NOVARTIS

Clinical Development

QVA149 (indacaterol maleate/glycopyrronium bromide)

CQVA149A3405 / NCT02516592

A 12-week treatment, multi-center, randomized, doubleblind, double-dummy, parallel group study to assess the efficacy and safety of switching from salmeterol/fluticasone to QVA149 (indacaterol maleate/glycopyrronium bromide) in symptomatic COPD patients

RAP Module 3 – Detailed Statistical Methodology

Author: [redacted]

Document type: RAP Documentation

Document status: Amendment 2

Release date: 13-Jun-2017

Number of pages: 26

Document History – Changes compared to previous version of RAP module 3.

Version	Date	Changes
Amendment 1	3-Feb- 2017	Combined assessment of COPD (according to GOLD 2014 and 2017) included in baseline characteristics and as subgroup. Added FEV1 reversibility as categorized variable. Additional subgroup analysis added for the primary endpoint. Included ECG analysis as described in protocol. Clinical notable criteria for QTcF were included Compliance calculation was adapted since number of capsules/doses taken is not available on eCRF. Compliance >100% can not be detected. Further categories for COPD related medications were included. Baseline definition for CAT score was changed. Added more detail for derivation of rescue medication use in Appendix 16.1.9.4.
Amendment 2	13-Jun-2017	Correction: Removed minor protocol deviations (those not excluding the patient from any analysis) from Section 16.1.9.2. Updated Section 16.1.9.3: Added the exclusion for FAS, PP and SAF for patients which are not treated, as per definition of the analysis sets.

Table of contents

1 Introduction

This document contains details of the statistical methods which will be used in the phase IIIb/IV clinical trial CQVA149A3405. The purpose of this study is to investigate whether switching symptomatic COPD patients from a fixed-dose combination of salmeterol/fluticasone 50/500 µg b.i.d. to a fixed dose combination of QVA149 110/50 µg o.d. leads to improved lung function and airflow, and to assess the effect on symptom burden, breathlessness, and use of rescue medication.

Data will be analyzed according to the data analysis section 9 of the study protocol. Important information is given in the following sections and details are provided, as applicable, in Appendix 16.1.9 of the CSR.

1.1 Changes to RAP from protocol specified analysis:

The following changes from the planned analyses according to the Clinical Trial Protocol (release date 17-Jun-2015) will be performed.

1) **Compliance**

Compliance is not calculated as percentage of days but as percentage of doses taken, since only the number of doses are recorded on the eCRF.

2) **Logistic regression model**

According to the protocol clinically significant improvement in TDI after 12 weeks of treatment should be analyzed using a logistic regression. A logistic regression model for repeated measures will be performed in accordance with the analysis of the other efficacy variables.

2 Detailed statistical methodology

2.1 Study objectives

The study protocol section 2 lists the following primary and secondary objectives.

2.1.1 Primary objective

To demonstrate superiority of QVA149 110/50 µg o.d. compared to salmeterol/fluticasone 50/500 µg b.i.d. in trough pre-dose FEV_1 at week 12.

2.1.2 Secondary objectives

To evaluate the effect of QVA149 110/50 µg o.d. compared to salmeterol/fluticasone 50/500 µg b.i.d. on:

- Total score of TDI at week 12
- FVC at week 12
- COPD symptoms at week 12 as measured by the COPD Assessment Test (CAT)
- The mean number of puffs per day of rescue medication use, and percentage of days

without rescue medication use over the 12 week treatment period

- Safety and tolerability during the 12 week treatment period

2.2 Analysis sets

- The randomized set (RAN) will include all randomized patients, regardless of whether or not they actually received randomized study medication. Patients in the RAN will be analyzed according to the treatment they were randomized to.
- The full analysis set (FAS) will include all randomized patients who received at least one dose of randomized study medication. Following the intent-to-treat principle, patients will be analyzed according to the treatment they were assigned to at randomization.
- The per-protocol set (PPS) will include all patients in the FAS without any major protocol deviations. Major protocol deviations will be defined prior to database lock and the unblinding of the study. The list of major protocol deviations is available in Appendix 16.1.9. The PPS will be used for supportive analysis to assess robustness of the primary analysis. Patients will be analyzed according to the randomized treatment. Patients who receive another than their randomized treatment because of a dispensing error will be excluded from the PPS.
- The safety set will include all patients who received at least one dose of randomized study treatment, whether or not being randomized. Patients will be analyzed according to the treatment they received during the randomized treatment period. If a patient took both treatments the patient will be analyzed according to the randomized treatment. The safety set will be used for all safety endpoints.

2.3 Patient disposition, background and demographic characteristics

2.3.1 Patient disposition

The RAN will be used for all patient disposition outputs, unless stated otherwise below.

The number of patients randomized will be summarized, by region ("Asia" including Malaysia, Taiwan and Philippines, "Australia/South Africa", "India" and "Middle East" including Egypt, Lebanon, Saudi Arabia, Turkey and Israel), country and center. The overall number of patients who were screened, randomized, completed the study and discontinued from the study will be summarized with reasons for discontinuation.

Number of patients with protocol deviations will be tabulated by deviation. Patients with protocol deviations will be listed with date and study day of occurrence, deviation and severity codes.

The number of patients included in each analysis set will be tabulated. Reasons for exclusions from analysis sets will be tabulated for all patients. Patient exclusion from analysis sets will be listed for all patients with reasons for exclusion, including both protocol deviations and other exclusion criteria.

2.3.2 Demographics and baseline characteristics

Demographics and baseline characteristics will be summarized using the FAS. Note that some of the parameters covered in this section (e.g. baseline FEV_1 and FVC, BDI and rescue medication use, baseline ECG) will not appear in summaries of demographics and baseline characteristics, but will be included in the applicable efficacy or safety summaries.

Age, sex, race, ethnicity, height, weight, body mass index (BMI), severity of air flow limitation (GOLD stages), severity of combined assessment of COPD (GOLD 2014 and 2017), duration of COPD, smoking status (ex-/current smoker) at screening, estimated number of pack years, COPD exacerbation history (number of COPD exacerbations that occurred during the previous year prior to screening), other relevant medical history, concomitant medications (COPD related and non-COPD related) discontinued prior to start of randomized study drug, spirometry at screening (predicted FEV_1, FEV_1 pre and post use of bronchodilator, and % of predicted FEV_1 prior to and after inhalation of ipratropium bromide and salbutamol or albuterol, FEV_1 reversibility, FVC pre and post use of bronchodilator, % of FEV_1/FVC prior to and after inhalation of ipratropium bromide and salbutamol or albuterol), baseline FEV1, baseline FVC, CAT score at screening, BDI score and baseline rescue medication will be summarized.

In addition, the following categorizations will be done:

- Age into <65 years, 65 - < 75 years, and ≥ 75 years;
- BMI into ≤ 30.0 kg/m^2 and > 30.0 kg/m^2;
- Duration of COPD into < 1 year, 1 - 5 years, > 5 - 10 years, > 10 - 15 years, > 15 - 20 years, and > 20 years;
- Post-bronchodilator FEV_1 into < 30 % of predicted FEV_1, 30 - < 50 % of predicted FEV_1, 50 - < 80 % of predictive FEV_1, and ≥ 80 % of predicted FEV_1;
- Post-bronchodilator FEV_1/FVC ratio into < 0.70, and ≥ 0.70;
- FEV_1 reversibility (reversible [≥12% and ≥200ml increase]/not reversible)
- CAT scores into 0 - 10, 11 - 20, 21 - 30 and 31 - 40 representing mild, moderate, severe or very severe clinical impact of COPD on patients.

Baseline ECG interpretation and vital signs results (radial pulse rate, systolic and diastolic blood pressure) will be summarized for the safety set. Clinical notable vital signs are defined in Appendix 16.1.9.

Demographic data, inclusion/exclusion criteria, AEs and SAEs of screening failures will be listed only.

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

Details regarding the definition of baseline measurements are given in Appendix 16.1.9.

2.3.3 Medical history

Medical history will be coded using the Medical Dictionary for Regulatory Activities terminology (MedDRA). History/conditions will be summarized for the FAS by primary system organ class and preferred term, and overall. Verbatim recorded history/conditions will be listed together with the coded terms, date of diagnosis/surgery and whether the problem was ongoing at start of the study.

2.4 Study medication

2.4.1 Study drug administration

The extent of exposure will be examined to determine the degree to which safety can be assessed for QVA149. The extent of exposure to study drug will be characterized according to the duration of exposure and the number of subjects exposed.

Duration of exposure to a treatment will be calculated as the number of days between the 1st 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 study drug – Date of first dose of double-blind study drug + 1).

The duration of exposure (in days) will be summarized, by standard descriptive statistics, by treatment for the safety set.

2.4.2 Compliance

Study drug compliance will be assessed by the investigator and/or center personnel at designated visits according to the procedures defined in the study protocol. The overall compliance will be calculated as the percentage of number of doses expected-number of doses missed/number of doses expected between Day 1 and the date of last dose..

Compliance will be calculated separately for the SDDPI and the Accuhaler/Diskus and will be summarized by treatment for the safety set. In addition the following categories will be summarized: <80%, 80% - 100%. Compliance >100% can not be detected.

2.5 Concomitant medication

Each medication will have the start and end dates recorded. Concomitant medications are defined as those medications that were taken on or after the first dose of study drug but not prior to the first dose of study drug.

Concomitant medications will be summarized by treatment for the safety set. Concomitant COPD related medications will be summarized by pre-specified categories, route of administration and preferred term. Concomitant medications not related to COPD will be summarized by preferred term.

The following pre-specified categories will be presented:

- Leukotriene modifier
- Short acting B2 agonist

- Long acting anti-cholinergic
- Short acting anti-cholinergic
- Antibiotics
- Long-acting B2 agonist
- Xanthine
- Anti-Histamines
- Vaccine
- Corticosteroid
- Mast cell stabilizers
- PDE-4 inhibitor
- Anti-IGE
- Other
- LAMA/LABA
- ICS/LAMA
- ICS/LABA
- SABA/SAMA

2.6 Efficacy evaluation

The primary objective of this study is to investigate whether switching symptomatic COPD patients from a fixed-dose combination of salmeterol/fluticasone 50/500 µg b.i.d. to a fixed dose combination of QVA149 110/50 µg o.d. leads to improved lung function and airflow, and to assess the effect on symptom burden, breathlessness, and use of rescue medication.

2.6.1 Analysis of the primary efficacy variable

2.6.1.1 Variable

Trough pre-dose FEV_1 at Week 12 is defined as the average of the measurements taken -45min and -15min pre study medication dose in the clinic after 12 weeks of treatment (Day 84). The baseline measurement is defined in Appendix 16.1.9.

Trough pre-dose corresponds to 24 hours after the last randomized study drug dose (i.e. last dose prior to the FEV_1 measurement) and 12 hours after the last dose of salmeterol/fluticasone. Spirometry is performed in the morning prior to the administration of the morning dose of study medication.

2.6.1.2 Statistical model, hypothesis, and method of analysis

The following null hypothesis (Ho) versus the alternative hypothesis (Ha) will be tested at a significance level of 0.05 in the FAS:

- Ho: There is no difference in trough pre-dose FEV_1 after 12 weeks of treatment between QVA149 100/50 µg o.d. and salmeterol/fluticasone 50/500 µg b.i.d.
- Ha: There is a significant difference in trough pre-dose FEV_1 after 12 weeks of treatment between QVA149 100/50 µg o.d. and salmeterol/fluticasone 50/500 µg b.i.d.

The treatment difference between QVA149 110/50 µg o.d. and salmeterol/fluticasone 50/500 µg b.i.d. in terms of trough pre-dose FEV_1 after 12 weeks of treatment will be estimated using ANCOVA for repeated measures. The model will include terms of treatment, baseline FEV_1, smoking status at baseline, region, visit, treatment*visit interaction, baseline FEV_1*visit interaction and an unstructured variance-covariance structure assuming normal distribution of the FEV_1 data, where visit will be treated as a categorical variable. Center nested within region may be added as random effect depending on model convergence. Missing values of trough pre-dose FEV_1 at Day 42 and Day 84 will not be imputed. The SAS code for the ANCOVA model is displayed in Appendix 16.1.9.

The estimated treatment difference of QVA149 110/50 µg o.d. – salmeterol/fluticasone 50/500 µg b.i.d. will be displayed along with the associated 95% confidence interval and p-value. Superiority of QVA149 110/50 µg o.d. to salmeterol/ fluticasone 50/500 µg b.i.d. will be demonstrated if the p-value is less than 0.05 and the 95% confidence interval lies entirely to the right of 0 L.

Trough pre-dose FEV1 will be summarized descriptively.

Subgroup analyses will be done by gender, age (<65, 65-<75, >=75 years), race (Caucasian, Asian and Other), region (Asia, Australia/South Africa, India, Middle East), BMI (<=30, >30), severity of airflow limitation, combined assessment of COPD (2014 and 2017), FEV1 reversibility at screening (not reversible, reversible), smoking status at screening (ex-smoker, current smoker) and exacerbation history (0 exacerbations/1 exacerbation).

The same analysis models as those without subgroup analysis will be used. If the subgroup variable is not already included in the covariates of the model, then the terms for subgroup and treatment by subgroup interaction will be added to the model; otherwise only the treatment by subgroup interaction will be added. There are exceptions: In the subgroup analysis for race, the fixed effect of region will be removed from the model and as random effect center will be included only.

2.6.1.3 Handling of missing values/censoring/discontinuations

Only spirometry measurements that are deemed acceptable (see Clinical Trial Protocol Appendix 1: Spirometry Guidance for more details) will be included in the analyses. If any of the values contributing to trough pre-dose FEV_1 or baseline FEV_1 are collected within 6 hours of rescue medication or less than 7 days after systemic corticosteroid use then the individual FEV_1 value will be set to missing. Also if a value is taken post-dose instead of pre-dose it will not be used for calculation of trough pre-dose or baseline values. Since dosing times are only

recorded in the diary for the days prior to scheduled visits no further time window will be applied for deriving pre-dose values. For the primary analysis, values will not be carried forward since ANCOVA for repeated measures will be applied. If all values contributing to the baseline FEV_1 are missing (or are not confirmed to be pre-dose), then the pre-bronchodilator measurements taken at Screening will be used as the baseline.

2.6.1.4 Supportive analyses

The following supportive analyses may be conducted for trough pre-dose FEV_1 at Week 12:

(1) Imputation of missing Week 12 data by Last Observation Carried Forward (LOCF), i.e., using the value from Day 42. A linear mixed model including the terms of treatment, baseline FEV_1, smoking status at baseline, and region as fixed effects and center nested within region as random effect will be applied.

(2) Analysis based on pattern mixture model: An additional method for handling missing data will be implemented to explore the sensitivity to the Missing At Random (MAR) assumption. A pattern-mixture model based on nonignorable missing, i.e., Missing Not At Random (MNAR) will be used, where all drop-outs are assumed to switch back to salmeterol/fluticasone 50/500 μg b.i.d. treatment. This approach will be implemented by multiple imputations, where missing values are repeatedly replaced by sampling from the Bayesian posterior predictive distribution based on the switching model. Each of the completed datasets is then analyzed using the primary analysis, and then the estimates are combined using Rubin's rules (Biometrika, 1976) to get the final inference.

(3) Applying the ANCOVA for repeated measurements using the PPS.

2.6.2 Analysis of secondary variables

The FAS will be used for the analysis of all secondary efficacy variables.

2.6.2.1 Dyspnea

TDI

Dyspnea is measured at baseline using the baseline dyspnea index (BDI), at week 6 and at the end of the treatment period (Week 12) using the transition dyspnea index (TDI), which captures changes from baseline.

The BDI and TDI each have three domains: functional impairment, magnitude of task and magnitude of effort. The BDI domains are rated from 0 (severe) to 4 (unimpaired) and the rates are summed for the baseline focal score ranging from 0 to 12; the lower the score the worse the severity of dyspnea. The TDI domains are rated from -3 (major deterioration) to 3 (major improvement) and the rates are summed for the transition focal score ranging from -9 to 9; minus scores indicate deterioration. A TDI focal score of 1 is considered to be a clinically significant improvement from baseline.

The TDI focal score at Week 6 and Week 12 will be analyzed using ANCOVA for repeated measures. The model will be similar to the model of the primary analysis with the baseline BDI instead of baseline FEV_1 included in the model.

Additionally the percentage of patients with a clinically significant improvement after 12 weeks of treatment in TDI (of >=1 change from baseline) will be analyzed using a logistic regression model for repeated measures which includes the same terms as used for the analysis of the TDI focal score.

BDI and TDI will be summarized descriptively.

2.6.2.2 Spirometry

Trough pre-dose FVC

The same analyses as for trough pre-dose FEV1 will be done for trough pre-dose FVC with the baseline FVC instead of baseline FEV1 included in the model.

FEV1 at individual post-baseline time points

A similar method as for the primary endpoint will be applied for the analysis of FEV1 at Week 6 (Visit 3) and Week 12 (EOS/Visit 4), i.e., separate ANCOVA for repeated measures will be performed for each time point using visit as repeated variable. FEV1 will be summarized descriptively by visit and time point.

FVC at individual post-baseline time points

The same analyses as for FEV1 at Week 6 (Visit 3) and Week 12 (EOS/Visit 4) will be done for FVC.

2.6.2.3 Rescue medication

Use of rescue medication (number of puffs taken in the previous 12 hours) is recorded (once in the morning and once in the evening) by the patient, in the patient diary. The total number of puffs of rescue medication since last study visit and the total number of days without rescue use since last visit will be reported on the eCRF.

Daily use of rescue medication (number of puffs)

The mean daily number of puffs will be summarized descriptively and analyzed using a linear mixed model. The model will contain treatment, baseline mean daily number of puffs, smoking status at baseline, and region as fixed effects with center nested within region as a random effect.

Percentage of days without rescue medication usage over 12 weeks

The percentage of days without rescue medication usage will be calculated for each patient over the 12 week treatment period. This endpoint will be summarized descriptively and analyzed using a similar model as for rescue medication use, but with baseline percentage of days without rescue medication usage instead of baseline mean number of puffs included in the model.

Rescue medication data prior to the first intake of study drug will be used to calculate the baseline. Details on how to calculate baseline and post-baseline measurements are given in Appendix 16.1.9.

2.6.2.4 COPD assessment test (CAT)

The CAT is a short instrument used to quantify the symptom burden of COPD and is used to assess the health status of patients in this study. It is completed by the patient at the beginning of the study visit before any other assessment to avoid influencing the responses. The CAT is completed by the patient at the investigator's site at Visits 1, 2, 3 and 4 or at the time of discontinuation for patients who prematurely withdraw from the study. It consists of eight items, each presented as a semantic 6-point differential scale, providing a total score out of 40. A higher score indicates a worse health status. The result is immediately available without the need for any calculation, apart from summing the scores on individual items. Scores of 0 - 10, 11 - 20, 21 - 30 and 31 – 40 represent a mild, moderate, severe or very severe clinical impact of COPD upon the patient. If one or two items are missing, they will be replaced with the mean of the completed items. If three or more items are missing, the CAT total score will be missing.

A similar method as for the primary endpoint will be applied for the analysis of CAT with the baseline CAT result instead of baseline FEV1 included in the model. The CAT score will be summarized descriptively, in addition the proportions of patients with different levels of clinical impact, i.e., total CAT scores of 0-10 (mild), 11-20 (moderate), 21-30 (severe) and 31-40 (very severe), will be summarized descriptively by treatment for each visit.

2.7 Safety evaluation

The safety set will be used in the analysis of all safety variables.

2.7.1 Adverse Events

All adverse events including COPD exacerbations, coded using MedDRA, will be listed. Treatment emergent adverse events are defined as those events starting on or after the time of the first administration of randomized study drug but not later than 7 days (30 days in the case of serious adverse events) after the last administration of study medication.

Summaries of treatment emergent adverse events will be produced, overall by system organ class and preferred term for the following types of adverse events:

(1) All adverse events

The number and percentage of patients who reported adverse events will be summarized by primary system organ class, preferred term, and treatment. Unless otherwise specified, primary system organ classes will be sorted alphabetically and, within each primary system organ class, the preferred terms will be sorted in descending order of frequency in the investigational study drug arm. If a patient reported more than one adverse event with the same preferred term, the adverse event will be counted only once. If a patient reported more than one adverse event within the same primary system organ class, the patient will be counted only once at the system organ class level.

(2) Adverse events by maximum severity

All adverse events will be summarized by maximum severity, primary system organ class, preferred term, and treatment. If a patient reported more than one adverse event within the same primary system organ class, only one adverse event will be counted for that patient at

the highest severity level in the total row for each primary system organ class. Missing severity will be assumed to be severe in the summary table.

(3) Adverse events suspected to be study drug related

The adverse events suspected to be related to study drug (according to the investigators) will be summarized by primary system organ class, preferred term, and treatment.

(4) Adverse events leading to permanent discontinuation of study treatment

Adverse events leading to permanent study drug discontinuation, regardless of study drug relationship, will be summarized by primary system organ class, preferred term, and treatment.

(5) Adverse events leading to dose adjustment / temporary interruption of study treatment

Adverse events requiring dose adjustment / temporary interruption, regardless of study drug relationship, will be summarized by primary system organ class, preferred term, and treatment.

(6) Adverse events requiring additional therapy

Adverse events requiring additional therapy (i.e., concomitant medication or non-drug therapy), regardless of study drug relationship, will be summarized by primary system organ class, preferred term and treatment.

(7) Serious adverse events (SAEs)

Number and percentage of patients with treatment-emergent serious adverse events, regardless of study drug relationship, will be presented by primary system organ class, preferred term and treatment.

Number and percentage of patients with onset of serious adverse events between visit 1 and first dose of randomized drug will be presented by primary system organ class, preferred term and treatment.

(8) Fatal adverse events

Fatal adverse events, regardless of study drug relationship, will be presented by primary system organ class, preferred term and treatment.

(9) Investigator-reported cause of death

Number and percentages of deaths will be presented by primary system organ class, preferred term, and treatment group, based on investigator-reported principal cause of death, regardless of study drug relationship.

2.7.2 COPD exacerbations during the 12 week treatment period

COPD exacerbations are defined in the study protocol section 6.6.8 as worsening of two or more major symptoms (i.e., dyspnea, sputum volume, sputum purulence), or worsening of any 1 major symptom together with any 1 of minor symptoms (i.e., sore throat, colds, fever without other cause, cough, wheeze) for at least 2 for consecutive days. The start date for a COPD exacerbation recorded in the eCRF should be the first day of symptom worsening of two or more major symptoms or of one major and one minor symptom, as defined above. The

end of a COPD exacerbation episode is marked by the return to pre-exacerbation symptom status.

A COPD exacerbation is considered of moderate severity if treatment with systemic corticosteroids or antibiotics or both was required and severe, if hospitalization was required. An emergency room (ER) visit of longer than 24 hours will be considered a hospitalization.

A worsening of symptoms that either do not meet the above symptom definitions but is treated by the investigator with systemic corticosteroids or antibiotics, or that meets the symptom definition but does not receive antibiotics and/or systemic corticosteroids, is captured as mild exacerbation on the eCRF.

In patients with multiple exacerbations, if the start date of an exacerbation is less than 7 days after the end date of a previous episode then this will be assumed to be one continuous exacerbation with the start date taken from the first episode and the end date from the second or last episode. The worst severity of these episodes will be taken as the severity of the combined exacerbation.

COPD exacerbations, as recorded on the eCRF, will be included in all adverse event analyses described above.

In addition, the number of collapsed COPD exacerbations reported during the 12 week treatment period will be summarized descriptively by severity (total (also including mild), moderate and/or severe, mild, moderate, severe). COPD exacerbations will be considered on treatment, if they started on or after the time of the first administration of randomized study drug but not later than 1 day after the last administration of study medication.

A subgroup analysis based on the exacerbations history at baseline (0 exacerbations/1 exacerbation) will be performed on the number of COPD exacerbations during treatment, by severity.

2.7.3 Vital signs and ECG

Vital signs (blood pressure, pulse rate, and body weight) and ECG data will be summarized by treatment, scheduled visit and for the minimum and maximum post-baseline value for vital signs and the maximum QTc and ventricular rate for ECG data, including changes from baseline.

In addition, frequencies of notable vital signs and Fridericia's QTc (QTcF) will be summarized by treatment, scheduled visit and for the minimum and maximum post-baseline value for vital sign and the maximum post-baseline value for Fridericia's QTc.

Data measured more than 7 days after last inhalation of study drug are regarded as post-treatment data and will not be summarized, only listed. For the derivation of the minimum and maximum post-baseline value all post-baseline data from scheduled, unscheduled and premature discontinuation visits will be considered.

Listings of all patients with notable vital signs and notable Fridericia's QTc will be provided.

Notable absolute values and notable changes from baseline for each vital sign parameter and for Fridericia's QTc are defined in Appendix 16.1.9.

The overall ECG interpretation (normal, clinically significant abnormal) will be shown by visit and for the worst post-baseline measurement during treatment (considering all post-baseline data from scheduled, unscheduled and premature discontinuation visits). In addition a shift table from baseline to the worst interpretation during treatment will be presented.

2.8 Interim analyses

No interim analysis will be performed.

2.9 Determination of sample size

For the endpoint trough pre-dose FEV_1 change from baseline, a SD of 220mL was used based on a review of the results of Novartis phase III COPD studies. Discontinuation rate = 15% was assumed.

Sample size calculation for trough pre-dose FEV_1 change

alpha=0.05, 2-sided			
Power(%)	90	90	90
Difference to detect	65	70	75
SD	220	220	220
n per arm	242	209	182
with 15% drop-out rate	285	246	215
Total	570	492	430

Based on these calculations, a sample size of 418 completed patients is selected in order to detect a trough pre-dose FEV_1 change of 70mL with 90% power. Assuming a 15% drop out rate, 492 patients will be randomized.

Clinical Study Report - Appendix 16.1.9 Documentation of statistical methods

16.1.9.1 Introduction

This appendix gives details about statistical methods in addition to the report text. All analyses will be performed by using SAS Version 9.3.

16.1.9.2 Major protocol deviations

Patients fulfilling any of the following criteria are not eligible for inclusion in this study. No additional exclusions may be applied by the investigator, in order to ensure that the study population will be representative of all eligible patients.

- Informed consent not signed prior to assessments being performed
- Age < 40 years
- Smoking history < 10 pack years
- Post-bronchodilator FEV1 <30% & >=80% of normal value or post-bronchodilator $FEV_1/FVC >= 0.70$
- Treated with salmeterol/fluticasone 50/500 μ g b.i.d. for less than 3 months prior to visit 1
- CAT score < 10
- Treatment with any LAMA in the 2 weeks prior to visit 1
- Prior or current diagnosis of asthma
- Clinically significant co-morbidity that could interfere with the assessment of the safety and efficacy of the study medication.
- Malignancy within the past 5 years (except localized basal cell carcinoma of the skin)
- Not achieved acceptable spirometry results at visit 1 or 2 in accordance with the American Thoracic Society (ATS) or European Respiratory Society (ERS) criteria for acceptability.
- More than one COPD exacerbation requiring treatment with antibiotics and/or systemic corticosteroids and/or hospitalization in the year prior to Visit 1
- COPD exacerbation of any severity within the 6 weeks before the screening (Visit 1) or between screening (Visit 1) and start of treatment (Visit 2)
- Respiratory tract infection within 4 weeks prior to Visit 1
- Respiratory tract infection between Visit 1 and 2
- Requiring oxygen therapy prescribed for >12 hours per day
- Onset of respiratory symptoms, including a COPD diagnosis prior to age 40 years
- Concomitant pulmonary disease
- Diagnosed with α -1 anti-trypsin deficiency
- Active pulmonary tuberculosis
- Pulmonary lobectomy or lung volume reduction surgery or lung transplantation

- Participating in or planning to participate in the active phase of a supervised pulmonary rehabilitation program during the study
- Receiving prohibited medication
- Use of investigational drugs/devices (approved or unapproved) at the time of enrollment, or within 30 days or 5 half-lives of Visit 1, whichever is longer
- Unable to use a dry powder inhaler device, Metered Dose Inhaler (MDI) or a pressurized MDI (rescue medication) or comply with the study regimen.
- Patient received study drug but not randomized into study
- Use of prohibited indication medication during the study
- Patient randomized in error
- Major GCP deviation

16.1.9.3 Subject classification

Protocol deviations will lead to patient classification into the analysis sets as follows:

Analysis set exclusion based on population codes

Analysis set	PD severity codes that cause a subject to be excluded	Other criteria that cause a subject to be excluded
RAN	**NA**	
Safety	**5, 8**	Patient randomized but did not receive any double-blind study treatment
FAS	**0, (5*), 8**	Patient randomized but did not receive any double-blind study treatment
Per Protocol	**0, 1, (5*), 8**	Patients who receive another than their randomized treatment because of a dispensing error. Patient randomized but did not receive any double-blind study treatment

* Note: given that the Per Protocol population is nested within the FAS, which is itself nested within the Safety population, severity code 5 implies exclusion from all three of these analysis populations (i.e. SAF, FAS and PP).

Population code text

Population Code	Population code - text
0	Exclude from all efficacy analysis (FAS, PPS)
1	Exclude from Per-Protocol analysis
5	Exclude from all safety analyses (SAF)
8	Exclude from all analysis (FAS, PPS, and SAF)
49	Report relevant protocol deviation - include in all analyses

Unless otherwise stated, summary tables, figures and listings will be on all subjects included in the analysis set under consideration.

16.1.9.4 Baseline and post-baseline definitions, missing data handling

16.1.9.4.1 Baseline definitions

In general, baseline is defined as the last measurement before the first dose of study drug at Day 1.

1) For all **FEV_1 and FVC** endpoints, the baseline value is defined as the average of the values taken at -45 and -15 min prior to first dose of study drug at Day 1. Checks will be performed to ensure values were indeed taken prior to the first dose of study drug. If one of the values is missing (or is not confirmed to be pre-dose) then the remaining non-missing value will be taken as baseline. If both values are missing (or are not confirmed to be pre-dose), then the pre-bronchodilator measurements taken at the screening visit 1 can be used as the baseline. If the FEV_1 or FVC measurements are missing both on Day 1 and at screening visit, the respective baseline values will be set to missing.

2) **Baseline Dyspnea Index (BDI)** and baseline **CAT** are defined as the assessments taken right before the first dose of the study drug on Day 1 (Visit 2). If the CAT assessment on Day 1 is missing, then the last CAT assessment taken prior to first dose will be used as baseline. Missing baseline values will not be imputed.

3) For **rescue medication use**, the baseline percentage of days without rescue use is defined as the difference between number of days between visit 1 and visit 2 (date of first dose of study drug) and the number of days with rescue medication use as entered on the eCRF at visit 2 divided by the total number of days between visit 1 and visit 2 multiplied by 100.

The baseline mean number of puffs is defined as the number of puffs between visit 1 and visit 2 divided by the number of days between visit 1 and visit 2. Missing diary data will not be imputed.

4) **Vital signs** include pulse rate (measured as radial pulse for 60 seconds) and systolic and diastolic blood pressures. Baseline vital signs are defined as the assessments taken pre-dose on Day 1. Checks will be performed to ensure the assessments were indeed taken prior to the first dose of study drug on Day 1. If this assessment is missing or not confirmed to be pre-dose, the last value taken prior to first dose will be used for baseline.

5) Baseline **height and weight** are defined as the measurements taken at visit 1. Missing baseline will not be imputed.

6) Baseline **ECG** is defined as ECG measured prior to the first dose of study drug on Day 1. Checks will be performed to ensure the ECG was indeed assessed prior to the first dose of study drug. If it is missing (or not confirmed to be pre-dose), then the last assessment taken prior to first dose will be used. Otherwise, the ECG baseline will be set to missing without imputation.

16.1.9.4.2 Post-baseline measurements

Post-baseline measurements are defined as those assessments after the first dose of study drug.

Trough pre-dose FEV_1 and trough pre-dose FVC are defined as the mean of the respective measurements made at -45 min and -15 min before the morning dose. If one of the -45 and -15 min values is missing (or is not confirmed to be pre-dose) then the remaining non-missing value will be taken as trough pre-dose FEV_1 value. If both values are missing (or are not confirmed to be pre-dose), the trough pre-dose value will be set to missing.

For **rescue medication use**, the post-baseline percentage of days without rescue medication use is defined as the difference between the sum of days between visit 2 (date of first dose of study drug) and visit 3 (date of dosing, and if missing visit date) and between visit 3 and 4 (date of dosing, and if missing visit date) and the sum of days with rescue medication use as entered on the eCRF at visit 3 and visit 4, divided by the total number of days between visit 2 and visit 4 multiplied by 100. The treatment period ends at visit 4 /EOS.

For **safety data**, post-baseline measurements comprise recordings up to the last dose of study drug (+ 7 days for patients who discontinued early) for ECG, vital signs, non-serious AE and up to the last dose of study drug + 30 days for SAEs and death.

Deaths (if any) that happened to occur after the clinical database lock will be manually described in the CSR main text, but not in the post-text table that will be done by programming.

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.

16.1.9.5 Derivation of demographics and baseline characteristics

Baseline values are as described in “Baseline measurements” section above.

- Years since patient stopped smoking is calculated using the formula: Years since patient stopped smoking = (Date of visit 1 – date stopped smoking)/365.25.
  If the date stopped smoking is missing in day and/or month, it will be imputed in the same way as described for the date of COPD diagnosis.
- 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/or month, it will be imputed as follows. If the year is before Visit 1, the missing days will be imputed as the first of the month and the missing months will be imputed as July. If the year is the

current year of Visit 1, the missing days will be imputed as the first of the month and the missing months will be imputed as January.

- The estimated number of pack years is defined as the total years of smoking multiplied by cigarette packs smoked per day (e.g.10 pack years = 1 pack /day x 10 yrs., or ½ pack/day x 20 yrs.). The number of pack years will be analyzed as recorded on the CRF.
- Spirometry at screening: Only acceptable values will be analyzed. Moreover values with
    - SABA use less than 6 hours
    - LABA use less than 24 hours
    - SAMA use less than 8 hours
    - LAMA use less than 14 days
    - Indacaterol use less than 3 days

  prior to spirometry will be set to missing.
- Airflow limitation is classified based on % of predicted FEV_1 and FEV_1/FVC post bronchodilation:
- 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 $FEV_1 < 30$ %
- Combined assessment of COPD (2014) is classified by using the stages of airflow limitation (see above), the number of COPD exacerbations, COPD exacerbations leading to hospital admission in the last year and CAT at baseline. When assessing risk (low or high), the highest risk according to GOLD grade or exacerbation history will be chosen in GOLD 2014. When assessing symptoms (less or more), the category according to the CAT score will be chosen.

  The following algorithm will be applied based on the standard in GOLD 2014:
- Group A (Low Risk and Less Symptoms):

  (GOLD 1-2 and history of exacerbation ≤ 1 not leading to hospital admission) and CAT < 10
- Group B (Low Risk and More Symptoms):

  (GOLD 1-2 and history of exacerbation ≤ 1 not leading to hospital admission) and CAT ≥ 10
- Group C (High Risk and Less Symptoms):

  (GOLD 3-4 or history of exacerbation ≥ 2 or ≥ 1 leading to hospital admission) and CAT < 10
- Group D (High Risk and more Symptoms):

  (GOLD 3-4 or history of exacerbation ≥ 2 or ≥ 1 leading to hospital admission) and CAT ≥ 10

- Combined assessment of COPD (2017) is classified by using the number of COPD exacerbations, COPD exacerbations leading to hospital admission in the last year and CAT at baseline. When assessing risk (low or high), the highest risk according to exacerbation history will be chosen in GOLD 2017. When assessing symptoms (less or more), the category according to the CAT score will be chosen.
  The following algorithm will be applied based on the standard in GOLD 2017:

– Group A (Low Risk and Less Symptoms):
(history of exacerbation ≤ 1 not leading to hospital admission) and CAT < 10
– Group B (Low Risk and More Symptoms):
(history of exacerbation ≤ 1 not leading to hospital admission) and CAT ≥ 10
– Group C (High Risk and Less Symptoms):
(history of exacerbation ≥ 2 or ≥ 1 leading to hospital admission) and CAT < 10
– Group D (High Risk and more Symptoms):
(history of exacerbation ≥ 2 or ≥ 1 leading to hospital admission) and CAT ≥ 10

- % of predicted FEV_1 is the percentage of FEV_1 relative to the predicted normal value.
  Quanjer equations to calculate predicted FEV_1 (L)
  The following equations will be used by third party vendors to give predicted FEV_1 (L):
  Male: (4.30 x Height in meters) – (0.029 x Age in years) – 2.49
  Female: (3.95 x Height in meters) – (0.025 x Age in years) – 2.60
  If Race = Black or Ethnicity = Indian then the predicted normal given by the formulae above was multiplied by 0.9.
- FEV_1 reversibility is calculated as percentage increase of FEV_1 values after sequential inhalation of 84 µg of ipratropium bromide and 400 µg of salbutamol (or 360 µg albuterol) relative to FEV_1 values prior to the inhalation.

16.1.9.6 Vital signs – definition of clinically notable values

The following table shows the clinical notable criteria for vital signs.

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 range 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

Clinical notable criteria for QTcF (Friderica's formula)

ECG parameter (unit)	Clinically notable range
Notable value considering newly occurring or worsening cases	
QTc (ms)	> 450 (both male and female)
QTc (ms)	> 480
QTc (ms)	< 500
Notable change from baseline	
QTc	30 - 60
QTc	> 60

16.1.9.8 Statistical methodology and assumptions

- **Mixed Model Repeated Measures (MMRM)**

The following MMRM ANCOVA model will be used for analysis of trough pre-dose FEV_1:

Trough pre-dose FEV_1 = intercept + treatment + baseline FEV_1 + baseline smoking status + region + visit + treatment*visit + baseline FEV_1*visit + center(region) as random effect + error.

The random effect in the mixed model will be represented by the use of an unstructured covariance matrix for the within-patient error.

The SAS procedure MIXED will be used with the following code:

```
proc mixed  data=....  order=internal;
  where vis_1n in (3,4);
  class trtn1 smh1c region vis_1n sid1a ctr1n;
  model val_1n = trtn1 bsval_1n smh1c region vis_1n trtn1*vis_1n bsval_1n*vis_1n/ ddfm=kr;
   random ctr1n(region);
  repeated vis_1n / subject=sid1a type=un;
  lsmeans trtn1*vis_1n / cl diff;
run;
```

where val_1n = trough pre-dose FEV_1
trtn1 = Treatment
bsval_1n = baseline FEV_1 value
smh1c = Smoking status at baseline (ex-, current smoker)
region = region (Australia/South Africa, Asia, India, Middle East)
vis_1n = visit (include the scheduled visits only, no unscheduled or premature discontinuation visits)
ctr1n = center
sid1a = subject identifier

The estimated treatment difference of QVA149 110/50 µg o.d. – salmeterol/fluticasone 50/500 µg b.i.d. will be displayed along with the associated 95% confidence interval and p-value.

Other variables (i.e., TDI, trough pre-dose FVC, $FEV_1$and FVC at individual post-baseline time points, CAT) will be analyzed with the same MMRM model as used for the difference in treatments in trough pre-dose FEV_1, however, by including appropriate baseline values.

If the analysis fails to converge with an unstructured covariance matrix first choice is to remove the random effect of center nested within region and the second choice is to use a different covariance structure; either a compound symmetry (first choice) or first order autoregressive (AR1) (second choice).

- **Linear Mixed Model**

A linear mixed ANCOVA model will be used for trough pre-dose FEV_1 with imputation of missing week 12 data by LOCF:

Trough pre-dose FEV_1 = intercept + treatment + baseline FEV_1 + smoking status + region + center(region) as a random effect + error.

The SAS procedure MIXED will be used with the following code:

```
proc mixed  data=....  order=internal;
  class trtn1 smh1c region ctr1n;
  model val_1n = trtn1 bsval_1n smh1c region / ddfm=kr;
  random ctr1n(region) / type=vc;
  lsmeans trtn1 / cl diff;
run;
```

where val_1n = trough pre-dose FEV_1 at Week 12
trtn1 = Treatment
bsval_1n = baseline FEV_1
smh1c = Smoking status at baseline (ex-, current smoker)
region = region (Australia/South Africa, Asia, India, Middle East)
ctr1n = center

Results will be presented with least squares mean and standard error for treatment effects and least squares mean, standard error, associated two-sided 95% confidence interval, and two-sided p-value for the treatment contrast.

Other variables will be analyzed with the same mixed model (i.e. daily use of rescue medication, percentage of days without rescue medication over 12 weeks), however, by including appropriate baseline values.

- **Pattern mixture model**

An additional method for handling missing data will be implemented to explore the sensitivity to the Missing At Random (MAR) assumption. A pattern-mixture model based on nonignorable missing, i.e., Missing Not At Random (MNAR) will be used, where all drop-outs are assumed to switch back to salmeterol/fluticasone 50/500 µg b.i.d. treatment. This approach will be implemented by multiple imputations, where missing values are repeatedly replaced by sampling from the Bayesian posterior predictive distribution based on the

switching model. Each of the completed datasets is then analyzed using the primary analysis, and then the estimates are combined to get the final inference.

In general, 4 programming steps will be done using the SAS procedures MI, MIXED, and MIANALYZE:

1. Impute missing intermediate visit (i.e., Day 42 is missing but Week 12 is non-missing) to obtain a monotone missing data pattern.

```
proc mi data=... out=_imp1 nimpute=100 seed=.....;
  by trtn1;
  mcmc chain=multiple impute=monotone;
  var VIS3 VIS4;
run;
```

   with VIS3 = Day 42 value and VIS4 = Week 12 value.

   The resulting output dataset _imp1 will have 100 copies of the original dataset where each copy will have a monotone missing pattern.

2. Impute missing Week 12 data by using the dataset obtained from Step 1 and including all salmeterol/fluticasone subjects and only the subjects of QVA149 that have missing values at Week 12. Thus missing Week 12 values will be imputed based on the observed outcomes of completers on salmeterol/fluticasone. A regression model including preceding values will be used for imputation.

```
proc sort data=_imp1 out=_imp2;
  where trtn1=Salm/Flut or (trtn1=QVA149 and vis4=.);
  by _imputation_;
run;
proc mi data=_imp2 out=_imp3 nimpute=1 seed=...;
  by _imputation_;
  var VIS3 VIS4;
  monotone reg(VIS4);
run;
```

   Pool the above imputed dataset with the dataset obtained from Step 1 but including only the subjects of the QVA149 group that have values at Week 12.

```
data _imp4;
  set _imp1(where=(trtn1=QVA149 and vis4 gt .))
      _imp3;
run;
proc sort data=_imp4;
  by _imputation_;
run;
```

3. Analyze Week 12 data by using the pooled dataset derived from Step 2.

```
proc mixed data=_imp4 order=internal;
  by _imputation_;
  class trtn1 smh1c region;
  model vis4 = trtn1 bsval_1n smh1c region / solution covb ddfm=kr;
  lsmeans  trtn1 / cl diff;
  random ctr1n(region);
  ods output lsmeans=ls diffs=con;
```

```
run;
```

4. Combine the results from Step 3analysis on 100 imputed datasets to derive an overall result.

```
proc mianalyze parms=ls [parms=con in a second step]
  clas trtn1;
  modeleffects trtn1;
run;
```

Parms=ls gives the least squares means of treatments and parms=con of treatment contrasts.

- **Logistic regression analysis for repeated measures**

The percentage of patients with a clinically significant improvement after 12 weeks of treatment in TDI (of >=1 change from baseline) will be analyzed using a logistic regression model for repeated measures which includes the same terms as used for the analysis of the TDI focal score.

Response = intercept + treatment + baseline BDI + smoking status at baseline + region + visit + treatment*visit + baseline BDI*visit + error.

The SAS procedure GENMOD will be used with the following code:

```
proc genmod data=.... order=internal rorder=internal descending;
   where vis_1n in (3,4);
   class trtn1 bsval_1n smh1c region sid1a;
   model response = trtn1 bsval_1n smh1c region vis_1n trtn1*vis_1n
                    bsval_1n*vis_1n / dist=bin link=logit;
   repeated subject=sid1a / type=un;
   lsmeans trtn1*vis_1n / cl diff oddsratio;
run;
```

where response = improvement TDI focal score (= 1, or 0 if no response)
trtn1 = Treatment (QVA, Salm/Flut)
bsval_1n = baseline BDI
smhgr1 = Smoking status (ex-, current smoker), as measured at baseline
region = region (Australia/South Africa, Asia, India, Middle East)
vis_1n = visit (include the scheduled visits only, no unscheduled or premature discontinuation visits)
sid1a = unique subject identifier

The DESCENDING option of the PROC GENMOD statement causes the last ordered category (here = improvement) as the event.

Odds ratios will be presented along with 95% confidence intervals and p-values for the treatment contrast of QVA vs. Salm/Flut.

If the analysis fails to converge with an unstructured covariance matrix first choice is to use a different covariance structure; either a compound symmetry (first choice) or first order autoregressive (AR1) (second choice).