AstraZeneca

Statistical Analysis Plan

Study Code	PT003019
NCT #	NCT02937584
Date	16APRIL2018

A Randomized, Double-Blind, Two Treatment, Two Period, Chronic Dosing (2 Weeks), Cross-Over, Multi-Center Study to Evaluate the Effects of PT001 and PT005 on Specific Image Based Airway Volumes and Resistance in Subjects With Moderate to Severe COPD

Statistical Analysis Plan

Study Code	PT003019
Edition Number	1.0
Date	16 Apr 2018

A Randomized, Double-Blind, Two Treatment, Two Period, Chronic Dosing (2 Weeks), Cross-Over, Multi-Center Study to Evaluate the Effects of PT001 and PT005 on Specific Image Based Airway Volumes and Resistance in Subjects With Moderate to Severe COPD.

A Randomized, Double-Blind, Two Treatment, Two Period, Chronic Dosing (2 Weeks), Cross-Over, Multi-Center Study to Evaluate the Effects of PT001 and PT005 on Specific Image Based Airway Volumes and Resistance in Subjects With Moderate to Severe COPD.

Study Statistician [redacted]

A Randomized, Double-Blind, Two Treatment, Two Period, Chronic Dosing (2 Weeks), Cross-Over, Multi-Center Study to Evaluate the Effects of PT001 and PT005 on Specific Image Based Airway Volumes and Resistance in Subjects With Moderate to Severe COPD.

SAP Author [redacted]

A Randomized, Double-Blind, Two Treatment, Two Period, Chronic Dosing (2 Weeks), Cross-Over, Multi-Center Study to Evaluate the Effects of PT001 and PT005 on Specific Image Based Airway Volumes and Resistance in Subjects With Moderate to Severe COPD.

Global Product Statistician [redacted]

TABLE OF CONTENTS PAGE

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
AE	Adverse Event
AT	Air Trapping
ATC	Anatomic Therapeutic Class
ATS	American Thoracic Society
BID	Bis In Die, twice daily
CFD	Computational Fluid Dynamics
COPD	Chronic Obstructive Pulmonary Disease
CSR	Clinical Study Report
CT	Computed Tomography
ECG	Electrocardiograms
eCRF	electronic Case Report Form
ERS	European Respiratory Society
FEF25-75	Forced Expiratory Flow between 25% to 75% of FVC
FEV_1	Forced Expiratory Volume In 1 Second
FRC	Functional Residual Capacity
FRI	Functional Respiratory Imaging
FVC	Forced Vital Capacity
GFF	Glycopyrronium and Formoterol Fumarate
GOLD	Global initiative for chronic Obstructive Lung Disease
HFA	Hydrofluoroalkane
IAD	Internal Airflow Distribution
IC	Inspiratory Capacity
IPD	Important Protocol Deviation
iRaw	Image based Airway Resistance
ITT	Intent-To-Treat
iVaw	Image based Airway Volume
iVlobe	Image based Lobar Volume
LUL	Left Upper Lobe
LL	Lower Lobe
LLL	Left Lower Lobe
MDI	Metered Dose Inhaler
mITT	Modified Intent-To-Treat
PFT	Pulmonary Function Test
Raw	Airway Resistance
RLL	Right Lower Lobe

Abbreviation or special term	Explanation
RML	Right Middle Lobe
RUL	Right Upper Lobe
RV	Residual Volume
sGaw	Specific Airway Conductance
siRaw	Specific Image based Airway Resistance
siVaw	Specific Image based Airway Volume
sRaw	Specific Airway Resistance
TEAE	Treatment Emergent Adverse Event
TLC	Total Lung Capacity
TP	Treatment Period
UA	Upper Airway
UL	Upper Lobe
WHO-DD	World Health Organization Drug Dictionary

AMENDMENT HISTORY

Date	Brief description of change
	N/A

1. STUDY DETAILS

1.1 Study objectives

1.1.1 Primary objectives

- To assess the effect of treatment with Glycopyrronium (GP) MDI administered twice daily (BID) and Formoterol Fumarate (FF) MDI administered BID on specific image-based airway volumes and resistance in subjects with moderate to severe chronic obstructive pulmonary disease (COPD) following chronic dosing after approximately two weeks treatment.

1.1.2 Secondary objectives

- To assess the effects of GP MDI and FF MDI on various Functional Respiratory Imaging (FRI) parameters.
- To assess the effects of GP MDI and FF MDI on lung function parameters.

1.1.3 Safety Objective

To assess the safety of GP MDI and FF MDI in subjects with moderate to severe COPD based on adverse events (AEs), and any clinically relevant findings from vital sign measurements, electrocardiograms (ECGs), physical examination findings, and clinical laboratory evaluations.

1.2 Study design

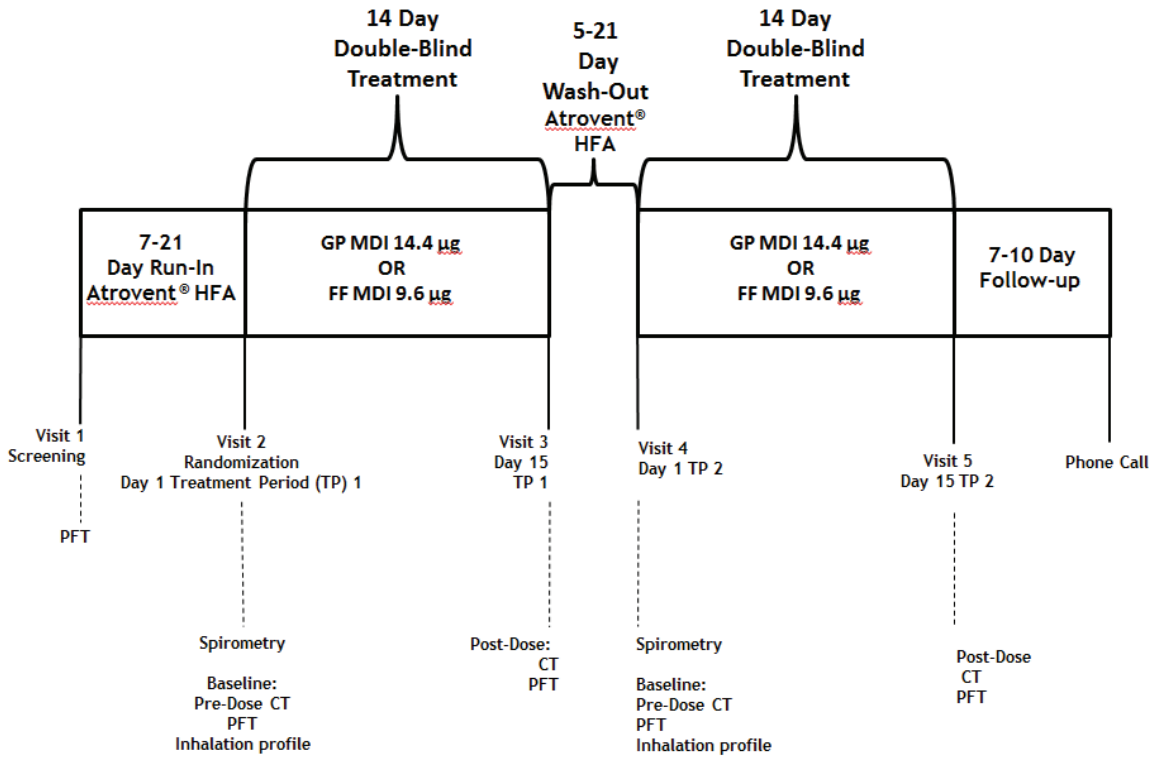
This is a multi-center, double-blind, two-treatment, two-period, full cross-over, chronic dosing (2 weeks) study to assess the effects of GP MDI (14.4 µg BID) and FF MDI (9.6 µg BID) on specific image-based airway volumes and resistance in subjects with moderate to severe COPD. In this study, airway dimension parameters will be calculated for each of the two active compounds.

This multi-center study will be conducted at two sites. Across these sites, it is planned that approximately 20 subjects with moderate to severe COPD will be randomized into this study.

Study Duration:

Each subject will receive in total study treatment for approximately four weeks (two separate Treatment Periods (TP) of approximately two weeks, separated by a washout period of 5-21 days). The entire study period is scheduled to take approximately 13 weeks for each individual subject from the time of screening through follow-up.

Study design chart



Computed Tomography (CT)-scans during study:

On Day 1 of each Treatment Period (TP [Visit 2 and Visit 4]), baseline measurement inspiratory scan (total lung capacity [TLC] scan) and expiratory scan (functional residual capacity [FRC] scan) will be conducted. During Visit 2, an additional scan of the upper airway (UA) will be taken. Post-dose measurement inspiratory scan (total lung capacity [TLC] scan) will be taken after approximately 2 weeks of treatment with either GP MDI or FF MDI on Day 15 ± 5 days (Visit 3 and Visit 5). Post-dose activities should be started 1 hour after dosing on Visit 3 and Visit 5, and should be concluded within 2.5 hours after dosing. Between the Treatment Periods there will be a washout period of 5-21 days.

1.3 Number of subjects

Approximately 20 subjects with moderate to severe COPD will be randomized into the study.

In a double-blind, cross-over study where 10 COPD Global Initiative for Chronic Obstructive Lung Disease (GOLD) III subjects received a budesonide/formoterol combination (De Backer 2012); FRI was able to quantify the effects of the active compound relative to baseline. The percentage change from baseline in airway volume was 6.48% (SD=7.46%), giving an effect

size of 0.869. Assuming a similar effect size for the improvements from baseline in airway volume parameters with active treatments in this study, a sample size calculation (power goal 90%, alpha 0.05) revealed that in order to have a well-powered study with change in imaging based airway volume as primary outcome parameter, a total of 16 subjects would be required. If the significance level for the volume is set to 0.025 by the Hochberg's step-up procedure, 20 subjects would be required. It can be assumed that sufficient power will be obtained when including 20 subjects.

2. ANALYSIS SETS

2.1 Definition of analysis sets

The following analysis populations are defined in this study:

2.1.1 Intent-To-Treat (ITT) population

The **ITT Population** is defined as all subjects who are randomized to treatment. Patients will be analysed according to the treatment they were assigned to at randomization, regardless of the treatment actually received.

2.1.2 Modified Intent-To-Treat (mITT) population

The **modified Intent-to-Treat (mITT Population)** is a subset of the ITT Population including subjects who received treatment and have post-treatment efficacy data from both Treatment Periods. Data judged to be impacted by major protocol deviations will be determined prior to unblinding and excluded. Statistical tabulations and analyses will be by randomized treatment, but data obtained after subjects receive an incorrect treatment will be excluded from the affected periods.

2.1.3 Safety population

The **Safety Population** is defined as all subjects who are randomized to treatment and receive at least one dose of the study treatment. Statistical analyses and tabulations will be by the treatment actually received.

2.1.4 Population for analyses

Analyses will be performed as follows:

> Given that the ITT, mITT and Safety Populations are expected to only differ slightly, demographics and other characteristics will be summarized only for the ITT population. Extent of exposure will be summarized for the Safety Population. The Safety Population will be used to summarize safety data.
>
> Efficacy analyses will be performed for the ITT Population. The mITT Population is a subset of the ITT Population, but if the data contributing to the mITT analyses differs from the data contributing to the ITT analyses, then the mITT Population will be used to conduct sensitivity analyses for the co-primary endpoints.

2.2 Violations and deviations

Only Important Protocol Deviations (IPDs) will be listed or tabulated. These are defined as those protocol deviations that may significantly affect the completeness, accuracy, and/or reliability of the study data or that may significantly affect a patient's rights, safety, or well-being. These will be identified based upon blinded data, before database lock.

The following will be considered important protocol deviations:

- Patient randomized despite not meeting key inclusion criteria, including:
  - Patients not meeting COPD diagnosis (as defined by ATS/ERS) and severity criteria (FEV_1/FVC ratio of <0.70 and post-bronchodilator FEV_1 >30% and <80% of predicted normal) as per inclusion criteria 5 and 7.
  - Patients meeting exclusion criteria for other respiratory disease as per exclusion criteria 4, including diagnoses of asthma.
- Patient received incorrect study drug
- Prohibited concomitant COPD medication taken during the study, defined as the classes of medications listed in table 5-1 of the protocol, received after visit 2, other than those provided by the sponsor.
- Developed discontinuation criteria but not withdrawn from study or discontinued investigational product

Randomization errors

If a subject is given a treatment pack for a different patient or is randomized out of chronological order, they will be included in the statistical analysis. If a subject would receive the same treatment twice, this will be considered an important protocol deviation and sensitivity analyses will be conducted using the mITT population.

3. PRIMARY AND SECONDARY VARIABLES

3.1 Efficacy Endpoints

3.1.1 Definition of baselines

Efficacy baseline assessments are taken pre-dose at Day 1 of each Treatment Period (TP) (Visits 2 and 4). The primary efficacy analyses involve a paired comparison to baseline, for each treatment separately, and the relevant period-dependent baseline (defined below) will be used as the Day 1 value for a subject.

Where linear mixed models are applied, the mean of the available pre-dose assessments across both visits will be taken as the **subject-average baseline.** As well, the mean of the available pre-dose assessments at each of Visits 2 and 4 respectively will be taken as the **period-dependent baseline** for that respective TP.

For the analysis of spirometry and plethysmography endpoints, the subject-average baseline will be derived as follows. For spirometry, the mean of the pre-dose -60 and -30-minute values will be obtained, at each of Visits 2 and Visit 4, then averaged over visits. In patients missing either of these assessments, the single value from that visit will be used. For plethysmography, the -30-minute assessments at Visits 2 and 4 will be obtained and averaged over these visits. Period-dependent baselines will be derived similarly based upon either Visit 2 or Visit 4 respectively for spirometry and plethysmography analyses, and used for paired tests of baseline to Day 15 within treatment.

FRI parameters allow for estimates by lobe. Subject-average baseline for individual lobes is derived for each subject as the mean of the -30-minute lobe-specific baseline (Day 1) across periods (Visits 2 and 4), and these will be used in the individual lobe descriptive statistics. Subject-average baseline "across-lobes" is derived for each subject as the mean of all available baselines across all lobes and periods, and will be used in descriptive statistics, and statistical analyses that include a baseline.

Period-dependent baseline for individual lobes is defined for each subject as the lobe-specific value on Day 1 of the respective treatment period (Visit 2 or 4). Period-dependent baseline "across lobes" is defined for each subject as the mean over all lobes on the first day of the respective treatment period (Visit 2 or 4).

For FRI parameters that allow estimates by generation, subject-average baselines will be derived as described in the preceding paragraph (using "generation" instead of "lobe"), and included in descriptive statistics. Period-dependent baselines will not be required for analyses involving generation.

3.1.2 Treatment periods and visit windows

There will be two treatment periods in this study. The first day on which a study treatment is received is Treatment Day 1 (for that study treatment and period). Treatment day is numbered sequentially thereafter until washout is begun or a subsequent study treatment is received. Each treatment period is intended to last until Day 15. Visit windows during each Treatment Period are relative to Day 1 of that Treatment Period. Washout period between Visit 3 and Visits 4 is a minimum of 5 days to a maximum of 21 days.

No study time windows will be derived for the reporting of data. Data will be reported according to the protocol-scheduled day. Efficacy data obtained during unscheduled visits after Day 1 will be allocated to Day 15 within the same treatment period if the Day 15 visit is missing. For analysis of spirometry data by time point, if multiple values are collected the last available assessment will be used for a given nominal time point.

3.1.3 Primary Efficacy Endpoints

FRI Parameters:

The PT003019 Data Transfer Agreement describes the physiology of the lung in terms of the FRI endpoints described here, and should be read in conjunction with this (PT003019) SAP.

The co-primary endpoints of this study are as follows:

- Specific Image based airway volume (siVaw): Absolute value in Airway Volume (iVaw) relative to the lobar volume at Day 1 and Day 15.
- Specific Image based airway resistance (siRaw): Absolute value in Airway Resistance relative to the lobar volume at Day 1 and Day 15.

The primary efficacy analyses involve a paired comparison to baseline (Day 1), for each treatment separately, of the Day 15 siVaw average across lobes and the siRaw average across lobes parameters.

Two breathing levels are used to generate FRI parameters at period baseline Visits 2 and 4 - TLC and FRC - while only TLC is used at Visits 3 and 5. Only data generated from the scans taken at TLC will be used for statistical analyses. Data generated at FRC will be listed and summarized for selected parameters.

For FRI parameters, data is generated within each of five lung lobes: right upper lobe, right middle lobe, right lower lobe, left upper lobe and left lower lobe. Each lobe comprises several segments (branches) that are further subdivided into typically 5-10 visible airway generations. Generation-level data (within segment within lobe) will be provided by FLUIDDA for iVaw and Mass of deposited particles. Lobe-level data can be produced by summing over all data points within a lobe for a patient. Lobe-level data will be provided for the other FRI parameters. Across-lobe summaries will consist of an average of available lobe-level data.

Because the number of generations visible on a CT scan can vary over time, a subject's airway generations will be "**trimmed**" so that the generations are the same over a subject's visits. For iVaw, trimming will be done by segment, so that for each segment, the generations for a subject's four visits (Visits 2-5) are identical. For all other FRI parameters where trimming occurs, trimming will be done within period, such that for each segment the generations for Visits 2 and 3 are identical, and the generations for Visits 4 and 5 are identical. For the analysis of iVaw itself, an "untrimmed" analysis in addition to a trimmed analysis will be undertaken (as noted later), while only the trimmed iVaw values will be used for the derivation of other FRI parameters.

As described in the Data Transfer Agreement, FRI parameters siVaw and siRaw are derived from iVaw and iRaw by dividing iVaw by the image-based lobe volume (iVlobe) for siVaw, and by multiplying iRaw by iVlobe for siRaw, respectively. iVaw and siRaw are derived for each lobular region, where: RUL (right upper lobe), RML (right middle lobe), RLL (right lower lobe), LUL (left upper lobe), LLL (left lower lobe). For example:

- siVaw LLL = iVaw LLL / iVlobe LLL,
- siRaw LUL = iRaw LUL * iVlobe LUL.

siVaw and siRaw are also derived for the larger, LL (lung, lower lobes), UL (lung, upper lobes), DISTAL (all 5 lobes), CENTRAL, and TOTAL lung regions by summing parameters using the formulae below:

Note, for airflow and resistance parameters, a result from the CENTRAL region of the lung is also provided, and this combined with the DISTAL total gives an overall TOTAL. For deriving resistance totals, parallel sums of iRaw should be used as described above, by summing the inverse of iRaw values and then inverting again (i.e., a harmonic mean).

siVaw
- iVaw LL = iVaw RLL + iVaw LLL
- iVaw UL = iVaw RUL + iVaw RML + iVaw LUL
- iVaw DISTAL = iVaw LL + iVaw UL
- iVaw TOTAL = iVaw CENTRAL + iVaw DISTAL

- siVaw LL = iVaw LL / iVlobe LL
- siVaw UL = iVaw UL / iVlobe UL
- siVaw CENTRAL= iVaw CENTRAL / iVlobe TOTAL
- siVaw DISTAL= iVaw DISTAL / iVlobe TOTAL
- siVaw TOTAL = iVaw TOTAL / iVlobe TOTAL

siRaw
- iRaw LL = 1 / (1/iRaw RLL + 1/iRaw LLL)
- iRaw UL = 1 / (1/iRaw RUL + 1/iRaw RML + 1/iRaw LUL)
- iRaw DISTAL = 1 / (1/iRaw RLL + 1/iRaw LLL + 1/iRaw RUL + 1/iRaw RML + 1/iRaw LUL)
- iRaw TOTAL = 1 / (1/iRaw CENTRAL + 1/iRaw RLL + 1/iRaw LLL + … + 1/iRaw LUL)

- siRaw LL = iRaw LL * iVlobe LL
- siRaw UL = iRaw UL * iVlobe UL
- siRaw CENTRAL = iRaw CENTRAL * iVlobe TOTAL
- siRaw DISTAL = iRaw DISTAL * iVlobe TOTAL
- siRaw TOTAL = iRaw TOTAL * iVlobe TOTAL

iVlobe
- iVlobe LL = iVlobe RLL + iVlobe LLL
- iVlobe UL = iVlobe RUL + iVlobe RML + iVlobe LUL
- iVlobe TOTAL = iVlobe RLL + iVlobe LLL + iVlobe RUL + iVlobe RML + iVlobe LUL

3.1.4 Secondary Efficacy Endpoints

3.1.4.1 FRI Parameters

- Airway volume (iVaw): Absolute value at Day 1 and Day 15

  In the CT scans, the airways can be segmented up to the point where no distinction can be made between the intraluminal and alveolar air. This is where the airway diameter is around 1 – 2 mm, typically around the 5th to 10th bifurcation, depending mainly on the disease state of the individual patient. From the resulting model, the central and distal airway volumes can be assessed at individual airways or in different regions.

The distal airway volume is defined as the segmented airway volume starting from the third bifurcation that is from the 5 lung lobes without the central region. It is this distal region that will be used in the analyses in this study.

Note that iVaw is provided at the generation-level within segment within lobe. Lobe-level data can be derived by summing over all data points (generations) within a lobe for a subject. The total for a given generation across lobes can be derived by summing over all data points (segments) within a generation for a subject.

- As mentioned previously, the visible airway generations that comprise the lobes can vary over time. The iVaw measurements can thus be performed in two ways: 1) for each segment use all the generations visible at the particular study visit scan ("untrimmed"), or 2) for each segment use only generations of airways that are visible in the scans of all four visits (Visits 2-5) ("trimmed"). Both iVaw measures will be analysed.

- Airway resistance (iRaw): Absolute value at Day 1 and Day 15

  The airway resistance (iRaw) is determined using Computational Fluid Dynamics (CFD). During the CFD calculations, the outflow to each lobe is adjusted iteratively for each subject to match the internal flow rate distributions obtained from the segmentation of the CT scans. Hence, iRaw accounts for the patient-specific internal airflow distribution which might be greatly altered by the lung disease. Hence, the airflow distribution in the CFD calculation reflects the airflow distribution as derived from the expansion of the lung lobes from FRC to TLC. The iRaw is defined as the total pressure drop over an airway, divided by the flow rate through that airway.

  For iRaw, the airways are trimmed until they are the same length for Visits 2 and 3, and for Visits 4 and 5, respectively, for a subject. We thus only assess the generations of airways that are visible in the scans of both Visits 2 and 3, and both Visits 4 and 5, respectively.

3.1.4.2 Spirometry Parameters

- Forced expiratory volume in one second (FEV_1) at Day 1 and Day 15. Change from baseline to Day 15 in post-dose FEV_1.

  FEV_1 is the volume of air exhaled under forced conditions in the first second. Post-dose results will be used. Change from baseline is calculated relative to the subject-average baseline.

3.1.4.3 Body Plethysmography Parameters

- Functional residual capacity (FRC) at Day 1 and Day 15. Change from baseline to Day 15 in post-dose FRC.

- FRC is the volume remaining in the lungs at the end-expiratory position. Change from baseline is calculated relative to the subject-average baseline.

3.1.5 Other Efficacy Endpoints

3.1.5.1 FRI Parameters

The following parameters will be analysed formally using scans at TLC with the exception of the mass of deposited particles, Internal Airflow Distribution and Air Trapping, which will be summarized descriptively only because they are dependent on FRC, which is collected only on Day 1.

- Image-based Lobe volumes (iVlobe): Absolute value at Day 1 and Day 15

  By identifying and grouping the voxels that represent the air in the lungs the lung volume (L) can be determined from the scans. During segmentation, identifying the fissure planes on the CT images and using these surfaces as cutting objects can separate lung lobes. This means that not only the total lung volume is determined, but also the volume of each lobe individually.

  - For each treatment, only scans at the TLC level are taken after the treatment. That means that analyses can be performed only on TLC lobar volumes but not for FRC. For the FRC level, only descriptive summary of baseline values will be presented.

- Percent predicted image-based lobe volume [based upon age, sex and height] (PP iVlobe) This will also be derived and analysed similarly to iVlobe using scans at the TLC level and summarized at baseline using scans at the FRC level.

- Air trapping (AT): Descriptive summary of baseline results

  Air trapping is defined as all the intrapulmonary voxels with Hounsfield Units between -1024 and -850 using the expiratory scans at FRC. This parameter is given as percentage of iVlobe.

  Air Trapping is only valid on FRC scans, so because there are no FRC scans after treatment, no analyses can be done, only descriptive summary of baselines values are provided. The air trapping will also be derived for the upper lobes, lower lobes and total by summing data from relevant individual lobes.

- Internal lobar airflow distribution (IAD): Descriptive summary of baseline results

  By segmenting the lobes at FRC and TLC for each patient, the patient-specific airflow distribution can be established by assessing lobar volume expansion.

Internal Airflow Distribution requires both FRC and TLC scans, so because there are no FRC scans after treatment, no analyses can be done. Only descriptive summary of baselines values will be produced.

- Low attenuation or emphysema score (LAS): Absolute value at Day 1 and Day 15

  Low attenuation score is defined as all the intrapulmonary voxels with Hounsfield Units between -1024 and -950 using the inspiratory scans at TLC. This parameter is given as percentage of iVlobe.

- Blood vessel density or fibrosis score (iVbv): Absolute value at Day 1 and Day 15

  Blood vessel density can be determined through segmentation and three-dimensional reconstruction of the blood vessels. The segmentation is based on a Hounsfield unit (HU) threshold between -600 and 600 and is performed on the TLC scan. The blood vessel density can be considered a surrogate for perfusion. This parameter is given as percentage of iVlobe.

- Airway wall thickness (iVaww): Absolute value at Day 1 and Day 15

  The airway wall volume (also called airway wall thickness) consists of all visible tissue in the CT scan that encompasses the airway wall. The airway wall volume can typically be described to the same generation level as the volume description of the airway lumen; this is where the airway diameter is around 1 - 2 mm. Visible airway generations can be different over repeated scans. For airway wall thickness, the airways are trimmed until they are the same length for Visits 2 and 3, and for Visits 4 and 5, respectively, for a subject. We thus only assess the generations of airways that are visible in the scans of both Visits 2 and 3, and both Visits 4 and 5, respectively.

- Mass of deposited particles per defined airway section: Descriptive summary of baseline results

  Regional aerosol deposition is determined by simulating the flow in the patient specific geometries using patient specific boundary conditions by means of CFD. While solving the flow equations, simultaneously particles are released in the flow and the force mass balance of the individual particles is determined through additional discrete phase computations. When a calculated particle trajectory intersects with the airway wall, the particle is trapped in that location. This allows determining the regional concentration of inhaled aerosols and consequently the effective lung dose of inhaled medication. These are measured in the full airway tree.

  Mass of deposited particles is a useful way to see the particle deposition in each region of the lung. Data from the baseline TLC scan will be summarized. Distal and peripheral data will be summed. Upper lobe and lower lobe totals will also be derived. However, for the analysis of mass of deposited particles by generation number, any distal or peripheral data corresponding to a given generation number will be employed.

A given generation within a segment within a lobe will be either distal or peripheral. Thus, no summing is needed of distal and peripheral values for the by-generation analyses.

FRI parameters - Low attenuation or emphysema score (LAS), Air Trapping (AT), and Blood vessel density or fibrosis score (iVbv), will be provided by FLUIDDA, while Internal Airflow Distribution (IAD) and some of iVaww - will need to be calculated. Further detail can be found in the Data Transfer Agreement.

Internal Airflow Distribution (IAD)

- IAD LLL = 100 * (iVlobe LLL TLC - iVlobe LLL FRC) / (iVlobe TOTAL TLC - iVlobe TOTAL FRC)
- IAD LUL = 100 * (iVlobe LUL TLC - iVlobe LUL FRC) / (iVlobe TOTAL TLC - iVlobe TOTAL FRC)
- IAD RML = 100 * (iVlobe RML TLC - iVlobe RML FRC) / (iVlobe TOTAL TLC - iVlobe TOTAL FRC)
- IAD RUL = 100 * (iVlobe RUL TLC - iVlobe RUL FRC) / (iVlobe TOTAL TLC - iVlobe TOTAL FRC)
- IAD RLL = 100 * (iVlobe RLL TLC - iVlobe RLL FRC) / (iVlobe TOTAL TLC - iVlobe TOTAL FRC)
- IAD LL = IAD RLL + IAD LLL
- IAD UL = IAD RUL + IAD RML + IAD LUL

Airway wall thickness (iVaww)

- iVaww LL = iVaww RLL + iVaww LLL
- iVaww UL = iVaww RUL + iVaww RML + iVaww LUL
- iVaww DISTAL = iVaww RLL + iVaww LLL + iVaww RUL + iVaww RML + iVaww LUL
- iVaww TOTAL = iVaww CENTRAL + iVaww DISTAL

For FRI parameters the values for each lobe, the lower lobe total, upper lobe total, and overall total will be listed.

3.1.5.2 Spirometry Parameters

The following parameters will be assessed at baseline (Day 1) and Day 15. The change from baseline at Day 15 post-dose will be calculated relative to the subject-average baseline.

- Forced vital capacity (FVC): the determination of the vital capacity from a maximally forced expiratory effort
- Tiffeneau index (FEV_1/FVC ratio): the proportion of a person's vital capacity that they are able to expire in the first second of forced expiration
- Forced expiratory flow 25%-75% (FEF25-75): forced expiratory flow at 25–75% of forced vital capacity

- Inspiratory capacity (IC): the amount of air that can be inhaled after the end of a normal expiration. The average of the available Day 15 post-dose assessments will be used.

3.1.5.3 Body Plethysmography Parameters

The following parameters will be assessed at baseline (Day 1) and Day 15. The change from baseline at Day 15 will be calculated relative to the subject-average baseline.

- Residual volume (RV): the volume of air remaining in the lungs after a maximal exhalation
- Total Lung Capacity (TLC): the volume in the lungs at maximal inflation
- Airway resistance (Raw): the resistance of the respiratory tract to airflow during inspiration and expiration
- Specific airway resistance (sRaw): airway resistance corrected for differences in lung volume.
- Specific airway conductance (sGaw): derived as the mathematical inverse of airway resistance corrected for differences in lung volume.

3.2 Safety Endpoints

The safety assessments include Adverse Events (AEs) and Serious Adverse Events (SAEs) during the study period.

AEs experienced by the subjects will be collected throughout the entire study and will be coded using the latest version of the Medical Dictionary for Regulatory Activities (MedDRA).

Only treatment emergent adverse events from the time of randomization onwards will be summarized. AEs will be assigned to the treatment period in which they first occurred. AEs reported as starting during a Washout Period or follow-up period will be assigned to the last treatment received. If an AE has a missing or partial onset date, then unless the stop date of the AE indicates otherwise, this will be considered an on-treatment event.

Clinically relevant findings from laboratory, ECG and vital sign parameters will be reported as adverse events. As such these assessments will not be further tabulated.

3.3 Handling of missing data

The lung comprises 5 lobes which are each divided into further segments. The airway continues to divide into smaller airways and each branching point defines a new generation of the airway. Data for FRI parameters will be provided for each subject for each lobe of the lung, and, for some parameters, for each generation within each segment within each lobe of

the lung. Descriptive statistics and analyses may require that a subject's data are summarized by lobe, across lobes, by generation, or across generations.

There may be missing data for certain lobes, if for example the subject is missing that lobe, or no airways lead to the lobe (giving a volume of zero and infinite airway resistance). Similarly, there may be missing segments for a lobe, or generations within a segment, as not all airway generations are observable in each CT scan. Some generations may not be present, or their diameter may be too small to be evaluable. It is assumed that in such cases the contribution of this generation to the total airway volume of the segment or to the total for that generation across lobes is negligible.

A segment's value will be calculated by summing its available generation values. For the calculation of the lobe's values, if there is a missing segment, the calculation will consist of the sum of the other present segments. Data from all individually evaluable lobes will be used in the analysis to allow for the estimation of effects at the subject-level for each lobe, and across all lobes.

Similarly, data from individually evaluable generations will be summed in the generational analysis to allow for estimation for each generation, and across generations. Sensitivity analyses are described in section 4.2.4 for by-generation analyses imputing small values for patients with no evaluable data at later generations.

If, for a subject, a particular lobe has missing data, then summed totals of lobes will not be computed for that subject (for regions to which the missing lobe would belong).

Where scans are not evaluable for a single assessment or period, this data is assumed to be missing at random. Where Day 15 assessments are missing, data from unscheduled post-baseline assessments will be used in their place, if available. No further imputation for entirely missing assessments is planned.

Combined or indistinguishable lobes

If fissures between lobes cannot be distinguished, then only a combined estimate across multiple lobes may be produced. Such results from combined lobes will not be used in the derivation of averages (Baseline, Day 15, or across lobes) or in analyses that require the use of individual-lobe values (e.g. repeated measures analyses for which lobe is the repeated measure, and any within-lobe analyses). This will be consistent for both periods – i.e. if for a subject, certain lobes are combined for one period, then they must be combined for the other period too.

Such combined lobe results may however enable totals and sub-totals to be calculated for a subject, if the lobes which are combined coincide with those required for relevant regions (for which a summed total of the affected lobes would be needed).

For subjects with lobes that cannot be distinguished, data for that subject's lobe will be considered unevaluable and will not contribute to the lobe analysis. This rule will be applied for both periods.

4. ANALYSIS METHODS

The primary objective of this study is to assess the effect of treatment with Glycopyrronium (GP) MDI administered twice daily (BID) and Formoterol Fumarate (FF) MDI administered BID on specific image-based airway volumes and resistance in subjects with moderate to severe chronic obstructive pulmonary disease (COPD) following chronic dosing after approximately two weeks treatment.

Additional analyses will compare the distribution of regional effects between the two treatments for the different FRI endpoints. The magnitude of differences in effects between treatments overall, locally (by lobe), and in terms of central vs. peripheral effects (i.e. by generation) will be estimated. As such, treatment effects will also be estimated between treatments overall ("across lobes"), locally (by lobe), and by generation (where available), and it will be investigated whether there is any evidence for differential patterns in the effects of each treatment.

4.1 General principles

4.1.1 Treatment Effect Assessment

For the efficacy analyses, GP MDI and FF MDI will be assessed for Day 15 effects. P-values for differences from baseline, and between treatments, as well as for tests of treatment interaction, will be reported as two-sided. For the co-primary endpoints, the other FRI parameters, as well as spirometry and plethysmography parameters, the Day 15 summary statistics for each treatment period will be reported and analysed.

The Intent-to-Treat (ITT) Population will be considered the primary analysis population for efficacy. Supportive analyses will be conducted for the modified Intent-to-Treat (mITT) Population, if the mITT Population is significantly different from the ITT Population.

Continuous efficacy variables will be summarized with descriptive statistics n (the number of non-missing values, mean, standard deviation, median, minimum and maximum). Where data have been logarithmically transformed for analysis, the summary statistics on the back-transformed data will include the geometric mean and the coefficient of variation (calculated as $100x\sqrt{[\exp(s^2)-1]}$, where s is the standard deviation of the data on the log scale. Categorical variables will be summarized with frequency counts and percentages, by treatment.

4.1.2 Hypothesis Testing

For the efficacy analyses, the individual effects of each of GP MDI and FF MDI respectively, will be assessed. The effect is defined as the difference between the value at Day 15 and the value at Day 1. Effect is thus the change from baseline to Day 15, for a GP MDI treatment, and the change from baseline to Day 15, for a FF MDI treatment.

Denote the FRI value on Day 1 by FRI_1 and FRI on Day 15 by FRI_{15}.The difference between FRI_{15} and FRI_1 for each patient will be calculated as $FRI_{DIFF} = FRI_{15} - FRI_1$ for the within-treatment comparison between Day 1 and Day 15. The hypotheses for the respective "per-treatment" comparisons are:

H_o: $FRI_{DIFF} = 0$, for GP MDI (no effect of treatment)
H_1: $FRI_{DIFF} \neq 0$, for GP MDI

H_o: $FRI_{DIFF} = 0$, for FF MDI (no effect of treatment)
H_1: $FRI_{DIFF} \neq 0$, for FF MDI.

The hypothesis that $FRI_{DIFF} = 0$ will be tested using a paired t-test and a two-sided p-value $\leq$ 0.05 will be regarded statistically significant.

Additional analyses will compare the distribution of regional effects between the two treatments for the different FRI endpoints. These treatment comparisons will not be considered to be formal hypothesis testing, but as estimation and investigation of regional treatment effects.

4.1.3 Control of Type I error

For the primary efficacy endpoints, Hochberg's step-up procedure will be used as a multiplicity adjustment to control the type I error within each product at 5%. Specifically, Hochberg's procedure will be applied once for the siVaw and siRaw endpoints, for GP MDI, and then applied separately again for the same endpoints for FF MDI. No correction will be performed for the secondary or other endpoints or for between treatment comparisons. P-values outside of the Hochberg procedures will be interpreted at the two-sided 5% level.

4.1.4 Software

All statistical analysis will be conducted using SAS version 9.2 or higher, or other validated software as appropriate.

4.2 Analysis methods

4.2.1 Disposition, demographics and baseline characteristics

Descriptive summaries of disposition, demographics and baseline characteristics will be produced. Disposition tables will summarize the number of patients who received each treatment, the number who completed 1 or 2 treatment periods and the number of early discontinuations. The number of patients in each analysis set will be summarized along with any reasons for exclusion.

Demographics and baseline characteristics at screening will be summarized descriptively, including COPD disease duration, symptom severity scores. Baseline pre and post bronchodilator spirometry parameters (and their percent predicted values) and reversibility to Ventolin HFA will be summarized using descriptive statistics based upon the baseline assessments within each treatment period. Percent predicted residual volume (RV) and percent

predicted functional residual capacity (FRC) will also be summarized. Predicted values for FEV1 will be calculated using The Third National Health and Nutrition Examination Survey (NHANES III) (Hankinson 1999) reference equations. Predicted values for RV and FRC will be calculated using Quanjer (1993).

4.2.2 Concomitant medications

Concomitant medications will be assigned a preferred term and an ATC (Anatomic Therapeutic Class) term using the latest version of the World Health Organization Drug Dictionary (WHO-DD) available. Medications will be summarized according to whether they are COPD related or non-COPD related. Summaries will be presented for those medications started prior to the beginning of study treatment, those started during a treatment period and those started during the washout period.

4.2.3 Exposure

The number of days of exposure to each treatment will be defined as ((End date of treatment – Date of first dose of treatment) + 1).

Exposure to IP will be summarized by treatment using the safety analysis set.

4.2.4 Analysis of FRI data

Primary endpoint analyses (siVaw, siRaw)

Within-Treatment Comparison of Day 1 and Day 15

The primary efficacy analyses will consist of a within-treatment comparison of Day 1 and Day 15 using a paired t-test for each primary endpoint and for each treatment. A paired t-test will compare the Day 15 siVaw in the GP MDI treatment period to the corresponding Day 1 value for that period; the same will be done for siRaw. Paired t-tests will be similarly applied for FF MDI.

Linear Mixed Effect Models

As secondary analyses, the Day 15 value for each parameter in each period will be analyzed using a linear mixed effect model. It is anticipated that the data will be logarithmically transformed prior to analysis. A multi-level model will be used to incorporate the repeated measurements from the lobes for each subject.

The model will include fixed effects for period, treatment, lobe, and treatment-by-lobe interaction. The model will not include treatment sequence unless that term is determined to be important ($p<0.10$). Lobe will be included as a random effect, within each subject. Heterogeneity across lobes (within subject) will be modelled using an unstructured variance-covariance matrix, with independence assumed between subjects.

If this model fit fails to converge, a compound symmetry covariance structure will be considered to model correlation across lobes from the same patient.

Estimates will be produced of the difference between treatments, for each lobe and across all lobes. Where logarithmic transformations have been made, estimates will be exponentiated and treatment effects presented as ratios. Estimates and standard errors will be reported along with 2-sided 95% confidence intervals and p-values. The Kenward-Roger approximation for degrees of freedom and correction of downward bias in the standard error of fixed effect parameters will be used. The p-values for the treatment effect and the treatment-by-lobe interaction will be reported across-lobe.

Line-plots will be produced displaying Day1 and Day 15 absolute values by treatment and sequence at the individual subject level using the mean of data from available single lobes for each subject.

Secondary and other efficacy endpoint analyses for FRI

Similar analyses to the primary endpoint analysis will be conducted for secondary and other efficacy endpoints, except for Mass of deposited particles, Internal Airflow Distribution (IAD), and Air Trapping (AT), whose analyses are described under generation-level analyses, Mass of Deposited Particles, and IAD and AT analyses respectively.

Sensitivity Analyses

If the mITT population differs from the ITT population, the mITT population will be used to conduct sensitivity analyses of the mixed models for the co-primary endpoints. The mITT population will only include patients with data from both periods without important protocol deviations. Similar mixed models will be used.

For the primary endpoints only, a further model will be produced which includes covariates for baseline. The model will include fixed effects for period, treatment, lobe, and treatment-by-lobe interaction, subject-average baseline by lobe and period-dependent baseline by lobe. Both versions of baseline are continuous covariates. The use of both subject-average and period-dependent baselines allows for different coefficients for the between and within subject covariate regressions (Kenward and Roger 2010). Period-dependent baselines will be summarized for the primary endpoints. The p-values for the treatment effect and the treatment-by-lobe interaction will be reported across-lobes.

IAD and AT Analyses

For Internal Airflow Distribution (IAD) and Air Trapping (AT) endpoints, no statistical analyses will be undertaken. Descriptive summary statistics will be produced for these endpoints in each lobe of the subject's lungs at baseline (Day 1).

Mass of Deposited Particles

Mass of deposited particles is a useful way to see the particle deposition in each lobe of the lung. Summaries will be provided for the deposition in each lobe of the patient's lungs at

baseline TLC, as well as for the upper lobe and lower lobe totals. Distal and peripheral data will be summed. Only descriptive summaries will be produced.

Generation-level Analyses

Two endpoints, iVaw and Mass of deposited particles, will have data produced by generation-level and not lobe-level data. The number of generations per subject may differ, and a typical airway model includes 5-10 generations, depending mainly on the disease state of the subject. For each segment, the generations available for each subject-visit will be reviewed and agreed upon in a blinded manner prior to unblinding of the database. [For further details regarding the numbering of the generations, please refer to the data transfer specifications.]

Mixed models of iVaw by generation will be conducted using the total across all segments for a given generation number so that each subject has at most one data point per generation. Several analyses (1, 2, 3, and 4 below) will be conducted to apply different methods to handle the fact that each subject potentially has a different number of generations with evaluable data.

If the model fit fails to converge, a compound symmetry covariance structure will be considered to model correlation across generations from the same patient.

In an analysis called "*repeated measures through generation H and beyond*" (**Analysis 1**), for each subject-visit, the number of generations available for each segment within each lobe will be determined and numbered separately, from 1, 2, 3, etc. The maximum generation number will then be determined across the lobe-specific segments, for each subject-visit. Using only the maximum generation numbers from Visits 3 and 5, the smallest value will be determined across all subjects, and will be denoted as H. H is the highest generation number common to all subjects (amongst Visits 3 and 5). Where a subject has generations higher than H, the parameter values will be aggregated (summed) across those generations and added to the value for generation H, and treated as generation "≥H".

In an analysis called "*repeated measures through generation H*" (**Analysis 2**), parameter values from generations beyond generation H will not be employed, and the final generation included in the analysis will be generation H and will be called generation "H".

Let K denote the largest generation number observed for at least half of the subjects (again, using only Visits 3 and 5). Let m-iVaw denote the smallest iVaw that is observed for any subject for any visit for any generation. For a log-transformed analysis, impute any unobservable generation for a subject with the value of m-iVaw. The analysis of such data will be called "*repeated measures through generation K: transformed*" (**Analysis 3**).

For an untransformed analysis, impute any missing generation for a subject with the value of zero for iVaw. The analysis of such data will be called "*repeated measures through generation K: untransformed*" (**Analysis 4**). Analysis 4 will only be conducted if the data distribution allows.

Note the values of "H" and "K", once determined prior to data unblinding, will be used in table titles, rather than "H" and "K" themselves.

For each repeated-measures iVaw analysis (Analyses 1, 2, 3, and 4), a repeated measures model will include fixed effects for period, treatment, generation and treatment-by-generation interaction. The model will not include treatment sequence unless that term is determined to be important ($p<0.10$). Generation will be included as a random effect, within each subject. Heterogeneity across generations (within subject) will be modelled using a Toeplitz variance-covariance matrix, with independence assumed between subjects. The p-values for the treatment effect and the treatment-by-generation interaction will be reported across-generations. Analyses 1, 2, and 3 will have logarithmic transformation. Analysis 4 will not. Analyses 1-4 will be conducted separately for trimmed and untrimmed data.

A graphical display of model-based estimates for iVaw by generation for each treatment will be implemented. The estimated geometric mean Untrimmed Airway Volume (iVaw) will be plotted along with 95% confidence interval bars by generation number (on the horizontal axis). The estimated geometric mean Trimmed Airway Volume (iVaw) will also be plotted, in a separate display.

For mass of deposited particles, no statistical analyses will be undertaken. Descriptive summary statistics will be produced showing the deposition in each generation of the subject's lungs at baseline (Day 1).

4.2.5 Spirometry / Body Plethysmography data

For spirometry and body plethysmography parameters, data are acquired at the subject-level and not by lobe. Paired t-tests will be used for within-treatment comparisons of Day 1 and Day 15. Associated estimates of the change from period-dependent baseline to Day 15 will be reported for each treatment, with a 95% confidence interval and p-value.

For comparisons between treatments, the change from baseline to Day 15 for each endpoint will be analysed using a linear mixed effect model including subject-average baseline as a continuous covariate and treatment and period as fixed effects. The model will not include treatment sequence unless that term is determined to be important ($p<0.10$). Subject will be modelled as a random effect. The ITT population will be used. It is expected that plethysmography endpoints may require logarithmic transformation, but that spirometry endpoints will not. Estimates of the difference between treatments, with 95% confidence interval and p-value, will be reported.

See Appendix 8.1 for a consolidated list of efficacy analyses and their essential differences for the FRI parameters, spirometry, and plethysmography.

4.2.6 Assumptions Checks and Removal of Outliers in Sensitivity Analyses

It is expected that logarithmic transformations will be required for most FRI and plethysmography endpoints, with the exception of the percent predicted lobe volume. Data will be transformed via the natural logarithm prior to analysis and the adjusted mean estimates

and confidence intervals from the linear models will be exponentiated back for presentation. As such the treatment effect for these endpoints will be presented as a ratio.

Hypotheses may be tested by means of the Wilcoxon signed rank test, in addition to the paired t-test, if the distribution of the data warrants it.

Under certain circumstances, (eg, during a COPD exacerbation unrelated to treatment) extreme and atypical values can arise. Such values may disproportionately affect model-based estimates of the fixed effect and variance parameters. Depending on the plots of the standardized residuals versus fitted values of the mixed-effect models, other data transformations may be considered. If necessary, this may include the fifth root, adapted for positive and negative values: fifth root (x) = sign (x) * (abs (x)) ^ (1/5).

For resistance measures, outliers can occur that are real values but affect the model estimates extremely. In these cases, values that are larger than the third quantile + 3 times the interquartile range may be removed. Infinite values for resistance are possible and these will be listed, but will be removed from the data before statistical analysis. If erroneous values are detected, every effort will be made to correct them prior to database lock. However, if these values cannot be corrected, they will be considered for removal from the analysis.

4.2.7 Exploratory analyses

To gain insights into the mode of action of the product, additional exploratory analyses may be executed using (robust) linear regression or mixed models, and these may be reported outside of the CSR. These may entail post-hoc tests on the previously generated mixed-effect model. A regression of airway volume (iVaw), or change from baseline in airway volume, on Mass of deposited particles may be considered, to assess any correlation.

Pearson or Spearman's rank correlations between variables may be produced (depending upon behaviour of distributions). If appropriate, exploratory analyses of mixed-effect regressions or linear regression between parameters may be conducted to investigate associations between parameters, but these will be reported outside of the CSR.

The results of the PT003019 study can be combined with the results of the PT003018 (EudraCT: 2015- 001743-36) for further examination. However, this would be the subject of a separate SAP.

4.2.8 Safety data

Adverse events will be summarized by the number and percentage of subjects experiencing an event. Tables will show the overall incidence of AEs, and the incidence for each treatment. Treatment emergent adverse events from the time of randomization onwards will be summarized. AEs reported as starting during a Washout Period or follow-up period will be assigned to the last treatment received.

AEs will be presented according to MedDRA preferred term and system organ class. Summaries will be produced by severity, seriousness, AEs leading to discontinuation, and by causality assessment to study drug. No hypothesis tests will be performed.

Clinically relevant findings from laboratory, ECG and vital sign parameters will be reported as adverse events. As such, these assessments will not be further tabulated.

5. INTERIM ANALYSES (NOT APPLICABLE)

No interim analyses are planned in this study.

6. CHANGES OF ANALYSIS FROM PROTOCOL

Similar to the effect described in section 9.3.1 of the protocol, a paired comparison to baseline will be implemented. However, for FRI parameters, regional (lobe and generation) models to compare treatments will use Day 15 values.

For spirometry and plethysmography, models to compare treatments will use the change to Day 15 (either as change from baseline, or ratio to baseline).

Given that the ITT, mITT and Safety Populations are expected to only differ slightly, demographics and other characteristics will only be summarized for the ITT population. The mITT population will purely be used for sensitivity analyses of the primary endpoints should the ITT and mITT populations differ.

Statistical analyses will be done primarily with SAS (not R).

7. REFERENCES

De Backer L, Vos W, De Backer J, Van Holsbeke C, Vinchurkar S, De Backer W. The acute effect of budesonide/formoterol in COPD: a multi-slice computed tomography and lung function study. Eur Respir J. 2012 Aug 1;40(2):298–305.

Kenward MG, Roger JH. The use of baseline covariates in crossover studies. Biostatistics 2010, 11, 1, 1–17

Hankinson JL, Odencrantz, JR, et al. Spirometric reference values from a sample of the general U.S. population. Am J Respir Crit Care Med 1999; 159: 179-187

Quanjer, Ph., Tammeling, et al. Lung volumes and forced ventilatory flows. European Respiratory Journal - Eur Resp J. 1993; 6. Suppl. 16 5-40.