NOVARTIS

Clinical Development

QVM149B (Indacaterol acetate / Glycopyrronium bromide / Mometasone furoate)

CQVM149B2302 / NCT02571777

A multicenter, randomized, 52-week, double-blind, parallel-group, active controlled study to compare the efficacy and safety of QVM149 with QMF149 in patients with asthma

RAP Module 3 – Detailed Statistical Methodology

Author: [redacted], CRO Trial Statistician

Document type: RAP Documentation

Document status: Amendment 2

Release date: 28-Sep-2018

Number of pages: 47

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

Version	Date	Changes
Final 2.0	25-Nov-2015	
Amendment 1	30-May-2018	The amendment is largely based on the changes made in the Clinical Trial Protocol Amendment 6 (released 18-Dec-2017). More details about the rationale of the major changes are given in the section "Amendment 1". In summary, the following changes are implemented: • The primary analysis will be conducted after all patients have completed the Week 26 assessments or discontinued from the study; changes are made in multiple sections with details given in Section 2.9. • Sample size has been updated in Section 2.10 based on new drop-out rate. • Modified the handling of missing items when calculating the ACQ-7 score and clarified the handling of missing ACQ-7 score for ACQ-responder analysis in Section 2.6.1.1.2. • Added ACQ-5 in Section 2.6.2.2. • Modified the handling of missing items in the AQLQ overall and domain scores in Section 2.6.2.7. • Added exposure-adjusted analyses of adverse events in Section 2.7.2. • Changed the analysis plan for glucose and potassium data. • Removed the logistic regression analysis for percentage of patients with at least one asthma exacerbation by exacerbation category and patients who permanently discontinued study medication due to asthma exacerbations in Section 2.6.2.7. • Extended the post-morning dose time window restriction in handling of spirometry measurements from 10-13 hours to 8-13 hours for b.i.d. regimen salmeterol/fluticasone 50/500 µg in Section 2.6.1.1. • Modified the multiple imputation method in tipping point analysis from monotone method to fully conditional specification method in Section 2.6.1.4 and Section 16.1.9.6. • Combined the imputation steps into one step for the ACQ responder analysis in Section 2.6.2.2 and Section 16.1.9.6. • Removed SAS code from SAP in Section 16.1.9.6. and included them in TFLs. • Removed the definitions of AEs of special interest from Section 2.7.2. • Updated criteria for Hy's law as per current guideline. • Added Section 3 for specifying analyses that will be reported outside of CSR. • Added Japanese and Chinese patients in subgroup analysis in Section 3.1. [redacted] • Added the analysis of add-on effect of glycopyrronium to QMF149 treatment for trough FEV_1 in Section 2.6.2.1, ACQ in Section 2.6.2.2, and asthma exacerbations in Section 2.6.2.6.

Version	Date	Changes
		• Re-organized Section 2.6.2.6 to place the detailed analysis under each parameter of interest. • Added the AE summary based on incidence rates per 100 patient-years in Section 2.7.2. • Deleted several safety endpoints for time to event analyses in Section 2.7.2. • Added major protocol deviations list in Section 16.1.9.2. • Some other minor changes. • Added notable criterion for Magnesium.
Amendment 2	28-Sep-2018	• Updated action plan if model fails to converge in Section 2.6.1.2 and 16.1.9.6. • Added non-parametric test exploration in case of non-convergence or highly skewed data in Section 2.6.2.4 and 2.6.2.5. • Outlined the sensitivity analyses to be done with and without two sites which were identified to have lack of quality issues in Section 1. • Correction, removed "The responder analysis will use the same imputed datasets as that of ACQ-7", as for ACQ-5 same analyses is done as for ACQ-7 but not same imputed dataset as multiple imputation is for each score separately in Section 2.6.2.3. • Correction of PD-ID in Section 16.1.9.2. • Changed notable criteria for Magnesium to SI unit for consistency in Section 16.1.9.5. • Adjusted categories for adjudicated CCV events in Section 2.7.2. • Added the details of least squares mean computation in Section 16.1.9.6. • Added additional analyses on potassium excluding all flagged data in Section 2.7.3. • Added the definition of implausible value for FEV_1 in Section 2.6 and 16.1.9.3.

Table of contents

Amendment 1

This amendment of the statistical analysis plan (SAP) is largely based on the changes made in the Clinical Trial Protocol Amendment 6 (released 18-Dec-2017). The rationale of the major changes in protocol is provided in the following.

Amendment Rationale

Conduct primary analysis after all patients have completed at least 26 weeks treatment:

An interim database freeze will be conducted to perform the primary analysis. The database for primary analysis will contain all data up to the Data Lock Point (DLP), which is defined as the date when all patients complete Week 26 (V207) assessments or withdraw from the study. For patients who are ongoing at the time of interim database freeze, the database will include all data available until the last completed visit.

Primary analysis includes primary and key secondary endpoints as well as other pre-specified endpoints at Week 26. [redacted]
Safety data will also be summarized.

In terms of reporting, two separate Clinical Study Reports (CSRs) will be written. CSR 1 will base on the interim database freeze and CSR 2 on the final database lock. A dedicated unblinded team will be involved in CSR 1 related activities. In order to maintain the integrity of the study data, a separate blinded team will continue the study until its completion.

1 Introduction

This document contains details of the statistical methods that will be used in the phase III clinical trial CQVM149B2302. This study is designed to evaluate the efficacy and safety of two different doses of QVM149 (QVM149 150/50/80 µg and QVM149 150/50/160 µg via Concept1) over two respective QMF149 doses (QMF149 150/160 µg and QMF149 150/320 µg via Concept1) in poorly controlled asthmatic patients as determined by pulmonary function testing and effects on asthma control.

Data will be analyzed according to Section 9 of the study protocol. The primary analysis will be conducted using all data up to the Data Lock Point (DLP), which is defined as the date when all patients have completed the assessments after at least 26 weeks of treatment (Visit 207) or prematurely withdrawn from the study. In particular for efficacy analysis, each patient's data until the last available study visit (scheduled or unscheduled) prior to the DLP will be included. For safety analysis, data up the DLP will be included too. This will include all planned analyses presented in this SAP.

Two separate CSRs will be published:

- Primary Analysis CSR: To support analyses once all patients have completed the assessments after at least 26 weeks of treatment (Visit 207) or prematurely withdrawn from the study.
- Full CSR: To support analyses once all patients have completed 52 weeks of treatment (Visit 214) plus follow-up (Visit 301) or prematurely withdrawn from the study.

This SAP is valid for analyses of both CSRs.

Important information is given in the following sections and details are provided, as applicable, in Appendix 16.1.9.

Subgroup analyses of primary, key and other secondary, [redacted] objectives will be performed for specific countries (e.g. Japan). These analyses will be identified in the tables, figures, and listings shell document.

In consideration of the lack of quality control applied by the PI on study source data, a sensitivity analysis will be performed by assessing for primary, key secondary endpoints and AEs/SAEs both with and without patient data from the following two sites in order to evaluate whether the data from these sites have an impact on overall results:

1271 (India), 1760 (Switzerland).

2 Detailed statistical methodology

2.1 Study objectives

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

2.1.1 Primary objective

The primary objective of the study is to demonstrate superiority of either QVM149 150/50/80 µg o.d. to QMF149 150/160 µg o.d. or QVM149 150/50/160 µg o.d. to QMF149 150/320 µg

o.d., all delivered via Concept1, in terms of trough FEV_1 after 26 weeks of treatment in patients with asthma.

2.1.2 Key secondary objective

The key secondary objective is to demonstrate the superiority of either QVM149 150/50/80 µg o.d. to QMF149 150/160 µg o.d. or QVM149 150/50/160 µg o.d. to QMF149 150/320 µg o.d., all delivered via Concept1, in terms of asthma control as assessed by the Asthma Control Questionnaire (ACQ-7) after 26 weeks of treatment in patients with asthma.

2.1.3 Secondary objectives

Secondary objectives are to compare:
1) QVM149 150/50/80 µg o.d. to QMF149 150/160 µg o.d. both delivered via Concept1,
2) QVM149 150/50/160 µg o.d. to QMF149 150/320 µg o.d. both delivered via Concept1,
3) QVM149 150/50/80 µg o.d. delivered via Concept1 to salmeterol xinafoate/fluticasone propionate 50/500 µg b.i.d. via Accuhaler®,
4) QVM149 150/50/160 µg o.d. delivered via Concept1 to salmeterol xinafoate/fluticasone propionate 50/500 µg b.i.d. via Accuhaler®,

in terms of:

- Trough FEV_1 at Week 52
- Pre-dose FEV_1 (defined as the mean of -45 min and -15min FEV_1 values pre-evening dose) at Week 4 and Week 12
- FEV_1, Forced Vital Capacity (FVC) and Forced Expiratory Flow between 25% and 75% of FVC (FEF_{25-75}) over 52 weeks
- Morning and Evening Peak Expiratory Flow Rate (PEF) over 26 and 52 weeks of treatment
- Asthma control as assessed by the ACQ-7 at Weeks 4, 12 and 52
- Percentage of patients achieving the minimal important difference (MID) in ACQ-7 of ≥ 0.5 at Week 26 and Week 52
- Percentage of days with no symptoms, the percentage of nights with no night-time awakenings, and the percentage of mornings with no symptoms on rising over 52 weeks of treatment, as recorded by e-diary
- Percentage of days without rescue medication usage (salbutamol/albuterol) as recorded by e-diary over 26 and 52 weeks of treatment
- To evaluate the efficacy of in terms of asthma exacerbation-related parameters described here further during 52 weeks of treatment. The analysis will be performed by exacerbation category wherever specified. The exacerbation categories are: mild, moderate, severe, and moderate or severe:
  - Time to first asthma exacerbation by exacerbation category
  - Time to first hospitalization for asthma exacerbation
  - Annual rate of asthma exacerbations excluding measurements in patients requiring chronic corticosteroid use after an exacerbation (beyond permitted steroid taper for exacerbation) by exacerbation category

  - Duration in days of asthma exacerbations by exacerbation category
  - Percentage of patients with at least one asthma exacerbation by exacerbation category
  - Time to permanent discontinuation of study medication due to asthma exacerbations
  - Percentage of patients who permanently discontinued study medication due to asthma exacerbations
  - Total amounts of systemic corticosteroids (in doses) used to treat asthma exacerbations
- Quality of life as assessed by Asthma Quality of Life Questionnaire (AQLQ) over 52 weeks
- Furthermore, as additional secondary comparisons, QVM149 150/50/80 µg o.d. and QVM149 150/50/160 µg o.d. delivered via Concept1 will be compared with salmeterol xinafoate /fluticasone propionate 50/500 µg via Accuhaler® (comparisons 3) and 4)) for the primary and key secondary endpoints:
  - Trough FEV_1 measured after 26 weeks of treatment
  - Asthma control as assessed by the ACQ-7 after 26 weeks treatment
- The following safety and tolerability endpoints will be evaluated for all treatment comparison groups:
  - Cumulative incidence of the composite endpoint of adjudicated serious asthma outcomes (i.e. asthma-related hospitalization, asthma-related intubation, or asthma-related death),adjudicated cardiovascular and cerebrovascular (CCV) events, and adjudicated atrial fibrillation/flutter over 52 weeks of treatment
  - Adverse events, vital signs, ECG and laboratory analysis (hematology, blood chemistry including glucose and potassium, urinalysis) over 52 weeks of treatment
  - Plasma evening cortisol over 52 weeks of treatment

2.1.5 Changes in the conduct of the study or planned analyses

The restriction on trough FEV_1 measurement (at Visit 201, 207, and 214) time that it has to fall into 10 - 13 hours post-morning dose time window has been extended to 8 - 13 hours and only for b.i.d. regimen salmeterol/fluticasone 50/500 μg. This is considered clinically justified based on the Seretide® dosing interval and expected duration of the bronchodilator effect. The original 10-13 hours post-morning dose time window is considered conservative based on Seretide® dosing regimen from the evidence in available literature showing that the bronchodilator effect starts to decrease after 8 hours (Pearlman et al. 1992).

2.2 Analysis sets

The following analysis sets are defined:

- The randomized (RAN) set will consist of all patients who were assigned a randomization number, regardless of whether or not they actually received study medication. Patients in RAN will be analyzed according to the treatment they were randomized to. The RAN set will be used for a summary of patient disposition, demographics and baseline characteristics.
- The Full Analysis Set (FAS) will consist of all patients in the RAN set who received at least one dose of study medication. Following the intent-to-treat principle, patients will be analyzed according to the treatment they were assigned to at randomization. The FAS will be used in the analysis of all efficacy variables.
- The Per-Protocol set (PPS) will include all patients in the FAS who did not have any major protocol deviations. Major protocol deviations will be defined in the validation analysis plan prior to database lock and the un-blinding of the study. The list of major protocol deviations

is available in Appendix 16.1.9. Patients will be analyzed according to the treatment they received. The PPS will be used for supportive analysis of the primary analysis only.

- The Safety Set will consist of all patients who received at least one dose of study medication including non-randomized patients who received study drug in error. Patients will be analyzed according to the treatment they received. If patients switch double-blind treatment during the study, they will be counted and analyzed only once according to their initial treatment. The safety set will be used in the analysis of all safety variables.

[redacted]

2.3 Patient disposition, background and demographic characteristics

2.3.1 Patient disposition and protocol deviations

The number of patients will be summarized by region (i.e., Eastern Europe, Western Europe, India, Japan, Asia (other than India and Japan), rest of world), country and treatment group for the RAN set. Further, for each study epoch (i.e., screening, run-in, treatment phase, post treatment follow-up), the overall number of patients who entered, completed, and discontinued that phase will be summarized including the reasons for discontinuation of that phase.

The number of patients who completed the 52-week study treatment and who discontinued prematurely will be shown including the reasons for discontinuation of study treatment. In the Primary Analysis CSR, the number of patients who are ongoing at the time of interim database freeze will be presented. Time to premature treatment discontinuation will be displayed graphically for each treatment group using a Kaplan-Meier curve for the safety set. The date of premature treatment discontinuation is defined as the date of last dose of study medication. Patients who completed the study treatment will be censored at the date of last dose of study medication. In the interim analysis, patients who are ongoing at the time of interim database freeze will be censored at the Data Lock Point (DLP), which is defined as the date when all patients have completed at least Week 26 assessments or prematurely withdrawn from the study.

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

The number of patients included in each analysis set will be tabulated, as well as the reasons for exclusions from analysis sets.

2.3.2 Demographics and baseline characteristics

Demographics and baseline characteristics will be summarized by treatment group using the RAN set. Summaries will include age, gender, race, ethnicity, height, weight, body mass index (BMI), relevant medical history, pre- and post-bronchodilator screening spirometry parameters: (FEV_1 in L and % of predicted FEV_1, FVC, FEV_1/FVC and FEF_{25-75}), FEV_1 reversibility (in L and %), duration of asthma, history of asthma exacerbations, prior asthma treatment, and smoking history.

Continuous variables will be summarized using descriptive statistics (number of non-missing data, mean, standard deviation, median, first and third quartiles, minimum, and maximum) and categorical variables will be summarized in terms of the number and percentage of patients in each category including a category for missing data if any.

Baseline is defined as the last available measurement before first dose of study drug.

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

In addition, the following categorizations will be done:

- Age into 18 - 64 years, and ≥ 65 years;
- BMI into ≤ 30.0 kg/m^2 and > 30.0 kg/m^2;
- Duration of asthma into < 1 year, 1 - 5 years, > 5 - 10 years, > 10 - 15 years, > 15 - 20 years, and > 20 years;
- Number of asthma exacerbations in the 12 months prior to the start of the study that required medical care into 1, 2, 3, ≥ 4
- pre-bronchodilator FEV_1 into < 40 %, 40 - < 60%, 60% - < 80% of predicted FEV_1
- ACQ-7 into 1.5- < 2, 2 – < 2.5, ≥ 2.5 (< 1.5 will be added in case of protocol deviations)
- prior asthma treatment into medium and high dose ICS/LABA

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 the RAN set by primary system organ class and preferred term. 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 Duration of exposure

Duration of exposure to study treatment will be calculated as the number of days starting from the first dose date up to and inclusive the last dose date (ignoring the duration when study treatment may have been interrupted). The duration of exposure 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 for each 1 - 29 days, 30 – 86 days, 87 – 183 days, 184 – 254 days, 255 – 365 days, and > 365 days.

In the interim analysis, patients who are ongoing at the time of interim database freeze will be considered as exposed until the DLP.

2.4.2 Compliance

Compliance will be calculated by counting the days where study drug was administered "As per protocol" according to the records on the Dosage Administration Record (DAR) Summary eCRF. The percentage of days divided by the days of exposure will be analyzed. To be counted as a day of compliance with study treatment, all three administrations, i.e., the Accuhaler both

in the morning and evening and the Concept 1 inhaler in the evening, have to be taken as per protocol. Compliance will be calculated by study treatment.

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

2.5 Prior and concomitant medication

Each medication, either an asthma or non-asthma medication, will have the start and end dates recorded on the eCRF. Separate tables will be provided for medications which were started and stopped prior to the first dose of study drug and medications which were taken concomitantly to the study drug (regardless of whether continued or started after the first dose of study drug).

Asthma medications will be summarized by the route of administration, the recorded pre-specified drug subcategories (including types of combination) and the coded preferred terms. The summary will be repeated by showing ingredients instead of preferred terms.

Non-asthma medications will be summarized by route of administration and preferred term.

Surgical and medical procedures (non-drug therapies) will be coded using MedDRA and presentations will be done by MedDRA primary system organ class and preferred term, separately for prior procedures and those after start of study drug.

Patients taking prohibited concomitant medications will be noted in the summary of protocol deviations.

All summarizations will be by treatment group in the Safety Set.

2.6 Efficacy evaluation

All efficacy analyses will be performed on the FAS and selected analyses will be repeated on the PPS. Measurements made after patients discontinue randomized treatment (off-treatment measures) will not be used for any efficacy evaluation.

2.6.1 Analysis of the primary and key secondary variables

The primary objective is to demonstrate the superiority of either QVM149 150/50/80 µg o.d. to QMF149 150/160 µg o.d. or QVM149 150/50/160 µg o.d. to QMF149 150/320 µg o.d. in terms of trough FEV_1 after 26 weeks of treatment. The key secondary objective is to demonstrate the same in terms of ACQ-7 after 26 weeks of treatment.

2.6.1.1 Variables

2.6.1.1.1 Primary variable trough FEV_1

Trough FEV_1 is defined as the average of the two FEV_1 measurements taken 23 hr 15 min and 23 hr 45 min post-evening dose. Trough measurements will be done at Day 2, Day 184 (Week 26, the primary endpoint) and Day 365.

If any of the 23 hr 15 min and 23 hr 45 min values contributing to the trough FEV_1 are collected within 3 months of a single depot corticosteroid injection, or within 7 days of systemic corticosteroid use (except for patients who were on stable systemic corticosteroids as

background therapy), or within 6 hours of rescue medication use, or if actual measurement times are outside of the 22 - 25 hour post-evening dose time window for o.d. regimens (or 8 – 13 hour post-morning dose time window for b.i.d. regimen salmeterol/fluticasone 50/500 μg), then the individual FEV_1 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 FEV_1. If both values are missing, then trough FEV_1 will be missing.

2.6.1.1.2 Key secondary variable ACQ-7 score

The ACQ-7 measures asthma symptom control and consists of 7 items: 5 on symptom assessment, 1 on rescue bronchodilator use and 1 on airway calibre (FEV_1 % predicted). Patient recall is considered 1 week. All 7 questions of the ACQ-7 are equally weighted. Items 1-5 are scored along a 7-point response scale, where 0 = good controlled and 6 = poorly controlled. Item 6 is scored between 0 = no rescue medication and 6 = More than 16 puffs/inhalations most days. The 7th item will be scored by the investigator based on the FEV_1 % predicted from the masterscope at the site (i.e., Score = 0 means > 95% of predicted FEV_1, 1 = 90 – 95%, 2 = 80 – 89%, 3 = 70 – 79%, 4 = 60 – 69%, 5 = 50 – 59%, and Score = 6 means < 50% of predicted FEV_1). The total score is calculated as the mean of all questions. The mean will be calculated as the sum of scores divided by the number of questions that were answered, as long as there were at least 6 questions answered and the missing question is neither Question 1 ('On the average, during the past week, how often were you woken by your asthma during the night?') nor Question 7. In case of one missing response (not Question 1 or Question 7), that missing response will be imputed as per the method given in the following paragraph and then the ACQ score will be calculated.

For a missing individual item while on-treatment, the interpolation method using either previous or subsequent completions of the questionnaire (ACQ User Guide, 2004) will be used. For instance,

Item	Visit 1	Visit 2
1	4	6
2	3	5
3	0	0
4	4	4
5	0	0
6	2	Missing
7	5	6

Total Visit 1 score for items answered on both visits is 4+3+0+4+0+5 = 16 (A).

Total Visit 2 score for items answered on both visits is 6+5+0+4+0+6 = 21 (B).

Item 6 score at Visit 1 = 2.

Item 6 score at Visit 2 = B/A * 2 = 21/16*2 = 2.63.

The ACQ score for Visit 1 is (4+3+0+4+0+2+5)/7 = 2.57

The ACQ score for Visit 2 is (6+5+0+4+0+2.63+6)/7 = 3.38

For baseline, missing item at Day 1 will be imputed by using Day -14 data or data from any unscheduled run-in visit, whichever is later. For post-baseline visits, missing items will be

imputed by using data from subsequent visit first, if these are not available, data from previous visit will be used for imputation of the missing item. To clarify, for imputation of the first post-baseline value, if the second post-baseline assessment is not available, the value of the baseline assessment will be used. If there are two assessments from the same day, the first will be used only and the second ignored even for imputation.

ACQ-7 measurements are scheduled for Days -14, 1, 30, 86, 183, and 364.

2.6.1.2 Statistical model, hypothesis, and method of analysis

The two comparisons of QVM149 150/50/80 µg o.d. versus QMF149 150/160 µg o.d. and QVM149 150/50/160 µg o.d. versus QMF149 150/320 µg o.d. will be evaluated by testing the following null hypotheses (H_0) versus the alternative hypotheses (H_a) for both the primary variable trough FEV_1 after 26 weeks of treatment and key secondary variable ACQ-7 after 26 weeks of treatment:

H_{01}: QVM149 150/50/80 µg is equal to QMF149 150/160 µg

versus

H_{a1}: QVM149 150/50/80 µg is not equal to QMF149 150/160 µg

and

H_{02}: QVM149 150/50/160 µg is equal to QMF149 150/320 µg

versus

H_{a2}: QVM149 150/50/160 µg is not equal to QMF149 150/320 µg

The primary variable, trough FEV_1 at Week 26, will be analyzed using a mixed model for repeated measures (MMRM) on the FAS. The model will contain treatment, region (Eastern Europe, Western Europe, India, Japan, Asia (other than India and Japan), rest of world), visit (Days 2, 184, and 365), and treatment-by-visit interaction as fixed effects with baseline FEV_1 measurement, baseline-by-visit interaction, FEV_1 prior to inhalation and FEV_1 within 15 to 30 min post inhalation of salbutamol/albuterol (components of short acting beta-2 agonist (SABA) reversibility) as covariates, and center nested within region as a random effect. The within-patient correlation will be modeled using an unstructured covariance matrix in the mixed model. The Kenward-Roger approximation will be used to estimate denominator degrees of freedom (Kenward and Roger, 1997).

For the primary analysis, if the model does not converge, data only up to week 26 (i.e., Days 2 and 184) will be used (with unstructured covariance matrix). If still the model fails to converge, compound symmetry covariance matrix will be used. If model still does not converge, the random effects of center will be removed. For the final analysis, if the model does not converge with unstructured covariance matrix, the compound symmetry covariance matrix will be used in the mixed model. If still the model does not converge, the random effects of center will be removed. For both analyses, if the model converges after removing the random center effects, the original model without random center effects will be re-assessed. If the final model for the primary analysis converges without Week 52 data, only summary statistics will be provided for Week 52 data analysis. Restricted maximum likelihood method will be used.

The MMRM is based on the missing at random (MAR) assumption where missingness is independent of the unobserved data provided that relevant observed data are accounted for in the statistical model.

Baseline FEV_1 is defined as the mean of the 45 min and 15 min pre-evening dose FEV_1 measurements at Day 1 (V201). If one of the two measurements is missing, the non-missing measurement will be taken as the baseline FEV_1. If both are missing, the pre-bronchodilator reversibility measurement (at V101 or repeated V101) will be used as baseline FEV_1.

Each between-treatment comparison will be carried out using the adjusted mean (least-squares mean) difference based on the treatment main effect and the coefficient for the treatment-by-visit interaction factor corresponding to Day 184. The estimated adjusted treatment difference (QVM149 – QMF149) will be displayed along with the associated standard error (SE), 2-sided 95% confidence interval (CI), and p-value (2-sided).

In addition, to estimate the add-on effect of glycopyrronium to QMF149, the average of QVM149 doses (150/50/80 µg and 150/50/160 µg) vs. the average of QMF149 doses (150/160 µg and 150/320 µg) will be computed in terms of FEV_1. QVM149 and QMF149 doses will be pooled using appropriate contrasts within the MMRM specified above.

The key secondary variable, ACQ-7 after 26 weeks of treatment, will be analyzed using the same MMRM (including all scheduled visits with ACQ-7 data) on the FAS as used for the primary analysis but will include baseline ACQ-7 score instead of baseline FEV_1.

2.6.1.3 Multiplicity adjustment

To control the family-wise type-I error rate at the two-sided 5% significance level, a testing procedure based on the generalized Simes test as described in Maurer et al (2011) will be used. The family for the overall type-I error rate control contains four hypotheses: two hypotheses for the primary endpoint trough FEV_1 (H1 and H2 for comparing QVM149 150/50/80 µg vs. QMF149 150/160 µg and QVM149 150/50/160 µg vs. QMF149 150/320 µg) and two hypotheses for the key secondary endpoint ACQ-7 (H3 and H4 for the same treatment comparisons as H1 and H2).

A description of the testing procedure is provided in Appendix 16.1.9.

Other than the 4 treatment comparisons mentioned above for the primary and the key secondary endpoint, all other analyses will be performed at the nominal 2-sided 0.05 level (2-sided) without multiplicity adjustment.

2.6.1.4 Supportive analyses for the primary objective

As supportive analysis, the same MMRM used in the primary analysis will be also performed on the PPS to assess the treatment effect in protocol adherers.

In addition, the same primary MMRM on the FAS will be performed including all spirometric measures irrespective of single depot corticosteroid injection, other systemic corticosteroid use or rescue medication use. However, measures taken outside of the 22 – 25 hour post-evening dose window for o.d. regimens (or 8 - 13 hour post-morning dose window for salmeterol/fluticasone 50/500 µg b.i.d. regimen) will be excluded.

As a further sensitivity analysis, a so-called 'tipping point analysis' will be performed for the primary endpoint to evaluate the impact of a deviation of the MAR assumption. The delta-adjusting approach described in Ratitch et al (2013) will be used to find the tipping point, in a spectrum of conservative missing not at random (MNAR) assumptions, at which conclusions change from being favorable to QVM149 to being unfavorable. After such a tipping point is determined, clinical judgment can be applied as to the plausibility of the assumptions underlying this tipping point. This methodology would provide a good picture of what it would take to overturn study conclusions on the basis of varying assumptions about missing data. The basic idea is to first impute the missing Week 26 values using the multiple imputation (MI) method based on the missing at random (MAR) assumption. The imputed Week 26 values for the QVM149 arms are then adjusted by a delta value. A detailed description of the analysis steps is provided in Appendix 16.1.9.

The following exploratory subgroup analyses for trough FEV_1 at Week 26 using MMRM will be performed (using the appropriate interaction term in the model and additional covariate as a fixed effect if necessary) for the FAS to explore the treatment effect in:

- Race (Caucasian, Asian, Black and other)
- Sex (male, female)
- History of asthma exacerbation in the 12 months prior to screening (1, 2, 3, ≥ 4)
- Patients' prior therapies before run-in period (mid and high dose ICS/LABA; see Appendix 10 of the study protocol for the definition of ICS dose levels)
- Pre-bronchodilator FEV_1 in % of predicted FEV_1 at run-in visit 101 (< 40%, 40% - < 60%, 60% - < 80%)
- ACQ-7 at baseline (1.5 - < 2, 2 - < 2.5, ≥ 2.5)

The subgroup analyses for patient's prior therapies before run-in period (mid and high dose ICS/LABA) will also be performed for endpoints ACQ-7 and AQLQ at Week 26.

For ACQ-7, the same MMRM used for key secondary endpoint will also be performed on the FAS excluding measurements in patients requiring chronic corticosteroid use after an exacerbation (beyond permitted steroid taper for exacerbation of approximately 7-10 days). That means, if an asthma exacerbation requires either an initiation of systemic corticosteroids or an increase in the maintenance dose of oral corticosteroids (OCS) which is extended to more than 10 days, then all ACQ-7 measurements during that period from start to stop date will be excluded.

2.6.2 Analysis of secondary variables

In addition to the two primary treatment comparisons QVM149 150/50/80 µg vs. QMF149 150/160 µg and QVM149 150/50/160 µg vs. QMF149 150/320 µg, each dose of QVM149 will be compared vs. salmeterol xinafoate /fluticasone propionate 50/500 µg b.i.d. for all secondary variables (and for the primary and key secondary variables too).

The estimation of the add-on effect of glycopyrronium to QMF149 (i.e., the average of QVM149 150/50/80 µg and QVM149 150/50/160 µg vs. the average of QMF149 150/160 µg and QMF149 150/320 µg) will be done for trough FEV_1, ACQ (mean change and responder analysis), and asthma exacerbations.

2.6.2.1 Spirometry data other than the primary endpoint

Trough FEV_1 at Day 2 and at Day 365 will be analyzed using the same MMRM as specified for the primary analysis and between-treatment comparison will be carried out using the adjusted mean (least-squares mean) difference based on the treatment main effect and the coefficient for the treatment-by-visit interaction corresponding to the respective visit. Adjusted mean (LS mean) will be displayed for each treatment group (for both the absolute FEV_1 value and the change from baseline value) along with the estimated treatment differences and the 95% confidence intervals and the two-sided p-values by visit.

In addition to trough FEV_1, pre-dose trough FEV_1 is defined as the average of the two FEV_1 measurements taken 45 min and 15 min pre-evening dose. Measurements will be done at Days 30, 86, 183, and 364. As for all spirometry data, measurements taken within 6 hours of rescue medication use, or within 7 days of systemic corticosteroid use (except for patients who were on stable systemic corticosteroids as background therapy), or within 3 months of a single depot corticosteroid injection will be set to missing. However, the post-evening dose time window of 22 – 25 hours or the post-morning dose time window of 8 – 13 hours cannot be applied here since the dosing data of the previous day or the morning of the day of pre-dose measurements will not be recorded. The same MMRM as for trough FEV_1 will be applied, however, with other visits included.

Post-dose FEV_1 measurements (5 min, 15 min, 30 min, 1 hr) will be analyzed separately by visit using the same MMRM as for trough FEV_1 with post-dose time point as repeat variable (i.e., in the MMRM, visit will be replaced by time point).

Pre- and post-dose FVC and FEF_{25-75} will be analyzed similar to FEV_1.

For all spirometry measurements, we exclude the implausible values. Specifically, if a patient has implausible FEV_1 value (defined as > 7 L) at an assessment, all spirometry measurements related to that assessment will be excluded. The threshold of 7 L was chosen because according to ERS/ECCS 1993 regressions, the maximum normal predicted FEV_1 is 5.17 L, we set 7 L as the implausible cutoff to allow for some fluctuation around the maximum mean.

2.6.2.2 ACQ at Weeks 4, 12 and 52

ACQ-7 at Weeks 4, 12 and 52 will be analyzed using the same MMRM as specified for the key secondary endpoint ACQ-7 at Week 26. Change from baseline in the ACQ-7 will be also analyzed using the same MMRM model.

In addition, the proportion of patients who achieve an improvement of at least 0.5 in ACQ-7 (i.e., a decrease of ACQ-7 score of at least 0.5 from baseline) at Week 4, 12, 26, and 52 visits will be analyzed using a logistic regression model. First missing continuous ACQ-7 values will be imputed using a MI method under MAR assumption. Then continuous ACQ values will be converted into binomial responses and generalized estimating equation (GEE) method will be used to analyze the imputed datasets and then the results from GEE method will be combined for the inference. The model will include the same terms as used for the analysis of ACQ-7 but no random effect for center nested within region. The estimated adjusted odds ratios will be displayed along with the associated 95% (two-sided) confidence intervals and p-values.

Additionally, the number (%) of patients will be summarized by visit with the following categories: Response (decrease from baseline of ≥ 0.5 units in ACQ-7), no change (-0.5 < change from baseline in ACQ-7 < 0.5), worsening (increase from baseline of ≥ 0.5 units in ACQ-7). Net response is expressed as the number of patients with response – the number of patients with worsening.

As already mentioned above, subgroup analyses for patient's prior therapies before run-in period (mid and high dose ICS/LABA) will be performed for ACQ-7 at Week 26 (using the appropriate interaction term in the model and additional covariate as a fixed effect).

All the analyses described above will be repeated for ACQ-5, including the responder analysis with the same threshold of 0.5.

The ACQ-5 consists of five items on symptom assessment, which is a subset of ACQ-7. All 5 questions of the ACQ-5 are equally weighted. The total score is calculated as the mean of all five questions, which is calculated as the sum of scores divided by the number of questions that were answered, as long as there were at least 4 questions answered and the missing question is not Question 1 ('On the average, during the past week, how often were you woken by your asthma during the night?').

2.6.2.3 Peak Expiratory Flow Rate (PEF)

All the patients are instructed to record PEF twice daily using a mini Peak Flow Meter device, once in the morning (before taking the morning dose) and once approximately 12 h later in the evening (before taking the evening dose), from screening and throughout the study.

PEF (liters/min) will be analyzed separately for morning and evening values. The baseline values are defined as the average from all non-missing records taken over the last 14 days of the run-in period between run-in Visit 101 and up to one day before first dose of study drug. Mean values will be calculated over the first 26 weeks and over the whole 52 weeks of double-blind treatment.

Separate ANCOVA will be performed to investigate treatment differences in the change from baseline in mean morning/evening PEF during the first 26 weeks and during the whole 52 weeks of double-blind treatment. The ANCOVA model will contain treatment and region as fixed effect factors with center nested within region as a random effect, and baseline morning/evening PEF, FEV_1 prior to inhalation and FEV_1 within 15 to 30 min post inhalation of salbutamol/albuterol (components of SABA reversibility) as covariates. LS means and associated 95% confidence intervals will be presented for treatments and treatment differences.

In addition, the mean morning/evening PEF will be summarized by 4 weekly (28 days) intervals and analyzed using a similar MMRM as specified for the primary analysis with baseline FEV_1 value replaced with the appropriate baseline PEF.

2.6.2.4 Rescue medication

The number of puffs of rescue medication during the past 12 hours is recorded twice (morning/evening) by the patient in the e-Diary. The baseline values are defined as the average from all non-missing records taken over the last 14 days of the run-in period between run-in Visit 101 and up to one day before first dose of study drug.

The mean daily number of puffs of rescue medication use will be calculated for each patient over the first 26 weeks, over the whole 52 weeks of double-blind treatment, and over 4 weekly intervals as described for other data of the e-Diary. This will be done separately for morning (night-time), evening (daytime), and daily (night-time plus daytime) rescue medication use. Similar ANCOVA and MMRM as used for the analysis of PEF with baseline PEF value replaced with the appropriate baseline mean number of puffs.

The percentage of 'rescue medication free days' (defined as any day where the patient did not use any puffs of rescue medication during daytime and night-time) will be summarized by treatment and analyzed in the same way as described for the number of puffs of rescue medication use with baseline mean number of puffs replaced with the baseline percentage of rescue medication free days as a covariate. In case of non-convergence or evidence of highly skewed data, non-parametric tests such as Wilcoxon Rank Sum Test will be explored.

2.6.2.5 Asthma symptom based on e-Diary

Daily e-Diary recordings will be used to derive

- the mean daytime asthma symptom score i.e., the mean score over the 5 evening questions with respect to shortness of breath, wheeze, cough, chest tightness, and 'Did your respiratory symptoms stop you from performing your usual daily activities?', each with scores from 0 (no problems) to 4 (very severe problems)
- the total daily symptom score, defined as the sum of the nighttime score ('How did you sleep last night?' with scores from 0 - 4), the morning awakening score ('Did you have asthma symptoms upon awakening in the morning?' with scores from 0 - 4) and the mean daytime asthma symptom score. The sum will be in the range of 0 – 12.
- day with no daytime symptoms, i.e., all 5 evening questions must have a score = 0
- night with no night-time awakenings, i.e., the question 'How did you sleep last night?' must be answered with 'I did not wake up because of any breathing problems'
- morning with no symptoms on rising, i.e., the question 'Did you have asthma symptoms upon awakening in the morning?' to be answered with 'None'
- asthma symptoms free days, i.e., days with no daytime symptoms and no night-time awakenings and no symptoms on rising

The baseline values are defined as the average from all non-missing records taken over the last 14 days of the run-in period between run-in Visit 101 and up to one day before first dose of study drug. Daily scores will be averaged for each patient over Weeks 1-26, the complete 52 weeks double-blind period, and over 4 weekly intervals (from Weeks 1-4 up to Weeks 49-52). Days with no daytime symptoms, nights with no night-time awakenings, mornings with no symptoms on awakening, and days with no asthma symptoms will be expressed in terms of percentages. The percentages will be calculated as 100 * total number of days without symptoms (daytime, night-time, morning, etc.) / total number of days in the interval (i.e., 28 days for 4 week interval). The same intervals as used for average scores will be used for analyzing percentage of days. In case of non-convergence or evidence of highly skewed data, non-parametric tests such as Wilcoxon Rank Sum Test will be explored. For post-baseline periods only the days on double-blind treatment will be considered. ANCOVA and MMRM will be performed on changes from baseline similar to the analysis of PEF with baseline PEF value replaced with the appropriate baseline percentage.

2.6.2.6 Asthma exacerbations

All analyses will base on data as reported on the "Asthma Exacerbation Episodes" eCRF.

In patients with multiple exacerbations, if the start date of an exacerbation will be 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. Collapsing of exacerbation episodes will only be done for efficacy analyses but not for safety analyses of adverse events which includes asthma exacerbations.

Asthma exacerbations starting after the first dose and not later than one day after the date of last dose will be included in the analyses of efficacy.

As already mentioned above, the add-on effect of glycopyrronium to QMF149 (i.e., the average of QVM149 150/50/80 µg and QVM149 150/50/160 µg vs. the average of QMF149 150/160 µg and QMF149 150/320 µg) will also be estimated for asthma exacerbations.

The following asthma exacerbation-related parameters will be summarized considering the whole 52 weeks of double-blind treatment. The analysis will be performed by exacerbation category wherever specified. The exacerbation categories are: All (mild, moderate, severe), the combination of moderate or severe, and severe.

The annual rate of asthma exacerbations, by exacerbation category

The annual rate of asthma exacerbations will be analyzed using a generalized linear model assuming the negative binomial distribution including treatment and region as fixed-effect factors, and history of asthma exacerbation in the 12 months prior to screening (i.e., the number of asthma exacerbations in the 12 months prior to screening), FEV_1 prior to inhalation and FEV_1 15 to 30 min post inhalation of salbutamol/albuterol (components of SABA reversibility) as covariates. The log exposure in years will be used as an offset variable in the model (see Section 2.4.1 for the definition of exposure). The estimated rate ratio along with two-sided 95% interval and corresponding p-value will be provided.

An additional analysis of the annual rate of asthma exacerbations will be done by excluding events/days in patients requiring corticosteroid use after an exacerbation (beyond permitted steroid taper for exacerbation). If an asthma exacerbation requires either an initiation of systemic corticosteroids or an increase in the maintenance dose of OCS extended to more than 10 days treatment, all data from start to stop date of that systemic corticosteroid use will be excluded.

Time to first asthma exacerbation, by exacerbation category

Time-to-event variables will be analyzed using a Cox regression model stratified by region. The model will include treatment as fixed-effect factor, and history of asthma exacerbation in the 12 months prior to screening (i.e., the number of asthma exacerbations in the 12 months prior to screening), FEV_1 prior to inhalation and FEV_1 15 to 30 min post inhalation of salbutamol/albuterol (components of SABA reversibility) as covariates. For treatment comparisons the estimated adjusted hazard ratio will be displayed along with the associated two-sided 95% confidence interval and corresponding p-value. Kaplan-Meier analysis stratified by treatment group will be also presented and displayed graphically. Patients without an event

will be considered as censored at the earliest date of (last treatment + 1, death, last visit). In the interim analysis, patients without an event who are ongoing at the time of interim database freeze will be censored at the DLP. For patients having an event, the start date of the exacerbation will be considered to calculate the time to event (i.e., the number of days from start of treatment up to the event start date).

Time to first hospitalization for asthma exacerbation (only for severe exacerbations possible)

The analysis is similar to time to first asthma exacerbation.

Time to permanent discontinuation of study drug due to asthma exacerbation

The analysis is similar to time to first asthma exacerbation.

Duration of asthma exacerbations in days, by exacerbation category

The duration of asthma exacerbations is defined as the sum of days per patient over all exacerbations and the respective lengths (from start to end date). This will be analyzed for treatment group differences using the van Elteren test stratified for region and history of asthma exacerbation in the 12 months prior to screening (1, 2, $\geq$ 3).

Total amounts (in doses) of oral corticosteroids used to treat asthma exacerbations

Total amount (in prednisone-equivalent mg doses) of oral corticosteroid used to treat asthma exacerbations during the 52 week treatment period will be summarized descriptively by treatment group.

2.6.2.7 Quality of life as assessed by Asthma Quality of Life Questionnaire (AQLQ)

AQLQ is a 32-item disease specific questionnaire designed to measure functional impairments that are most important to patients with asthma, with a recall time of two weeks and each question to be answered on a 7-point scale (1-totally limited/problems all the time, 7-not at all limited/no problems). It consists of 4 domains:

- Symptoms = Mean of Items 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 29, 30 (12 items)
- Activity limitation = Mean of Items 1, 2, 3, 4, 5, 11, 19, 25, 28, 31, 32 (11 items)
- Emotional function = Mean of Items 7, 13, 15, 21, 27 (5 items)
- Environmental stimuli = Mean of Items 9, 17, 23, 26 (4 items)
- Overall Score = Mean of Items 1 to 32 (32 items)

Dr. Juniper, developer of AQLQ questionnaire suggests no more than 10% of missing data. This means that for a questionnaire of 32 items, no more than 3 items should be missing. Further, for the activity and symptom domains, the recommendation is no more than 1 missing value per domain, and for the emotional function and environmental stimuli domains, no missing responses at all. Missing individual items will be imputed similarly to missing items for ACQ-7.

For the overall score and each respective domain score, treatment group comparisons will be performed using the same MMRM model as specified for the primary analysis with baseline AQLQ as covariate. The Day 1 assessment constitutes the baseline value. The post-baseline visits are scheduled for Weeks 4, 12, 26, 36, and 52.

As mentioned above, subgroup analyses for patient's prior therapies before run-in period (mid and high dose ICS/LABA) will be performed for AQLQ at Week 26 (using the appropriate interaction term in the model and additional covariate as a fixed effect).

The proportion of patients who achieve an improvement in the AQLQ (i.e., an increase in the AQLQ overall score of ≥ 0.5 from baseline) at post-baseline visits will be analyzed using the same logistic regression model via GEE as specified for the ACQ-7 analysis except that baseline AQLQ will be used instead of the baseline ACQ-7. The handling of missing data will also be the same as for ACQ-7, i.e., first missing continuous AQLQ values will be imputed using a MI method under MAR assumption. Then continuous AQLQ values will be converted into binomial responses and GEE method will be used to analyze the datasets and then the results from GEE method will be combined for the inference.

2.7 Safety evaluation

All safety evaluation will be based on the safety set.

The study will be reviewed for safety issues by an independent, external data safety monitoring board (DMC), as described in Section 2.9 Interim analysis.

2.7.1 Serious asthma outcomes

A composite endpoint of serious asthma outcomes is defined as a) asthma-related hospitalization, b) asthma-related intubation, or c) asthma-related death. All serious asthma outcomes and deaths occurring from the time of randomization until the 30 days after permanent discontinuation of study drug will be adjudicated by an independent external committee to determine their asthma relatedness.

The composite endpoint as well as each single component of it will be analyzed for the number of patients with the event If a sufficient number of events ($\geq$ 10% of patients) occurs, the composite endpoint and its components will be also analyzed for the time to event and the annual rate of events; similar inferential analyses as described for asthma exacerbations will be performed.

2.7.2 Adverse events

All adverse events (AEs) including asthma exacerbations, coded with MedDRA using the most actual version at the time of database lock, will be listed. In general, summaries will include treatment-emergent adverse events only, i.e., those 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 serious adverse event, SAE) after the last administration.

The number, percentage, and exposure-adjusted rate of patients who reported treatment-emergent adverse events will be summarized by primary system organ class, preferred term, and treatment for

- all adverse events
- all adverse events by maximum severity
- adverse events suspected to be related to study drug
- all adverse events with time at onset ≤ 26 weeks
- adverse events leading to permanent study drug discontinuation
- all adverse events by standardized MedDRA query (SMQ) level, and by SMQ level and preferred term
- serious adverse events
- serious adverse events suspected to be related to drug
- serious adverse events leading to permanent study drug discontinuation
- fatal adverse events
- adjudicated serious cardiovascular and cerebrovascular (CCV) events
- adjudicated new onset of atrial fibrillation/ flutter
- adverse events of special interest (by risk and preferred term)
- serious adverse events of special interest (by risk and preferred term)

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 QVM149 150/50(160 µg treatment group. 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.

In addition, the most frequent adverse events will be presented by preferred term in descending order of frequency in the QVM149 150/50/160 µg treatment group.

To control for possible differences in exposure between the treatment groups, AEs will also be presented with incidence rates per 100 patient-years, for first onset of event where the exposure is only counted up to the time of the first event. The incidence rate per 100 patient years will be calculated as 100*(the total number of patients with the event of interest divided by the total number of exposure in patient-years). In the denominator, patients with the event of interest will be counted up to the time of the first event while patients without the event of interest will be counted with their full exposure (see Section 2.4.1 for the definition of exposure). Total patient years will be computed as sum of days of exposure over all patients divided by 365.25).

Serious CCV events will be adjudicated as Major Adverse Cardiovascular Event (MACE) or non-MACE. The MACE events will be categorized into: Coronary revascularization (CABG or PCI), heart failure requiring hospitalization, non-fatal unstable angina, , non-fatal myocardial infarction, and non-fatal stroke. Non-MACE events will not be categorized.

Atrial fibrillation and atrial flutter will be adjudicated according to their onset into: new onset, recurrent/persistent, or unknown.

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

AEs and SAEs of special interest and adjudicated serious CCV events will also be analyzed for the following selected subgroups:

- Sex (male, female)
- Pre-bronchodilator FEV_1 in % of predicted FEV_1 at run-in visit 101 (< 40%, 40% - < 60%, 60% - < 80%)

All deaths in the clinical database will be listed with both the investigator-reported principal cause and the adjudicated primary cause presented side by side. If adverse events (including asthma exacerbations) leading to death start between the first dose and the last dose + 30 days, then those deaths will be included in summary tables. The adjudicated primary cause of death will be summarized with pre-defined categories (cardiovascular, respiratory, cancer, other, intermediate causes) and subcategory.

In addition, time to event analyses will be done for the following endpoints:

- adverse event by the 5 most frequent preferred terms,
- serious adverse event (if there are more than 10% of patients with at least one SAE)
- adverse event of special interest by risk category (if there are more than 10% of patients with events in the respective risk category)
- adjudicated serious asthma outcome (composite endpoint and by single component – if there are more than 10% of patients with events)
- adjudicated serious CCV event (composite endpoint, MACE, non-MACE – if there are more than 10% of patients with events)
- new onset of adjudicated atrial fibrillation/flutter (if there are more than 10% of patients with events)

The adverse event onset date will be considered as the date of event. Patients who did not experience the event of interest will be taken as censored. For those patients who either completed or discontinued study treatment, the censoring time will be the minimum of (the last treatment day + 30 days, day of death, day of follow-up day Visit 301). For patients who are ongoing at the time of interim database freeze, the DLP date will be used for censoring.

For the legal requirements of ClinicalTrials.gov and EudraCT, two required tables on treatment emergent adverse events which are not serious adverse events with an incidence greater than 3% and on treatment emergent serious adverse events and SAE suspected to be related to study treatment will be provided by system organ class and preferred term on the safety set.

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

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

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

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

2.7.3 Laboratory data

All laboratory samples will be processed through the central laboratory. Laboratory data consist of hematology, biochemistry and urinalysis measurements, including plasma cortisol. All data will be listed with abnormal values flagged.

Laboratory 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. All data will be included in the analyses regardless of rescue medication use.

The following analyses will be performed by treatment:

- absolute values and change from baseline summarized for continuous laboratory parameters by visit and time point
- frequency table of results for categorical laboratory parameters by visit and time point
- shift tables relative to the normal reference ranges summarizing the change from baseline to most extreme post-baseline value for each continuous laboratory parameter
- shift tables from baseline to the worst post-baseline value for categorical laboratory parameters
- the number and percentage of patients with newly occurring or worsening laboratory abnormalities meeting the clinically notable criteria (see Appendix 16.1.9 for definition of notable values) summarized by laboratory parameter, scheduled post-baseline visit and time point and additionally at any time on treatment considering all post-baseline data from scheduled, unscheduled and premature discontinuation visits

The by-visit/time point summaries will include the worst case post-baseline values (determined from all post-baseline data even if from unscheduled and premature discontinuation visits, if not measured more than 7 days after last dose). To determine the worst case, the direction of interest (maximum and/or minimum value) is tabulated for hematology and biochemistry parameters in Appendix 16.1.9. For continuous urinalysis parameters the direction of interest is always "High" (maximum value).

For a patient to meet the criterion of a newly occurring clinically notable value, the patient needs to have a baseline value that 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 that is clinically notable and also have a worse post-baseline value. For patients with missing baseline value, any post-baseline notable value will be considered as newly occurring. A listing of all patients with notable laboratory values will be provided.

The maximum change from baseline will be summarized for liver parameters (ALT, AST, alkaline phosphatase, total bilirubin).

Furthermore, the number and percentage of patients with newly occurring or worsening abnormalities in liver function tests (LFT) will be summarized by treatment and scheduled post-baseline visit and time point and additionally at any time on treatment considering all post-baseline data from scheduled, unscheduled and premature discontinuation visit (if not measured more than 7 days after last dose) based on the following criteria:

Notable liver function test values

Criterion
ALT > 3 x the upper limit of normal range (ULN) ALT > 5 x ULN ALT > 8 x ULN ALT > 10 x ULN ALT > 20 x ULN
ALT or AST > 3 x ULN ALT or AST > 5 x ULN ALT or AST > 8 x ULN ALT or AST > 10 x ULN ALT or AST > 20 x ULN
Total bilirubin > 1 x ULN Total bilirubin > 1.5 x ULN Total bilirubin > 2 x ULN Total bilirubin > 3 x ULN
ALP > 1.5 x ULN ALP > 2 x ULN ALP > 3 x ULN ALP > 5 x ULN
ALT or AST > 3 x ULN and total bilirubin > 1.5 x ULN ALT or AST > 3 x ULN and total bilirubin > 2 x ULN ALT or AST > 5 x ULN and total bilirubin > 2 x ULN ALT or AST > 8 x ULN and total bilirubin > 2 x ULN ALT or AST > 10 x ULN and total bilirubin > 2 x ULN ALT or AST > 20 x ULN and total bilirubin > 2 x ULN
ALP > 3 x ULN and total bilirubin > 2 x ULN ALP > 5 x ULN and total bilirubin > 2 x ULN
ALT or AST > 3 x ULN and Total Bilirubin > 2 x ULN and ALP < 2 x ULN (Hy's law)
ALT or AST > 3 x ULN and (nausea or vomiting or fatigue or general malaise or abdominal pain or (rash and eosinophilia))*

ALT = alanine aminotransferase, AST = aspartate aminotransferase, ALP = alkaline phosphatase
* Based on the signs/symptoms information as recorded on the liver events eCRF, not the adverse events eCRF.

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

Listings of patients with clinically notable LFT values will be provided.

During the trial, there has been an unexpected high reporting rate of elevated potassium by Central Lab, which was investigated and in the end considered clinically irrelevant and no risk

to patient safety. Central Lab was notified to flag all the elevated potassium data. As additional analyses, potassium data without flagged values will be summarized.

2.7.4 Vital signs

Vital signs measurements include systolic and diastolic blood pressure (SBP and DBP), pulse rate (all measured at Visit 101, 102, 201, 202, 204, 207, 210, 214, EOT – early treatment discontinuation), and body weight (measured at Visit 101, 207, 214, EOT – early treatment discontinuation).

Vital signs 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. All data will be included in the analyses regardless of rescue medication use.

The following analyses will be performed by treatment:

- absolute values and change from baseline summarized by parameter, visit and time point
- the number and percentage of patients by parameter, visit and time point with
  - pulse rate < 40 bpm, 40 – 90 bpm, and > 90 bpm
  - SBP < 90 mm Hg, 90 – 140 mm Hg, and > 140 mm Hg
  - DBP < 50 mm Hg, 50 – 90 mm Hg, and > 90 mm Hg
- the number and percentage of patients with newly occurring or worsening notable vital signs values (see Appendix 16.1.9 for definition of notable values) summarized by parameter, scheduled post-baseline visit and time point and additionally at any time on treatment considering all post-baseline data from scheduled, unscheduled and premature discontinuation visits

The by-visit/time point summaries will include the maximum and minimum post-baseline SBP, DBP, and pulse rate values (even if from post-baseline unscheduled and premature discontinuation visits, if not measured more than 7 days after last dose).

The same approach as for notable laboratory values will be used to define a newly occurring notable vital sign value and a worsening notable vital sign value.

A listing of all patients with notable vital sign values and changes will be provided.

2.7.5 Electrocardiogram (ECG)

ECG measurements include ventricular rate, QT interval, RR interval, PR interval, QRS duration, and Fridericia's QTc (calculated as QTcF = QT / $3\sqrt{}$ RR(in seconds), where $3\sqrt{}$ denotes the cube root). Furthermore, an overall interpretation of the central cardiologist will be provided as well as a specification of abnormal findings.

ECG 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. All data will be included in the analyses regardless of rescue medication use.

The following analyses will be performed by treatment:

- absolute values and change from baseline summarized by parameter, visit and time point
- the number and percentage of patients with newly occurring or worsening notable QTcF values (see Appendix 16.1.9 for definition of notable values) summarized by scheduled

post-baseline visit and time point and additionally at any time on treatment considering all post-baseline data from scheduled, unscheduled and premature discontinuation visits

- frequency table of results for overall ECG interpretation (normal, abnormal) by visit and time point
- the number and percentage of patients with ECG abnormalities summarized by evaluation type, abnormality finding, visit and time point. In addition, the number and percentage of patients with newly occurring or persistent/recurrent ECG abnormalities at any time point over the treatment period (considering all post-baseline data from scheduled, unscheduled and premature discontinuation visits) will be summarized by evaluation type and abnormality finding

The by-visit/time point summaries will include the maximum QTcF and maximum ventricular rate (even if from post-baseline unscheduled and premature discontinuation visits, if not measured more than 7 days after last dose).

The same approach as for notable laboratory values will be used to define a newly occurring notable QTc value and a worsening notable QTc value.

A listing of all patients with notable QTc values and changes will be provided.

[redacted]

2.9 Interim analysis

2.9.1 Data Safety Monitoring Board (DMC)

An independent, external data safety monitoring board (DMC), comprising experts (as defined in the Charter) will be set up to review all serious adverse events (including deaths and all hospitalizations) and pneumonia. DMC members will review this data generated externally and independently of Novartis, at predetermined intervals. If significant safety issues arise in between scheduled meetings, ad-hoc meetings will be scheduled to review the data. Based on the safety implications of the data, the DMC may recommend modification or termination of the study.

It is planned that the independent DMC will review semi-blinded (i.e., treatment group named as A, B, C, D, or E) safety data. The details of the information flow, confidentiality and specific analyses required for the safety monitoring analysis will be documented in the DMC Charter. The Charter will be finalized prior to semi-blinding the data for the safety monitoring analysis. Since the purpose of the DMC is not based on efficacy for stopping rule, there will be no alpha spent for the safety monitoring analysis. All analyses will be considered exploratory.

2.9.2 Primary analysis after all patients have completed at least 26 weeks treatment

The primary and key secondary endpoints of CQVM149B2302 study are the trough FEV_1 and ACQ-7 after 26 weeks of treatment, respectively while the entire study treatment period is 52 weeks. In order to support a marketing authorization filing once all patients have completed 26 weeks of treatment (V207) or prematurely withdrawn from the study, an interim database freeze will be conducted to perform the Primary Analysis.

The database for Primary Analysis will contain data up to 26 weeks for all patients and data up to 52 weeks for patients who have completed the study. For patients who are ongoing at the time of interim database freeze, the database will include data available until last completed visit.

Primary analysis includes primary and key secondary endpoints as well as other pre-specified endpoints at Week 26. [redacted]

At the time of interim database freeze for Primary Analysis, a limited number of pre-specified Sponsor members will be unblinded to the study data to work on study and submission related activities which require knowledge of these data.

The study will continue under the management of a separate blinded team who will be responsible for study conduct from the time of the 26 week interim database freeze until end of study. In order to maintain the integrity of the study data, the blinded team members will not

have access to any of the unblinded data or results. In addition, all patients and investigators will remain blinded.

Two separate CSRs will be published:

- Primary Analysis CSR: To support analyses once all patients have completed the assessments after 26 weeks of treatment (Visit 207) or prematurely withdrawn from the study. This will include analysis of primary and key secondary as well other pre-specified endpoints which are evaluable at this time.
- Full CSR: To support analysis once all patients have completed the assessments after 52 weeks of treatment (Visit 214) plus follow up (Visit 301) or prematurely withdrawn from the study.

Since the analysis of primary and key secondary objectives will be performed only for the Primary Analysis CSR, the interim analysis will not require any adjustment to the overall type I error rate.

2.10 Sample size calculation

The sample size calculation takes into account the following consideration:

1. To achieve at least 90% power (with multiplicity adjustment) for primary endpoint trough FEV_1 with a treatment difference of 90 mL between QVM149 vs. QMF149 at the corresponding doses, assuming standard deviation of 380 mL based on internal studies QMF149A2210, QMF149E2201, QMF149E2203 and Kerstjens (2012);
2. To achieve at least 80% power (with multiplicity adjustment) for key secondary endpoint ACQ-7 with a treatment difference of 0.15 between QVM149 vs. QMF149 at the corresponding doses, assuming standard deviation of 0.80 based on studies QMF149A2210, QMF149E2201, QMF149E2203 and Kerstjens (2012).

If 10% dropout rate is assumed, then calculation shows that the sample size of 2980 patients (i.e., 596 per arm) will provide 97% power for item 1 and 82% power for item 2, with multiplicity adjustment.

The sample size calculation was performed in R 3.1.2 with package gMCP (Rohmeyer, 2012).

3 Analyses to be reported outside of CSR

3.1 Subgroup analyses

Subgroup analyses for Japanese, Chinese, and Asian patients will be performed for the FAS to explore the treatment effect in respective sub-populations.

Asian subgroup contains patients from the following countries:

- China
- Japan
- Philippines
- Thailand
- Vietnam

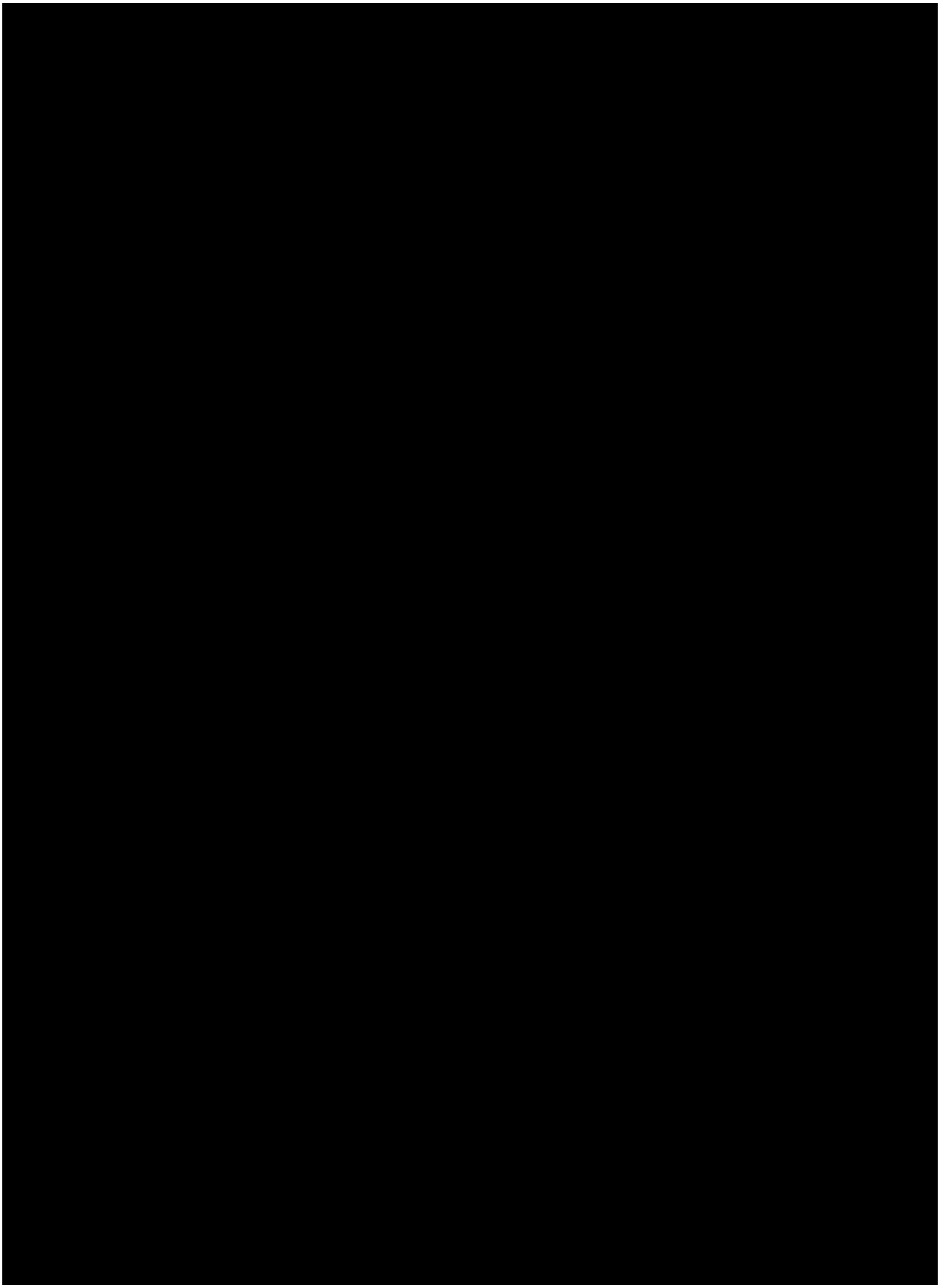
The following endpoints will be summarized and/or analyzed for all three sub-populations using similar methods as described in previous sections, if not marked with special symbol.

- Patient disposition
- Protocol deviations
- Analysis sets
- Demographics
- Disease characteristics
- Spirometry at run-in visits and at baseline
- Trough FEV1
- ACQ-7
- Pre-dose trough FEV1*
- Trough FVC*
- Pre-dose trough FVC*
- Mean morning/evening PEF*
- Asthma symptom based on e-diary *
- Rescue medication*
- Asthma exacerbations*
- Asthma Quality of Life Questionnaire (AQLQ-S+12)*
- [redacted]
- Exposure
- Concomitant asthma medications
- Laboratory data
- Vital signs
- ECG
- Adverse events

*: only needed for Japanese subgroup analyses

The specific outputs for each sub population under each endpoint will be specified in the TFLs.

[redacted]



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 9.4.

16.1.9.2 Major protocol deviations and other exclusion criteria

The following protocol deviations will be considered as major and will lead to exclusion of patients from analysis sets:

Exclusions from per-protocol set:

- INCL01: Age < 18 years or > 75 years or missing
- INCL03: No current diagnosis of persistent Asthma of at least 12 months prior to V1
- INCL04.A5: Patients are not treated with medium or high doses of ICS/LABA combination for at least 3 months and/or not at stable dose 30 days prior to Visit 1
- INCL05: ACQ-7 score < 1.5 at V101 and/or V102
- INCL05.01: Patient used invalidated ACQ due to translation issue at Visit 101 and/or Visit 102
- INCL06: Patients without at least one documented asthma exacerbation in the 12 months prior to Visit 1 that required physician care/ER/hospitalization and systemic corticosteroid
- INCL07: Pre-bronchodilator FEV_1 at V101 and/or V102 $\geq$ 80% of predicted normal at Visit 101
- INCL07.01: Patient has not achieved assessments according to ATS/ERS criteria
- INCL08: Patient has not achieved an acceptable spirometry result and does not have historical reversibility or bronchoprovocation
- INCL08.01: Patient without Historical reversibility OR bronchoprovocation according to ATS/ERS as per source documentation
- INCL 08.02: Increase in FEV_1 < 12% or < 200 mL of pre-bronchodilator value in the reversibility test without historical data
- EXCL01: Patient has a smoking history of > 10 pack years or missing
- EXCL01.01: Patient smoked or inhaled tobacco within the 6 month period prior to V1
- EXCL01.02: Patient who use smoking nicotine inhalers (e.g., e-cigarette) at Visit 1
- EXCL02: Asthma attack/exacerbation requiring SCS or hospitalization or ER visits within 6 weeks prior to V1
- EXCL03: History of severe asthma exacerbation requiring intubation prior to Visit 1
- EXCL05.A4: Patients treated with a LAMA for asthma within 3 months prior Visit 1
- EXCL07: Asthma exacerbation within 4 weeks prior to V1
- EXCL07.01: Respiratory tract infection or asthma worsening between screening and run-in Visit 102
- EXCL07.02: Respiratory tract infection within 4 weeks prior to V1

- EXCL08: Visual clinical significant oropharyngeal candidiasis between Visit 101 and Visit 102
- EXCL09: Chronic condition affecting upper respiratory tract
- EXCL10: Patient has a history of chronic lung diseases other than asthma
- EXCL13: Other investigational drugs used within 5 half lives or until return of baseline pharmacodynamics effect whichever is longer
- EXCL27: Patient is not on maintenance immunotherapy for allergies for at least 3 months prior to V101
- EXCL30: Unable to use the dry powder inhaler device at V101 and/or V201
- EXCL31: Patient has history of alcohol or other substance abuse
- OTH15: Patient received incorrect dose of run-in medication (Visit 101 to Visit 102)
- COMD07: Use of parenteral corticosteroids less than 4 weeks prior to Visit 101 not in stable dose (OCS)
- COMD08: Intra-muscular corticosteroid less than 3 months prior to Visit 101
- COMD09: Use of biologic less than 3 months prior to Visit 101 not in stable dose
- COMD10: Xanthine's use less than 30 days prior to Visit 101 not in stable dose
- COMD11: Leukotriene antagonists and leukotriene synthesis inhibitors taken less than 30 days prior to Visit 101 not in a stable dose
- COMD12: ICS and LABA, not discontinued before required washout period to Visit 101 and/or Visit 102
- COMD13: Last LABA wash out more than 4 days before visit 101 or last ICS wash out more than 7 days before visit 101
- OTH06: Compliance failure of the patient influencing the safety and efficacy data as per investigator judgement
- OTH04: Patient receiving investigational drug other than study drug during the course of the study
- COMD06: Use of SABA other than rescue medication, not discontinued at Visit 1
- COMD01: Banned asthma related Concomitant medication
- WITH01: More than 5 exacerbation and still on treatment

[redacted]

Exclusions from full analysis set and per-protocol set:

- OTH05: Patient randomized more than one trial
- OTH18: Patient randomized more than once in this trial

Exclusions from full analysis set, per-protocol set, safety set, [redacted]:

- OTH02: Patient was randomized but no study drug was taken

Exclusions from randomized set, full analysis set, per-protocol set, [redacted]:

- Patient who received study medication without randomization

16.1.9.3 Assessment windows, baseline and post-baseline definitions, missing data handling

16.1.9.3.1 Data pooling and assessment windows

Data from unplanned or unscheduled visits or the early treatment/study discontinuation visits will be listed. For patients who do not complete the study, the treatment discontinuation visit will be an unscheduled visit.

Clinical laboratory measurements, vital signs and ECG data from unplanned or unscheduled visits will only be included in the summaries of the notable values and extreme values. All efficacy data (including spirometry data) from these visits will not be used for missing data imputation unless specified otherwise.

Laboratory, vital signs, and ECG values that 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. If a measurement scheduled as pre-dose is actually performed post-dose, or vice versa, the data will not be used for by time point assessments.

16.1.9.3.2 Baseline measurements

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

1) For all **FEV_1, FVC, and FEF_{25-75}** 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 the evening of Day 1. Checks will be performed to ensure both values were indeed taken prior to the first dose of study drug. 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 the baseline. If both values are missing (or not confirmed to be pre-dose), then the last out of the pre-bronchodilator measurements taken at the run-in Visits 101 and 102 or at unscheduled run-in visits will be used as the baseline. If the FEV_1, FVC or FEF_{25-75} measurements are missing both on Day 1 and at run-in visits, 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 (except for patients who were on stable systemic corticosteroids as background therapy), or within 3 months of single depot corticosteroid injection will be set to missing.

For all baseline spirometry measurements, we exclude the implausible values. Specifically, if a patient has implausible FEV_1 value (defined as > 7 L) at baseline assessment, all spirometry measurements related to that assessment will be excluded.

2) For the **ACQ-7** score, the baseline value is defined as the score obtained at Day 1 (V102). If the value is missing, the last measurements from run-in visit 101 or unscheduled run-in visits will be used.

3) For patient **diary data (PEF, symptoms, and rescue medication use**), the baseline values are defined as the average from all non-missing records taken in the run-in period over 14 days from Visit 101 up to one day before first dose of study drug (i.e., from Day -14 to Day -1 inclusively). Missing diary data will not be imputed.

Baseline PEF will be calculated separately for morning and evening values.

The night-time symptom score consists of one question 'How did you sleep last night?' which has to be answered with scores from 0 = 'I did not wake up because of any breathing problems.' up to 4 = 'I had difficulty sleeping because of my breathing problems even though I used my rescue medication.' The morning score consists of the question 'Did you have asthma symptoms upon awakening in the morning?' which has to be answered with scores from 0 = 'None' up to 4 = 'Very Severe'. There are 5 questions including the today's severity of shortness of breath, wheeze, cough, and chest tightness during the past 12 hours, and 'Did your respiratory symptoms stop you from performing your usual daily activities?' which are part of the daytime symptom score. All have to be answered with a score from 0 = 'None / Not at all' up to 4 = 'Very Severe / Completely'. The average of the 5 scores is defined as the daytime symptom score. The baseline scores will be derived by averaging over the days of the run-in period with non-missing data, separately for each score, the night-time, morning, and daytime symptom score.

The baseline total daily symptom score will be obtained in the first step by adding the night-time symptom score, the morning score, and the mean daytime symptom score for each single day. In the second step, these daily symptom scores will be averaged over the days of the run-in period with non-missing data. The score is in the range from 0 to 12. For the derivation of the total daily symptom score, all 3 scores have to be available at that day. Otherwise a total daily symptom score will not be calculated for that day.

Rescue medication use will be analyzed separately for the number of puffs during daytime, night-time, and over the complete 24 hours. Baseline is defined as the average number of puffs per day in the run-in period. If for the value over the whole day (24 hours) the number of puffs is missing for part of the day (either day-time or night-time) then a half day will be used in the denominator to calculate the average value. Any values > 30 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) **Laboratory data** include hematology, biochemistry, urinalysis, and plasma cortisol. Baseline laboratory data is defined as the assessment taken 20 minutes pre-dose at 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 at the run-in Visit 101 or at an unscheduled visit before the first administration of study drug will be used for baseline. Otherwise, the baseline laboratory data will be set to missing.

5) **Vital signs** include pulse rate and systolic and diastolic blood pressures. Baseline vital signs are defined as the assessment taken 25 minutes pre-dose at 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 at the run-in Visit 101 or at an unscheduled visit before the first administration of study drug will be used for baseline.

6) Baseline **height and weight** are defined as the measurements taken at the run-in Visit 101. Missing baseline will not be imputed.

7) Baseline **ECG** is defined as ECG measured 35 minutes 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 at the run-in Visit 101 or at an unscheduled visit before the first administration of study drug will be used. Otherwise, the ECG baseline will be set to missing without imputation.

16.1.9.3.3 Post-baseline measurements

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

FEV_1, FVC and FEF_{25-75} measurements taken within 6 hours of rescue use or within 7 days of systemic corticosteroid use (except for patients who were on stable systemic corticosteroids as background therapy) or within 3 months of single depot corticosteroid injection will be set to missing. The trough value is defined as the average of the two measurements taken 23 hr 15 min and 23 hr 45 min post-evening dose. If the actual measurement times will be outside of the 22 - 25 hour post-evening dose time window for o.d. regimens (or outside the 8 - 13 hour post-morning dose time window for salmeterol xinafoate/fluticasone propionate 50/500 µg b.i.d. regimen) then the individual 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 the trough value. If both values are missing, then the trough value will be missing. In a similar way pre-dose trough is defined as the average of the two measurements taken 45 min and15 min pre-evening dose. Pre-dose measurements will be set to missing when they are actually taken post-dose. However, the post-evening dose time window of 22 – 25 hours or the post-morning dose time window of 8 – 13 hours can not be applied here since the dosing data of the previous day or the morning of the day of pre-dose measurements will not be recorded. 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 used as average pre-dose value.

Post-dose spirometry measurements (i.e., at 5, 15, 30, 60 min) which were actually taken pre-dose will be set to missing and not used for by time point assessments.

For all post-baseline spirometry measurements, we exclude the implausible values. Specifically, if a patient has implausible FEV_1 value (defined as > 7 L) at an assessment, all spirometry measurements related to that assessment will be excluded.

For **diary data (PEF, symptom scores and rescue medication use)**, post-baseline measurements start with the morning recordings at Day 2 and end with the evening recordings one day after the last evening dose. Summary data (i.e., mean values and percentage of days with) will be calculated for 4-weekly intervals (defined as 28 days), for Weeks 1 – 26 (defined up to Visit 207, Day 183), and for Weeks 1 – 52 (defined as the whole treatment period up to Visit 214, Day 365). Similar calculations as for baseline diary data will be done. If there will be less than 14 days with data in a 4-week period, then the data should collapse with the previous 4-week period. For Weeks 1 – 26 and 1 – 52, summary values will only be calculated if a patient has at least 30% of their diary days and at least 20 diary days with evaluable data for that variable in the period of interest.

For **safety data**, post-baseline measurements comprise recordings up to the last dose of study drug + 7 days for laboratory, ECG, vital signs, non-serious AE and up to the last dose of study drug + 30 days for SAEs and death. ECG, vital signs and laboratory 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 and time point are non-missing. If a measurement scheduled as pre-dose is actually performed post-dose, or vice versa, the data will not be used for by time point assessments but will be included in the summaries of the notable values and extreme values.

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.

16.1.9.4 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.

Direction of interest for worst case value for laboratory parameters

Laboratory Parameter	Direction of interest for worst case value
A. Hematology	
Hemoglobin	Low
Hematocrit	Low
Erythrocytes	Low
Leukocytes	Low and high
Basophils	High
Eosinophils	High
Lymphocytes	Low and high
Monocytes	High
Neutrophils	Low and high
Platelets	Low and high
B. Chemistry	
Albumin	Low
Alkaline Phosphatase	High
ALT/SGPT	High
AST/SGOT	High
Bilirubin Total	High
Blood Urea Nitrogen (BUN)	High
Creatinine	High
Glucose	Low and high
Gamma GT	High
Potassium	Low and high
Magnesium	Low and high
Uric acid	High
Plasma cortisol	Low and high

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

Clinical notable criteria for selected laboratory tests

Laboratory parameter (unit)	Lower bound of clinically notable range	Upper bound of clinically notable range
Hematology		
Hematocrit (v/v))		
Male	0.37	-
Female	0.32	-
Hemoglobin (g/L)		
Male	115	-
Female	95	-
Platelets (x10E^9/L)	75	700
Leukocytes (x10^9/L)	2.8	16.0
Chemistry		
Albumin (g/L)	25	-
Alkaline Phosphatase (U/L)	-	3 x ULN
ALT/SGPT (U/L)	-	3 x ULN
AST/SGOT (U/L)	-	3 x ULN
Bilirubin Total (µmol/L)	-	34.2
BUN (mmol/L)	-	9.99
Creatinine (µmol/L)	-	176.8
Glucose (mmol/L)	2.78	9.99
Gamma GT (U/L)	-	3 x ULN
Potassium (mmol/L)	3	6
Magnesium (mmol/L)	0.51	1.07

v = volume, ULN = upper limit of normal

16.1.9.5 Vital signs and ECG – 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 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

The following table shows the clinical notable criteria for QTcF.

Clinical notable criteria for QTcF (Fridericia's formula)

ECG parameter (unit)	Clinically notable range
Notable value considering newly occurring or worsening cases	
QTc (ms)	> 450 (male)
QTc (ms)	> 460 (female)
QTc (ms)	> 500
Notable change from baseline	
QTc	30 – 60
QTc	> 60
Combination of QTc above a threshold and QTc increase from baseline above a threshold	
QTc (ms)	> 500 ms and increase from baseline > 60 ms

16.1.9.6 Statistical methodology and assumptions

In this section, the least squares mean (LSM) is estimated by the expected response for an "average" patient (with covariates set as weighted average of the covariates value in the population) based on the model (using OM option in PROC MIXED in SAS).

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

The following MMRM ANCOVA will be used for trough FEV_1, ACQ-7, and other data:

Dependent variable = intercept + treatment + region + baseline value + FEV_1 prior to inhalation + FEV_1 15 to 30 min post inhalation + visit + treatment*visit + baseline value*visit + random effect of center nested within region + error.

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

For the primary analysis, if the model does not converge, data only up to Week 26 (i.e., Days 2 and 184) will be used (with unstructured covariance matrix). If still the model fails to converge, compound symmetry covariance matrix will be used. If the model still does not converge, the random effects of center will be removed. For the final analysis, if the model does not converge with unstructured covariance matrix, the compound symmetry covariance structure will be used. If still the model does not converge, the random effect of center will be removed. For both analyses, if the model converges after removing the random center effects, the original model without random center effects will be re-assessed. If the final model for the primary analysis converges without Week 52 data, only summary statistics will be provided for Week 52 data analysis.

Results will be presented with LSM and standard error (SE) for treatment effects and LSM, SE, associated two-sided 95% confidence interval, and two-sided p-value for all relevant treatment contrasts.

The combined effects of QVM149 (150/50/80 & 150/50/160 µg) and QMF149 (150/160 & 150/320 µg) will be derived by weighting treatment groups equally.

[redacted]

- **Tipping point analysis for trough FEV_1 at Week 26**

As a sensitivity analysis to evaluate the impact of missing data, a tipping point analysis will be performed for the primary endpoint trough FEV_1 at Week 26. The delta-adjusting approach described in Ratitch et al (2013) will be used to find the tipping point, in a spectrum of conservative missing not at random (MNAR) assumptions, at which conclusions change from being favorable to QVM149 to being unfavorable. The basic idea is to first impute the missing Week 26 values using the multiple imputation (MI) method based on the MAR assumption. The imputed Week 26 values for the QVM149 arms are then adjusted by a delta value.

The endpoint is assumed to have a normal distribution at Day 2, Week 26 and Week 52. The primary analysis is performed on the delta-adjusted datasets to see if the conclusions change. If not, a larger delta is chosen and the process is repeated until the conclusion is overturned. The tipping point is the delta value which will cause the QVM149 vs. QMF149 comparison on the primary endpoint to be no longer statistically significant. Different delta values are possible for QVM149 150/50/80 µg o.d. versus QMF149 150/160 µg o.d. and QVM149 150/50/160 µg o.d. versus QMF149 150/320 µg o.d. A series of analyses on a range of delta values will be performed to find the tipping point. For each delta value, the following steps will be performed.

1. Impute missing covariates (baseline FEV_1, FEV_1 prior to inhalation and FEV_1 15 to 30 min post inhalation value) and missing values at Day 2 and/or Week 26 using the MI approach based on the fully conditional specification (FCS) method for 100 times and obtain 100 imputed datasets. Missing baseline FEV_1, FEV_1 prior to inhalation and FEV_1 15 to 30 min post inhalation value will be imputed separately for each treatment group using a model with geographic region and the respective other covariates (baseline FEV_1, FEV_1 prior to inhalation and FEV_1 15 to 30 min post inhalation value) as predictors. Missing Day 2 values will be imputed separately for each treatment group using a model with the imputed values of age group, geographic region, baseline FEV_1, FEV_1 prior to inhalation and FEV_1 15 to 30 min post inhalation value as predictors. Missing Week 26 values will be imputed separately for each treatment group using a model with the imputed values of geographic region, baseline FEV_1, FEV_1 prior to inhalation, FEV_1 15 to 30 min post inhalation, and Day 2 value as predictors.
2. After the imputation obtained in above step, for subjects in the two QVM149 treatment arms only, make the imputed value at Week 26 worse by a value of delta (i.e., subtract delta from the imputed value at Week 26). Note: delta = 0 represents the standard missing at random (MAR) based MI.
3. The final multiply-imputed dataset where all missing values are filled will be analyzed (by imputed dataset) using a linear model with treatment, geographic region, baseline FEV_1, FEV_1 prior to inhalation, and FEV_1 15 to 30 min post inhalation as predictors of trough FEV_1 at Week 26.
4. The results for the treatment effect from the 100 datasets will then be combined using Rubin's rule. If the combined results does not change the conclusion (i.e., the QVM149 treatment effect is still statistically significantly superior to QMF149), a larger delta will be chosen and Steps 1-4 will be repeated using the datasets from Step 1 until the tipping point is found (when $p > 0.05$).

- **Logistic regression analysis**

The proportions of patients who achieve a clinically relevant improvement in ACQ-7 score (i.e., decrease of ACQ-7 score of at least 0.5 from baseline) at the scheduled post-baseline visits will be analyzed using logistic regression. The following steps will be performed for imputation of missing ACQ-7 data and following logistic regression analysis:

1. First, missing ACQ-7 values (continuous) at Visit 202, 204, 207, and 214 will be imputed sequentially using the MI method under MAR assumption. Missing values will be imputed separately for each treatment group using a model with region, baseline ACQ value, FEV_1 prior to inhalation and FEV_1 within 15 to 30 min post inhalation of salbutamol/albuterol and previous post-baseline ACQ-7 values as predictors for 100 times.
2. Continuous ACQ values will be converted into binomial responses (1 = an improvement of at least 0.5, 0 = no such improvement) and generalized estimating equation (GEE) method will be used to analyze the dataset (by imputed dataset). The GEE model will include terms for treatment, region, visit, and treatment-by-visit interaction as fixed effects, and baseline ACQ-7, baseline-by-visit interaction, FEV_1 prior to inhalation and FEV_1 within 15 to 30 min post inhalation of salbutamol/albuterol (components of SABA reversibility) as covariates.

Combining results from 100 copies of original dataset:

3. In the next step, results from each copy, obtained from GEE method will be combined for the inference using Rubin's rule. The SAS procedure PROC MIANALYZE will be used for combining the results considering the linear scale and then we will take exponential of combined result to obtain the odds ratio.

The combined effects of QVM149 (150/50/80 & 150/50/160 µg) and QMF149 (150/160 & 150/320 µg) will be derived by weighting treatment groups equally.

The proportions of patients who achieve a clinically relevant improvement in AQLQ (i.e., decrease of AQLQ score of at least 0.5 from baseline) at the scheduled post-baseline visits (Visit 202, 204, 207, 210, and 214) will be analyzed using the same procedures as mentioned above for ACQ-7, except that baseline ACQ-7 is replaced by baseline AQLQ in all steps.

- **Linear Mixed Model Analysis of Covariance (ANCOVA)**

The following linear mixed ANCOVA model will be used for diary data summarized over 26 or 52 weeks:

Dependent variable = intercept + treatment + region + baseline value + FEV_1 prior to inhalation + FEV_1 15 to 30 min post inhalation + random effect of center nested within region + error.

Results will be presented with LSM and SE for treatment effects and LSM, SE, associated two-sided 95% confidence interval, and two-sided p-value for all relevant treatment contrasts.

- **Cox regression analysis**

A Cox proportional hazards regression model will be applied in time-to-event analyses to test the null-hypothesis H_0: $\lambda_{QVM}(t) / \lambda_{Control}(t) = 1$, where $\lambda(t)$ is the hazard function for the failure time of patients treated with QVM and the control group, respectively. The Cox regression

model will be stratified by region and will include terms for treatment, history of asthma exacerbation in the 12 months prior to screening (i.e., the number of asthma exacerbations in the 12 months prior to screening), FEV_1 prior to inhalation and FEV_1 within 15 to 30 min post inhalation of salbutamol/albuterol (components of SABA reversibility).

Results will be presented with adjusted hazard ratios for treatment group comparisons and associated 95% confidence intervals and two-sided p-values. P-values 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).

The combined effects of QVM149 (150/50/80 & 150/50/160 µg) and QMF149 (150/160 & 150/320 µg) will be derived by weighting treatment groups equally.

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

- **Generalized linear model assuming a negative binomial distribution**

The annual rate of asthma exacerbations (by exacerbation category) will be analyzed using a generalized linear model assuming a negative binomial distribution. The model will include terms for treatment, region, history of asthma exacerbation in the 12 months prior to screening (i.e., the number of asthma exacerbations in the 12 months prior to screening), FEV_1 prior to inhalation and FEV_1 within 15 to 30 min post inhalation of salbutamol/albuterol (components of SABA reversibility).

The time at risk for a patient is defined as the duration of exposure in days + 1 day and the log(time at risk in years) will be used as the offset variable in the model.

Treatment group ratios of exacerbation rates will be presented together with 95% confidence intervals and two-sided p-values.

The combined effects of QVM149 (150/50/80 & 150/50/160 µg) and QMF149 (150/160 & 150/320 µg) will be derived by weighting treatment groups equally.

- **Van Elteren test**

The duration of asthma exacerbations (= the sum of days with an exacerbation, summarized by exacerbation category) will be analyzed for treatment group differences using the van Elteren test stratified for region and history of asthma exacerbation in the 12 months prior to screening ($1, 2, \geq 3$).

Modified ridit scores (scores=modridit) represent the expected values of the order statistics of the uniform distribution on (0,1). Modified ridit scores are derived from rank scores. The 'Row Mean Scores Differ' statistic gives the relevant p-value of the test. To obtain p-values for pairwise treatment comparisons, the code above will be applied separately on datasets which include only the two treatments to be compared.

16.1.9.7 Multiplicity adjustment

To control the family-wise type-I error rate at the two-sided 5% significance level, a testing procedure based on the generalized Simes test as described in Maurer et al (2011) will be used. The family for the overall type-I error rate control contains four hypotheses: two hypotheses for the primary endpoint trough FEV_1 and two hypotheses for the key secondary endpoint ACQ-7. Denote the two hypotheses for the primary endpoint as H1 and H2 for comparing QVM149

150/50/80 µg vs. QMF149 150/160 µg and QVM149 150/50/160 µg vs. QMF149 150/320 µg respectively. Similarly, denote the two hypotheses for the key secondary endpoint ACQ-7 as H3 and H4, for comparing QVM149 150/50/80 µg vs. QMF149 150/160 µg and QVM149 150/50/160 µg vs. QMF149 150/320 µg respectively.

Below is a brief description of the testing procedure based on the generalized Simes test in Maurer et al (2011).

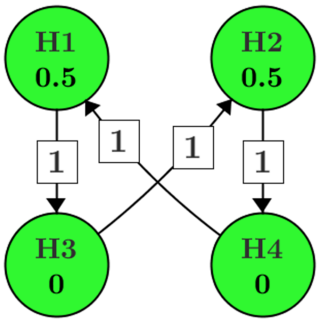
Let p1, p2, p3, p4 be the corresponding p-values (2-sided) of the four hypotheses of H1, H2, H3, and H4.

Step 1: Retain all four hypotheses if $p_i \leq 0.05$ **AND** the observed treatment difference for the corresponding p_i is in the wrong direction (i.e. MF is performing better than QMF) for **ANY** i= 1, 2, 3, 4, stop here; otherwise go to Step 2.

Step 2: Reject all four hypotheses if $p_i < 0.05$ for **ALL** i=1, 2, 3, 4 and stop here; otherwise go to Step 3.

Step 3: If neither Step 1 or 2 applies, perform a closed successive weighted Bonferroni test as given in Figure 9-1. The initial weights of 0.5 to H1 (corresponding to 0.025 alpha level, 2-sided) and 0.5 to H2 (corresponding to 0.025 alpha level, 2-sided) were assigned. Here is a brief summarization of the successive weighted Bonferroni test based on Bretz et al (2011): If H1 is rejected at the initial significance level of 0.025, then H3 can be tested at the significance level of 0.025. Similarly, if H2 is rejected at the initial significance level of 0.025, then H4 can be tested at the significance level of 0.025. If neither primary null hypothesis can be rejected at the initial significance levels, then the testing stops and efficacy cannot be claimed for neither of the doses and endpoints. Otherwise the graph is sequentially updated with reallocated weights after each hypothesis is rejected. In addition, if efficacy can be shown for one of the doses on both the primary and key secondary endpoints at the initial significance level, the associated weight is passed on to the other dose for further testing.

Figure 9-1 Testing Procedure



For each of the four hypotheses, the corresponding testing statistics (estimated least square mean difference) follows normal distribution. Hence for any two out of the four hypotheses, their corresponding testing statistics follow jointly bivariate normal distribution. Therefore this testing procedure controls the overall type I error rate at the 2-sided 0.05 level in the strong sense regardless if the bivariate normal distributions have positive or negative correlations as shown in Maurer et al (2011).

References

Maurer W, Glimm, and Bretz F (2011). Multiple and repeated testing of primary, coprimary, and secondary hypotheses. Statistics in Biopharmaceutical Research; 3; 336-352.

Ratitch, B., O'Kelly, M., and Tosiello, R. (2013). Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. Pharmaceutical Statistics, 12, 337-347.

Pearlman, D., Chervinsky, P., LaForce C., et al. (1992). A comparison of salmeterol with albuterol in the treatment of mild-to-moderate asthma. The New England Journal of Medicine, 327(20), 1420-1425.