial treatment or initiated rescue medication, missing week 26 data for these subjects will be imputed based on subjects from all four treatment arms, and the model will in addition include treatment arm as a factor. If data is still too sparse to fit this imputation model, the model will be reduced by removing factors in the following order:

    - Region

    - Treatment arm.

Analysis used for superiority versus sitagliptin at week 26:

For each of the 1000 (now complete) imputed data sets, the change in HbA_{1c} from baseline to week 26 will be analysed using an ANCOVA with treatment and region as categorical fixed effects and baseline HbA_{1c} as covariate. The results obtained from analysing the data sets will be combined using Rubin's rule[4] to draw inference.

Analysis used for non-inferiority versus sitagliptin at week 26:

Prior to analysing the data using the same model and approach as used for superiority (see above), a value of 0.3% (the non-inferiority margin) will be added to imputed values at week 26 for the oral semaglutide treatment arms only[5]. For evaluating non-inferiority versus sitagliptin unadjusted two sided p-values for testing no difference from the non-inferiority margin will be presented.

2.3.3 Sensitivity analyses

To investigate the sensitivity of the primary analysis results, a complementary and separate analysis will be performed for the primary and secondary estimand. In line with European Medicines Agency (EMA) recommendations[6] and with a report from the US National Research Council[7], these analyses will primarily evaluate the sensitivity of the results due to the impact of missing data. Since conservatism, i.e. avoiding bias in favour of oral semaglutide, depends on the context, separate sensitivity analyses will be made for non-inferiority and superiority testing.

The evaluation of the robustness of the primary analysis results will be based on a pattern mixture model approach using multiple imputation. The pattern mixture model sensitivity analyses aim to

stress-test the primary HbA_{1c} results by changing the assumptions for the missing data in the oral semaglutide treatment arms, while maintaining the MAR data assumption for the sitagliptin Additionally a sensitivity analysis for the primary analysis will be described that is not based on the pattern mixture model approach.

Sensitivity analyses for the primary estimand

The estimation of the primary estimand will be repeated using a tipping-point multiple imputation analysis based on FAS using the on-treatment without rescue medication observation period:

In this sensitivity analysis, missing data will first be imputed using a sequential multiple imputation approach assuming MAR. The imputation will be done as described below:

- Intermittent missing values in the on-treatment without rescue medication observation period will be imputed using a Markov Chain Monte Carlo (MCMC) method, in order to obtain a monotone missing data pattern. This imputation is done for each treatment group separately and a 1000 copies of the data set will be generated.
- A sequential regression approach for imputing monotone missing values at planned visits will be implemented starting with the first visit after baseline and sequentially continuing to the planned end of treatment visit. For each treatment group an analysis of covariance model will be used to impute missing values at each planned visit. The model will include region as categorical effect and baseline and post-baseline values prior to the visit in question as covariates.

Secondly, for all oral semaglutide treatment arms a penalty will be added to the imputed values at week 26. The approach is to gradually increase this penalty until all confirmed HbA_{1c} conclusions from the primary analysis are reversed. For each hypothesis tested the specific value of the penalty that reverses the conclusion will be used to evaluate the robustness of the primary analysis results.

Sensitivity analyses for the secondary estimand

The estimation of the secondary estimand will be repeated using a tipping-point multiple imputation analysis based on FAS using the in-trial observation period. For all oral semaglutide treatment arms a penalty will be added to the imputed values at week 26 from the primary analysis for the secondary estimand. The approach is to gradually increase this penalty until all confirmed HbA_{1c} conclusions from the primary analysis are reversed. For each hypothesis tested the specific value of the penalty that reverses the conclusion will be used to evaluate the robustness of the primary analysis results.

2.3.3.1 Other sensitivity analyses

Per-protocol analysis: This sensitivity will be based on the per-protocol analysis set. Data from the on-treatment without rescue medication observation period will be analysed using the

primary MMRM analysis approach for the primary estimand. This sensitivity analysis will be used to evaluate the robustness of the HbA1c non-inferiority conclusions.

2.3.3.2 Chinese Subpopulation Analyses

The primary endpoint, secondary endpoints, and other analyses defined in this SAP will be performed on region China subpopulation. The region will no longer be included in the model. The statistical analyses will be the same as that performed on the overall population. The results will be included in China subpopulation CTR.

2.4 Secondary endpoints

2.4.1 Confirmatory secondary endpoints

Change from baseline to week 26 in body weight (kg) will be a confirmatory secondary endpoint.

The primary and secondary estimands will be estimated using the same approaches as described for the primary HbA_{1c} endpoint. Body weight will only be tested for superiority. Baseline body weight will be used as a covariate instead of baseline HbA_{1c} in both the MMRM and the multiple imputation analysis models.

Superiority will be considered confirmed if the mean treatment difference is supporting the corresponding hypothesis and the two-sided p-value from the analysis of the primary estimand is strictly below its local two-sided significance level resulting from the closed testing procedure in Figure 2–1. Sensitivity analyses similar to the ones pre-specified for testing superiority for the primary HbA_{1c} endpoint will be made to evaluate the robustness of the body weight results.

2.4.2 Supportive secondary endpoints

2.4.2.1 Efficacy endpoints

The below supportive secondary efficacy endpoints will be evaluated for:

- the primary estimand based on FAS using the on-treatment without rescue medication observation period. For endpoints where the first planned visit falls later than 8 weeks after randomisation, the baseline will not be carried forward
- the secondary estimand based on FAS using the in-trial observation period

No sensitivity analyses are planned for these.

Continuous efficacy endpoints

Change from baseline to week 26 in:

- FPG
- Body weight (%)

- BMI
- Waist circumference
- Fasting lipid profiles (total cholesterol, LDL cholesterol, VLDL cholesterol, HDL cholesterol, triglycerides, free fatty acids)
- Patient-reported outcome
  Short Form-36 version 2 (SF-36v2™) (acute version) health survey

BMI will be calculated based on body weight and height based on the formulae:
BMI kg/m^2 = body weight (kg)/(height (m) × height (m)) or (kg/m^2 = [lb/in^2 × 703])

Change from baseline to week 26 in the below derived endpoints from the 7-point SMPG profile:
- Mean of the 7-point profile, defined as the area under the profile, calculated using the trapezoidal method, divided by the measurement time
- Mean post prandial increment (over all meals)

The above continuous endpoints will be analysed separately using similar model approaches as for the primary endpoint with the associated baseline response as a covariate. Fasting lipid profile endpoints will be log-transformed prior to analysis with the associated log-transformed baseline value as a covariate.

Binary efficacy endpoints

If a subject after week 26 achieves (yes/no):
- $HbA_{1c} < 7.0$ % (53 mmol/mol) (ADA target)
- $HbA_{1c} \leq 6.5$ % (48 mmol/mol) (AACE target)
- HbA1c reduction ≥ 1 %-point (10.9 mmol/mol)
- Body weight loss ≥ 3 %
- Body weight loss ≥ 5 %
- Body weight loss ≥ 10 %
- $HbA_{1c} < 7.0$ % (53 mmol/mol) without hypoglycaemia (severe or BG confirmed symptomatic hypoglycaemia) and no body weight gain
- HbA_{1c} reduction ≥ 1 %-point (10.9 mmol/mol) and body weight loss ≥ 3 %

When addressing the treatment-policy estimand the ‘without hypoglycaemia’ component of the composite endpoint will also include non-treatment-emergent events of severe or BG-confirmed symptomatic hypoglycaemia as data collected regardless of discontinuation of trial product or initiation of rescue medication(s) is used.

Missing data for the above six binary endpoints will be accounted for using multiple imputation techniques. For the treatment-policy estimand the binary endpoints will be calculated as dichotomisations of the 1000 multiple imputations underlying the primary MI analysis. For the

hypothetical estimand the model will be implemented using a sequential imputation approach as in the tipping-point sensitivity analysis. The binary endpoints will be derived as dichotomisations of the 1000 imputed complete data sets from the sequential imputation.

The imputed complete data sets will be analysed using a logistic regression model with treatment and region as categorical fixed effects and baseline response as covariate (i.e. baseline HbA_{1c} for binary HbA_{1c} endpoints, baseline weight for binary weight endpoints and both baseline HbA_{1c} and body weight for the binary endpoints that combines both parameters). Inference comparing treatments will be drawn using Rubin's rule[4].

For the composite endpoints involving both HbA_{1c} and body weight the imputed data sets will be combined by imputation number.

Time to event endpoint

- Time to additional anti-diabetic medication (to support the treatment policy estimand)
- Time to rescue medication (to support the hypothetical estimand)

Definition of additional anti-diabetic medication: New anti-diabetic medication and/or intensification of anti-diabetic medication initiated at or after randomisation and before (planned) end-of-treatment.

Definition of rescue medication: New anti-diabetic medication and/or intensification of anti-diabetic medication initiated at or after randomisation and before last date on trial product. This is a subset of the additional anti-diabetic medication.

The following rules will be applied based on the concomitant medication data reported by the investigator, to determine whether or not the recorded anti-diabetic medication is 1. *New anti-diabetic medication* or 2. *Intensification of anti-diabetic medication*

1. **New anti-diabetic medication:** Anti-diabetic medication (4th-level ATC code) that is initiated at or after randomisation and is new compared to the anti-diabetic background medication at randomisation (see above) and with a dosing duration of more than 21 days

2. **Intensification of anti-diabetic medication**: A more than 20% increase in the dose of anti-diabetic medication at or after randomisation as compared to the anti-diabetic medication dose at randomisation (5th-level ATC code not changed) and with a dosing duration of more than 21 days.

More than 21 days is chosen as a minimum duration for the medication to be considered as 'anti-diabetic medication'. This threshold is set to ensure that the short-term durations (i.e. ≤ 21 days) of

anti-diabetic medication (e.g. in connection with concurrent illnesses) are not included because such intensifications are not likely to affect the effect endpoints.

Treatment policy estimand: Time to additional anti-diabetic medication

The treatment policy estimand is addressed for the FAS using the in-trial observation period and additional anti-diabetic medication will be considered an event regardless of whether or not subjects prematurely discontinued treatment. Time from randomisation to additional anti-diabetic medication will be analysed using a Cox proportional hazards model with treatment and stratification factor as categorical fixed effects and baseline HbA_{1c} as a covariate. From this analysis the estimated Hazard ratios between each of the oral semaglutide dose levels and placebo together with associated two-sided 95% CIs and unadjusted two-sided p-values will be presented. The analysis aims to address the need of additional anti-diabetic medication regardless of this is due to lack of effect or tolerability. Switch to other anti-diabetic treatment is therefore also considered an event and withdrawn subjects or subject lost to follow-up will be considered as having an event on the day of withdrawal. Subjects will be censored on the day before planned end of treatment visit.

Hypothetical estimand: Time to rescue medication

The hypothetical estimand is addressed for the FAS using the on-treatment without rescue medication observation period. Time from first dose of trial product to initiation of rescue medication will be analysed using the same model as described above. The analysis aims to address lack of effect and only initiation of rescue medication as add-on to randomised treatment is considered an event. Switch to other anti-diabetic treatment is not considered an event and as a consequence subjects will be censored on the day before date of last trial product. Potential events occurring between randomisation and first date on trial product will be included in the analysis as events at day 0, in order to count all events of rescue medication.

2.4.2.2 Safety endpoints

The safety endpoints will be evaluated based on SAS using the on-treatment observation period and based on SAS using the in-trial observation period unless otherwise stated. The following endpoints are used to support the safety objectives:

Adverse events

- Number of TEAEs during exposure to trial product, assessed up to approximately 31 weeks

All AEs will be coded using the most recent version of the Medical Dictionary for Regulatory Activities (MedDRA) coding.

A TEAE is defined as an AE with onset in the on-treatment observation period (see definition of observation periods in Section 2.2).

TEAEs will be summarised in terms of the number of subjects with at least one event (N), the percentage of subjects with at least one event (%), the number of events (E) and the event rate per 100 patient-years of observation time (R) for the on-treatment observation period. Supportive summaries of AEs will be made for the in-trial observation period. The development over time in GI AEs will be evaluated by the use of graphical methods.

Other safety endpoints

Change from baseline to week 26 in:
- Amylase
- Lipase
- Pulse rate
- Systolic blood pressure
- Diastolic blood pressure

The above safety endpoints will be evaluated using the primary analysis for the primary estimand based on SAS using the on-treatment observation period and using the primary analysis for the secondary estimand based on SAS using the in-trial observation period. Endpoints will be analysed separately as described above for continuous efficacy endpoints. Results at week 26 will be presented. Amylase and lipase endpoints will be log-transformed prior to analysis with the associated log-transformed baseline value as a covariate.

Change from baseline to week 26 in:
- ECG category
- Physical examination category
- Eye examination category

Safety assessments

Change from baseline to week 26 in:

- Haematology
- Biochemistry (except for amylase and lipase)
- Calcitonin

The above safety endpoints and assessments will be summarised descriptively by treatment arm and visit. Categorical safety endpoints will be summarised as counts and relative frequencies. Calcitonin will also be presented by gender.

Hypoglycaemia

- Number of treatment-emergent severe or BG confirmed symptomatic hypoglycaemic episodes during exposure to trial product, assessed up to approximately 31 weeks
- Treatment-emergent severe or BG confirmed symptomatic hypoglycaemic episodes during exposure to trial product, assessed up to approximately 31 weeks (yes/no)

Classification of Hypoglycaemia:

Hypoglycaemic episodes will be summarised for the SAS and the on-treatment observation period only.

Treatment-emergent: hypoglycaemic episodes will be defined as treatment-emergent if the onset of the episode occurs within the on-treatment observation period (see definition of observation periods in Section 2.2).

Nocturnal hypoglycaemic episodes: episodes occurring between 00:01 and 05.59 both inclusive.

Hypoglycaemic episodes are classified according to the Novo Nordisk classification of hypoglycaemia and the ADA classification of hypoglycaemia (see Figure 2–2).

Novo Nordisk classification of hypoglycaemia

In normal physiology, symptoms of hypoglycaemia occur below a PG level of 3.1 mmol/L (56 mg/dL)[8]. Therefore, Novo Nordisk has included hypoglycaemia with PG levels below this cut-off point in the definition of BG confirmed hypoglycaemia.

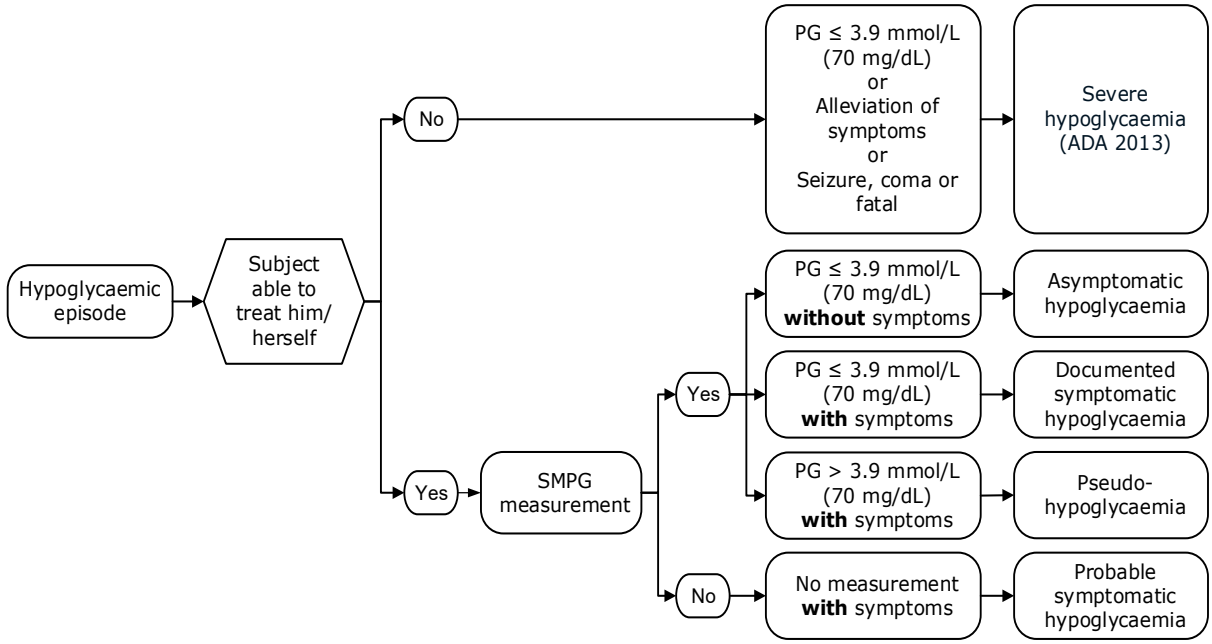
Novo Nordisk uses the following classification in addition to the ADA classification:

- Severe or BG confirmed symptomatic hypoglycaemia: An episode that is severe according to the ADA classification[9] or BG confirmed by a PG value <3.1 mmol/L (56 mg/dL) **with** symptoms consistent with hypoglycaemia.

ADA classification[9] of hypoglycaemia

- Severe hypoglycaemia: An episode requiring assistance of another person to actively administer carbohydrate, glucagon, or take other corrective actions. PG concentrations may not be available during an event, but neurological recovery following the return of PG to normal is considered sufficient evidence that the event was induced by a low PG concentration.
- Asymptomatic hypoglycaemia: An episode not accompanied by typical symptoms of hypoglycaemia, but with a measured PG concentration ≤ 3.9 mmol/L (70 mg/dL).
- Documented symptomatic hypoglycaemia: An episode during which typical symptoms of hypoglycaemia are accompanied by a measured PG concentration ≤ 3.9 mmol/L (70 mg/dL).

- Pseudo-hypoglycaemia: An episode during which the person with diabetes reports any of the typical symptoms of hypoglycaemia with a measured PG concentration > 3.9 mmol/L (70 mg/dL) but approaching that level.
- Probable symptomatic hypoglycaemia: An episode during which symptoms of hypoglycaemia are not accompanied by a PG determination but that was presumably caused by a PG concentration ≤ 3.9 mmol/L (70 mg/dL).



Note: Glucose measurements are performed with capillary blood calibrated to plasma equivalent glucose values

PG: plasma glucose SMPG: Self-measured plasma glucose

Figure 2–2 ADA classification of hypoglycaemia

Data on treatment emergent hypoglycaemic episodes will be presented in terms of the number of subjects with at least one episode, the percentage of subjects with at least one episode (%), the total number of episodes and the episode rate per 100 patient years of observation time.

Analysis of severe or BG confirmed symptomatic hypoglycaemic endpoints

The number of treatment-emergent severe or BG confirmed symptomatic hypoglycaemic episodes in the on-treatment period will be summarized by descriptive statistics.

2.5 Interim analysis

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

2.6 Patient-reported outcomes

PRO endpoints

Change from baseline to week 26 in:

- SF-36v2 (acute version) health survey: Scores from the 8 domains, the physical component summary score and mental component summary score.

The PRO endpoints will be evaluated using the primary analysis for the primary estimand based on FAS using the on-treatment without rescue medication period and using the primary analysis for the secondary estimand based on FAS using the in-trial observation period. Scores will be analysed separately as the other continuous efficacy endpoints with the associated baseline response as a covariate.

2.6.1 SF-36v2® (acute version) health survey

The SF-36v2® Health Survey (SF-36v2) (acute version) instrument is a commonly used generic instrument measuring health-related quality of life (HRQoL)/general health status across disease areas including diabetes. The SF-36v2 (acute version) for adults with a 1-week recall period contains 36 items.

A total of 35 items measure eight domains of functional health and well-being as well as two component summary scores: physical functioning (10 items), role limitation due to physical health problems (4 items), bodily pain (2 items), general health perceptions (5 items), vitality (4 items), social functioning (2 items), role limitations due to emotional problems (3 items) and general mental health (5 items), mental component summary (MCS) score, physical component summary (PCS) score. There is an additional single item giving information on health change over the past week.

Domain scores

Norm-based scores (NBS) will be derived using the QualityMetric Health Outcomes™ Scoring Software1 including the 2009 US general population norm. The most recent version of the QualityMetric Health Outcomes™ Scoring Software available at time of licensing was used for the specific trial. Table 2-3 provides an overview of the domains. NBS standardises domain and component scores into T-scores using the means and standard deviations from the US general population. Higher scores on all domains and component summary measures (PCS and MCS) indicate better HRQoL/general health status. Item 2 (i.e. Question 2 in CRF) is not included in any score.

Table 2–3 Overview of domains for SF-36v2 (acute version)

Domain	Items numbers of items included in domain	Comment
Physical Functioning (PF)	Items 3a-j	
Role Limitations Due to Physical Health (Role-Physical; RP)	Items 4a-d	
Bodily Pain (BP)	Items 7, 8	Both item scores reversed
General Health Perceptions (General Health; GH)	Items 1, 11a-d	Item scores 1, 11b and 11d reversed
Vitality (VT)	Items 9a, 9e, 9g, 9i	Item scores 9a and 9e reversed
Social Functioning (SF)	Items 6, 10	Item score 6 reversed
Role Limitations Due To Emotional Problems (Role-Emotional; RE)	Items 5a-c	
Mental Health (MH)	Items 9b, 9c, 9d, 9f, 9h	Item scores 9d and 9h reversed
Physical component summary (PCS)	NA	The PCS score is a weighted average of the 8 domain scores.
Mental component summary (MCS)	NA	The MCS score is also a weighted average of the 8 domain scores. Weights differ from PCS to MCS.

Missing data at instrument level will be handled using the Maximum Data Recovery method: The method applies a value to a domain item rendered missing if at least one of the items in that domain has valid data. A domain score is considered missing if all item values in the domain are missing. PCS and MCS are calculated when at least seven of the eight domains have valid data, either actual or estimated. However, to calculate PCS, the PF domain must be one of the seven domains having valid data. Also, to calculate MCS, the MH domain must be one of the seven domains having valid data.

Responder threshold values

The responder threshold values, in terms of T-score points for change from baseline are defined in Table 2-4.

Table 2–4 Responder thresholds for SF-36v2 (acute version)

Domain	Responder threshold
Physical Functioning (PF)	4.3
Role Limitations Due to Physical Health (Role-Physical; RP)	4.0
Bodily Pain (BP)	5.5
General Health Perceptions (General Health; GH)	7.0
Vitality (VT)	6.7
Social Functioning (SF)	6.2
Role Limitations Due To Emotional Problems (Role-Emotional; RE)	4.6
Mental Health (MH)	6.7
Physical component summary (PCS)	3.8
Mental component summary (MCS)	4.6

Responder analyses will be based on the responder threshold values and are described in Section 2.6.2.

2.6.2 Responder analyses

Responder analyses will be conducted for both estimands and separately for each score.

For descriptive statistics the following subject responder categorization is applied for all relevant time points and scores:

- Responder - improvement: Individual change from baseline in score ≥ positive responder threshold
- Non-responder - no change: Individual change from baseline in score > negative responder threshold value and < positive responder threshold value
- Non-responder - worsening: Individual change from baseline in score ≤ negative responder threshold value

3 Changes to the statistical analyses planned in the protocol

The SAP specifies several changes and clarifications to the statistical analyses that were described in the trial protocol. All the changes were finalised and approved before unblinding of trial data at DBL. The changes and clarifications are listed below (Table 3-1):

Table 3–1 Change to planned statistical analyses

Change to planned statistical analysis	Rationale for change and possible implications
Insert description on handling sparse data in Section 2.3.2 Primary analysis for the secondary estimand.	Propose alternative statistical model when sparse data encountered.
Remove negative binomial regression and logistic regression on severe or BG confirmed symptomatic hypoglycaemic endpoints that described in protocol Section 17.3.2.2 Safety endpoints.	Too few hypoglycaemic episodes were observed in the study.
Remove statistical modelling on SF-36 responder binary endpoint analyses in Section	Statistical testing result from SF36 is not informative in other Pioneer trials

Change to planned statistical analysis	Rationale for change and possible implications
2.6.1, keep descriptive analysis instead.	
Chinese subgroup analyses described in protocol Section 17.2.4 changed to SAP Section 2.3.3.2 as Chinese subpopulation analyses.	More detailed analyses on Chinese subpopulation has performed to assess the efficacy and safety on subpopulation and document in separate CTR.

4 References

1. Merck Sharp & Dohme Limited. JANUVIA - European Summary of Product Characteristics Annex 1. 26 May 2015.
2. Bretz F, Posch M, Glimm E, Klinglmueller F, Maurer W, Rohmeyer K. Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. Biometrical Journal. 2011;53(6):894-913.
3. Rohmeyer K, Klinglmueller F. gMCP: Graph Based Multiple Test Procedures. R package version 0.8-8. 3 Oct 2014.
4. Little RJA, Rubin DB. Statistical analysis with missing data: New York: John Wiley & Sons. 1987.
5. Koch GG. Comments on 'Current issues in non-inferiority trials' by Thomas R. Fleming, Statistics in Medicine, DOI: 10.1002/sim.2855. Stat Med. 2008;27(3):333-42.
6. European Medicines Agency, Committee for Medicinal Products for Human Use (CHMP). Guideline on Missing Data in Confirmatory Clinical Trials (EMA/CPMP/EWP/1776/99 Rev. 1). 2 Jul 2010.
7. National Academy of Sciences (NAS). The Prevention and Treatment of Missing Data in Clinical Trials. Washington D.C.: The National Academies Press. 2010.
8. Schwartz NS, Clutter WE, Shah SD, Cryer PE. Glycemic thresholds for activation of glucose counterregulatory systems are higher than the threshold for symptoms. J Clin Invest. 1987;79(3):777-81.
9. Seaquist ER, Anderson J, Childs B, Cryer P, Dagogo-Jack S, Fish L, et al. Hypoglycemia and diabetes: a report of a workgroup of the American Diabetes Association and the Endocrine Society. Diabetes Care. 2013;36(5):1384-95.