NOVARTIS

Clinical Development

QBW251

CQBW251B2201 / NCT04072887

A 24-week multi-center, double-blind, placebo controlled dose-range finding study to investigate the efficacy and safety of oral QBW251 in COPD patients on triple inhaled therapy (LABA/LAMA/ICS)

Statistical Analysis Plan (SAP)

Author:	Trial statistician [redacted]

Document type:	SAP Documentation
Document status:	Amendment 1
Release date:	26-Apr-2021
Number of pages:	56

[redacted]

All Sections

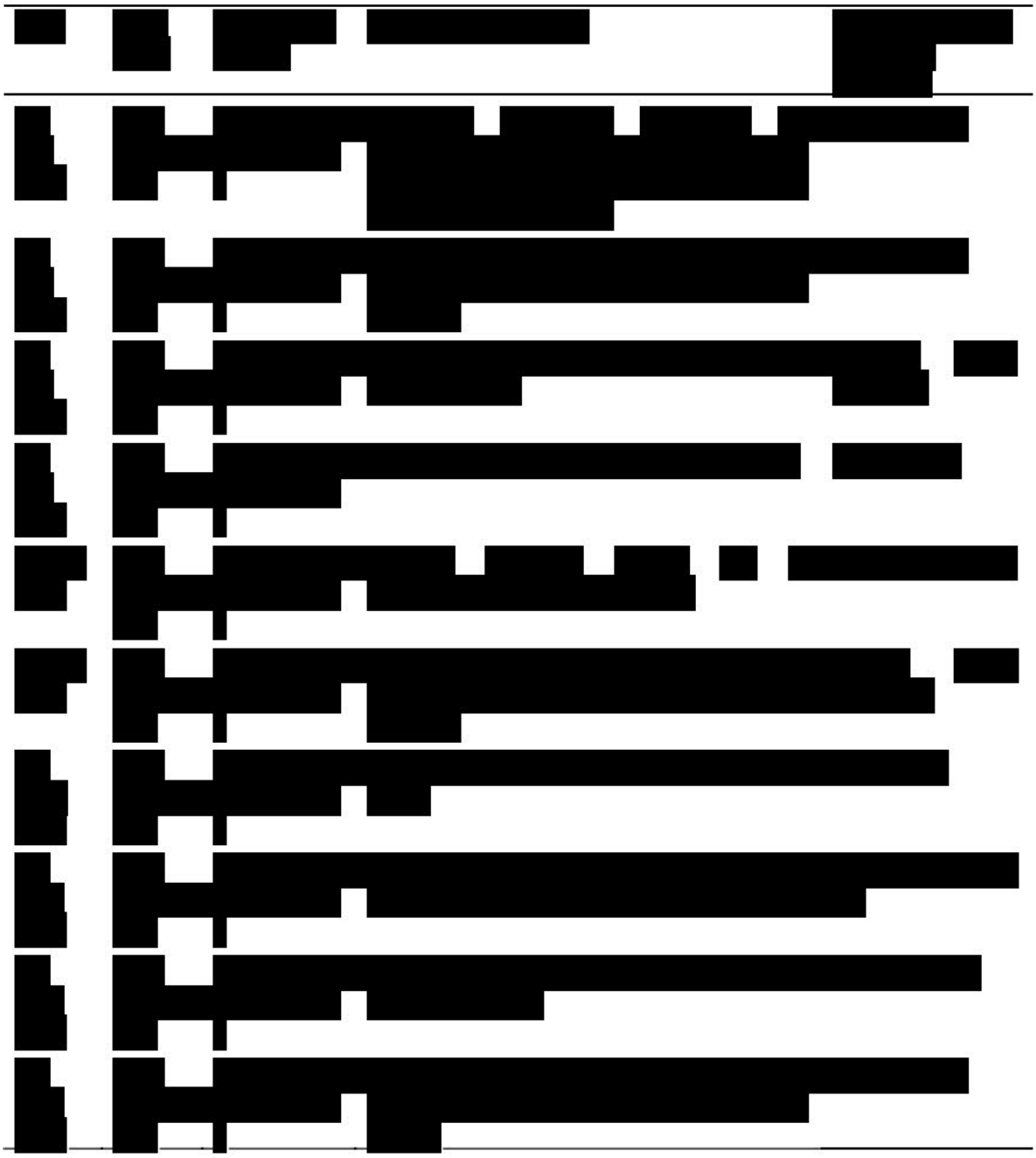


Table of contents

List of abbreviations

AE	adverse event
AIC	Akaike information criterion
Alb	Albumin
ALP	alkaline phosphatase
ALT	alanine aminotransferase
AST	aspartate aminotransferase
AUC	Area under the plasma concentration-time curve
$AUC_{0\text{-}Xh}$	Area under the plasma concentration curve over X hours
bid	bis in diem/twice a day
b.i.d.	twice a day
BMI	Body Mass Index
BUN	blood urea nitrogen
CASA-Q	Cough and sputum assessment questionnaire
[redacted]	[redacted]
[redacted]	[redacted]
C_{max}	Maximum concentration
C_{min}	Minimum concentration
COPD	Chronic Obstructive Pulmonary Disease
CSR	Clinical Study report
[redacted]	[redacted]
DMC	Data Monitoring Committee
[redacted]	[redacted]
DR	Dose response
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
ED50	efficacious dose in 50% of subjects
E-RS	Evaluating Respiratory Symptoms in COPD
[redacted]	[redacted]
FAS	Full analysis set
FDA	Food and Drug Administration
FEV_1	Forced expiratory volume in one second
FVC	Forced vital capacity
GOLD	The Global Initiative for Chronic Obstructive Lung Disease
GLM	generalized linear model
h	Hour
[redacted]	[redacted]
IA	Interim analysis
ICS	Inhaled CorticoSteroid
J2R	jump-to-reference
LABA	Long-Acting β2-Agonist
LAMA	Long-Acting Muscarinic receptor Antagonist

LFT	Liver function test
LLOQ	Lower limit of quantification
MAR	Missing at random
MCP-Mod	Multiple comparison procedure-modelling
MED	Minimum effective dose
MedDRA	Medical dictionary for regulatory activities
MNAR	Missing not at random
mg	milligram(s)
mL	milliliter(s)
MMRM	Mixed-effect linear model for repeated measures
ng	nanogram
PD	pharmacodynamic(s)
PDS	Programming dataset specifications
PGI-S	Patient Global Impression of Severity
PK	pharmacokinetic(s)
PT	Preferred term
QTc	corrected QT interval
QTcF	QT interval with Fridericia's correction
RAS	Randomized analysis set
SAE	serious adverse event
SAP	Statistical analysis plan
SCS	Systemic corticosteroids
SD	standard deviation
SGRQ	St George's Respiratory Questionnaire
SOC	System Organ Class
TFLs	Tables, Figures, Listings
TBL	total bilirubin
T_{max}	time to reach maximum (peak) plasma concentration following drug administration
ug	microgram
ULN	upper limit of normal
TFLs	Tables, Figures, Listings

1 Introduction

This document contains details of the statistical methods that will be used in the Phase 2b clinical trial CQBW251B2201. The purpose of this Phase 2b study is to support the dose selection for future studies by evaluating efficacy and safety of different QBW251 doses in COPD patients with chronic bronchitis and a history of exacerbations, compared to placebo, when added to a triple combination therapy of LABA/LAMA/ICS. Lung function improvement will be the primary endpoint: trough FEV_1 change from baseline after 12 weeks of treatment.

Data will be analyzed according to Section 12 of the study protocol.

Important information is given in the following sections and details are provided, as applicable, in Section 5.

1.1 Study design

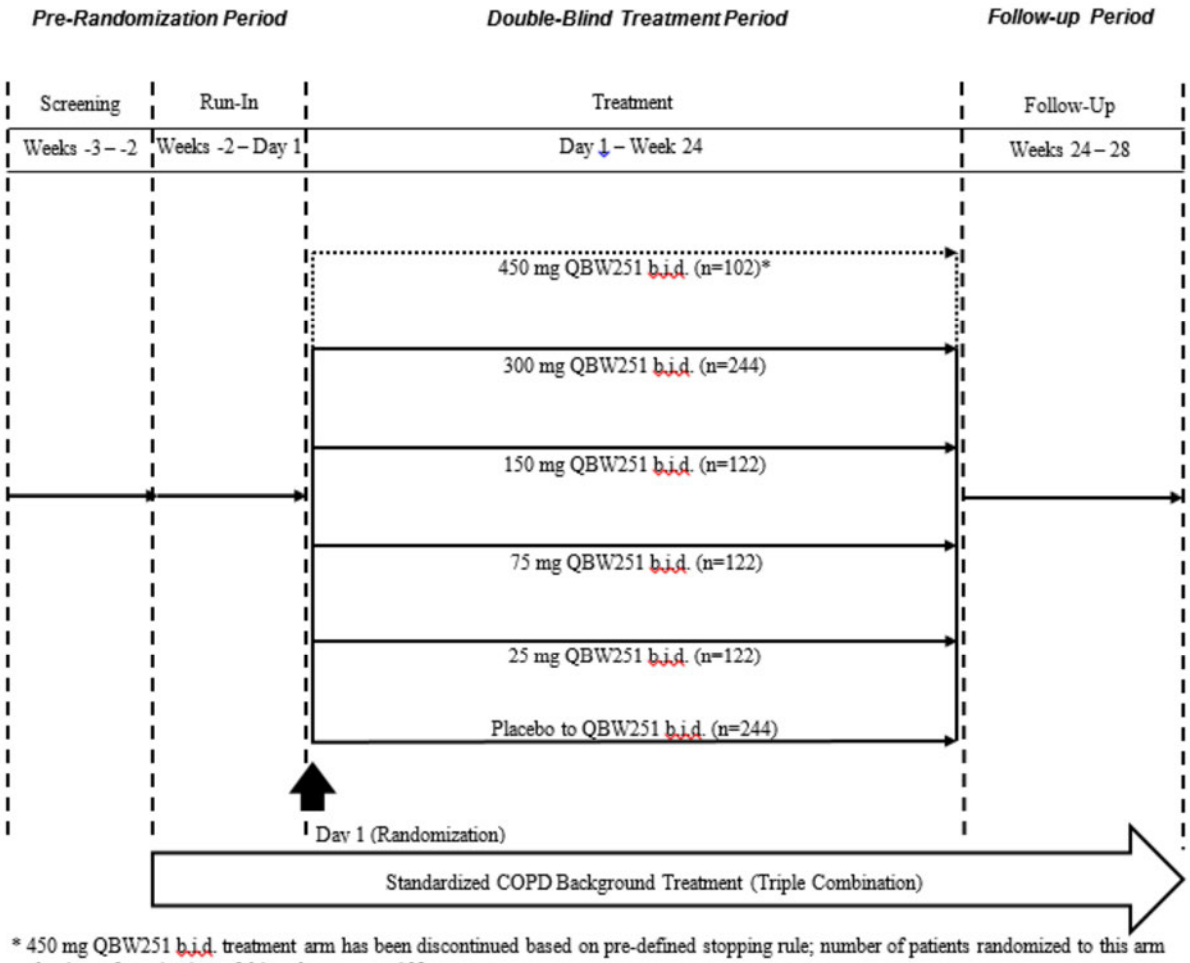
This study uses a 6 treatment arm, parallel-group, randomized, double-blind study design. One treatment arm (450 mg b.i.d.) was discontinued after reaching a pre-defined stopping criterion (see Section 12.7.1 of the protocol) at the first DMC interim analysis.

Patients initially undergo a screening period of 1 week. This screening period (Weeks -3 to -2) serves to assess eligibility and to taper patients off disallowed medications. Subsequently, patients enter the run-in period (2 weeks; Days -14 to 1) to establish baseline values for symptom assessments, to standardize the COPD background therapy (triple combination LABA/LAMA/ICS), and to complete eligibility assessments. At visit Run-In 2 final eligibility will be assessed. Eligible patients will move into the Day 1 visit where they will be randomized. Day 1 also generates additional baseline data and study treatment is initiated. The treatment period consists of 24 weeks, during which the patient returns to the site for regular visits (Day 1 – Week 24). Upon completion of the treatment period, patients are followed up for safety assessments during the Follow-Up period (Weeks 25 – 28).

Approximately 956 male and female COPD patients are randomized into 1 of 6 treatment arms with a randomization ratio of 2:2:1:1:1:2 (450 mg b.i.d. (discontinued), 300 mg b.i.d., 150 mg b.i.d., 75 mg b.i.d., 25 mg b.i.d., placebo). Detailed information regarding sample size calculation is provided in Section 3. After reaching the pre-defined exposure-based stopping criterion, patients were no longer randomized to the 450 mg b.i.d. treatment, resulting effectively in a 2:1:1:1:2 randomization ratio (300 mg b.i.d.:150 mg b.i.d.: 75 mg b.i.d.: 25 mg b.i.d.: placebo) from the day of 450 mg b.i.d. discontinuation (20-Apr-2020) onward.

Randomization will be stratified by smoking status (current or ex-smoker) and severity of airflow limitation (FEV1 $\geq$ 30% to < 50% and $\geq$ 50% to < 80%).

Figure 1-1 Study design



* 450 mg QBW251 b.i.d. treatment arm has been discontinued based on pre-defined stopping rule; number of patients randomized to this arm at the time of termination of this cohort was n=102.

The primary endpoint will be assessed after 12 weeks of treatment.

Interim analyses (IA) on safety data are planned for an independent external safety data monitoring committee (DMC) to evaluate the results of pre-specified interim PK and safety data. Up to 5 DMC IAs are proposed to occur. Detailed information regarding DMC analysis will be provided in the DMC charter and a separate DMC SAP.

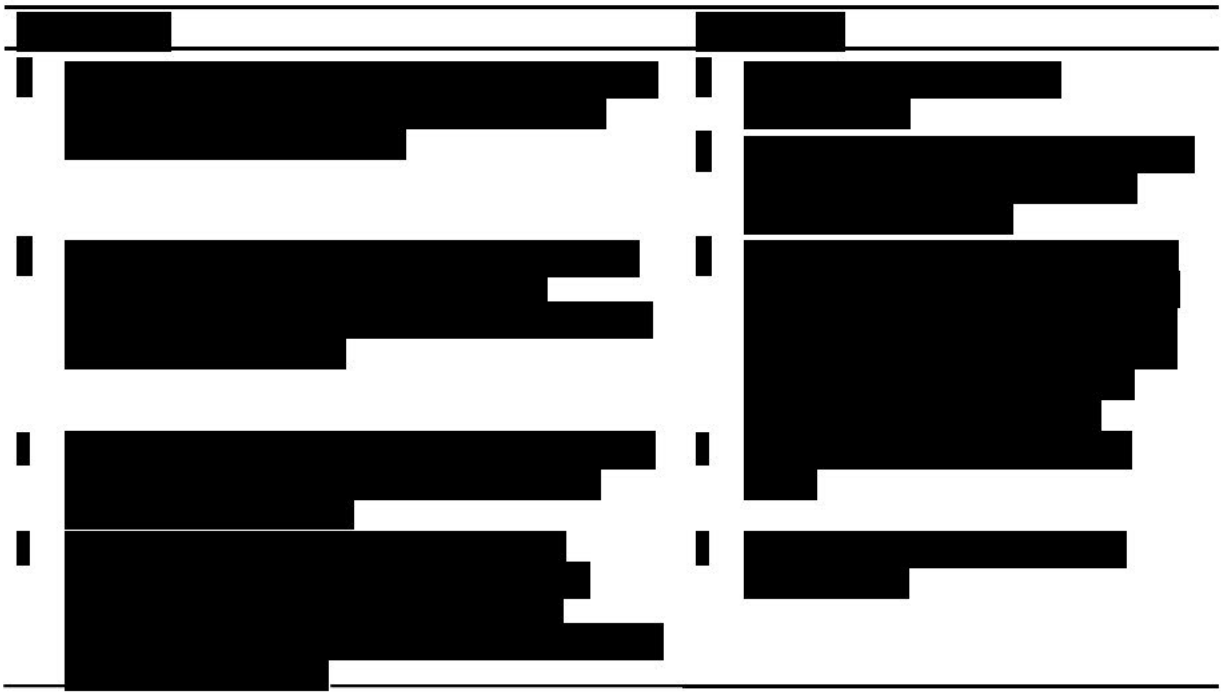
Up to 2 interim analyses (IAs) may be conducted on efficacy data. The purpose of the IA(s) is to have an early assessment of the efficacy data for planning future studies in the QBW251B program. The first IA will be triggered when approximately the first 645 to 680 randomized patients have either discontinued or completed 12 weeks of treatment. If the first IA does not lead to a clear decision for the planning of the QBW251B program, a second IA may be performed when all randomized patients have either discontinued or completed 12 weeks of treatment. The efficacy IA results will not trigger any design adaptations.

1.2 Study objectives and endpoints

Section 2 of the study protocol lists the following primary, key secondary, other secondary [redacted] objectives.

Table 1-1 Objectives and related endpoints

Objective(s)	Endpoint(s)
Primary objective(s)	**Endpoint(s) for primary objective(s)**
• Characterize the dose-response relationship of QBW251 administered orally over 12 weeks on lung function, compared to placebo when added to inhaled triple combination therapy (long-acting β2-agonist/long-acting muscarinic receptor antagonist/inhaled corticosteroid; LABA/LAMA/ICS).	• Trough FEV_1 change from baseline after 12 weeks of treatment.
Secondary objective(s)	**Endpoint(s) for secondary objective(s)**
• Evaluate symptoms (overall COPD symptoms, cough and sputum) across various dose levels of QBW251 administered orally over 24 weeks, compared to placebo at Weeks 12 and 24.	• Change from baseline in the Evaluating Respiratory Symptoms in COPD (E-RS) weekly mean scores (total and subscale scores). • Change from baseline in Patient Global Impression of Severity (PGI-S) score. • Change from baseline in the Cough and Sputum Assessment Questionnaire (CASA-Q) domain scores - cough symptoms, cough impact, sputum symptoms, and sputum impact.
• Evaluate health-related quality of life across various dose levels of QBW251 administered orally over 24 weeks, compared to placebo, at Weeks 12 and 24.	• Change from baseline in St. George's Respiratory Questionnaire (SGRQ) total and domain scores at Weeks 12 and 24.
• Evaluate lung function across various dose levels of QBW251 administered orally over 24 weeks, compared to placebo, over 4, 8, 16, 20 and 24 weeks.	• Trough FEV_1 change from baseline after 4, 8, 16, 20 and 24 weeks of treatment, respectively.
• Evaluate safety and tolerability across various dose levels of QBW251, administered orally over 24 weeks, compared to placebo.	• Assessment of ECGs, laboratory tests, vital signs, and adverse events per treatment group.
• Assess the pharmacokinetics of QBW251 in COPD patients. Additionally this information may be used to understand the relation between drug exposure and efficacy and/or safety.	• Assessment of drug exposure (trough concentration; C_{min}) on all visits and around C_{max} on Days 1, 15 and 169. AUC and C_{max} on Days 1 and 15 in a sub-group of patients.
[redacted]	[redacted]



2 Statistical methods

2.1 Data analysis general information

All analyses will be performed by [redacted]. The most recent version of SAS available in the statistical programming environment of [redacted] will be used for the analysis.

Detailed information regarding DMC analysis will be provided in the DMC charter and a separate DMC SAP.

2.1.1 General definitions

The terms 'double-blind treatment' will be used in this document and refer to the double-blind QBW251 doses and Placebo.

2.1.1.1 Study day

Study day will be defined as the number of days since the date of first dose of double-blind treatment. The date of first dose of double-blind treatment will be defined as Day 1 and the day before the first dose of study drug will be defined as Day -1.

Therefore, for a particular date, study day will be calculated as follows:

for dates on or after the first date of double-blind treatment,

Study day = Assessment date – Date of first dose of double-blind treatment + 1;

for dates prior to the first date of double-blind treatment,

Study day = Assessment date – Date of first dose of double-blind treatment.

2.1.1.2 Baseline definition

In general, baseline is defined as the last measurement before the first dose of double-blind treatment. Details on calculation of baseline will be provided in the latter sections.

2.1.1.3 Post-baseline measurement

Post-baseline measurements are defined as those assessments after the first dose of double-blind treatment.

When change from baseline is of interest the following formula will be used for each scheduled visit and time-point where baseline and post-baseline values are both available:

Change from baseline = post-baseline value – baseline value.

If not stated otherwise for efficacy analyses on-treatment values are defined as values taken post-baseline but no later than 1 day after last dose of double-blind treatment. Off-treatment values are defined as post-baseline values taken more than 1 day after last dose of double-blind treatment.

For safety analyses other than adverse events (AEs), e.g. laboratory, ECG, vital signs, on-treatment values are defined as values taken post-baseline but no later than 30 days after last dose of double-blind treatment.

AEs will be considered as on-treatment if the event is starting on or after the time of first dose of double-blind treatment but not later than 30 days after last dose of double-blind treatment.

Details on calculation of post-baseline values are provided in the latter sections.

If not stated otherwise, 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. When visit windows are used, all visits will be re-aligned, i.e., they will be mapped into one of the visit windows. Details on calculation of visit windows will be provided in Section 5.

2.1.1.4 Handling early termination of the 450mg b.i.d arm

If not stated otherwise, all efficacy inferential analyses will not include the 450 mg dose due to early termination of the 450 mg b.i.d arm, resulting in very few or no data in some of the strata (covariates/factors in the model) during the later visits. All descriptive summaries will include the 450 mg b.i.d. dose.

2.2 Analysis sets

The Randomized Analysis Set (RAS) consists of all randomized patients. Patients will be analyzed according to the treatment they were assigned to at randomization.

The Full Analysis Set (FAS) will include all randomized patients who received at least one dose of randomized treatment. Following the intent-to-treat principle, patients will be analyzed

according to the treatment they were assigned to at randomization. It is reasonable to require that patients took randomized treatment for inclusion in the FAS because the decision on whether or not randomized treatment is started will not be influenced by the treatment group assignment due to effective treatment blinding procedures.

The Safety Set will include all patients who received at least one dose of double-blind treatment. Patients will be analyzed according to the treatment they received. If a patient received more than one double-blind treatment, e.g. QBW251 300mg and QBW251 25mg, the patient will be analyzed according to the treatment taken for the majority of days.

Note that the FAS and Safety Set are the same except that the Safety Set allows the inclusion of non-randomized patients who receive double-blind treatment in error. In addition, analyses based on the FAS assigns randomized treatment while analyses based on the Safety Set assigns received treatment.

The PK set will include all patients with at least one evaluable drug concentration data sample. The PK serial sub-group will include all patients with at least one evaluable drug concentration data sample and who consented to participate in the PK serial sub-group. Patients will be analyzed according to the treatment received for both sets.

[redacted]

2.2.1 Subgroup of interest

Subgroup analyses of primary endpoint and selected secondary endpoints (E-RS cough & sputum scores) will be performed.

For tables presenting results from statistical models, the treatment effects in the subgroup will be derived using the appropriate interaction terms in the model and additional covariate as a fixed effect if necessary. For subgroup analyses by severity of airflow limitation, run-in post-bronchodilator FEV_1 will be removed from the model.

The following subgroups will be used for supporting analyses:

- Smoking status (current, former)
- Severity of airflow limitation (moderate, severe)

2.3 Patient disposition, demographics and other baseline characteristics

2.3.1 Patient disposition

The RAS will be used for the summary and listing of patient disposition. The screening disposition and the analysis sets table will be based on all screened patients.

The number of patients in the RAS will be summarized by region, country, center and treatment group. Further, the overall number of patients who entered completed, and discontinued study will be summarized including the reasons for discontinuation for each period: pre-randomization, double-blind treatment and follow-up.

Number of patients with protocol deviations will be tabulated by category (e.g., selection criteria not met, subject not withdrawn as per protocol, treatment deviation, prohibited concomitant medication, other) and deviation.

The number of patients included in each analysis set will be tabulated. Patients exclusion from analysis sets will be listed for all patients with reasons for exclusion (i.e. including both protocol and non-protocol deviations).

2.3.2 Patient demographics and other baseline characteristics

Demographics and baseline characteristics will be summarized and listed using the RAS.

Demographic and baseline characteristics including age, gender, race, ethnicity, region (Latin and South America, Eastern Europe, Western Europe, Asia and Australia, North America), height, weight, body mass index (BMI), duration of COPD (as time span between date of diagnosis of COPD to date of enrollment), run-in spirometry (FEV_1, FVC, FEV_1/FVC and % of predicted FEV_1 pre- and post-bronchodilator, FEV_1 reversibility), severity of COPD (GOLD 2018), COPD exacerbation history, COPD assessment test (CAT) score, smoking status and history, vaping history, assigned background medication (Relvar+Incruse, Trelegy), , and cardiovascular risk factors at baseline will be summarized.

Summaries of continuous variables like age will include mean, standard deviation (SD), first and third quartile, median, minimum and maximum. Summaries of categorical variables will show absolute (n) and relative (%) frequencies including a category for missing data if any.

No statistical analyses will be provided for baseline comparability among the treatment groups.

In addition, the following categorizations of continuous variables will be done:

- Age into 40 - 64 years, and ≥ 65 years;
- BMI into ≤ 30.0 kg/m2 and > 30.0 kg/m2;
- Duration of COPD into < 1 year, 1 - 5 years, > 5 - 10 years, > 10 - 15 years, > 15 – 20 years, and > 20 years;
- [redacted]
- Severity of airflow limitation: pre-bronchodilator FEV1 % predicted values into <30%, 30 - < 50%, 50% - < 80%, and ≥ 80% (< 30% and ≥ 80% will only be included in case of protocol deviation).
- CAT into (1 - 9 mild, 10 - 20 moderate, 21 - 30 severe, 31 - 40 very severe) (mild will only be included in case of protocol deviation)

2.3.3 Medical history/current medical condition

Medical history will be coded with the Medical Dictionary for Regulatory Activities terminology (MedDRA) using the most recent version at the time of database lock. History/conditions, including pre-specified protocol solicited events (pulmonary diseases and cardiovascular events), will be summarized for the RAS by primary system organ class and preferred term. In addition pulmonary diseases and cardiovascular events will be summarized be pre-specified medical history term.

2.4 Treatments (study treatment, rescue medication, concomitant therapies, compliance)

All summaries of treatments will be performed on the Safety Set.

2.4.1 Study treatment / compliance

The duration of exposure in days to each treatment group (QBW251 dose or placebo) will be summarized by means of descriptive statistics.

Duration of exposure to double-blind treatment will be calculated as the number of days between the first dose date and the last dose date exposed to that treatment over the specified period (expressed as: Duration of exposure = Date of last known dose of double-blind treatment – Date of first dose of double-blind treatment + 1).

The number of patients who permanently discontinued from double-blind treatment and the reasons will be summarized by treatment group. Patients who permanently discontinued from double-blind treatment will be listed including reason and date of discontinuation.

Compliance will be calculated as the percentage of the number of days where double-blind treatment was administered as per protocol (i.e., one dose in the morning and one dose in the evening) divided by the duration of exposure (i.e., the number of days between first and last dose).

Compliance will be categorized by < 80 % and 80 % - 100 % and summarized by treatment group.

2.4.2 Prior, concomitant and post therapies

The following summaries for COPD-related medications will be performed separately for medications prior to run-in (medications starting prior to screening and ending prior to or at the run-in visit) and for concomitant medications (medications which were taken anytime between the first dose and last dose of randomized treatment, inclusive). COPD-related medications will be summarized by pre-specified drug categories, route of administration, preferred term, and treatment group.

Non-COPD related medication prior to and after the start of randomized treatment will be summarized by route of administration, preferred term, and treatment group.

Surgical and medical procedures (non-drug therapies) are coded using MedDRA. Presentations will be done by MedDRA primary system organ class and preferred term, separately for prior and concomitant procedures.

2.5 Analysis of the primary objective

The primary objective of this study is to characterize the DR (dose-response) efficacy relationship among QBW251 doses (25, 75, 150, and 300 mg b.i.d.) and placebo with regards to the change from baseline in trough FEV_1 after 12 weeks of treatment. The goals associated with this objective are below.

- To confirm an overall DR signal

- To estimate the dose(s) that corresponds to the target effect over placebo based on the estimated DR curve

Due to early termination of the 450 mg, analysis of the primary endpoint will not include the 450 mg dose.

2.5.1 Primary endpoint

The primary estimand, defined below, quantifies a hypothetical on-treatment Week 12 effect during stable periods (i.e. outside episodes of COPD worsening that require rescue medication or systemic corticosteroids). This estimand targets the maximum treatment effect for QBW251 and allows selection of the best dose for those who take it for 12 weeks.

- **Population**: patients with moderate to severe COPD enriched for presence of chronic bronchitis and a history of exacerbations and treated with LABA/LAMA/ICS background therapy
- **Variable**: trough FEV_1 (average of the 2 values taken 15 and 45 min pre-dose) change from baseline after 12 weeks of treatment
- **Intervention effect of interest**: Effect of interventions initiated at randomization during stable periods (i.e. outside episodes of COPD worsening that require rescue medication or systemic corticosteroids) and that would have been observed had all patients remained on their assigned treatment for 12 weeks, with the following post-randomization events accounted for by assessing:
  - Intake of rescue medication or systemic corticosteroids (SCS): the effect outside of periods of worsening disease that necessitate rescue medication or SCS
  - Discontinuation of study treatment or study participation: hypothetical on-treatment trough FEV_1 value.
  - Missing data due to any reason prior to completion or discontinuation of treatment: on-treatment trough FEV_1 values.
- **Summary measure**: mean difference between treatment groups (QBW251 compared with placebo)

The baseline value is defined as the average of the FEV_1 values taken approximately 45 and 15 minutes prior to first dose of randomized treatment at Day 1. If one of the 2 values is missing (or is not confirmed to be pre-dose) then the remaining non-missing value will be taken as the baseline. If both values are missing (or are not confirmed to be pre-dose), then the measurements taken at the Run-In 1 visit will be used as the baseline.

2.5.2 Statistical hypothesis, model, and method of analysis

The Multiple Comparison Procedure – Modelling (MCP-Mod) methodology (see Bretz et al 2005 and Pinheiro et al 2014) will be employed to assess the primary objective. An overview of the steps for the MCP-Mod methodology is given below.

Step 1 (Testing an overall dose-response signal - MCP part):

The (covariate) adjusted mean responses at each individual dose will be obtained from a mixed-effect linear model for repeated measures (MMRM) with terms for baseline FEV_1, visit, region,

treatment, smoking status at screening, severity of airflow limitation, run-in post-bronchodilator FEV_1, treatment-by-visit interaction, and baseline FEV_1-by-visit interaction as fixed effects. To allow adjustment for correlations between time points within patients, an unstructured variance-covariance structure will be used. The estimated treatment difference and the associated 90% confidence intervals will be presented for the treatment contrast of each dose versus Placebo.

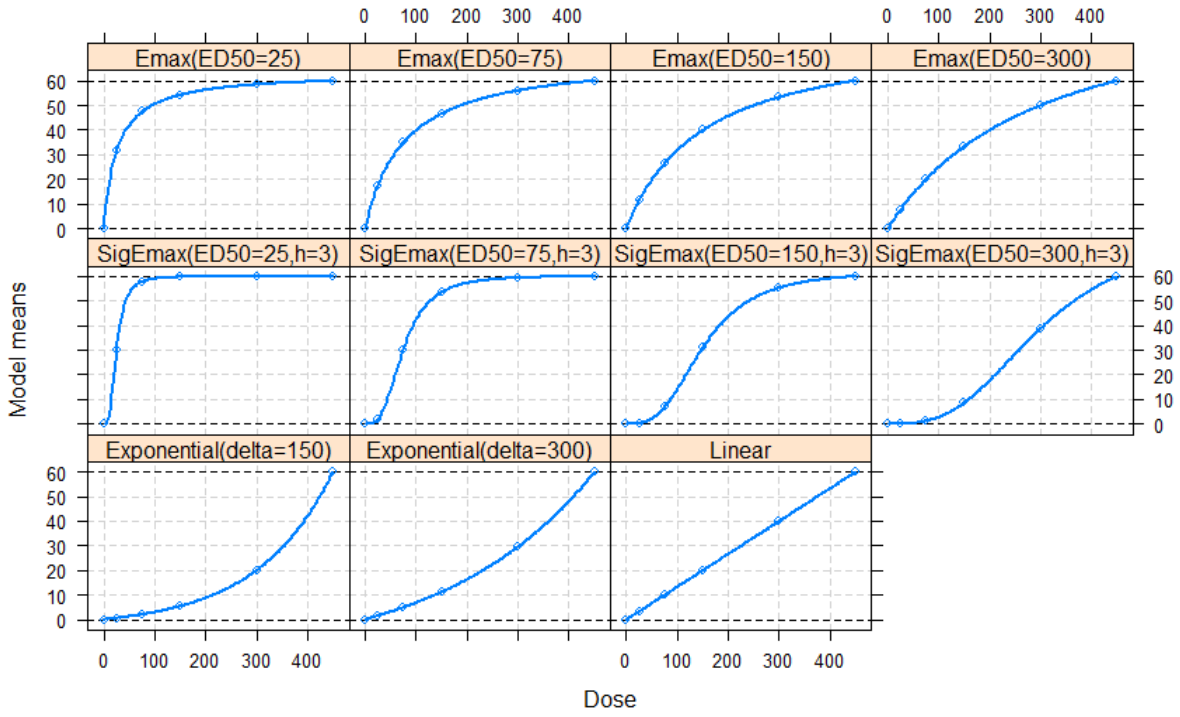
The adjusted treatment means from the MMRM will be used to test the null hypothesis of a flat DR relationship for the primary efficacy endpoint at a one-sided significance level of 5% against the alternative hypothesis of a non-constant DR curve. The testing will be performed with a multiple contrast test described in the MCP-Mod methodology.

A wide range of possible dose-response relationship will be considered to take model uncertainty into account (Figure 2-1). A monotonic DR is assumed, supported by preclinical data and the observed DR of another CFTR-potentiator, ivacaftor. Eleven candidate DR curves (4 Emax, 4 Sigmoid Emax, 2 Exponential, and 1 Linear) will be used to derive the optimal model contrasts for the multiple contrast tests. Emax and sigmoidal Emax models are considered based on the observed DR of ivacaftor (FDA 2012). The ED50 values (the dose at which half of the maximum effect is reached) for the Emax models will be 25, 75, 150, and 300. The parameters of the sigmoidal Emax models (ED50, h) will be (25, 3), (75, 3), (150, 3) and (300, 3) where h is the Hill parameter that determines the steepness of the dose-response shape. Exponential models are considered since there is an over-proportional increase of exposures at the high end of the dose range tested. The parameter controlling the convexity of the Exponential models will be 150 and 300.

For each of the 11 candidate DR curves, a contrast test statistic will be derived that maximizes the power assuming the true mean response is the one assumed by the candidate DR curve. The detection of a significant DR signal is based on the maximum of the 11 contrast test statistics. The overall null hypothesis of no DR relationship is rejected if the multiplicity adjusted p-value for at least one contrast test is < 0.05 (one sided).

For each candidate DR curve, the test statistics and corresponding adjusted p-values will be presented.

Figure 2-1 Candidate Dose Response Curves



Due to termination of the 450 mg b.i.d. arm candidate curves Emax(ED50=300), SigEmax(ED50=300) and Exponential(delta=300) may no longer be appropriate for MCP-Mod analyses that do not include the 450 mg b.i.d. arm. These candidate curves will be removed and the Exponential(delta=75) will be added.

Step 2 (Estimation of the dose-response curve and target dose – Mod part):

Once the DR signal is declared, the DR curve and the target dose(s) of interest will be estimated by model averaging. A large number of bootstrap samples from the multivariate normal distribution will be drawn with adjusted means from the MMRM and corresponding covariance matrix. For each sample:

- DR models from the candidate families (Emax, Sigmoid Emax, Exponential, Linear) will be fitted to the data and the best model according to the generalized Akaike information criterion (AIC) will be chosen.
- The predictions for dose-response will be obtained from the best model for each sample.

The final DR curve estimate is the median of these predictions while confidence intervals will be calculated from the quantiles. The final DR curve estimate with the model-based two-sided 90% confidence interval will be presented graphically. In addition, the plot will include the mean responses from the MMRM and the associated 90% confidence intervals for each of the studied dose groups.

The target dose(s) that corresponds to a clinically relevant effect over placebo can be estimated using inverse regression techniques (Bretz et al 2005). The target dose will be estimated for a delta of 50ml and presented graphically within the estimated DR curve.

All data related to safety, efficacy and other assessments including PK exposure will be taken into consideration to propose a dose for Phase 3.

2.5.3 Handling of missing values/censoring/discontinuations

Since the estimand is related to an effect outside of rescue med intake, spirometry measurements within 6 hours of rescue medication or within 7 days of intake of SCS use will be set to missing. If one of the values contributing to the trough FEV_1 variable is missing at a visit, the remaining non-missing value will be taken as trough FEV_1. If both values are missing, then their trough FEV_1 will be regarded as missing at that visit.

For the primary analysis, only on-treatment data (from date of first randomized dose up to 1 day after date of last randomized dose) will be used as the estimand specifies a hypothetical on-treatment effect. Missing on-treatment data will not be explicitly imputed as the MMRM model implicitly imputes missing data assuming the missing at random (MAR) mechanism.

The imputation procedure(s) related to the trough FEV_1 supportive analyses (Section 2.5.4) are "jump-to-reference" (J2R) and MAR (Carpenter et al 2013). The imputations will be based on all available data (i.e. from all scheduled time points) using all covariates as specified in the MMRM. For J2R, only placebo (reference) data will be used. For MAR, data from the same treatment arm will be used for building the imputation model. Imputation of intermittent missing observations before treatment discontinuation will be carried out following a MAR mechanism for all treatment arms. Additional details are provided in Section 5.7.

2.5.4 Supportive analyses

A supplementary analysis will be performed that quantifies the treatment effect in all randomized patients during stable periods (i.e. outside episodes of COPD worsening that require rescue medication or SCS) with an adherence to treatment like we would see in clinical practice in a world without COVID-19, with the following post-randomization events accounted for by assessing:

a. Intake of rescue medication or systemic corticosteroid: effect outside of periods of worsening disease that necessitate rescue med or systemic corticosteroid
b. Discontinuation of study treatment for non-COVID-19 related reason with follow-up: actual off-treatment value. If no data was retrieved after study treatment discontinuation, missing data will be multiply imputed based on placebo arm data: J2R assumption for the QBW251 arms and MAR assumption for placebo arm.
c. Discontinuation of study treatment due to a COVID-19 related reason: on-treatment value. Missing data will be multiply imputed based on the MAR assumption.
d. Discontinuation of study participation due to any non-COVID -19 related reason: off-treatment value. Missing data will be multiply imputed based on placebo arm data: J2R assumption for the QBW251 arms and MAR assumption for placebo arm.

e. Any other event leading to missing data prior to completion or discontinuation of study treatment: on-treatment value. Missing data will be multiply imputed based on the MAR assumption.

Results will be presented similarly to those of primary analyses.

Details are provided in Section 5.7.

A supportive analysis will be performed including the 450mg dose and only data up to week 12. This will be done similarly as the primary analysis, however in addition candidate curves Emax(ED50=300), SigEmax(ED50=300) and Exponential(delta=300) will be considered.

A supportive exposure-response analysis may be performed outside the CSR.

2.6 Analysis of the key secondary objective

There is no key secondary objective defined in the protocol.

2.7 Analysis of secondary efficacy objective(s)

2.7.1 Secondary endpoints

No multiplicity adjustment will be carried out for secondary analyses described below. In addition, the treatment effect of QBW251 compared to placebo that would have been observed had all patients remained on their assigned treatment will be estimated. Only on-treatment data will be used.

All analysis of secondary endpoints will be performed on the FAS.

Missing data for any reason will not be explicitly imputed and will be handled by the respective mixed effects model which implicitly imputes missing data assuming MAR.

If not stated otherwise, in the 450mg dose will be excluded from all inferential statistics. Descriptive analyses will include all doses.

Details on statistical models are provided in Section 5.7.

2.7.1.1 COPD symptoms

2.7.1.1.1 Respiratory symptoms from E-RS

The E-RS assesses both overall daily respiratory COPD symptoms (Total score) and specific respiratory symptoms using 3 subscales (Breathlessness, Cough & Sputum, and Chest Symptoms). Higher scores indicate more severe symptoms. Details on derivation of scores and handling of missing data are provided in Section 5.6.

The mean baseline E-RS Total and subscale scores will be the average of the corresponding daily scores from the run-in period. The last 14 days of the run-in period, i.e. the scheduled duration of run-in period, will be considered.

The daily scores post-randomization will be averaged for each week (Week 1, Week 2, etc.).

Monthly means over a 4-week period will also be calculated.

E-RS Cough & Sputum score

The change from baseline in the E-RS Week 12 Cough & Sputum mean scores will be analyzed using the same MCP-Mod approach described in for the primary variable in Section 2.5. The change from baseline in the E-RS Cough & Sputum weekly mean scores will be analyzed using a similar MMRM as described in Section 2.5 for the primary variable, with the appropriate baseline score replacing the baseline FEV_1 value in the model. The estimated treatment difference (QBW251 – placebo) at each week will be reported along with the associated 90% confidence interval. Results from MCP-Mod will be presented as described for primary variable in Section 2.5.

The proportion of patients who achieve a clinically important improvement (decrease) of at least 0.7 (Leidy et al 2014b) in the weekly mean score will be analyzed using a repeated measurements logistic regression. The model will include the same terms as for the MMRM analysis of the weekly scores. The estimated odds ratios will be displayed over time along with the associated 90% confidence intervals for treatment comparison of each dose versus Placebo.

The change from baseline in the monthly mean scores will be analyzed using a similar MMRM as described for weekly scores. The proportion of patients who achieve a clinically important improvement (decrease) in the monthly mean score will also be analyzed as described for weekly scores.

E-RS Cough & Sputum weekly mean and monthly mean scores will be summarized descriptively.

E-RS Total, Breathlessness, and Chest Symptoms scores

The change from baseline in the E-RS weekly mean Total, Breathlessness, and Chest Symptom scores will be analyzed using a similar MMRM as described in Section 2.5 for the primary variable with the appropriate baseline E-RS score replacing the baseline FEV_1 value. The estimated treatment difference (QBW251 – placebo) will be reported along with the associated 90% confidence interval.

E-RS mean weekly total score will be analyzed in addition using MCP-Mod as described for primary endpoint.

The minimum clinically important improvements (decrease) for the scores are defined below (Leidy et al 2014b).

- Total ≥ 2.0
- Breathlessness ≥ 1.0
- Chest Symptoms ≥ 0.70

The proportion of patients who achieve a clinically important improvement in the weekly mean scores will be analyzed using a repeated measurements logistic regression. The model will include the same terms as for the MMRM analysis of the weekly mean scores. The estimated odds ratios will be displayed over time along with the associated 90% confidence intervals for treatment comparison of each dose versus Placebo.

The change from baseline in the monthly mean scores will be analyzed using a similar MMRM as described above. The proportion of patients who achieve a clinically important improvement (decrease) in the monthly mean score will also be analyzed as described above.

E-RS Total, Breathlessness, and Chest Symptom weekly mean and monthly mean scores will be summarized descriptively.

2.7.1.1.2 Cough and sputum symptoms from CASA-Q

CASA-Q assesses cough and sputum symptoms and their impact with 4 domain scores (cough symptoms, cough impact, sputum symptoms, and sputum impact). Details on derivation of scores and handling of missing data are provided in Section 5.6.

The change from baseline in the domain scores will be analyzed using the MMRM described in Section 2.5 for the primary variable with the appropriate baseline domain score replacing the baseline FEV_1 value in the model. The estimated treatment difference (QBW251 – placebo) at each visit will be reported along with the associated 90% confidence interval.

CASA-Q domain scores will be summarized descriptively by visit.

2.7.1.1.3 Severity of symptoms from PGI-S

The PGI-S rates the severity of the respiratory symptoms and of cough and mucus. Shift tables will be used to compare baseline to Weeks 12 and 24 for each PGI-S value.

The PGI-S change from baseline will be analyzed using a proportional odds model for repeated measures with the same terms as for the MMRM analysis described in Section 2.5 for the primary variable with the baseline PGI-S value replacing the baseline FEV_1 value. The estimated odds ratios will be displayed over time along with the associated 90% confidence intervals. For the proportional odds analysis, the change in severity scores will first be grouped into the categories below, for cough and mucus categories -5 and 5 will be added. Further groupings of the change categories (such as better, no change, and worse) may be performed prior to the proportional odds analysis, if the sample size for some of the categories are too small.

Change from baseline in severity score	Change category
-4	Very much better
-3	Much better
-2	Better
-1	Slightly better
0	No change
1	Slightly worse
2	Worse
3	Much worse
4	Very much worse

2.7.1.2 Health-related quality of life from SGRQ

SGRQ measures health impairment. A total score is produced, along with the domain scores of symptoms, activity, and impacts. Details on derivation of scores and handling of missing data are provided in Section 5.6.

The Weeks 12 and 24 change from baseline in the SGRQ total and domain scores will be analyzed using the MMRM described in Section 2.5 for the primary variable with the appropriate baseline SGRQ score replacing the baseline FEV_1 value in the model. The estimated treatment difference (QBW251 – placebo) at each visit will be reported along with the associated 90% confidence interval.

The proportion of patients who achieve a clinically important improvement (decrease) of at least 4 in the total score will be analyzed using a repeated measurements logistic regression. The model will include the same terms as for the MMRM analysis of the total scores. The estimated odds ratios will be displayed over time along with the associated 90% confidence intervals.

SGRQ total and domain scores will be summarized descriptively by visit.

2.7.1.3 Lung function

The change from baseline in trough FEV_1 at visits other than Week 12 will be analyzed using the same MMRM as described for the primary analysis. The estimated treatment difference (QBW251 – placebo) at each visit will be reported along with the associated 90% confidence interval. The change from baseline in trough FEV_1 at Week 24 will also be analyzed using the same MCP-Mod approach. Supportive analyses as described for primary endpoint in Section 2.5 will be performed for change from baseline in trough FEV_1 at Week 24.

The change from baseline in trough FVC after 4, 8, 12, 16, 20, and 24 weeks of treatment will be analyzed using the MMRM model described in Section 2.5 for the primary variable with the appropriate baseline FVC value replacing the baseline FEV_1 value in the model. The estimated treatment difference (QBW251 – placebo) at each visit will be reported along with the associated 90% confidence interval.

Trough FEV_1 and trough FVC will be summarized descriptively by visit.

2.8 Handling of missing values for FVC will be similar to FEV_1.Safety analyses

Safety summaries include only data from the on-treatment period with the exception of baseline data which will also be summarized where appropriate (e.g. change from baseline summaries).

All safety analysis will be performed on the Safety Set.

2.8.1 Adverse events (AEs)

The on-treatment period for adverse events lasts from the date of first administration of double-blind study treatment to 30 days after the date of the last actual administration of double-blind

study treatment. Adverse events will be summarized by treatment group. Summary tables for AEs will summarize only on-treatment events (also known as treatment-emergent AEs).

The number (and percentage) of subjects with treatment-emergent adverse events will be summarized in the following ways:

- by treatment and preferred term (PT).
- by treatment, primary system organ class (SOC) and PT.
- by treatment, SOC, PT and maximum severity.

Separate summaries by SOC and PT will be provided for:

- AEs suspected to be study medication related
- fatal AEs
- leading to treatment discontinuation
- leading to dose interruptions and/or adjustments
- serious adverse events (SAEs)

Further SAEs will be summarized by PT.

In addition, treatment emergent AEs and SAEs will be presented by treatment, system organ class and preferred term, showing exposure adjusted event rates (i.e., the number of events expressed in rates per 100 patient years).

Unless otherwise specified, SOCs will be sorted alphabetically and, within each SOC, the PTs will be sorted in descending order of frequency in the highest QBW251 dose. A subject with multiple adverse events within a SOC or PT is only counted once towards the total of the SOC or PT.

Listings will be provided for all AEs, SAEs and fatal AEs.

2.8.1.1 AE reporting for CT.gov and EudraCT

For the legal requirements of clinicaltrials.gov and EudraCT, two required tables on treatment emergent adverse events which are not serious adverse events with an incidence greater than 2% and on treatment emergent serious adverse events and SAE suspected to be related to study treatment will be provided by system organ class and preferred term on the safety set population.

If for a same patient, several consecutive AEs (irrespective of study treatment causality, seriousness and severity) occurred with the same SOC and PT:

- a single occurrence will be counted if there is ≤ 1 day gap between the end date of the preceding AE and the start date of the consecutive AE
- more than one occurrence will be counted if there is > 1 day gap between the end date of the preceding AE and the start date of the consecutive AE

For occurrence, the presence of at least one SAE / SAE suspected to be related to study treatment / non SAE has to be checked in a block e.g., among AE's in a ≤ 1 day gap block, if at least one SAE is occurring, then one occurrence is calculated for that SAE.

The number of deaths resulting from SAEs suspected to be related to study treatment and SAEs irrespective of study treatment relationship will be provided by SOC and PT.

2.8.1.2 Adverse events of special interest / grouping of AEs

The number and percentage of patients who reported treatment-emergent adverse events of special interest will be summarized by risk name, PT and treatment group.

Risk names will be sorted alphabetically and, within each risk name, the PTs will be sorted in descending order of frequency in the highest QBW251 dose. If a patient reported more than one adverse event with the same PT, the AE will be counted only once. If a patient reported more than one AE within the same risk, the patient will be counted only once at that risk.

The Compound Case Retrieval Strategy (CRS) will be used to determine the MedDRA search criteria to be used to identify events of special interest. The most recent list of adverse events of special interest at the time of database lock will be used.

SAEs of special interest will be summarized by risk name, PT and treatment group.

2.8.2 In addition, treatment emergent AEs of special interest will be presented by treatment and risk, showing exposure adjusted event rates (i.e., the number of events expressed in rates per 100 patient years).Deaths

Fatal AEs will be summarized and listed as specified in Section 2.8.1.

2.8.3 Laboratory data

Summaries of laboratory data will include on-treatment measurements, which are defined as measurements taken post-baseline but no later than 30 days after last dose of double-blind treatment.

Absolute values and change from baseline will be summarized for continuous laboratory parameters by visit, including the worst on-treatment value. The direction of interest for worst on-treatment value for selected hematology and biochemistry parameters is shown in Section 5.4.

For selected laboratory tests, the number and percentage of patients with newly occurring or worsening laboratory abnormalities meeting the clinically notable criteria at any time on-treatment, considering all on-treatment data from scheduled, unscheduled and premature discontinuation visits, will be summarized by laboratory parameter. Notable criteria are defined in Section 5.4.

Furthermore, the number and percentage of patients with newly occurring or worsening abnormalities in liver function tests (LFT) will be summarized by treatment and at any time on-treatment considering all on-treatment data from scheduled, unscheduled and premature discontinuation visit. LFT criteria are defined in Section 5.4.

The baseline value is the last value prior to first dose of double-blind treatment.

For selected laboratory parameter boxplots of laboratory values relative to the normal ranges will be provided.

Laboratory data of patients with notable values will be listed. Urinalysis will be listed for patients with abnormal urine dipstick.

2.8.4 Other safety data

2.8.4.1 ECG and cardiac imaging data

Summaries of ECG data will include on-treatment measurements, which are defined as measurements taken post-baseline but no later than 30 days after last dose of double-blind treatment. Triplicate ECGs (3 ECGs are collected within about a five-minute window) will be performed. For the analysis of continuous ECG parameter the mean value of the 3 ECG measurements will be used, for ECG interpretation, all findings of the 3 ECGs will be considered.

Absolute values and change from baseline will be summarized for ECG parameters by visit and time point. The maximum on-treatment QTc value will be included in the summary.

The number and percentage of patients with newly occurring or worsening notable Fridericia's QTc values will be summarized at any time on-treatment, considering all on-treatment data from scheduled, unscheduled and premature discontinuation visits. For derivation of notable values, the mean of the triplicate ECGs will be used. Notable criteria are defined in Section 5.5.

The number and percentage of patients with ECG abnormalities will be summarized by evaluation type, abnormality finding, visit and time point.

The baseline value is the last value prior to first dose of double-blind treatment.

ECG data of patients with notable Fridericia's QTc values will be listed.

2.8.4.2 Vital signs

Summaries if vital signs data will include on-treatment measurements, which are defined as measurements taken post-baseline but no later than 30 days after last dose of double-blind treatment.

Absolute values and change from baseline will be summarized for vital sign parameters by visit including the minimum and maximum on-treatment value.

The number and percentage of patients with newly occurring or worsening notable vital sign values will be summarized at any time on-treatment, considering all on-treatment data from scheduled, unscheduled and premature discontinuation visits. Notable criteria are defined in Section 5.5.

The baseline value is the last value prior to first dose of double-blind treatment.

Vital sign data of patients with notable values will be listed.

2.9 Pharmacokinetic endpoints

PK measurements will be analyzed as reported; no visit windowing will be applied.

Descriptive summary statistics of QBW251B plasma concentration data will be provided by treatment and visit/sampling time point, including the frequency (n, %) of concentrations below the lower limit of quantification (LLOQ). Summary statistics will include mean (arithmetic and geometric), standard deviation, coefficient of variation (arithmetic and geometric), median, minimum and maximum. An exception to this is T_{max} where median, minimum and maximum will be presented. Concentrations below LLOQ will be treated as zero in summary statistics and for PK parameter calculations.

Pharmacokinetic parameters will be determined using non-compartmental method(s).

Summary of PK parameter of all patients in the PK set will include trough concentration C_{min} on all visits and C_{max} on Day 1, Week 2 and Week 24. Summary of PK parameter for patients in the PK serial sub-group will include C_{min}, C_{max}, $AUC_{0\text{-}8h}$, $AUC_{0\text{-}24h}$, and T_{max} on Day 1 and Week 2.

The number and percentage of patients in the PK serial sub-group above the exposure threshold of $AUC_{0\text{-}24h,ss}$= 91700 ng×h/mL will be presented overall and by visit.

The number and percentage of patients in the PK set above the exposure threshold $C_{min,ss}$=2942 ng/mL will be presented overall and by visit. This threshold is defined based on the exposure threshold $AUC_{0\text{-}24h,ss}$= 91700 ng×h/mL.

Details on derivation of exposure threshold and $AUC_{0\text{-}8h}$, $AUC_{0\text{-}24h}$ are provided in Section 5.6.

In addition, the plasma concentration data from this study may be combined with data from other studies to perform a population PK analysis, which will follow the principles outlined in the FDA Guidance for Industry 1999. These results, if performed, will be reported outside of the CSR.

[redacted]

[redacted]

[redacted]

[redacted]

2.13 Interim analysis

2.13.1 Safety

IAs are planned for the monitoring of PK and safety data and do not inflate the type I error for the primary efficacy hypothesis testing. Thus no adjustment for multiplicity is required. These analyses will be performed by an external Independent Statistician and an Independent Programmer for an independent external safety DMC. Persons directly involved in the conduct of the clinical trial will not be involved in performing the IA or reviewing the results. The DMC will be provided with reports which are semi-blinded. If necessary, the study treatment can be unblinded by the chair of the DMC. The DMC will review PK and safety data and make recommendations regarding the conduct of the trial and/or alteration of the current protocol, including termination of a treatment arm. As mentioned in Section 9.1.4.1 of the protocol, a treatment arm (along with all other treatment arms with higher doses administered) will be terminated if the observed proportion of patients above the exposure threshold ($AUC_{0\text{-}24h}$ = 91 700 ng×h/mL) is significant greater than expected. It is expected that < 5% of patients on 300 mg or 450 mg b.i.d. will exceed the threshold based on a conservative scenario. An α-spending function with Pocock type stopping boundary (as implemented in East 6.4) will be used to construct the stopping boundaries (Lan et al 1983) using the p-value scale such that the overall

significance level across all analyses is maintained at 0.05. At each interim analysis, an exact test on the proportion of subjects who exceed the exposure threshold will be performed to test the following hypotheses for these 2 treatment arms.

- Null hypothesis (Ho): the proportion of patients exceeding the exposure threshold < 5%
- Alternative hypothesis (Ha): the proportion of patients exceeding the exposure threshold ≥ 5%

The interim analyses are proposed to occur when approximately 20%, 40%, 60%, 80%, and 100% of subjects have completed at least 4 weeks of treatment, which should be long enough to determine if the patient's exposure will exceed the threshold or not. The corresponding stopping boundaries in the p-value scale are 0.015, 0.016, 0.017, 0.018, and 0.019 for the interim analyses. For example, when the first interim analysis is performed after 20% patients have completed 4 weeks of treatment in the 450 mg b.i.d. arm, the observed (i.e. nominal) p-value has to be equal to or smaller than 0.015 in order to reject Ho and terminate the treatment arm.

If the true proportion of patients in the 300mg or 450 mg b.i.d. treatment arm exceeding the exposure threshold is < 5%, then the probability of rejecting Ho would be very low (< 5%) with this group sequential test for the treatment arm. If the true proportion is 10% then the probability of rejecting Ho and stopping the treatment arm early would be high (>77%). If the timing of the actual interim analyses deviate from the proposed schedule (20%, 40%, 60%, 80%, and 100% completing at least 4 weeks of treatment), stopping boundaries will need to be recalculated using the pre-specified α-spending function and based on the actual sample size at each interim analysis. The observed p-values at the interim analyses will then be compared against the re-calculated stopping boundaries.

More details will be outlined in the DMC charter and a separate DMC SAP.

After reviewing the exposure data from the first interim analysis, the DMC informed the sponsor that the 450 mg b.i.d. arm had crossed the stopping boundary and recommended termination of the treatment arm. The DMC confirmed that the arm was well tolerated in this patient population and the stop was not based on adverse reactions to the 450 mg b.i.d. regimen. The sponsor followed the DMC's recommendation and discontinued investigational treatment from all patients randomized to the 450 mg b.i.d. arm.

2.13.2 Efficacy

Up to 2 IAs on efficacy data may be conducted and the timing will correspond to 1 or 2 safety DMC analyses. The purpose of these IA(s) is to have an early assessment of the data for planning future studies in the QBW251B program. The efficacy IA results will not trigger any change or decision in the conduct of this study.

An internal Novartis committee, external to the study team, will be set up to review the both the efficacy and safety data. Once team members are exposed to the unblinded IA data, they will no longer be involved in the further execution of the study. A charter will provide further details.

The first IA will be conducted when approximately 645 to 680 patients, including the ~102 patients randomized to the 450mg b.i.d. arm, have either discontinued or completed 12 weeks of treatment. This IA will have acceptable precision for the estimated trough FEV_1 dose-

response curve. The average half-length of the 90% confidence interval is around 26 mL and 32 mL for the estimated dose-response curve and the estimated placebo-adjusted dose-response curve, respectively. If the first IA does not lead to a clear decision for the planning of the QBW251B program, a second IA may be performed when all randomized patients have either discontinued or completed 12 weeks of treatment.

For efficacy, the interim analyses will include all data up to the cut-off date for the IA.

An analysis may be planned and reported separately on the relationship between measured plasma trough concentrations and trough FEV1 change from baseline as described in Section 2.10.

2.14 Additional analyses to assess the impact of COVID-19 pandemic

Number and percentage of patients enrolled (informed consent signed), discontinued double-blind treatment, completed double-blind treatment and exposed to double-blind treatment during the pre-COVID-19 pandemic period, the COVID-19 pandemic period and the post-COVID-19 pandemic period (if applicable) will be summarized by region and country using the FAS.

The pandemic periods are defined based on the start and end date of the pandemic in the respective region/country.

Demographics, baseline disease characteristics, and run-in spirometry will be summarized by time of enrollment (patients enrolled pre-COVID-19 pandemic, during COVD-19 pandemic, post- COVID-19 pandemic [if applicable]) to assess the impact of the COVID-19 pandemic on the study population.

Number and percentage of patients with COVID-19 related deviations will be summarized by deviation term and relationship to the COVID-19 pandemic. The PD process was used to identify COVID-19 related deviations even though these deviations were not true PDs.

To assess the impact of the COVID-19 pandemic on trough FEV1, the MMRM will be repeated for the following subgroups:

- by region (Latin and South America, Eastern Europe, Western Europe, Asia and Australia, North America)
- by time of enrollment (patients enrolled pre-COVID-19 pandemic, patients enrolled during COVD-19 pandemic)

The supplementary estimand considers differences between COVID-19 pandemic related and other discontinuations assuming missings related to COVID-19 pandemic are MAR.

To assess the impact of COVID-19 on safety, and to assess possible regional differences in reporting of AEs, the 10 most frequent AEs (in total) will be summarized by preferred term and region.

3 Sample size calculation

A total of 956 patients will be randomized to one of the 6 groups including 5 QBW251 dose groups (450, 300, 150, 75, 25 mg b.i.d.) and placebo. This includes the ~102 patients

randomized to the 450 mg b.i.d. arm prior to its discontinuation. Prior to the discontinuation of 450mg b.i.d., patients were randomized to QBW251 450, 300, 150, 75, 25 mg b.i.d., or placebo with an allocation ratio of 2:2:1:1:1:2. After discontinuation of 450mg b.i.d., patients were randomized to QBW251 300, 150, 75, 25 mg b.i.d., or placebo with an allocation ratio of 2:1:1:1:2. This sample size of 854 (956 total minus approximately 102 patients already randomized to 450 mg b.i.d.) achieves at least 80% power for dose-response signal detection on key endpoints for all models from the candidate set based on the MCP-Mod methodology and also to ensure an acceptable precision for the DR curve estimation. The calculations in Table 3-1 exclude 450 mg b.i.d. from the analysis and are based on a 20% dropout rate, a one-sided significance level of 5%, and the following standard deviation assumptions.

- Pre-dose trough FEV_1: A common standard deviation of 200 mL based on previous Novartis COPD studies.
- E-RS Cough & Sputum weekly mean score: A common standard deviation of 1.3 based on the FLAME study (Wedzicha et al 2016).

Table 3-1 Power and Precision

Variable	Maximum improvement over placebo	Power to detect a DR signal	Average half-length of 90% confidence interval	
			Estimated DR curve	Estimated placebo-adjusted DR curve
Trough FEV_1	60 mL	92%	21 mL	26 mL
	50 mL	81%		
E-RS Cough & Sputum	0.35	86%	0.14	0.17
	0.30	76%		

Power calculations were performed using ADDPLAN-DF-4.0 and the precision estimates (half-length of confidence interval) were obtained from simulations.

The randomization ratio was generally based on D-optimality criteria, which maximize the information on the dose-response curve, for the MCP-Mod analysis. The D-optimal allocation ratio randomizes approximately twice as many patients to placebo and the highest dose, compared to the middle doses.

4 Change to protocol specified analyses

Analysis of the primary endpoint:

Due to termination of the 450 mg b.i.d. arm, candidate curves Emax(ED50=300), SigEmax(ED50=300) and Exponential(delta=300) may no longer be appropriate for MCP-Mod analyses that do not include the 450 mg b.i.d. arm. These candidate curves will be removed and the Exponential(delta=75) will be added.

5 Appendix

5.1 Imputation rules

5.1.1 Study drug

Missing/partial start date or end date of double-blind treatment will not be imputed.

5.1.2 AE date imputation

Partial AE start and end dates will be imputed. If there is uncertainty whether an AE occurred on-treatment or not, imputation will be performed, such that AE will be considered as on-treatment. Rules for imputing AE end date or start date will be provided in Programming Dataset Specification (PDS) document in details.

5.1.3 Concomitant medication date imputation

Rules for imputing the CM end date or start date will be provided in Programming Dataset Specification (PDS) document in details.

5.2 Visit windows

In general, if two consecutive visits V_t and V_s are x days apart, the upper limit of the visit window for V_t will be V_t+x/2 and the lower limit for the visit V_s will be V_s-x/2 (if x is even, the lower limit for V_s will be V_s-x/2+1, and the upper limit for V_t will be V_t+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.

For ECG and spirometry assessments, unplanned visits will be ignored for the windowing as scheduled timing is not available. If there are multiple assessments within a visit window with different scheduled time points, the one with the planned scheduled time points of the target day will be used; if scheduled time points are the same, the visit closer to the target day will be used.

For all other assessments, unplanned visits will also be considered for assigning visit windows. If there are multiple assessments within a visit window the closest to the target day will be used.

Only on-treatment values will be considered for mapping of visit windows. For trough FEV_1 visit windowing will be performed twice, once for primary analysis considering on-treatment measurements, and once for supportive analysis considering all post-baseline measurements whether being on-treatment or not.

The below table presents the visit windows for the respective assessments.

Analysis visit	Scheduled Day	Visit window – assessments a)	Visit window – assessments b)	Visit window – assessments c)	Visit window – assessments d)	Visit window – assessments e)
Day 1 post dose	Day 1	Day 1 (post-dose) - Day 7				

Week 2	Day 15	Day 8 - Day 22	Day 1 (post dose) - Day 22			
Week 4	Day 29	Day 23 - Day 43	Day 23 - Day 43	Day 2 - Day 43		
Week 8	Day 57	Day 44 - Day 71	Day 44 - Day 71	Day 44 - Day 71		
Week 12	Day 85	Day 72 - Day 99	Day 72 - Day 99	Day 72 - Day 99	Day 2 - Day 127	
Week 16	Day 113	Day 100 - Day 127	Day 100 - Day 127	Day 100 - Day 127		
Week 20	Day 141	Day 128 - Day 155	Day 128 - Day 155	Day 128 - Day 155		
Week 24	Day 169	Day 156 - Day 183	Day 156 - Day 183	Day 156 - Day 183	>Day 127	> Day 1
Follow-up	Day 197	>= Day 184	>= Day 184	>= Day 184		
a) ECG b) Vital signs c) Spirometry, Hematology, Chemistry, Urinalysis d) SGRQ, PGI-S, CASA-Q e) Weight, smoking status, [redacted]						

5.3 AEs coding/grading

The MedDRA version which will be available at the time of database lock, will be used for the coding purpose of the adverse events.

5.4 Laboratory parameters derivations

The following table shows the direction of interest when analyzing worst case values in form of maximum and/or minimum post-baseline values. If the direction of interest is given as "High" the maximum value will be calculated and used as worst value, if the direction is given as "Low" the minimum value will be taken, and if it is given as "Low and high", both the minimum value and the maximum value will be calculated and presented in summary tables.

Table 5.4-1 Directions of interest for worst case value for laboratory parameters

Laboratory Parameter	Direction of interest for worst case value
A. Hematology	
Hemoglobin	Low
Hematocrit	Low
Erythrocytes	Low
WBC	Low and high
Basophils	High
Eosinophils	High
Lymphocytes	Low and high
Monocytes	High
Neutrophils	Low and high
Platelets	Low and high
B. Chemistry	
Albumin	Low
Alkaline Phosphatase	High
ALT/SGPT	High
AST/SGOT	High
Bilirubin Total	High
Blood Urea Nitrogen (BUN)	High
Creatinine	High
Gamma GT	High
Potassium	Low and high
Magnesium	Low and high
Calcium	Low and high
LDH	High
Phosphorus	Low and high
Sodium	Low and high
[redacted]	[redacted]
HbA1c	Low and high

The following table shows the criteria for clinically notable laboratory values. Not all parameters have notable criteria defined:

Table 5.4-2 Clinical notable criteria for selected laboratory tests

Laboratory parameter (unit)	Lower bound of clinically notable range	Upper bound of clinically notable range
Hematology		
Hematocrit (v/v)		
Male	0.37	
Female	0.32	
Hemoglobin (g/L)		
Male	115	
Female	95	
Platelets (x10E9/L)	75	700
WBC (x10E9/L)	2.8	16.0
Chemistry		
Albumin (g/L)	25	-
Alkaline Phosphatase (U/L)	-	3xULN
ALT/SGPT (U/L)	-	3xULN
AST/SGOT (U/L)	-	3xULN
Bilirubin Total (mcmol/L)	-	34.2
BUN (mmol/L)	-	9.99
Creatinine (mcmol/L)	-	176.8
Gamma GT (U/L)	-	3 x ULN
Potassium (mmol/L)	3	6
Magnesium (mmol/L)	0.51	1.07
Sodium (mmol/L)	125	160

v = volume, ULN = upper limit of normal

Table 5.4-3 Notable liver function test values

Criterion
ALT > 3 x the upper limit of normal range (ULN) ALT > 5 x ULN ALT > 8 x ULN ALT > 10 x ULN ALT > 20 x ULN
ALT or AST > 3 x ULN ALT or AST > 5 x ULN ALT or AST > 8 x ULN ALT or AST > 10 x ULN ALT or AST > 20 x ULN
Total bilirubin > 1 x ULN Total bilirubin > 1.5 x ULN Total bilirubin > 2 x ULN Total bilirubin > 3 x ULN
ALP > 1.5 x ULN ALP > 2 x ULN ALP > 3 x ULN ALP > 5 x ULN
ALT or AST > 3 x ULN and total bilirubin > 1.5 x ULN ALT or AST > 3 x ULN and total bilirubin > 2 x ULN ALT or AST > 5 x ULN and total bilirubin > 2 x ULN ALT or AST > 8 x ULN and total bilirubin > 2 x ULN ALT or AST > 10 x ULN and total bilirubin > 2 x ULN ALT or AST > 20 x ULN and total bilirubin > 2 x ULN
ALP > 3 x ULN and total bilirubin > 2 x ULN ALP > 5 x ULN and total bilirubin > 2 x ULN
ALT or AST > 3 x ULN and Total Bilirubin > 2 x ULN and ALP < 2 x ULN (Hy's law)

ALT = alanine aminotransferase, AST = aspartate aminotransferase, ALP = alkaline phosphatase

When a criterion contains multiple laboratory parameters, the criterion will only be considered to have been met when all conditions occur within a 3-day window. A case where all criteria are met at a post-baseline time point will be considered as newly occurring if the criteria are not met at baseline and will be considered as worsening if the criteria are met at baseline and at least one component is worsening from baseline irrespective of whether the other(s) are better.

5.5 Vital signs and ECG – definition of clinically notable values

The following two tables show the clinical notable criteria for vital signs and QTcF respectively.

Table 5.5-1 Clinical notable criteria for vital signs

Vital sign parameter (unit)	Lower bound of clinically notable range	Upper bound of clinically notable range
Notable value considering newly occurring or worsening cases		
Systolic blood pressure (mmHg)	< 75	> 200
Diastolic blood pressure (mmHg)	< 40	> 115
Pulse rate (bpm)	< 40	> 130
Notable change from baseline		
Systolic blood pressure (mmHg)	≤ 90 and decrease from baseline by ≥ 20	≥ 180 and increase from baseline by ≥ 20
Diastolic blood pressure (mmHg)	≤ 50 and decrease from baseline by ≥ 15	≥ 105 and increase from baseline by ≥ 15
Pulse rate (bpm)	≤ 50 and decrease from baseline by ≥ 15	≥ 120 and increase from baseline by ≥ 15
Weight (kg)	Decrease > 7% from baseline	Increase > 7% from baseline

Table 5.5-2 Clinical notable criteria for QTcF (Fridericia's formula)

ECG parameter (unit)	Clinically notable range
Notable value considering newly occurring or worsening cases	
QTc (msec)	> 450
QTc (msec)	> 480
QTc (msec)	> 500
Notable change from baseline	
QTc (msec)	30 – 60
QTc (msec)	> 60
Combined criterion	
QTc (msec)	> 500 & increase > 60

5.6 Details on derivation of variables

5.6.1 Derivation of baseline disease characteristics

- Airflow limitation is classified based on % of predicted FEV_1 and FEV_1/FVC post bronchodilation (GOLD 2018):
  - Mild (GOLD 1): FEV_1/FVC < 70 % and % of predicted $FEV_1 \geq 80$ %
  - Moderate (GOLD 2): FEV_1/FVC < 70 % and 50% ≤ % of predicted $FEV_1 < 80$ %
  - Severe (GOLD 3): FEV_1/FVC < 70 % and 30% ≤ % of predicted $FEV_1 < 50$ %
  - Very Severe (GOLD 4): FEV_1/FVC < 70 % and % of predicted $FEV_1 < 30$ %

- Patients will be classified based on a combined assessment of COPD using the number of COPD exacerbations and COPD exacerbations leading to hospital admission in the last year, and CAT at Run-In 1 visit.

  The following algorithm will be applied based on the standard in GOLD 2018:
  - Group A:
    history of exacerbation ≤ 1 not leading to hospital admission and CAT < 10
  - Group B:
    history of exacerbation ≤ 1 not leading to hospital admission and CAT ≥ 10
  - Group C:
    (history of exacerbation ≥ 2 or ≥ 1 leading to hospital admission) and CAT < 10
  - Group D:
    (history of exacerbation ≥ 2 or ≥ 1 leading to hospital admission) and CAT ≥ 10
- Duration of COPD is calculated from the date of COPD first diagnosed as recorded on the eCRF until the date of Visit 1. If the date is missing in day and month, the missing date part was imputed as 01-July if the year is before Visit 1 and as 01-January if the year is the same as the year of Visit 1. If the date is missing in day only, the missing day was imputed as 15 if month/year is before Visit 1 and as 01 if month/year is the same as of Visit 1.
- Cardiovascular risk factors at baseline: Seven cardiovascular risk factors are defined.
  1. CCV history/condition = at least one out of: Myocardial infarction, Stroke, Peripheral arterial disease, Coronary artery bypass graft, or Percutaneous transluminal coronary angioplasty, as reported on the eCRF page of "Medical History – Protocol solicited events – Cardiovascular events".
  2. Hypertension, as reported on the eCRF page of "Medical History – Protocol solicited events – Cardiovascular events".
  3. Hyperlipidemia = at least one out of: Hyperlipidemia, Hypercholesterolemia, as reported on the eCRF page of "Medical History – Protocol solicited events – Cardiovascular events".
  4. Diabetes mellitus = Type 1 or Type 2 diabetes mellitus, as reported on the eCRF page of "Medical History – Protocol solicited events – Cardiovascular events".
  5. Obesity at baseline (i.e., BMI > 30 kg/m2).
  6. Age ≥ 65 years.
  7. Current smoker at screening.

5.6.2 E-RS total and subscale scores

The E-RS total symptom score will be derived as the sum of items 1 to 11 (items 12 – 14 will not be used for the E-RS), the RS-Breathlessness score as the sum of items 7 - 11, the RS-Cough & Sputum scores as the sum of items 2 - 4 and the RS-Chest Symptoms score as the sum of items 1, 5, 6.

The mean baseline E-RS Total and subscale scores will be the average of the corresponding daily scores from the run-in period (which is usually a 14 day period), the last 14 days from Run-In 1 visit up to a day before the first dose of double-blind treatment will be included. A patient needs to have at least 8 days of data from the run-in period in order for the mean baseline

score to be calculated. If there are less than 8 days of data available during the last 14 days from the run-in period, then use the last 7 days of data if there are at least 4 days of data available to calculate the mean.

For the derivation on-treatment mean weekly and mean monthly scores, only values from Day 1 up to the date of last dose of double-blind treatment will be considered. Missing diary data will not be imputed.

The daily scores will be averaged for each week (Week 1, Week 2, etc.). A patient needs to have at least 4 days of data in any week in order for the mean score to be calculated.

Monthly means over a 4-week period will also be calculated. A patient needs to have at least 3 weekly means in that month in order for the monthly mean score to be calculated.

5.6.3 CASA-Q domain scores

The CASA-Q is a validated questionnaire instrument used to measure cough and sputum production, and their impact in patients with chronic obstructive pulmonary disease (COPD) and/or chronic bronchitis (Crawford et al 2008). It contains a total of 20 items on a 5-step scale distributed in 4 domains: Cough symptoms (3 items, question 1-3), Cough impact (8 items, question 4-11), Sputum symptoms (3 items, question 12-14) and Sputum impact (6 items, question 15-20).

Each item is answered on a scale from ‘never’ to ‘always’ (for frequency) or from ‘not at all’ to ‘a lot/extremely’ (for intensity), each type using five categories (1-5) as applicable. All items are rescored from 1-5 to 0-4 and then reverse scored such that better responses have higher scores.

There are only domain scores and no overall score. The four domains are Cough symptoms, cough impact, sputum symptoms and sputum impact with a range of 0-100. Higher scores associated with fewer symptoms/less impact due to cough or sputum.

Within each domain, items are summed and rescaled using the following algorithm:

CASA-Q domain scores= (sum of rescored items) / (range of rescored items) ×100

A domain score will only be derived if data of all corresponding items are available.

Baseline CASA-Q domain score is defined as the last CASA-Q measurement before first dose of double-blind treatment.

5.6.4 SGRQ total and component scores

The algorithm below is based on SGRQ Manual Version 2.3, June 2009.

The St. George Respiratory Questionnaire (SGRQ) contains 50 items divided into three components: "Symptoms" concerned with respiratory symptoms, their frequency and severity; "Activity" concerned with activities that cause or are limited by breathlessness; and "Impacts" which covers a range of aspects concerned with social functioning and psychological disturbances resulting from airway disease. A score will be calculated for each component and a "Total" score will also be calculated. In each case the lowest possible value is zero and the highest 100. Higher values correspond to greater impairment of quality of life.

SGRQ scoring algorithm

Principle of calculation

Each response will be given a unique empirically derived weight between 0 and 100, the weights of all responses will be summed up and divided by the maximum possible score and expressed as a percentage.

Each component of the questionnaire is scored separately in three steps:

1. The weights for all items with positive responses are summed.
2. The weights for missed items are deducted from the maximum possible weight for each component. The weights for all missed items are deducted from the maximum possible weight for the Total score.
3. The score is calculated by dividing the summed weights by the adjusted maximum possible weight for that component and expressing the result as a percentage. The Total score is calculated in similar way.

Sum of maximum possible weights for each component and Total:

Symptoms	662.5
Activity	1209.1
Impacts	2117.8
Total	3989.4

(Note: these are the maximum possible weights that could be obtained for the worst possible state of the patient).

Symptoms component is calculated from the summed weights for the positive responses to questions 1-8 (Part 1).

Activity component is calculated from the summed weights for the positive responses to questions 11 – 17 (Part 2, Section 2) and 36 – 44 (Part 2, Section 6) on the questionnaire.

Impacts component is calculated from the summed weights for the positive responses to questions 9-10 (Part 2, Section 1), 18 – 35 (Part 2, Sections 3 - 5), and 45 – 50 (Part 2, Section 7) on the questionnaire.

Total score is calculated by summing all positive responses in the questionnaire and expressing the result as a percentage of the total weight for the questionnaire.

Handling of missing items

Symptoms: The Symptoms component tolerates a maximum of 2 missed items. The weight for the missed item is subtracted from the total possible weight for the Symptoms component (662.5) and from the Total weight (3989.4).

Activity: The Activity component tolerates a maximum of 4 missed items. The weight for the missed item is subtracted from the total possible weight for the Activity component (1209.1) and from the Total weight (3989.4).

Impacts: The Impacts component tolerates a maximum of 6 missed items. The weight for the missed item is subtracted from the total possible weight for the Impacts component (2117.8) and from the Total weight (3989.4).

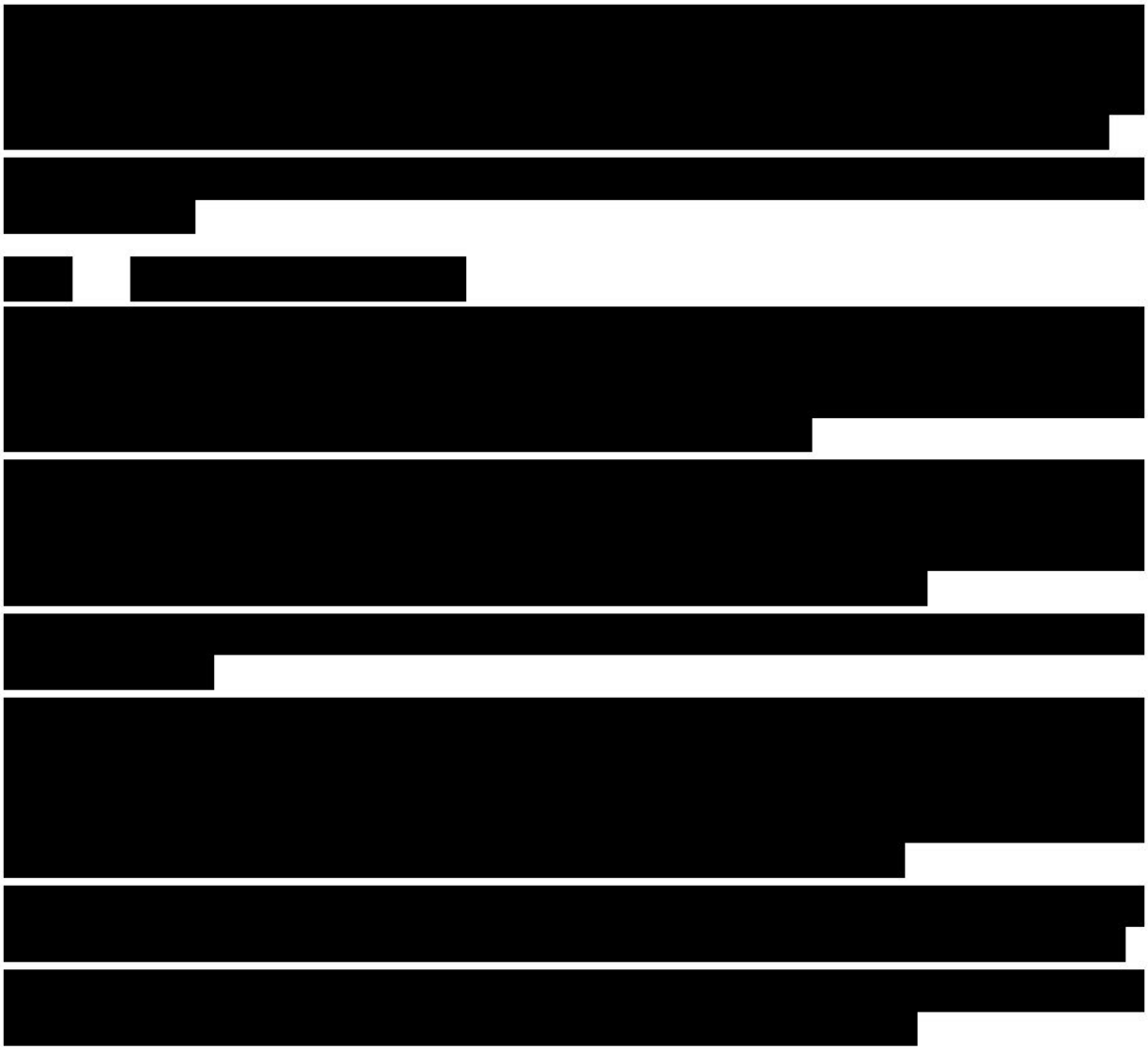
Total: The Total score requires that each of the three component scores has a tolerable number of missed items. The weight for the missed item is subtracted from the total possible weight (3989.4).

Baseline SGRQ total and domain scores are defined as the last measurements before first dose of double-blind treatment.

[redacted]

[redacted]

[redacted]



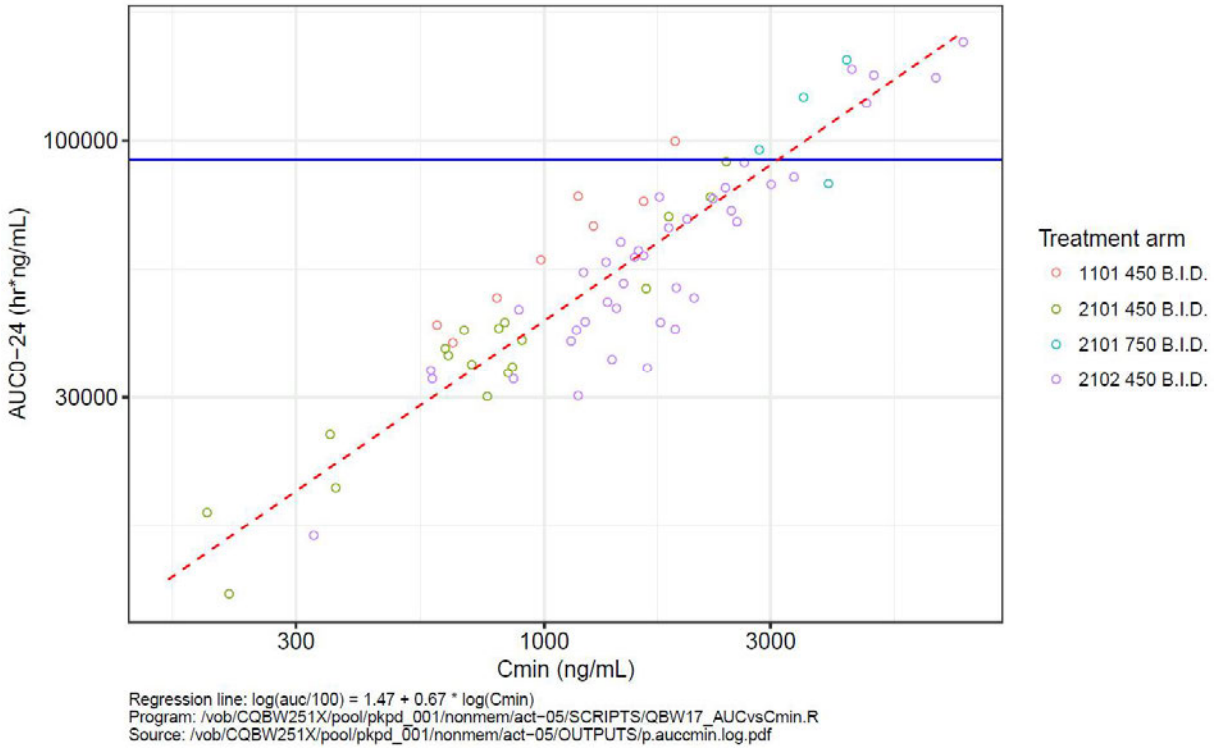
5.6.10 Pharmacokinetic parameter

The exposure threshold of $C_{min,ss}$=2942 ng/mL is defined based on the exposure threshold $AUC_{0\text{-}24h,ss}$= 91700 ng×h/mL using the following formula:

$\log(Cmin,ss) = (\log(AUC_{0\text{-}24h}, ss/100) - 1.47)/0.67$

The logarithm to the natural base is used.

The regression line was determined from subjects who received either 450 mg BID or 750 mg BID and for which AUC0-24h,ss could be estimated as two times AUC0-12h,ss. The data included PK from 8 subjects from study CQBW251B1101, 23 subjects from CQBW251X2101 and 39 subjects from study CQBW251X2102. For the CF patients in study CQBW251X2101, which had PK assessments up to 8h, AUC0-12h,ss was determined by imputing the steady state trough concentration at 12h to be equal to the steady state trough concentration at 0h.



AUC0-8h will be derived using a trapezoidal rule. If the 0h or the 8h concentration is missing the AUC0-8h will be set to missing. AUC0-12hr will be based on the observed concentrations from 0 to 8hr and extrapolation up to 12h (using trapezoidal rule). The daily drug exposure AUC0-24hr will then be estimated as 2 x AUC0-12hr.

5.7 Statistical models

5.7.1 Primary analysis

The following MMRM ANCOVA will be used for trough FEV_1:

Dependent variable = intercept + treatment + region + baseline FEV_1 + smoking status at screening + severity of airflow limitation + run-in post-bronchodilator FEV_1 + visit + treatment*visit + baseline value*visit + error.

The within-patient correlation will be modeled using an unstructured covariance matrix in the mixed model. The Kenward-Roger approximation will be used to estimate denominator degrees of freedom (Kenward and Roger, 1997).

If the model fails to converge with unstructured covariance matrix, either a compound symmetry (first choice) or first order autoregressive (AR1) (second choice) covariance structure will be used. For the efficacy IA, if the model still fails to converge only data up to week 12 will be included.

The SAS procedure PROC MIXED will be used for analysis. Results will be presented with LSM and standard error (SE) for treatment effects and LSM, SE, associated two-sided 90%

confidence interval, and two-sided p-value for treatment contrasts of each QBW251 dose versus Placebo.

- **MCP-Mod**

The Multiple Comparison Procedure – Modelling (MCP-Mod) methodology (see Bretz et al 2005 and Pinheiro et al 2014) will be employed to assess the primary objective. The following steps for the MCP-Mod methodology will be performed. The analysis will be implemented in SAS (used for MMRM) and R (MCP-Mod) using DoseFinding R package:

Step 1 (Testing an overall dose-response signal - MCP part):

a. Fit MMRM model. Extract estimates and variance-covariance matrix at Week 12 from fitted model.

b. Perform multiple contrast test with pre-defined model types and optimal contrasts, using variance-covariance matrix at Week 12 from the MMRM.

   Consider the following candidate models:

   - Emax (ED50=25)
   - Emax (ED50=75)
   - Emax (ED50=150)
   - sigEmax (ED50=25, h=3)
   - sigEmax (ED50=75, h=3)
   - sigEmax (ED50=150, h=3)
   - exponential (delta=75)
   - exponential (delta=150)
   - linear

   Present T-statistics and corresponding adjusted p-values for each candidate model.

   If there is at least one significant DR relationship, the DR signal is declared and Step 2 will be performed.

Step 2 (Estimation of the dose-response curve and target dose – Mod part):

a. Draw bootstrap samples (at least 1000) from a multivariate normal distributions N(μ, sigma) with parameter μ and sigma populated with the corresponding estimates and variance-covariance from the MMRM of step 1a. Fit each of the pre-defined models to the data of each sample. Select the model with the best fit based on generalized AIC and predict the dose-response curve with this model in each sample.

b. Derive median and other quantiles for predicted response over the dose range from the bootstrap samples. This will result in an averaged model.

c. Estimate the target dose(s): the smallest dose which shows a clinical relevant effect of delta for each of the bootstrap samples using the best model. The target dose will be

estimated for a delta of 50ml using inverse regression techniques (Bretz et al 2005). The averaged target dose is the median of the target doses over the bootstrap samples. The averaged target dose will only be derived if the estimated target dose of at least 50% of the bootstrap samples is within the investigated dose range.

- **Multiple imputation for supplementary analysis**

For supplemental analysis the following imputation steps need to be performed for missing data.

Off-treatment values will be considered for the analysis for patients discontinuing treatment not related to COVID-19, regardless of reason for discontinuing study.

a) The following table provides an overview of the imputations rules for data after discontinuation:

	Discontinuation of study	
Discontinuation of study treatment	COVID-19-related	Non-COVID-19 related
COVID-19-related	no off-treatment data, MAR	no off-treatment data, MAR during follow-up, J2R/MAR after study discontinuation
Non-COVID-19 related	off-treatment data, J2R/MAR	off-treatment data, J2R/MAR
J2R/MAR: Missing data after discontinuation of study treatment will be imputed using J2R assumption for the QBW251 arms and MAR assumption for placebo arm.		

b) Any other event leading to missing data prior to completion or discontinuation of study treatment: on-treatment value. Missing data will be multiply imputed based on the MAR assumption.

A patient can have missing data for both situations a and b.

1. For patients discontinuing double-blind treatment prematurely due to COVID-19 set off-treatment values to missing after COVID-related discontinuation.

2. Impute missing data using MAR:

   Select all patients, impute missing values at Week 4, 8, 12, 16, 20 and/or Week 24 using the MI approach based on the fully conditional specification (FCS) method for 100 time and obtain 100 imputed dataset. Missing Week 4 values will be imputed separately for treatment group using a model with the imputed values of baseline FEV1, region, smoking status, run-in pre-bronchodilator FEV_1 and severity of airflow as predictors. Missing Week 8 values will be imputed separately for each treatment group using a model with the imputed values of Week 4, baseline FEV1, region, smoking status, run-in pre-bronchodilator FEV_1 and severity of airflow as predictors. Repeat the same for subsequent visits.

   This results in 100 imputed datasets.

   For patients in the QBW251 treatment groups who discontinued double-blind treatment prematurely not related to COVID-19, set imputed values after discontinuation of treatment to missing, as these will be imputed using J2R in step 3.

For patients in the QBW251 treatment groups who discontinued double-blind treatment prematurely related to COVID-19 and discontinued study not related to COVID-19, set imputed values after discontinuation of study to missing, as these will be imputed using J2R in step 3.

3. Impute missing for patients in the QBW251 treatment groups using J2R:

   Select all patients, impute missing values at Week 4, 8, 12, 16, 20 and/or Week 24 using the MI approach, under assumption of missing not at random (MNAR) that patients in QBW251 treatment groups will behave as patients treated with Placebo, based on the fully conditional specification (FCS) method for 100 time and obtain 100 imputed dataset. Missing Week 4 values will be imputed using a model with the imputed values of baseline FEV1, region, smoking status, run-in pre-bronchodilator FEV_1 and severity of airflow as predictors. Missing Week 8 values will be imputed using a model with the imputed values of Week 4, baseline FEV1, region, smoking status, run-in pre-bronchodilator FEV_1 and severity of airflow as predictors. Repeat the same for subsequent visits.

   This results in 100 imputed datasets.

4. Merge datasets from step 2 and step 3.
   a. For patients in QBW251 treatment groups discontinuing double-blind treatment prematurely not related to COVID-19 take imputed values after discontinuation of treatment from step 3.
   b. For patients in the QBW251 treatment groups who discontinued double-blind treatment prematurely related to COVID-19 and discontinued study not related to COVID-19, take imputed values after discontinuation of study from step 3.
   c. For all other missing (including those for Placebo patients) values take imputed values from step 2.
5. Week 12 data of the final multiply-imputed dataset where all missing values are filled will be analyzed (by imputed dataset) using a linear mixed model with treatment, region, smoking status at screening, run-in pre-bronchodilator FEV_1, severity of airflow limitation and baseline FEV_1 as predictors.
6. The results for the treatment effect from the 100 datasets will then be combined using Rubin’s rule.

Step 1 to 4 will be implemented before MCP-Mod is used.

Step 5 and 6 will be used within MCP-Mod for step 1a Step 1b and 2a of the MCP-Mod will use the combined estimate and variance-covariance matrix of step 6..

5.7.2 Key secondary analysis

- **Repeated measurements logistic regression analysis**

The proportion of patients achieving a clinical important improvement, will be analyzed using a repeated measures logistic regression. The model is similar to the ANCOVA model:

Response = intercept + treatment + region + baseline + smoking status at screening + severity of airflow limitation + run-in post-bronchodilator FEV_1 + visit + treatment*visit + baseline value*visit + error.

The within-patient correlation will be modeled using an unstructured covariance matrix in the mixed model.

If the model fails to converge with unstructured covariance matrix, either a compound symmetry (first choice) or first order autoregressive (AR1) (second choice) covariance structure will be used. For the efficacy IA, if the model still fails to converge only data up to week 12 will be included.

The SAS procedure PROC GENMOD will be used for analysis. A binomial distribution and a logit link function will be used. Odds ratios will be presented with associated two-sided 90% confidence interval for treatment comparison of each QBW251 dose versus Placebo.

- **Proportional odds model for repeated measures**

PGI-S response change level will be analyzed using a proportional odds model for repeated measures. The model is similar to the ANCOVA model:

Response = intercept + treatment + region + baseline + smoking status at screening + severity of airflow limitation + run-in post-bronchodilator FEV_1 + visit + treatment*visit + baseline value*visit + error.

The within-patient correlation will be modeled using an unstructured covariance matrix in the mixed model.

If the model fails to converge with unstructured covariance matrix, either a compound symmetry (first choice) or first order autoregressive (AR1) (second choice) covariance structure will be used. For the efficacy IA, if the model still fails to converge only data up to week 12 will be included.

The SAS procedure PROC GLIMMIX will be used for analysis. A multinomial distribution and a cumulative logit link function will be used. Odds ratios will be presented with associated two-sided 90% confidence interval for treatment comparison of each QBW251 dose versus Placebo.

[redacted]

[redacted]

- **Cox regression analysis**

A Cox proportional hazards regression model will be applied in time-to-event analyses to test the null-hypothesis H0: λQBW251(t) / λPlacebo(t) = 1, where λ(t) is the hazard function for the failure time of patients treated with QBW251 dose and Placebo, respectively. The Cox regression model will be stratified by smoking status and severity of airflow limitation and will include terms for treatment, region, baseline E-RS total score, run-in post-bronchodilator FEV_1 and number of COPD exacerbations in the past year.

The SAS procedure PROC PHREG will be used for analysis. Results will be presented with adjusted hazard ratios for treatment group comparisons and associated 90% confidence intervals.

No check for the validity of proportional hazards assumptions will be done.

[redacted]

- **Logistic regression analysis**

The proportions of patients who achieve a clinically significant improvement will be analyzed using logistic regression. The model will include the same terms as used for the linear model ANCOVA.

The SAS procedure PROC GENMOD will be used. Odds ratios will be presented with associated two-sided 90% confidence interval for treatment comparison of each QBW251 dose versus Placebo.

5.8 Rule of exclusion criteria of analysis sets

The following table provides the protocol deviations (PD) and other criteria leading to partial or complete exclusion from analyses sets.

Deviation ID	Description of Deviation
Deviations leading to exclusion from all analyses sets	
INCL01A	Subject entered the study without signing the general patient informed consent.
Deviations leading to exclusion from RAS, FAS	
TRT06	The patient received investigational drug but was not randomized.
Deviations leading to exclusion from FAS, Safety and PK analysis set	
TRT07	Randomized but no study drug given
OTH03	Patient was randomized in error, but did not enter the treatment period
[redacted]	
[redacted]	[redacted]
Deviations leading to exclusion from PK serial sub-group	
INCL01E	Subject entered the additional pharmacokinetic data collection subgroup assessments without signing the additional PK patient informed consent.

6 Reference

Bornkamp, Bezlyak, Bretz (2015) Implementing the MCP-Mod procedure for dose-response testing and estimation, Modern Approaches to Clinical Trials Using SAS: Classical, Adaptive and Bayesian Methods, Chapter 7

Bretz F, Pinheiro JC, Branson M (2005) Combining multiple comparisons and modeling techniques in dose-response studies. Biometrics; 61(3):738-48.

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

Crawford B, Monz B, Hohlfeld J, et al (2008) Development and validation of a cough and sputum assessment questionnaire. Respir Med; 102(11):1545-55.

European Medicines Agency (EMA) (2018) Qualification Opinion on Proactive in COPD. https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/qualification-opinion-proactive-chronic-obstructive-pulmonary-disease-copd_en.pdf. Accessed 25-Jun-2019.

[redacted]

Food and Drug Administration (FDA) (1999) Guidance for industry: population pharmacokinetics. http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM072137.pdf. Accessed 11-Mar-2019.

Food and Drug Administration (FDA) (2012) FDA Center for Drug Evaluation and Research, Clinical Pharmacology and Biopharmaceutics Review Application Number: 203188Orig1s00. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2012/203188s000TOC.cfm. Accessed 11-Mar-2019.

Global Initiative for Chronic Obstructive Lung Disease (GOLD) Guidelines (2018) Global Strategy for the Diagnosis, Management and Prevention of Chronic Obstructive Pulmonary Disease - Global Initiative for Chronic Obstructive Lung Disease, p. 1-123. https://goldcopd.org/archived-reports/. Accessed 11-Mar-2019.

[redacted]

Kenward MG and Roger JH (1997). Small sample inference for fixed effects from restricted maximum likelihood. Biometrics 53(3):983-97

Lan, KKG and DeMets, DL (1983) Discrete sequential boundaries for clinical trials. Biometrika; 70(3):659–663.

[redacted]

Pinheiro J, Bornkamp B, Glimm E, et al (2014) Model-based dose finding under model uncertainty using general parametric models. Stat Med; 33(10):1646-61.

[redacted]

St George's Respiratory Questionnaire Manual (2009) http://www.healthstatus.sgul.ac.uk/SGRQ_download/SGRQ%20Manual%20June%202009.pdf/ Assessed on 11-Jul-2019.

Wedzicha JA, Banerji D, Chapman KR, et al (2016) Indacaterol-Glycopyrronium versus Salmeterol-Fluticasone for COPD. N Engl J Med; 374(23):2222-34.