NOVARTIS

Clinical Development

QVA149

CQVA149A2316

A 26-week, randomized, double blind, parallel-group multicenter study to assess the efficacy and safety of QVA149 (110/50 µg o.d.) vs tiotropium (18 µg o.d.) + salmeterol/fluticasone propionate FDC (50/500 µg b.i.d.) in patients with moderate to severe COPD

RAP Module 3 – Detailed Statistical Methodology

Author: [redacted]

Document type: RAP Documentation

Document status: Final version 3.0

Release date: 18AUG2017

Number of pages: 36

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

Version	Date	Changes
Amendment 1	07-Apr-2017	Removed missing data imputation for St. George Respiratory Questionnaire C (SGRQ-C) data. Added COPD GOLD 2011 and 2017 to demographic characteristics. Added subgroup analyses for trough FEV1 using the same model as the primary analysis. Time-to-first moderate or severe COPD exacerbation analysis is included in secondary efficacy analysis instead of exploratory efficacy analysis section. Removed summary and analysis for the COPD assessment test (CAT) individual item scores, not reported. Removed summary and analysis for patients who achieved a clinically significant difference or change in health status for at least 2 in the total CAT score after 26 weeks of treatment, not reported. Added subgroup analysis by region for the rate of protocol-defined COPD exacerbations. Summary of adjudicated MACE and/or CV deaths added. The rate of pneumonia will be analyzed using the same generalized linear model as for the rate of moderate or severe COPD exacerbations. Summary of proportion of patients with oral candidiasis added. Changed wording in Laboratory data section, 24h urinary cortisol as 24h urinary free cortisol corrected for creatinine. For biomarkers section, change from baseline is added. In Table 2-2, the QTc (msec) >480 notable value will not be presented separately. Added subgroup analyses for all secondary endpoints considering blood eosinophilia counts and history of prior COPD exacerbation. Information about BDI and TDI has been added in Section 2.7.2. Removed heart rate parameter from vital signs in Section 1.5.2. Removed '#' in Section 1.5.2.
Amendment 2	18-Aug-2017	Add protocol deviation EXCL10B and INCL04B to Table 2-5 in section 2.5

1 Statistical methods planned in the protocol and determination of sample size

1.1 Statistical and analytical plans

This report and analysis plan (RAP) module describes the statistical analysis methods specified in Section 9 of the clinical study protocol CQVA149A2316 along with any additional analyses, specifications or deviations before database lock. The purpose of this study is to determine whether the efficacy and safety of QVA149 (110/50 µg o.d.) and triple treatment with tiotropium (18 µg o.d.) + salmeterol/fluticasone propionate FDC (50/500 µg b.i.d.) are comparable in patients with moderate to severe COPD without a history of frequent exacerbations.

Data will be analyzed according to the data analysis Section 9 of the study protocol which is available in Appendix 16.1.1 of the CSR. Important information is given in the following sections and details are provided, as applicable, in Appendix 16.1.9 of the CSR.

1.2 Definitions and general strategy

1.2.1 Baseline definitions

Unless otherwise specified, baseline is defined as the last measurement before first dose of study treatment in the treatment epoch which will be from Visit 102 or an earlier visit.

1.2.2 Assessment windows

Data from unplanned or unscheduled visits or the early discontinuation visits will be listed.

For patients who complete the study, the end of treatment visit (Visit 208) will be remapped to Week 26. For patients who do not complete the study, the premature discontinuation visit will not be remapped. Unscheduled visits will be remapped using the repeat visit number and/or repeat page number. No other visits will be remapped.

All efficacy data from these visits will not be used for missing data imputation. Laboratory, vital signs and ECG data from these types of visits will only be included in the summaries of the notable values and extreme values.

Laboratory, vital signs, and ECG values which have complete data and time values will be slotted into pre or post-dose assessment based on the actual date/time. For values with missing date/time, scheduled visit date and time will be used. This rule will be applied to data from scheduled visits only.

1.2.3 Data collected after treatment discontinuation

Efficacy, laboratory, vital signs and ECG data collected on or after an early discontinuation visit will not be analyzed only listed. An exception however is a supportive analysis of the primary endpoint (see section 1.4.1.3).

1.3 Patients and treatments

1.3.1 Analysis sets

The Randomized set (RAN): The RAN will consist of patients who were assigned a randomization number; regardless if they actually received the study medication. It will be used for summaries of patient disposition and analysis sets, and listings of major protocol deviations and premature discontinuations. Patients in the RAN will be analyzed according to the treatment they were randomized to.

The Full analysis set (FAS): The FAS will consist of all patients in the RAN who received at least one dose of study medication. Following the intent-to-treat principle, patients in the FAS will be analyzed according to the treatment they are assigned to at randomization. The FAS will be used to analyze all efficacy endpoints, unless otherwise stated.

The Per-protocol set (PPS): The PPS will include all patients in the FAS without any major protocol deviations. Major protocol deviations will be defined in the data review plan prior to database lock and the unblinding of the study. Patients in the PPS will be analyzed according to the treatment group they were randomized to. Patients who receive treatment other than their randomized treatment because of a dispensing error will be excluded from the PPS. The PPS will be used for supportive analysis to assess robustness of the primary analysis. Major protocol deviations will be defined prior to database lock and unblinding of the study. Non-compliance to study drug will be regarded a minor protocol deviation.

The Safety set: The Safety set will include all patients who received at least one dose of study medication whether they were randomized or not. Patients will be analyzed according to the treatment they received. If patients receive more than one treatment during the study, they will be analyzed according to the treatment they were randomized to. The Safety set will be used in the analysis of all safety endpoints and in the listings of certain notable safety data.

1.3.2 Patient disposition, demographic and other baseline characteristics

1.3.2.1 Patient disposition

The number of patients screened who completed the screening epoch will be given and the reasons for not entering the run-in epoch will be summarized for all patients. A similar summary for completion of the run-in epoch will also be provided. Patients who screen-failed and were re-screened under a new patient number will be counted multiple times in the total number of screenings.

The number and percentage of patients in the RAN who completed and discontinued (including reasons for discontinuation) will be summarized by treatment group for the treatment and follow-up epochs.

Completion data for the treatment epoch (including date of last dose, completion status, and primary reason for discontinuation) will be listed for the RAN. The date of unblinding will be included if applicable. Completion data of the screening, run-in and follow-up epochs will also be listed.

The number of patients with major protocol deviations (see section 2.5) or other criteria which cause exclusion from the PPS will be tabulated by category and deviation in the RAN. Protocol deviations will be listed with date and study day of occurrence, deviation and population codes. Separate summaries for major and minor protocol deviations will be provided.

1.3.2.2 Demographics and other baseline characteristics

All demographics or baseline summaries will be done using the RAN. Demographics will be collected at Visit 1 and all other measurements will be collected on or prior to Visit 201 (date of randomization) unless stated otherwise.

Demographic and baseline characteristics include the following categorical variables:

- Gender (Male, Female)
- Race (Caucasian, Black, Asian, Native American, Pacific Islander, Unknown, Other)
- Child bearing potential (for females only) (able to bear children, post-menopausal, sterile-of child bearing age)
- Smoking status (Patient's usage status: current, former)
- Airflow limitation (GOLD 2011) (mild, moderate, severe, very severe)
- Severity of COPD (GOLD 2011) (Grade A, Grade B, Grade C, Grade D)
- Symptoms/risk of exacerbation (GOLD 2017) (Grade A, Grade B, Grade C, Grade D)
- Number of exacerbations in the previous year (0, 1, >1)

Categorical variables will be summarized based on the number and percentages of patients in each category.

Demographic and baseline characteristics include the following continuous variables:

- Age (in years)
- Height (in cm)
- Weight (in kg)
- Body mass index (BMI) (kg/m^2)
- Smoking history (Estimated amount consumed on average (pack years))
- Duration of COPD (in years)
- SGRQ-C score
- mMRC score
- CAT score
- BDI score
- Number of exacerbations in the previous year
- Baseline spirometry (FEV_1, FVC)
- Run-in spirometry (predicted FEV1, FEV1 , % of predicted FEV1 prior to and after inhalation of 400 µg of salbutamol, FEV1 reversibility, FVC, FEV1/FVC prior to and after inhalation of 400 µg of salbutamol)
- Vital signs

- Laboratory results
- ECG

Note: Duration of COPD is calculated as: (date of enrollment – date of diagnosis of COPD + 1)/365.25. BMI is calculated as: BMI (kg/m^2) = Weight (kg) / [Height (m) × Height (m)].

Continuous variables will be summarized by presenting mean, median, standard deviation, and, minimum and maximum.

1.3.2.3 Medical history/current medical conditions

Medical history will be coded with the Medical Dictionary for Regulatory Activities terminology (MedDRA) using the most actual version at the time of database lock. History/conditions will be summarized for both the RAN and Safety set by primary system organ class and preferred term. The summary will include pre-defined cardiovascular events and pulmonary diseases. 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.

1.3.3 Study medication

Unless otherwise stated, all analyses will be based on patients in the Safety set.

1.3.3.1 Study drug administration

The extent of exposure will be examined to determine the degree to which safety can be assessed for QVA149 (110/50 µg o.d.) and triple treatment with tiotropium (18 µg o.d.) + salmeterol/fluticasone propionate FDC (50/500 µg b.i.d.). The extent of exposure to study drug will be characterized according to the duration of exposure and the number of patients exposed.

Duration of exposure to a 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 study drug – Date of first dose of study drug + 1).

The duration of exposure (in days) will be summarized by treatment for the Safety set as a continuous variable with the standard descriptive statistics. In addition, the duration of exposure will be summarized with total patient years and the number (%) of patients who were exposed to study drug overall and for each ≥ 2, ≥ 4, ≥ 12, ≥ 26 weeks of treatment.

1.3.3.2 Compliance

Study drug compliance will be assessed by the investigator and/or site personnel at designated visits according to the procedures defined in the study protocol.

Compliance will be calculated as the percentage of the number of days where study drug was administered as per protocol divided by the duration of exposure (i.e. the number of days

between first and last dose). Compliant days will be calculated following instructions given in the Appendix Section 2.2.

Compliance will be categorized by < 80 % and 80 % - 100 % and summarized by treatment for the Safety set.

1.3.3.3 Prior and concomitant medication

Each medication will have the start and end dates recorded. Prior medications are defined as those medications which were taken and stopped prior to the first dose of double-blinded study drug. Concomitant medications are defined as those medications which were taken after the first dose of study drug. Prior concomitant medications are defined as those medications which were taken prior to and continued after the first dose of the study drug. All prior concomitant medications will be summarized as concomitant medications in one table. Medications started and stopped prior to study drug, and then taken again concomitantly will be included in both prior and concomitant summaries separately.

COPD-related medications will be summarized by pre-specified drug categories, route of administration, and preferred term, as recorded on the eCRF. The summary will be repeated by showing active ingredients instead of preferred terms. Non-COPD related medication will be summarized by route of administration and preferred term. Both COPD-related and non-COPD related medications will be summarized separately for prior and concomitant medications.

Short acting β2-agonist (SABA) usage (number of puffs) during the screening period will be summarized.

All summaries will be on the Safety set.

1.4 Efficacy evaluation

1.4.1 Primary efficacy analysis

The primary objective of this study is to demonstrate the non-inferiority of QVA149A (110/50 µg o.d.) on post-dose trough FEV_1 vs. tiotropium (18 µg o.d.) + salmeterol/fluticasone propionate FDC (50/500 µg b.i.d.) in terms of trough FEV_1 after 26 weeks of treatment in moderate-to-severe COPD patients.

The primary variable is the mean change from baseline in post-dose trough FEV_1 after 26 weeks of treatment. Trough FEV_1 is defined as the mean of the two FEV_1 values measured at 23 hr 15 min and 23 hr 45 min after the morning dose taken at site on Day 181 (Visit 207). Baseline FEV1 is defined as the average of the pre-dose FEV_1 measured at -45 min and -15 min at Day 1 (Visit 201).

1.4.1.1 Statistical model, hypothesis, and method of analysis

The comparison between QVA149 (110/50 µg o.d.) and tiotropium (18 µg o.d.) + salmeterol/fluticasone propionate FDC (50/500 µg b.i.d.) in terms of trough FEV_1 at Day 182

will be evaluated by testing the following null hypothesis (H_0) versus the alternative hypothesis (H_a) at one-sided 2.5% significance level:

H_0: mean change from the baseline in trough FEV_1 at Day 182 (QVA149 (110/50 µg o.d.)) is inferior to the mean change from the baseline in trough FEV_1 at Day 182 [tiotropium (18 µg o.d.) + salmeterol/fluticasone propionate FDC (50/500 µg b.i.d.)]

H_a: mean change from the baseline in trough FEV_1 at Day 182 (QVA149 (110/50 µg o.d.)) is non-inferior to the mean change from the baseline in trough FEV_1 at Day 182 [tiotropium (18 µg o.d.) + salmeterol/fluticasone propionate FDC (50/500 µg b.i.d.)]

The primary efficacy endpoint will be analyzed using a Mixed-Effect Model Repeated Measures (MMRM) model.

The model will include fixed, categorical effects of treatment and visit, region, and treatment-by-visit interaction as well as the continuous, fixed covariates of baseline and baseline-by-visit interaction. The within-patient correlation will be modeled using an unstructured covariance matrix. Restricted maximum likelihood method will be used. The Kenward-Roger approximation will be used to estimate denominator degrees of freedom.

If the analysis fails to converge, then spatial covariance, compound symmetry, or first order autoregressive model (AR1) structure will be used. The best model fit will be determined by the Akaike's Information Criterion.

The between-treatment comparison will be carried out using the adjusted mean difference between treatments at Day 182.

Non-inferiority of QVA149 (110/50 µg o.d.) from tiotropium (18 µg o.d.) + salmeterol/fluticasone propionate FDC (50/500 µg b.i.d.) will be demonstrated if the 95% confidence interval for the mean FEV1 difference of QVA149 (110/50 µg o.d.) minus tiotropium (18 µg o.d.) + salmeterol/fluticasone propionate FDC (50/500 µg b.i.d.) lies entirely to the right of (higher than) –50 mL.

1.4.1.2 Handling of missing values/censoring/discontinuation

If any of the -45 min and -15 min values contributing to the trough FEV1 are collected within 7 days of systemic corticosteroid use, 6 hr of rescue medication, or actual measurement times are outside the 22 - 25 hour post-morning dose time window then the individual FEV1 value will be set to missing.

If one of the two values is missing (or set to missing) then the remaining non-missing value will be taken as trough FEV1. If both values are missing, or if the patient withdrew from the study, regardless of the reason for discontinuation, then trough FEV1 will be regarded as missing in which case the missing value(s) of the patient at the particular visit(s) would not directly contribute to the primary analysis.

All timed-trough FEV1 data i.e. Day 29, Day 85, Day 181 and Day 182, recorded post-baseline will be included in the primary MMRM model and no imputation will be applied to missing data. Instead of imputation, in the model the profile of each patient is used to adjust the estimates of the parameters when data are not available (i.e. the post-withdrawal statistical behavior of a patient who discontinued is assumed to be the same as for a patient who

remained in the study and who shared the same measurement history and the same covariates, including treatment group).

1.4.1.3 Supportive analyses

The following supportive analyses for trough FEV1 will be performed:

1. The same MMRM model used in the primary analysis will be also performed on the PPS.
2. To assess the robustness of the primary results in the presence of missing data the following analyses are planned:
   - An analysis of covariance (ANCOVA) model (including treatment, region, and baseline FEV1) will be used to analyze complete datasets created using multiple imputations under varying assumptions. Missing FEV1 values after discontinuation for patients on QVA who discontinued for any reason will be imputed based on information from tiotropium + salmeterol/fluticasone propionate FDC patients only. This analysis assumes that QVA patients after discontinuation have a similar response to patients on triple therapy.
   - Trough FEV1 at Week 26 with missing data imputed with LOCF from Day 29 will be analyzed using the ANCOVA model as described above.

   If these analyses provide inferences consistent with the primary analysis method then the conclusions will not be considered highly sensitive to how missing data are handled.
3. All available trough FEV1 data at Week 26 (including retrieved data from patients who prematurely discontinue treatment but who attend subsequent scheduled visits) will be analyzed using the ANCOVA model described in point 2. To note this retrieved data may be affected by off-study medications.
4. Exploratory subgroup analyses for trough FEV1 using the same MMRM as the primary analysis will be performed for the following subgroups:
   - Blood eosinophils (< 2% vs ≥ 2% and < 150, 150 - <300, >=300 cells/microliter)
   - COPD exacerbation in the previous year (0 vs 1)
   - Gender (male vs female)
   - Airflow limitation - GOLD 2011 (moderate vs severe)
   - Age group (40-55 vs 56-64 vs >=65)

1.4.2 Secondary efficacy analysis

The secondary endpoints will be summarized by treatment for the FAS. Subgroup analyses considering blood eosinophils at baseline (< 2% vs ≥ 2% and < 150, 150 - <300, >=300 cells/microliter) and COPD exacerbation history (0 vs. 1 exacerbation) will be performed for each secondary endpoint described below.

Moderate or severe COPD exacerbations

In patients with multiple exacerbations, if the start date of an exacerbation was 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 collapsed exacerbation (the merging of multiple exacerbations into one continuous exacerbation as defined above).

The number of moderate or severe COPD exacerbations during the treatment period will be summarized by treatment groups, as continuous variables and as categorical variables classified into 0, 1, 2, 3, ≥ 4 events.

The rate of moderate or severe COPD exacerbations during the treatment period will be analyzed using a generalized linear model assuming a negative binomial distribution. The time at risk for a patient is defined as the length of time the patient is on double-blind treatment and the log (length of time on double-blind treatment) will be used as the offset variable in the model. The analysis model will include terms for treatment, region, and COPD exacerbation history (the number of moderate or severe COPD exacerbations in the year prior to screening). An estimate of the ratio of moderate or severe COPD exacerbation rates between the treatment groups, together with the 95% confidence interval, will be presented.

The above summary and analysis will be repeated for the rate of moderate or severe COPD exacerbations requiring

- systemic glucocorticosteroids and/or antibiotics during the treatment period (moderate exacerbations only)
- hospitalizations during the treatment period and re-hospitalization within 30 days during the treatment period (severe exacerbations only).

The time to the first moderate or severe COPD exacerbation during the treatment period will be analyzed using a Cox regression model. The model will include the same terms in the rate of moderate or severe COPD exacerbations analysis above.

Trough FEV1 and FVC, SGRQ-C, TDI

Change from baseline over 26 weeks of treatment in trough FEV1 will be analyzed using the same MMRM model described for the primary endpoint. Trough FEV1 is defined as the average of the pre-dose FEV1 measurements at -45 min and -15 min prior to dosing at each visit with baseline considered the Day 1 (Visit 201) average of pre-dose measurements.

The analyses of trough FVC change from baseline, SGRQ-C total score change from baseline, and Transition Dyspnea Index (TDI) total score over 26 weeks of treatment will also be done using a similar MMRM model with baseline FEV1 replaced by corresponding baseline for the endpoint. Baseline SGRQ-C is the last non-missing value prior to the first dose of double-blind treatment.

A graph of trough FEV1 will also be presented by treatment group over time.

Mean use of rescue therapy over 26 weeks

The mean daily number of puffs and percentage of days without rescue medication usage will be calculated for each patient over 26 weeks. Rescue use during the run-in epoch will be used to calculate the baseline. The mean daily number of puffs will be analyzed using a linear mixed model with fixed categorical effects of treatment and region, a random effect of center

nested within region and a fixed continuous covariate of baseline. Percentage of days without rescue medication usage will be analyzed using a similar model.

1.4.3 Exploratory efficacy analysis

Total CAT score following 26 weeks of treatment

COPD assessment test (CAT) scores following 4, 12, and 26 weeks of treatment will be summarized at each visit and analyzed at these time points using the same MMRM as the primary endpoint. The estimated treatment difference on the contrast between QVA149 and tiotropium+salmeterol/fluticasone at each visit will be reported along with the associated 95% confidence interval and p-value.

Percentage of patients with total CAT scores of 0-10 (mild), 11-20 (moderate), 21- 30 (severe) and 31-40 (very severe) clinical impact after 4, 12 and 26 weeks of treatment will be summarized descriptively by treatment groups at each visit.

mMRC at Week 26

mMRC will be summarized at each visit (Visit 101 and 207) and change in grade will also be presented. An exploratory analysis of the difference between groups in grade and change in grade at Day181 will be performed using the Mann-Whitney U test. Change in grade at Day 181 will be calculated as the change from Visit 101.

Rate of all protocol-defined exacerbations

The rate of protocol-defined exacerbations during the treatment period will be summarized by treatment groups, as continuous variables and as categorical variables classified into 0, 1, 2, 3, $\geq$ 4 events.

The rate of protocol-defined exacerbations during the treatment period will be analyzed using a generalized linear model assuming a negative binomial distribution. The time at risk for a patient is defined as the length of time the patient is on double-blind treatment and the log (length of time on double-blind treatment) will be used as the offset variable in the model. The analysis model will include terms for treatment, region, and COPD exacerbation history (the number of moderate or severe COPD exacerbations in the year prior to screening). An estimate of the ratio of moderate or severe COPD exacerbation rates between the treatment groups, together with the 95% confidence interval, will be presented.

Subgroup analysis considering region will be performed for the rate of protocol-defined COPD exacerbations. The same analysis model as that without subgroup analysis will be used. In this case the interaction term of treatment by region will be further added to the model.

Time to first exacerbation will also be performed and analyzed using a Cox regression model. The model will include the same terms in the rate of protocol-defined exacerbations analysis above.

Symptom scores

The mean total symptom score and mean individual symptom scores (i.e., scores for respiratory symptoms, cough, wheeze, amount of sputum, color of sputum, breathlessness, sore throat, cold, and fever) will be calculated for each patient over the baseline and treatment periods.

The change from baseline in mean daily symptom scores will be summarized by treatment and will be analyzed similarly as for rescue therapy use with the baseline rescue therapy use being replaced by the respective baseline symptom variables in the linear mixed model.

1.5 Safety evaluation

All safety endpoints will be summarized by treatment for patients in the Safety set.

1.5.1 Adverse events

All treatment emergent adverse events (TEAEs), defined as adverse events (AEs) starting on or after the time of the first administration of study drug but not later than 7 days (30 days in the case of a SAE) after the last administration, will be included in the summaries. All adverse events will be listed. Events of COPD exacerbation obtained from the COPD exacerbation episode eCRF will be combined with the AE data and reported under the preferred term of "chronic obstructive pulmonary disease" together with all other events of the same preferred term. As a result COPD exacerbations will be included in all AE outputs.

Any adverse events that started during the study before the time of the first inhalation of study drug will be summarized and listed as medical histories.

The following TEAE summaries will be produced: overall by system organ class and preferred term; overall by system organ class, preferred term and maximum severity; suspected treatment-related adverse events by system organ class and preferred term; SAEs by system organ class and preferred term; and adverse events leading to permanent discontinuation of study-treatment by system organ class and preferred term.

AEs of special interest will be summarized by category, preferred term and treatment. These are defined as:

- Bladder obstruction, urinary retention
- Atrial fibrillation/flutter
- Cardio- and Cerebro-Vascular events (CCV) events: Cardiac arrhythmias
- CCV events: Cardiac failure
- CCV events: Cerebrovascular events
- CCV events: Myocardial infarction
- CCV events: Ischaemic heart disease
- Dry mouth
- Constipation
- Glaucoma/increased intraocular pressure
- Paradoxical bronchospasm

- QT prolongation
- Diabetes mellitus/Hyperglycemia
- Pneumonia
- Lower respiratory tract infection
- Respiratory composite endpoint

Adjudicated Mortality

Causes of death will be listed by patient and summarized by treatment groups for (1) mortality of all causes, (2) mortality of probable cardiovascular cause, (3) mortality of probable respiratory cause, (4) mortality of probable cancer cause and (5) mortality of probable other causes using the Safety set based on (A) cases reported during the active treatment period and (B) cases reported during the active treatment period plus the following 30 days. Mortality of the above 4 causes will also be summarized between treatment groups.

Time to death will be displayed graphically for each treatment group using a Kaplan-Meier curve for the Safety set.

Adjudicated serious cardio- or cerebrovascular (CCV) events

All serious CCV events that occurred will be sent to the adjudication committee. Number and percentage of patients with serious cardiovascular (CCV) events as adjudicated by independent committee will be summarized by category and treatment group with and without adjustment by exposure. The CCV events will be adjudicated with categorized outcomes as Major adverse cardiovascular event (MACE) with subcategories

1) Non-fatal myocardial infraction (MI)
2) Unstable angina (USA) requiring hospitalization
3) Non-fatal stroke
4) Heart failure requiring hospitalization
5) Coronary revascularization (CABG or PCI),

or as a serious CCV but not MACE event.

Adjudicated MACE and/or CV deaths will be summarized as above including MACE subcategories.

Listing of the CCV adjudicated outcome will be provided. Adjudicated CCV events will be summarized as

1) on treatment (i.e., from the first dose up to the last known dose)
2) on treatment + 30 days follow up period whether patients completed the study treatment period or not.

Pneumonia

Time to first pneumonia will be summarized and analyzed for each treatment group using a Kaplan-Meier estimation and log rank test for the Safety set.

The rate of pneumonia will be analyzed using the same negative binomial model as for the rate of moderate or severe COPD exacerbations.

Serious/non-serious adverse events (SAE/NSAE) and deaths due to SAE

For the legal requirements of clinicaltrials.gov and EudraCT, two required tables on on-treatment/treatment emergent adverse events (TEAEs) which are not serious adverse events with an incidence greater than 1% and on TEAEs and SAEs 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.

1.5.2 Vital signs

Vital signs (blood pressure, radial pulse) will be summarized by treatment at each time point for each visit respectively. The maximum (systolic blood pressure, pulse rate) or minimum (diastolic blood pressure) post first dosing will also be summarized. Changes from baseline will be summarized by treatment.

All safety data will be included in the analysis regardless of rescue medication use.

Changes from baseline will also be summarized by treatment and visit. The baseline measurement will be the last measurement prior to first dose.

The number (%) of patients with pulse rate of < 40 bpm, 40 – 90 bpm, and > 90 bpm; systolic blood pressure of <90 mmHg, 90 – 140 mmHg, and >140 mmHg; diastolic blood pressure of <50 mmHg, 50 – 90 mmHg, and >90 mmHg will be summarized by treatment group.

Data measured more than 7 days after last inhalation of study treatment is regarded as post-treatment data and will not be summarized, only listed.

Notable values for vital signs and change from baseline will be summarized. A notable value is defined in Appendix 16.1.9.

1.5.3 ECG

Each patient has two consecutive ECGs at baseline (greater than 45 minutes pre-dose on Day 1 (Visit 102)) and one ECG at subsequent scheduled time points (1 hour post dose at Day 1 (Visit 201), both greater than 45 minutes pre-dose and 1 hour post dose at Week 12 and Week

26). The average of the values at baseline and the value at other scheduled time points will be calculated for the quantitative ECG assessments and the analysis.

ECG data will be summarized by treatment at each time point for each visit respectively. For ECG parameters, the baseline is defined as the average of the two measurements taken at greater than 45 min pre-dose on Day 1. The maximum QTc post first dosing will also be summarized. Changes from baseline will be summarized by treatment. All safety data will be included in the analysis regardless of rescue medication use.

Notable QTc values will be summarized for Fridericia's QTc. Data from unscheduled visits and from premature discontinuation visits will be included. A listing of all newly occurring or worsening abnormalities will be provided. The clinically notable ranges for selected ECG parameters and notable changes from baseline are shown in Appendix 16.1.9.

In addition, the ECG overall interpretation of the central interpretation will be presented with frequencies per category and time point.

Data measured more than 7 days after last inhalation of study treatment is regarded as post-treatment data and will not be summarized, only listed.

1.5.4 Oral fungal infections

Number and percentage of patients with oral fungal infections will be summarized by treatment groups. Oral fungal infections are adverse events identified by the NMQ (90000061). Oral and oropharyngeal candidiasis, i.e. all AEs with one of the following preferred terms:

- Oral candidiasis
- Oropharyngeal candidiasis
- Candida infection
- Fungal pharyngitis
- Oral fungal infection
- Oropharyngitis fungal
- Tongue fungal infection
- Tonsillitis fungal

1.5.5 Laboratory data

All the laboratory samples will be processed through a Central Laboratory. Laboratory data consist of hematology, biochemistry and urinalysis (including 24hr urinary free cortisol corrected for creatinine) measurements. Laboratory samples will also include oropharyngeal swabs which will be summarized separately. All data will be listed with abnormal values flagged. Laboratory data measured more than 7 days after last inhalation of study drug is regarded as post-treatment data and will not be summarized, only listed. The following sub-sections describe the methods of summary.

Summary of absolute values

For all continuous laboratory parameters, the absolute laboratory values, including the worst case post-baseline values (including values from post-baseline unscheduled and premature discontinuation visits), will be summarized with standard descriptive statistics by parameter, scheduled visit and timepoint, and treatment. The direction of interest for worst case post-baseline for selected hematology and biochemistry parameters is shown in table given in the Appendix 16.1.9. For continuous urinalysis parameters the direction of interest is always high.

For categorical urinalysis laboratory parameters, a frequency table of results will be produced by laboratory parameter, scheduled visit and time-point, and treatment. Worst-case post-baseline values (including values from post-baseline unscheduled and premature discontinuation visits) will also be included.

Summary of change from baseline

For continuous laboratory parameters, the change from baseline at each scheduled visit and time-point, and the change from baseline to the worst case post-baseline values (including values from post-baseline unscheduled and premature discontinuation visits) will be summarized by laboratory parameter, scheduled visit and time-point, and treatment with standard descriptive statistics.

Shift tables

Shift tables for laboratory parameters will be provided in order to compare a patient's baseline value to the value at each time point at each study visit, relative to the normal reference range for each laboratory parameter. For the shift tables, normal reference ranges provided by the central lab will be used to evaluate whether a particular laboratory test value for each time point at each visit is normal, low, high or non-available relative to the baseline value also categorized as normal, low, high, or non-available. These summaries will be presented by laboratory test, visit, time point, and treatment group.

In addition, shift tables relative to the normal reference ranges will be used to summarize the change from baseline to the most extreme post-dose value for each laboratory parameter. For each laboratory test, the patients will be classified into one of the four mutually exclusive groups (low, normal, high, and high + low), defined as follows:

- Low: at least one post-baseline value below the normal range and none above the normal range
- High: at least one post-baseline value above the normal range and none below the normal range
- Normal: all the post-baseline values within the normal range
- High + Low: at least one post-baseline value below the normal range and at least one above the normal range

Categorical parameters in the urinalysis panel will also be summarized with shift tables showing the shift from one categorical result to another.

Notable values

For selected laboratory tests, the number and percentage of patients with newly occurring or worsening laboratory abnormalities meeting the clinically notable criteria will be summarized

by laboratory parameter, post-baseline visit, time point and treatment. Data from unscheduled visits or from premature discontinuation visits will also be included.

For a patient to meet the criterion of a newly occurring clinically notable value, the patient needs to have a baseline value which is not clinically notable for that parameter. For a patient to meet the criterion of a worsening clinically notable value, the patient needs to have a baseline value which is clinically notable and also have a worse post-baseline value. For patients with missing value in baseline, any post-baseline notable value will be considered as newly occurring.

Guidelines for clinically notable criteria for laboratory tests are based on the FDA Guidelines for adults in SI units. For those parameters where ranges are available, the criteria for clinically notable results are presented in Appendix 16.1.9.

Listings of patients with notable laboratory values will be provided by laboratory parameter, treatment group, and patient number.

Liver and renal events

All data relating to liver and renal events will be listed.

1.6 Biomarkers

The analysis of biomarkers (serum WBC count, serum CRP and fibrinogen) will be exploratory. Mean, median, SD, minimum and maximum of the biomarkers at each visit (including change from baseline) will be displayed.

1.7 Interim analyses

Not applicable.

1.8 Other topics

Not applicable.

1.9 Determination of sample size

The primary objective is to demonstrate non-inferiority of QVA149 (110/50 µg o.d.) versus tiotropium (18 µg o.d.) + salmeterol/fluticasone propionate FDC (50/500 µg b.i.d.) with respect to post-dose trough FEV1 after 26 weeks of treatment.

For trough FEV1 it is assumed that the estimated treatment difference between QVA149 and tiotropium (18 µg o.d.)+salmeterol/fluticasone propionate FDC (50/500 µg b.i.d.) is 0 mL and the non-inferiority margin is assumed to be -50 mL. This non-inferiority margin is based on the treatment difference between ICS and placebo summarized in two Cochrane reviews (Nannini et al 2007 and Yang et al 2012) and the TIOSPIR spirometry sub-study (Wise et al. 2013). The selected trials were designed and conducted in COPD patients.

Based on the data pooled from two trials (Nannini et al 2007) investigating salmeterol/fluticasone versus placebo in terms of trough FEV1, the estimated treatment difference between salmeterol/fluticasone and placebo is 160 mL with a 95% confidence interval of 120 to 200 mL. Yang et al. provided summary data from 6 trials investigating ICS

doses of greater than 1000 g BDP equivalent/day versus placebo in terms of pre-dose FEV1, the estimated treatment difference between ICS and placebo is 80mL. A reasonable approach to establish the non-inferiority margin is to take one half of the lower bound of the confidence interval, which are -60 mL and -40mL respectively. The TIOSPIR spirometry sub-study in 1370 COPD patients used a non-inferiority margin of -50mL demonstrating the non-inferiority non-inferiority of the tiotropium Respimat 5 µg device versus the tiotropium HandiHaler 18 µg. The two. Two formulations of tiotropium are similar in clinical efficacy and safety.

A sample size of 375 evaluable patients in each treatment group provides 92% power for the testing of non-inferiority assuming a non-inferiority margin of -50mL and a SD of 200mL (2.5% significance level, one-sided). The standard deviation of 200mL is based on a review of the results of Phase III COPD studies run by the sponsor. Therefore assuming a dropout rate of 25% over 26 weeks a minimum sample size of 1000 patients (500 QVA149 (110/50 µg o.d.): 500 tiotropium (18 µg o.d.)+ salmeterol/fluticasone propionate FDC (50/500 µg b.i.d.) has been chosen.

nQuery (V7.0) has been used to calculate the sample sizes.

2 Clinical Study Report - Appendix 16.1.9 Documentation of statistical methods

This appendix gives details about statistical methods in addition to that provided in the main report text. SAS version 9.3 or higher will be used to perform all the statistical analyses in the report.

2.1 General data handling strategy

All categorical data will be summarized by frequencies and percentages. The frequencies and percentages will also be presented for missing observations.

Continuous data will be summarized with either standard descriptive statistics (i.e. the number of non-missing data points, arithmetic mean, standard deviation, minimum, 25% percentiles (Q25), median, 75% percentiles (Q75) and maximum), or will be collapsed into categorical data and be summarized as categorical data.

2.1.1 Output data precision

Decimal places to be used in displays of demographic, background and duration of exposure variables will be as follows:

- 3 decimal places (p-values; if p-value is less than 0.001, display < 0.001).
- 2 decimal places (standard errors and standard deviations).
- 1 decimal place (means, medians).
- 1 decimal place for min max
- 1 decimal place (percentages).
- If percentage = 100, no decimal is required.

Decimal places for efficacy and other safety summary tables and listings will be as follows:

- 4 decimal places (p-values; if p-value is less than 0.001, display < 0.001)
- data precision + 2 decimal places (standard errors and standard deviations)
- data precision + 1 decimal place (means, medians)
- same as data precision (minimums, maximums)
- 1 decimal place (percentages)
- If percentage = 100, no decimal is required.

In outputs containing estimates and p-values for inferential purposes, the means, SEs, SDs and confidence intervals will be output to 2 decimal places for all parameters. The p-values will be output to 3 decimal places (if a p-value is less than 0.001, it will be displayed as <0.001).

2.2 Compliance derivation

Compliance is presented as the percentage of days where study drug was administered as per protocol relative to the duration of exposure (i.e. the number of days between first and last dose). This will be derived using the following three eCRF summary pages:

- "*Dosage Administration Record – Summary – SDDPI (Concept 1)*"
- "*Dosage Administration Record – Summary – Accuhaler*"
- "*Dosage Administration Record – Summary – Handihaler*"

For a study day to be compliant there must be a summary page for SDDPI (Concept 1), Accuhaler and Handihaler with reason for dose recorded "AS PER PROTOCOL" on each summary page on that day.

It may occur that summary pages for a particular device overlap on dates (i.e. one summary page ends and the other starts on the same date or two kits are dispensed at a visit and it is unknown when one kit stopped and the other began). If this occurs then the rules below will be followed:

- If dates overlap between summary pages for the same device and the reason for dose is recorded "AS PER PROTOCOL" on both pages, then the day is considered "AS PER PROTOCOL" for that device and can be compared with other summary pages for different devices to decide if the study day is compliant overall.
- If dates overlap between summary pages for the same device and the reason for dose on at least one page is not recorded "AS PER PROTOCOL", then the device is not considered "AS PER PROTOCOL" and hence the study day is not considered compliant overall.

2.3 Vital signs and ECG – definition of clinically notable values

Table 2-1 Clinically 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 (mm	< 75	> 200

Hg)		
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 2-2 Clinical notable criteria for selected ECG parameters

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

2.4 Laboratory parameters – direction of interest and definition of clinically notable values

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 2-3 Directions of interest for worst case value for laboratory parameters

Laboratory parameter	Direction of interest for worst case value
Hematology	
Basophils	High
Eosinophils	High
Hematocrit	Low
Hemoglobin	Low
Lymphocytes	Low and High
Monocytes	High
Neutrophils	Low and High
Platelets	Low
RBC	Low

WBC total	Low and High
Clinical Chemistry	
Albumin	Low
Alkaline Phosphatase	High
ALT/SGPT	High
AST/SGOT	High
Bilirubin	High
Blood Urea Nitrogen (BUN)	High
Creatinine	High
Gamma GT	High
Glucose	Low and High
Sodium	Low and High
Potassium	Low and High
Total protein	Low and High

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

Table 2-4 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	
Thrombocytes (x10E^9/L)	75	700
WBC's (x10^9/L)	2.8	16.0
Chemistry		
Alkaline Phosphatase (U/L)	-	3xULN
Total Bilirubin (mcmol/L)	-	34.2
Creatinine (mcmol/L)		176.8
Potassium (mmol/L)	3	6
SGOT (U/L)	-	3xULN
SGPT (U/L)	-	3xULN
BUN/Serum Urea (mmol/L)		9.99
Sodium (mmol/L)	125	160
Gamma GT (U/L)		3xULN
Glucose (random) (mmol/L)	2.78	9.99
Protein(total) (g/L)	40	95
Albumin (g/L)	25	

v = volume, ULN = upper limit of normal

2.5 Protocol deviations and associated population codes

All protocol deviations defined at the start of the study are listed with associated population codes in Table 2-5. The list will be updated before CDBL.

Table 2-5 List of protocol deviations and associated population codes

Protocol Deviation ID	Protocol Deviation Description	Classification
INCL01	Patients have not signed Informed Consent Form prior to initiation of any study-related procedure	Exclude from everything
INCL02	Age criteria not met	Exclude from per-protocol
INCL03	Patient is not diagnosed of moderate to severe airflow obstruction with stable COPD according to 2014 GOLD guidelines	Exclude from per-protocol
INCL04	Patient not with post bronchodilator FEV1 greater than or equal to 40 percent and less than 80 percent of the predicted normal value or post bronchodilator FEV1/FVC less than 0.70 at Baseline Run-in.	Exclude from per-protocol
INCL05	The estimated amount consumed on average is not equal to or greater than 10 Pack years	Exclude from per-protocol
INCL06	Patients are not on triple treatment at least for the last 6 months (LAMA +LABA/ICS)	Exclude from per-protocol
EXCL01	Patients using other 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	Exclude from per-protocol
EXCL02	Patients contraindicated for treatment with, or having a history of reactions/ hypersensitivity to any of the following inhaled drugs, drugs of a similar class or any component thereof: - long and short anticholinergic agents - long and short acting beta-2 agonists - sympathomimetic amines - lactose or any of the other excipients of trial medication	Include in everything
EXCL03	Patient has history or current diagnosis of ECG abnormalities indicating significant risk of safety for patients participating in the study such as: - Concomitant clinically significant cardiac arrhythmias, e.g., sustained ventricular tachycardia, and clinically significant second or third degree AV block without a pacemaker - History of familial long QT syndrome or known family history of Torsades de Pointes	Include in everything
EXCL04	Patients with prolonged QTc interval (Frederica) is greater than or equal to 450 ms at Baseline Run-in Visit.	Include in everything
EXCL05	Concomitant use of agents known to significantly prolong the QT interval unless it can be permanently	Exclude from per-protocol

	discontinued for the duration of study	
EXCL06	Patients has a clinically significant laboratory abnormality at Visit 101 and would be at potential risk if enrolled into the study	Include in everything
EXCL07	Patients has clinically significant renal or cardiovascular or arrhythmia or neurological or endocrine, immunological or psychiatric, gastrointestinal or hepatic or hematological abnormalities which could interfere with the assessment of the efficacy and safety of the study treatment	Exclude from per-protocol
EXCL08	Patient has paroxysmal atrial fibrillation or un-controlled persistent atrial fibrillation	Include in everything
EXCL09	Patient has narrow angle glaucoma, symptomatic BPH/bladder neck obstruction, moderate/severe renal impairment or urinary retention	Include in everything
EXCL10	Reversibility test not performed to ATS/ERS standard	Exclude from per-protocol
EXCL11	COPD exacerbation occurred 12 months prior to study start that required treatment with Systemic corticosteroids, Systemic antibiotics or led to Hospitalization	Exclude from per-protocol
EXCL12	Patients developed COPD exacerbation between Screening and Randomization or re-screening is not as per protocol	Exclude from per-protocol
EXCL13	Patient has respiratory tract infection within 4 weeks prior to Screening (Visit 1)	Exclude from per-protocol
EXCL14	Respiratory tract infection occurred between Screening (Visit 1) and Randomization visit (Visit 201) or is re-screened within 4 weeks of resolution of Respiratory tract infection	Exclude from per-protocol
EXCL15	Long term oxygen therapy of > 12 h per day prescribed to patient	Exclude from per-protocol
EXCL16A	Patient has history of asthma	Exclude from per-protocol
EXCL16B	Patient has history of asthma	Exclude from per-protocol
EXCL17	Patients with a blood eosinophil count > 600/mm3 during screening (Visit 101)	Exclude from per-protocol
EXCL18	Patient has Allergic rhinitis and using either H1 antagonist or intranasal corticosteroid intermittently and is not on a stable dose or regimen	Include in everything
EXCL19	Patients with concomitant pulmonary disease (bronchiectasis and all non-solicited pulmonary diseases)	Exclude from per-protocol
EXCL20	Patient with diagnosis of alpha 1 anti-trypsin deficiency	Exclude from per-protocol
EXCL21	Patient with active tuberculosis	Exclude from per-protocol
EXCL22	Patient has undergone Pulmonary lobectomy or Lung volume reduction surgery or lung transplantation	Exclude from per-protocol
EXCL23	Patients plan to or are participating in the active phase of a supervised pulmonary rehabilitation programme	Exclude from per-protocol

EXCL24	Patient receiving prohibited medication	Exclude from per-protocol
EXCL25	Patients receiving any COPD related medications	Exclude from per-protocol
EXCL26	Patients receiving medications in the classes listed in Table 5-3 or not been stable on the medications for the specified period and stated conditions	Exclude from per-protocol
EXCL27	Patient unable to use an electronic patient diary	Exclude from per-protocol
EXCL28	Patient unable to use inhaler at Run-in and/or at Randomization or using Spacer device	Exclude from per-protocol
EXCL29	Patient has active cancer or has been cancer free for less than 5 years	Exclude from per-protocol
EXCL30	Subject is Pregnant or nursing (Lactating)	Exclude from per-protocol
EXCL31	Women of childbearing potential not using effective methods of contraception during dosing of study treatment	Include in everything
COMD01	Banned non-indication Conmed taken	Exclude from per-protocol
COMD02	Banned COPD Conmed taken	Exclude from per-protocol
OTH01	An Adverse Event for oral fungal infection is reported, however the presence of the infection was not confirmed by the central lab	Include in everything
OTH02	Non-compliance with SAE reporting timelines	Include in everything
OTH04A	Patient randomized in error	Exclude from per-protocol
OTH04B	Patient randomized in error	Exclude from per-protocol
OTH06	Patient randomized more than once into this study or also randomized into another study	Exclude from full analysis
TRT01	Patient administered at least one incorrect treatment	Exclude from per-protocol
TRT02	IMP administered instead of run-in medication	Exclude from per-protocol
TRT03	Evening dose from Accuhaler not administered at V207 (Visit 10) due to site instruction.	Exclude Visit 208 spirometry from per-protocol analysis if patient is in triple therapy arm
WITH01	Patient pregnant during study and not discontinued from study drug treatment	Include in everything
WITH02	Blind broken and study drug not permanently discontinued	Exclude from per-protocol
COMD03	Inactivated vaccination taken within 48 hrs prior to a study visit	Exclude from per-protocol
EXCL10B*	Reversibility test not performed to ATS/ERS standard	Exclude from per-protocol
INCL04B*	Patient not with post-bronchodilator FEV1 greater than or equal to 40 percent and less than 80 percent of the predicted normal value or post-bronchodilator FEV1/FVC less than 0.70 at Baseline Run-in.	Exclude from per-protocol

*protocol deviations EXCL10B and INCL04B will be created outside of clinical database during statistical analysis where deviations will be generated for:

- (EXCL10B) patients who do not have acceptable spirometry results, post-bronchodilator inhalation, at visit 101.

- (INCL04B) patients who do not have post-bronchodilator FEV1 greater than or equal to 40 percent and less than 80 percent of the predicted normal value or do not have post-bronchodilator FEV1/FVC less than 0.70 at visit 101.

2.6 COPD GOLD

2.6.1 COPD treatment: Gold 2011 Guidelines

Overall Severity

Gold 2011 (or 2014) guidelines grade patients (A-D) based on symptoms, airflow obstruction and exacerbation history as follows:

- A = Low risk, low symptom burden

  Low symptom burden (mMRCof 0-1 OR CAT score < 10) AND
  FEV1 of 50% or greater AND low exacerbation rate (0-1/year)
- B = Low risk, higher symptom burden

  Higher symptom burden (mMRC of 2 or more OR CAT of 10 or more) AND
  FEV1 of 50% or greater AND low exacerbation rate (0-1/year)
- C = High risk, low symptom burden

  Low symptom burden (mMRCof 0-1 OR CAT score < 10) AND
  FEV1 < 50% AND/OR high exacerbation rate (2 or more/year)
- D = High risk, higher symptom burden

  Higher symptom burden (mMRC of 2 or more OR CAT of 10 or more) AND
  FEV1 < 50% AND/OR high exacerbation rate (2 or more/year)

Airflow limitation:

Gold Grade	FEV1 (% predicted)
Gold 1 (Mild)	≥80
Gold 2 (Moderate)	50-79
Gold 3 (Severe)	30-49
Gold 4 (Very severe)	<30

2.6.2 COPD treatment: Gold 2017 Guidelines

Gold 2017 guidelines grade patients (A-D) based on symptoms and exacerbation history as follows:

- A= Low symptom burden (mMRCof 0-1 OR CAT score < 10) AND low exacerbation rate (0-1/year)
- B= Higher symptom burden (mMRC of 2 or more OR CAT of 10 or more) AND low exacerbation rate (0-1/year)

- C= Low symptom burden (mMRCof 0-1 OR CAT score < 10) AND high exacerbation rate (2 or more/year)
- D= Higher symptom burden (mMRC of 2 or more OR CAT of 10 or more) AND high exacerbation rate (2 or more/year)

2.7 Baseline and post-baseline definitions, missing data handling

2.7.1 Baseline measurements

1) For all FEV1 and FVC endpoints, the baseline value is defined as the average of the values taken at 45 and 15 minutes prior to first dose of double-blind drug at Day 1. Checks will be performed to ensure both values were indeed taken prior to the first dose of double-blind drug. If one of the values from 45 and 15 minutes prior to dose 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 pre-bronchodilator measurements taken at the run-in visit-day -28 can be used as the baseline. If the FEV1 or FVC measurements are missing both on Day 1 and at run-in visit, the respective baseline values will be set to missing. Measurements taken within 6 hours of rescue use or within 7 days of systemic corticosteroid use will be set to missing.
2) FEV1 reversibility is calculated as percentage increase of FEV1 values after sequential inhalation of 4 x 21 µg of ipratropium bromide and 4 x 100 µg puffs of salbutamol relative to FEV1 values prior to the inhalation. Patients are defined as "Not reversible" if post- bronchodilator FEV1 change from pre-bronchodilator value is < 12% or < 200 mL, and defined as "Reversible" if the increase of FEV1 is ≥ 12% and ≥ 200 mL.
3) For patient diary data (symptoms and rescue therapy use), the baseline values are defined as the average from all records taken in the run-in period, which is usually a 28-day period between run-in visit 101 and up to and including the morning assessments at Day 1. All days with data will be used for calculating the baseline value without limitation to the 28 days immediately prior to Day 1. For the baseline daily individual symptom score, the worst of the morning and evening assessments for the particular symptom on each of these days will be averaged (with denominator equal to the number of days featuring in the calculation). If either the morning or evening score is missing for any symptom then the non-missing value will be taken as the worst and used for the derivation of the baseline daily individual symptom scores. If both scores for an individual symptom are missing then the daily score for that individual symptom will be considered missing (i.e., no imputation performed).

   The baseline daily total symptom score will be obtained by adding the daily individual symptom scores and averaging over the number of days featuring in the calculation. Only the scores of the 6 COPD symptoms (i.e., respiratory symptoms, cough, wheeze, production of sputum, sputum color, and feeling of breathlessness) will be used to derive the total daily symptom score. For the derivation of the total daily symptom score, all 6 scores have to be available at that day. If one or more scores of the 6 COPD symptoms are missing, a total daily symptom score will not be calculated for that day.

For rescue medication, the baseline value is defined as the average number of puffs per day in the run-in period. If the number of puffs is missing for part of the day (either daytime or nighttime) then a half day will be used in the denominator. No imputation will be used for missing rescue therapy. Any values > 48 for the number of puffs of rescue medication in a 12 hour period will be set to missing. These high numbers are not realistic and could impact the analyses.

If a patient has less than 7 days with non-missing data, then the respective baseline value will be set to missing.

4) Baseline SGRQ-C, BDI and COPD assessment test (CAT) are defined as the assessments taken right before the first dose of the double-blind drug on Day 1.
5) Baseline vital signs (blood pressure, radial pulse) are defined as the assessments taken at > 45 minutes pre-dose on Day 1. Checks will be performed to ensure the assessments were indeed taken prior to the first dose of double-blind drug on Day 1. If this assessment is missing or not confirmed to be pre-dose, the last value taken at the run-in Visit 101 or at an unscheduled visit before the first administration of double-blind drug will be used for baseline.
6) Baseline height and weight are defined as the measurements taken at start of the run-in period (Visit 101). Missing baseline will not be imputed.
7) Baseline ECG is defined as the mean of two consecutive ECG (at 1 minute interval) readings taken prior to first dose of double-blind drug on Day 1. Checks will be performed to ensure the ECG was indeed assessed prior to the first dose of double-blind drug. If one of the two values is missing (or not confirmed to be pre-dose) then the remaining non-missing value will be taken as the baseline. If both values are missing (or not confirmed to be pre-dose), then the last value taken at the run-in Visit 101 or at an unscheduled visit before the first administration of double-blind drug will be used for baseline. Otherwise, the ECG baseline will be set missing without imputation. The worst interpretation and all findings associated with abnormal interpretation(s) of the two consecutive ECGs will be used as baseline.
8) Baseline laboratory data are defined as the assessment taken prior to first dose of double-blind drug on Day 1. Checks will be performed to ensure baseline laboratory values were indeed assessed pre-dose. If the values are missing (or not confirmed to be pre-dose) then the last value taken at the run- in Visit 101 or at an unscheduled visit before the first administration of double-blind drug will be used for baseline.
9) ICS use at screening will be derived from the concomitant medication eCRF data by selecting medications with route of administration = "Respiratory (inhalation)" and pre- specified medication subcategory = "Corticosteroid" or fixed combination type = "Inhaled corticosteroid / long-acting beta 2 agonist (ICS / LABA)" or "Inhaled corticosteroid / long- acting anticholinergic (ICS / LAMA)". Since ICS is a prohibited medication during the run-in period, the status at screening will be analyzed. If at least one of the above medications is taken at the screening, i.e. taken at the day of the screening visit (i.e. medications starting prior to visit 1 and ending at or after visit 1) , then ICS use at screening will be set to "yes", otherwise to "no". Note that this derivation of ICS use at screening will be used in all tables and statistical models rather than the stratified randomization data from the Interactive Response Technology

(IRT) vendor. The same is true for the other two stratified randomization variables smoking status (ex-smoker, current smoker) and airflow limitation severity (moderate, severe or worse).

10) LABA use at screening will be derived from the concomitant medication eCRF data by selecting medications with pre-specified medication subcategory = "Long-acting beta 2 agonist" or fixed combination type = "Inhaled corticosteroid / long-acting beta 2 agonist (ICS / LABA)" or "Long-acting beta 2 agonist / long-acting anticholinergic (LABA / LAMA)". Since LABA is a prohibited medication during the run-in period (same as ICS), the status at screening will be analyzed. If at least one of the above medications is taken at screening , i.e. taken at the day of the screening visit (i.e. medications starting prior to visit 1 and ending at or after visit 1)), then LABA use at screening will be set to "yes", otherwise to "no".
11) LAMA use at screening will be derived from the concomitant medication eCRF data by selecting medications with pre-specified medication subcategory = "Long-acting anticholinergic" or fixed combination type = "Inhaled corticosteroid / long-acting anticholinergic (ICS / LAMA)" or "Long-acting beta 2 agonist / long-acting anticholinergic (LABA / LAMA)". Since all patients will be given tiotropium during the run-in period, the status at screening will be analyzed. If at least one of the above medications is taken at screening, i.e. taken at the day of the screening visit (i.e. medications starting prior to visit 1 and ending at or after visit 1), then LAMA use at screening will be set to "yes", otherwise to "no".
12) ICS/LABA use at screening will be derived from the concomitant medication eCRF data by selecting medications with fixed combination type = "Inhaled corticosteroid / long-acting beta 2 agonist (ICS / LABA)". Since ICS/LABA is a prohibited medication during the run-in period, the status at screening will be analyzed. If at least one medication is taken at screening, i.e. taken at the day of the screening visit (i.e. medications starting prior to visit 1 and ending at or after visit 1), then ICS/LABA use at screening will be set to "yes", otherwise to "no".
13) 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.

2.7.2 Post-baseline measurements

1) FEV1 and FVC measurements taken within 6 hours of rescue use or within 7 days of systemic corticosteroid use will be set to missing.

   Post dose FEV1 and FVC values, where no morning dose was taken at the corresponding visit will be set to missing. If pre-dose measurements are performed one day after treatment end date, they will be set to missing if the last dose was not an evening dose.

   Scheduled pre-dose values which are performed post-dose and scheduled post-dose values which are performed prior to morning dose or after evening dose (serial spirometry set only) will be set to missing.

   The pre-dose trough value is defined as the average of values measured 45 and 15 minutes prior to the morning dose. If one of the two values is missing (or is not confirmed to be pre- dose) then the remaining non-missing value will be used as average pre-dose value.

2) The SGRQ-C contains 40 items divided into two parts covering three aspects of health related to COPD: Part I covers "Symptoms" and is concerned with respiratory symptoms, their frequency and severity; Part II covers aspects of "Activity" and "Impacts", "Activity" is concerned with activities that cause or are limited by breathlessness and "Impacts" covers a range of aspects concerned with social functioning and psychological disturbances resulting from airways disease. A score will be calculated for each of these three subscales 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 health status.

   The algorithm below is based on SGRQ-C Manual Version No. 1.2, April 2012, updated on 28 September 2012.

   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 and 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.

   Sum of maximum possible weights that could be obtained for the worst possible state of the patient for each component and Total:

Symptoms	566.2
Activity	982.9
Impacts	1652.8
Total	3201.9

Symptoms component is calculated from the summed weights for the positive responses to Questions 1-7 (Part 1). The Symptoms component tolerates a maximum of 1 missed item.

Activity component is calculated from the summed weights for the positive responses to Questions 9 and 12 (13 "true/false" questions, Part 2). The Activity component tolerates a maximum of 3 missed items.

Impacts component is calculated from the summed weights for the positive responses to Questions 8, 10, 11, 13, 14 (2 "select" questions and 18 "true/false" questions, Part 2). The Impacts component tolerates a maximum of 5 missed items.

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. The Total score requires that each of the three component scores has a tolerable number of missed items.

Scores for SGRQ-C, calculated as described above, need a small arithmetic adjustment to make them directly comparable to those obtained with the SGRQ.

The adjustment is:

Symptoms: SGRQ score = (SGRQ-C x 0.99) + 0.94 units

Activity: SGRQ score = (SGRQ-C x 0.87) +7.01 units

Impacts: SGRQ score = (SGRQ-C x 0.88) +2.18 units

Total: SGRQ score = (SGRQ-C x 0.90) + 3.10 units

3) CAT consists of eight items, each presented as a semantic 6-point differential scale (with scores from 0 to 5), providing a total score from 0 to 40 indicating the impact of the disease. It is completed by the patient. The total score will be derived as the sum of scores on individual items. If more than two responses are missing, a total score will not be calculated. When one or two items are missing their scores will be imputed by the average of the non-missing item scores for calculation of the total score (Jones et al, 2009). Scores of 0 - 10, 11 - 20, 21 - 30 and 31 - 40 represent mild, moderate, severe or very severe clinical impact.
4) Rescue medication use and other symptoms will be collected in the patient diary in the morning and evening at each day by assessing the past 12 hours. The post-baseline treatment period is defined as the period from the evening record of Day 1 up to the morning assessment following the day of the last inhalation of double-blind treatment. Missing diary data will not be imputed. Post-baseline values will be calculated only if a patient has at least 30% of their treatment days and at least 20 days with evaluable diary data for that variable in the post- baseline period. Daily individual and total symptom scores as well as daily rescue medication use will be derived as described in the section of baseline measurements. The mean daily, daytime and nighttime number of puffs of rescue medication and individual and total symptom scores will be calculated for the whole post-baseline treatment period.
5) Safety measurements include ECG, vital signs, laboratory results, 24 hour urinary cortisol (in a subset of patients) and adverse events. All safety data will be included in the analysis regardless of rescue medication usage. Post-baseline treatment-emergent measurements comprise recordings up to the last dose of double-blind drug + 7 days for laboratory, ECG, vital signs, and non-serious AEs and up to the last dose of

double-blind drug + 30 days for SAEs and death. ECG, vital signs, laboratory and 24 hour urinary cortisol values which have complete date and time values are assigned to pre or post-dose assessment based on the actual date/time. However, values with missing date/time are assigned to their respective scheduled visit date and time given the visit number is non-missing.

6) Dyspnea will be measured at baseline using the baseline dyspnea index (BDI) and during the treatment period using the transition dyspnea index (TDI), which captures changes from baseline. The BDI and TDI each has 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; negative scores indicate deterioration. A TDI focal score of 1 is considered to be a clinically significant improvement from baseline.

2.8 Statistical Analysis Outputs

List of statistical analysis outputs were outlined in RAP Module 7 – Section 16.1.9.

2.9 SAS code

Mixed Model for Repeated Measures (MMRM)

The mean change from baseline in post-dose trough FEV1 at Day 182 will be analyzed using a Mixed-Effect Model Repeated Measures (MMRM). The model includes the following factors:

Change from baseline = intercept + treatment + visit + region + baseline FEV1 + treatment*visit + baseline FEV1*visit + error.

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

```
Proc Mixed data=… order=internal;
where avisitn in (201 202 203 204 205 206 207 208);
class trt georgn avisitn usubjid;
model chg = trt base georgn avisitn trt*avisitn base*avisitn / DDFM=kr;
repeated avisitn / Subject= usubjid type=un rcorr;
lsmeans trt*avisitn / cl diff;
Run;
```

where

chg = change from baseline

trt = Treatment (QVA149, Tio+Salm/Flut)

base = baseline value

georgn = region (Africa, Asia, Eastern Europe, Western Europe, Latin and South

America, North America)

avisitn = visit (include the scheduled visits only, no unscheduled or premature

discontinuation visits)

usubjid = unique subject identifier

Appropriate baseline values and visits will be included for each analysis variable.

The default estimation method for the covariance parameters will be used, which is the residual (restricted) maximum likelihood.

Results will be presented with least squares means and standard errors (SE) for treatment effects and least squares mean, SE, associated two-sided 95% confidence interval, and one-sided p-value for treatment contrast QVA149 – Tio+Salm/Flut at week 26. Regarding treatment difference, the two-tailed p-value for the Wald test should be divided by 2 to get the approximate p-value for a one-tailed test. Degrees of freedom will be estimated based on Kenwards Rogers approximation (DDFM=kr).

If the analysis fails to converge with an unstructured covariance matrix (type=un), either a compound symmetry (first choice) or first order autoregressive (AR1) (second choice) covariance structure will be used.

ANCOVA analysis using Multiple Imputation (MI)

The mean change from baseline in post-dose trough FEV1 at Day 182 will be analyzed using ANCOVA with Multiple Imputation of the missing data. The model includes the following factors:

Change from baseline = intercept + treatment + region + baseline FEV1 + error.

The SAS PROC MI procedure carries out multiple imputations creating several "complete" versions of the dataset. A key indicator of the efficiency of the imputation is the DF for posterior parameter estimates. If the DF for a variable is less than 100, there is excessive variability among imputations for that variable, and more imputations are needed to improve the efficiency of imputation.

A multiple imputation will be performed based on missing at random (MAR) by treatment group for baseline and post-baseline values of trough FEV1 parameter for visits up to the primary time point (Week 26) using Markov Chain Monte Carlo (MCMC) method with EM algorithm.

Impute the missing values 100 times (NIMPUTE) with a seed=457<studycode> as shown below:

```
proc mi data=… out=imputed minmaxiter=10000000 nimpute=100 seed=4572316;
by trt;
var fev1_base fev1_week1-fev1_week26;
mcmc chain=multiple initial=em;
run;
```

Then PROC MIXED procedure will be used to analyze each imputed dataset as shown below:

```
proc mixed data=imputed;
ods output lsmeans=lsm;
```

```
where avisitn =208;
class trt georgn;
by _Imputation_;
model chg = trt base georgn;
lsmeans trt;
run;
quit;
```

where

chg = change from baseline

trt = Treatment (QVA149, Tio+Salm/Flut)

base = baseline value

georgn = region (Africa, Asia, Eastern Europe, Western Europe, Latin and South

America, North America).

At the last step PROC MIANALYZE procedure will combine data over imputations to produce the final results.

```
proc sort data=lsm;
by trt;
run;

proc mianalyze data=lsm;
ods output parameterestimates=ls_means;
by trt;
modeleffects estimate;
stderr stderr;
run;
```

Results will be presented with least squares means and standard errors (SE) for treatment effects and least squares mean, SE, associated two-sided 95% confidence interval, and one-sided p-value for treatment contrast QVA149 – Tio+Salm/Flut at Week 26. Regarding treatment difference, the two-tailed p-value for the Wald test should be divided by 2 to get the approximate p-value for a one-tailed test.

Generalized linear model assuming a negative binomial distribution

The rate of COPD exacerbations during the treatment period will be analyzed using a generalized linear model assuming a negative binomial distribution. The negative binomial distribution will often be used as an alternative to the Poisson distribution. It is especially useful for discrete data over an unbounded positive range whose sample variance exceeds the sample mean. If a Poisson distribution is used to model such data, the model mean and variance are equal. In that case, the observations are overdispersed with respect to the Poisson model. Since the negative binomial distribution has one more parameter than the Poisson, the second parameter can be used to adjust the variance independently of the mean.

The time at risk for a patient is defined as the exposure time and the log(exposure time in years) will be used as the offset variable in the model.

The explanatory variables are: treatment, region, COPD exacerbation history (the number of moderate or severe COPD exacerbations in the year prior to screening).

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

```
proc genmod data=.... order=internal;
class trt georgn;
model aval = trt georgn exprevyn /
dist=nb link=log offset=logrisk lrci type3 wald;
lsmeans trt / cl diff exp;
run;
```

where aval = number of COPD exacerbations

trt = treatment (QVA149, Tio+Salm/Flut)

georgn = region (Africa, Asia, Eastern Europe, Western Europe, Latin and South

America, North America)

exprevyn = number of moderate or severe COPD exacerbations during 12 months prior to study

logrisk = log (exposure time in years).

It should be noted that the regression model using GEE requires that in a particular level of a class factor at least one event in at least one patient must have been occurred to give an estimate for that level of a factor, otherwise the model does not converge.

Results will be presented with QVA149/ (Tio+Salm/Flut) ratio of rates and associated 95% confidence interval and two-sided p-value.

Cox regression analysis

Time to first COPD exacerbation will be analyzed with a Cox proportional hazards regression model. The null-hypothesis will be $H_0: \lambda_{QVA}(t) / \lambda_{SFC}(t) = 1$, where $\lambda(t)$ is the hazard function for the failure time of patients treated with QVA149 and Tio+Salm/Flut, respectively. The SAS procedure PHREG will be used with the following SAS code:

```
proc phreg data=.... ;
class trt(order=internal) georgn;
model aval*cnsr(1) = trt georgn exprevyn / rl ties=exact;
hazardratio trt;
contrast "QVA149 vs. Tio+Salm/Flut" trt 1 / estimate=exp;
run;
```

where aval = time to first COPD exacerbation

cnsr = data are censored if cnsr = 1

trt = treatment (QVA149, Tio+Salm/Flut)

georgn = region (Africa, Asia, Eastern Europe, Western Europe, Latin and South

exprevyn = number of COPD exacerbations during 12 months prior to study.

Results will be presented with the hazard ratio for QVA149/ (Tio+Salm/Flut) and associated 95% confidence interval and two-sided p-value. P-value will be obtained from the Wald chi-squared statistic testing the null-hypothesis that the parameter estimate for the respective treatment effect is 0 (then the hazard ratio is exp(0) = 1). No check for the validity of proportional hazards assumptions will be done.

Linear mixed model

A linear mixed ANCOVA model will be used for changes from baseline in daily number of puffs of rescue medication use averaged over the 26 weeks of treatment and other related variables derived from the patient diary.

Change from baseline = intercept + treatment + baseline + 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 trt georgn siteid;
model chg = trt base georgn / ddfm=kr;
random siteid(region)/ type=vc;
lsmeans trt / cl diff;
```

where chg = change from baseline

trt = treatment (QVA149, Tio+Salm/Flut)

base = baseline value

georgn = region (Africa, Asia, Eastern Europe, Western Europe, Latin and South

America, North America)

siteid = study site identifier.

Results will be presented with least squares mean and SE for treatment effects and least squares mean, SE, associated two-sided 95% confidence interval and two-sided p-value for the treatment contrast QVA149 - Tio+Salm/Flut.

Mann-Whitney U test

A Mann-Whitney U test will be performed to explore the difference between treatment groups (QVA149 and Tio+Salm/Flut) in mMRC grade and change in grade at week 26 (Day 181). The SAS procedure PHREG will be used with the following SAS code:

```
proc npar1way data=… Wilcoxon median;
class trt;
var mmrc;
run;
```

where mmrc = mMRC Dyspnea Scale (Grade 0, Grade 1, Grade 2, Grade 3, Grade 4)

trt = treatment (QVA149, Tio+Salm/Flut)

The WILCOXON option requests the Wilcoxon test for the difference between the groups, and the MEDIAN option requests the median test additionally. Results will be presented with mean scores and SE for treatment effects, and two-sided p-value for the treatment contrast QVA149 - Tio+Salm/Flut.