



---

**Statistical Analysis Plan**

Study Code D5970C00002

Edition Number 2.0

Date 08 Jun 2018

---

---

**A Randomised, Double-Blind, Double-Dummy, Multicentre, Parallel Group Study to Assess the Efficacy and Safety of Glycopyrronium/Formoterol Fumarate fixed-dose combination relative to Umeclidinium/Vilanterol fixed-dose combination over 24 Weeks in Patients with Moderate to Very Severe Chronic Obstructive Pulmonary Disease (AERISTO)**

---

---

**A Randomised, Double-Blind, Double-Dummy, Multicentre, Parallel Group Study to Assess the Efficacy and Safety of Glycopyrronium/Formoterol Fumarate fixed-dose combination relative to Umeclidinium/Vilanterol fixed-dose combination over 24 Weeks in Patients with Moderate to Very Severe Chronic Obstructive Pulmonary Disease (AERISTO)**

---

Study Statistician

REDACTED

REDACTED

REDACTED

Date

---

**A Randomised, Double-Blind, Double-Dummy, Multicentre, Parallel Group Study to Assess the Efficacy and Safety of Glycopyrronium/Formoterol Fumarate fixed-dose combination relative to Umeclidinium/Vilanterol fixed-dose combination over 24 Weeks in Patients with Moderate to Very Severe Chronic Obstructive Pulmonary Disease (AERISTO)**

---

Global Product Statistician

REDACTED

REDACTED

Date

<b>TABLE OF CONTENTS</b>	<b>PAGE</b>
TITLE PAGE.....	1
SIGNATURE OF STUDY STATISTICIAN.....	2
SIGNATURE OF GLOBAL PRODUCT STATISTICIAN.....	3
TABLE OF CONTENTS.....	4
LIST OF ABBREVIATIONS.....	7
AMENDMENT HISTORY.....	9
1. STUDY DETAILS.....	11
1.1 Study objectives.....	11
1.1.1 Primary objective.....	11
1.1.2 Secondary objectives.....	11
1.1.3 Safety objectives.....	12
1.1.4 Exploratory objectives.....	12
1.2 Study design.....	12
1.3 Number of subjects.....	13
2. ANALYSIS SETS.....	14
2.1 Definition of analysis sets.....	14
2.1.1 Efficacy analysis sets.....	14
2.1.2 Safety analysis sets.....	15
2.1.3 Other analysis sets.....	15
2.2 Violations and deviations.....	15
2.2.1 Important protocol deviations.....	15
2.2.2 Other deviations.....	17
3. PRIMARY AND SECONDARY VARIABLES.....	18
3.1 General Definitions.....	18
3.1.1 Definition of baseline.....	18
3.1.2 Visit windows.....	19
3.2 Pulmonary Function Outcomes.....	21
3.2.1 Morning pre-dose trough FEV <sub>1</sub> .....	22
3.2.2 Peak FEV <sub>1</sub> and Inspiratory Capacity post-dosing.....	22
3.2.3 Increase in FEV <sub>1</sub> at 5 minutes post-dosing.....	23
3.3 Patient Reported Outcomes.....	23
3.3.1 Transition Dyspnea Index (TDI) focal score.....	23
3.3.2 Early Morning Symptoms COPD Instrument (EMSCI) and Night Time Symptoms COPD Instrument (NiSCI).....	24

3.3.3	COPD Assessment Test (CAT) score .....	24
3.3.4	Daily rescue medication use .....	25
3.3.5	EuroQol 5 Dimensions 5 Levels Questionnaire (EQ-5D-5L) .....	25
3.3.6	Healthcare Resource Utilisation .....	26
3.3.7	Work Productivity and Activity Impairment – General Health (WPAI-GH) .....	27
3.4	Safety and Tolerability .....	28
4.	ANALYSIS METHODS .....	28
4.1	General principles .....	28
4.1.1	Statistical considerations .....	29
4.2	Analysis methods .....	30
4.2.1	Patient disposition .....	30
4.2.2	Demography data and patient characteristics .....	30
4.2.3	Prior and concomitant medications .....	31
4.2.4	Extent of exposure .....	31
4.2.5	Compliance .....	32
4.2.6	Analysis of the primary variables .....	33
4.2.6.1	Trough FEV <sub>1</sub> .....	33
4.2.6.2	Peak FEV <sub>1</sub> .....	35
4.2.6.3	Hierarchical testing strategy – primary endpoints .....	36
4.2.7	Analysis of the secondary variables .....	37
4.2.7.1	Hierarchical testing strategy – secondary endpoints .....	38
4.2.7.2	Pulmonary Function over 24 weeks .....	39
4.2.7.3	Pulmonary Function at Day 1 .....	40
4.2.7.4	Transition Dyspnea Index (TDI) Focal Score .....	40
4.2.7.5	Early Morning Symptoms COPD Instrument (EMSCI) and Night Time Symptoms COPD Instrument (NiSCI) .....	41
4.2.7.6	COPD Assessment Test (CAT) score .....	42
4.2.7.7	Daily Rescue Medication Use .....	42
4.2.8	Safety and tolerability .....	43
4.2.9	Exploratory analysis .....	44
5.	INTERIM ANALYSES .....	44
6.	CHANGES OF ANALYSIS FROM PROTOCOL .....	44
7.	REFERENCES .....	45
8.	APPENDIX .....	46

## LIST OF TABLES

Table 1	Visit windows for assessments conducted at every visit of Visit 4- Visit 7 .....	19
Table 2	Time windows for FEV <sub>1</sub> and FVC assessments .....	20

Table 3	Time windows for IC assessments.....	20
Table 4	Time period for symptom scores and rescue medication usage .....	21
Table 5	Statistical analyses and analysis populations.....	37

## **LIST OF FIGURES**

Figure 1	Study flow chart.....	13
Figure 2	Testing hierarchy for primary endpoints.....	36
Figure 3	Sequential testing hierarchy for secondary endpoints.....	39

## LIST OF ABBREVIATIONS

<b>Abbreviation or special term</b>	<b>Explanation</b>
AE	Adverse Event
ATC	Anatomical Therapeutic Chemical
ATS/ERS	American Thoracic Society/European Respiratory Society
AUC	Area Under Curve
BDI	Baseline Dyspnea Index
CAT	COPD Assessment Test
CRF	Case Report Form
CI	Confidence Interval
COPD	Chronic Obstructive Pulmonary Disease
CSP	Clinical Study Protocol
DAE	Adverse Event Leading to Treatment Discontinuation
EMSCI	Early Morning Symptoms COPD Instrument
ePRO	electronic Patient Reported Outcomes device
EQ-5D-5L	EuroQol 5 Dimensions 5 Levels Questionnaire
FEV <sub>1</sub>	Forced Expiratory Volume in 1 Second
FVC	Forced Vital Capacity
GFF	Glycopyrronium/ Formoterol Fumarate
GOLD	Global Initiative for Chronic Obstructive Lung Disease
IC	Inspiratory Capacity
ICS	Inhaled Corticosteroids
IP	Investigational Product
IVRS	Interactive Voice Response System
IWRS	Interactive Web Response System
LABA	Long-Acting $\beta$ 2-Agonist
LAMA	Long-Acting Muscarinic Antagonist
MAR	Missing At Random
MCID	Minimum Clinically Important Difference
MDI	Metered Dose Inhaler
MedDRA	Medical Dictionary for Regulatory Activities
NiSCI	Night Time Symptoms COPD Instrument
MCMC	Monte Carlo Markov Chain
MI	Multiple Imputation
MMRM	Mixed effects Model Repeated Measures

---

<b>Abbreviation or special term</b>	<b>Explanation</b>
MNAR	Missing Not at Random
PN	Predicted Normal
PT	Preferred Term
SAE	Serious Adverse Event
SD	Standard Deviation
SOC	System Organ Class
TDI	Transition Dyspnea Index
TEAE	Treatment Emergent Adverse Event
UV	Umeclidinium/Vilanterol
WHO-DD	World Health Organisation Drug Dictionary
WPAI-GH	Work Productivity and Activity Impairment - General Health

---



## AMENDMENT HISTORY

Date	Brief description of change
08 Jun 2018	<p>Section 2.2.1: Small edits to the definition of important protocol deviation 2; small edits to add a location for information detailing data removal due to severe site misconduct or duplicated patients.</p> <p>Section 2.2.2: Additional data exclusion conditions added to exclude pre-dose spirometry assessments not within required window since GFF (active or placebo) dosing in the evening of the day prior to visit or UV (active or placebo) dosing in the morning of the day prior to visit from per-protocol analyses.</p> <p>Section 3.1.1: ePRO baseline definition updated to account for delays between ePRO and spirometry completion at Visit 3.</p> <p>Sections 3.1.1 and 3.1.2: Updated to set the aggregated value as missing if at least half of the daily measures/scores within a time period are missing, for consistency with Section 3.3.2 and the CSP.</p> <p>Section 3.1.2: Small change to refer to Week 4 visit instead of first post baseline visit when defining visit windows as post-baseline spirometry is conducted on Day 1. Added clarification of how to define time period for symptom score parameters and rescue medication usage, if a scheduled visit is missing for a patient.</p> <p>Section 3.3.6: Added clarification of the timeframe used to define the number of exacerbations in the rate of COPD exacerbations calculation and time to first moderate or severe COPD exacerbation endpoint</p> <p>Section 4.1: P-value reporting changed to 4 decimal places for consistency with previous Bevespi studies.</p> <p>Section 4.2.1: The reason for partial data exclusion from the Per Protocol analysis set will not be tabulated. This is not applicable, as the Per Protocol analysis set is defined at the subject-level. The number of patients randomised by region (in addition to country and centre) will be summarised in the full analysis set to support the subgroup analyses.</p> <p>Section 4.2.3: Rescue albuterol/salbutamol MDI will be included in medication tables, so reference to omission from tables removed. Medications at study entry updated to include medications that stopped on the day of visit 1, or within 7 days prior to visit 1. Added clarification that the March 2017 version of WHO-DD will be used.</p> <p>Section 4.2.4: Exposure categories updated.</p>

---

<b>Date</b>	<b>Brief description of change</b>
	<p>Section 4.2.5: Definition of treatment compliance in ePRO data updated. Expected number of inhalations on the last day of treatment redefined for visit days. Clarification added so that denominators adjust for half days where ePRO data is collected. Added calculations for morning and evening compliance.</p> <p>Section 4.2.6.1: Subgroup analyses added for baseline CAT score and smoking status.</p> <p>Section 4.2.7 and 4.2.7.5: NiSCI non-inferiority margin added.</p> <p>Section 4.2.7.4: Subgroup analyses updated to be consistent with Section 4.2.6.1 - smoking status subgroup added.</p> <p>Section 4.2.7.4 and 4.2.7.6: Cumulative distribution function plots wording updated to be produced at week 24 instead of over week 24</p> <p>Section 4.2.7.5 and 4.2.7.6: References to Section 4.2.7.4 updates to refer to Section 4.2.6.1 for subgroup analyses.</p> <p>Section 4.2.8: Definitions of event rate updated for consistency with previous Bevespi studies. Most common AEs percentage threshold reduced.</p> <p>Section 4.2.9: Clarified the 1-sided p-value corresponding to superiority assessment.</p>

---

## 1. STUDY DETAILS

This statistical analysis plan outlines the analyses to be generated for the global clinical study report.

### 1.1 Study objectives

#### 1.1.1 Primary objective

<b>Primary Objective:</b>	<b>Outcome Measure:</b>
To assess the effects of Glycopyrronium/Formoterol Fumarate (GFF) relative to Umeclidinium/Vilanterol (UV) on lung function as measured by trough forced expiratory volume in 1 second (FEV <sub>1</sub> ) and peak FEV <sub>1</sub> in patients with moderate to very severe Chronic Obstructive Pulmonary Disease (COPD)	<b>Co-primary endpoints:</b> <ul style="list-style-type: none"> <li>• Change from baseline in morning pre-dose trough FEV<sub>1</sub> over 24 weeks</li> <li>• Peak change from baseline in FEV<sub>1</sub> within 2 hours post-dosing over 24 weeks</li> </ul>

#### 1.1.2 Secondary objectives

<b>Secondary Objective:</b>	<b>Outcome Measure:</b>
To further assess the effects of GFF relative to UV on lung function.	<b>Secondary endpoints:</b> <ul style="list-style-type: none"> <li>• Onset of action on day 1; proportion of patients with increase of FEV<sub>1</sub> of <math>\geq 100</math> mL from baseline at 5 minutes</li> <li>• Peak change from baseline in Inspiratory Capacity (IC) within 2 hours post-dosing over 24 weeks</li> </ul>
To assess the effects of GFF relative to UV on dyspnea	<b>Secondary endpoint:</b> Transition Dyspnea Index (TDI) focal score over 24 weeks
To assess the effects of GFF relative to UV on symptoms of COPD	<b>Secondary endpoint:</b> Change from baseline in Early Morning Symptoms COPD Instrument (EMSCI) over 24 weeks <b>Other endpoints:</b> <ul style="list-style-type: none"> <li>• Change from baseline in Night Time Symptoms COPD Instrument (NiSCI) over 24 weeks</li> <li>• Change from baseline in daily rescue (albuterol/salbutamol metered dose inhaler [MDI]) use over 24 week</li> </ul>
To assess the effects of GFF relative to UV on health-related quality of life	<b>Other endpoint:</b> Change from baseline in COPD Assessment Test (CAT) score over 24 weeks

### 1.1.3 Safety objectives

<b>Safety Objective:</b>	<b>Outcome Measure:</b>
To assess the safety of GFF	Adverse events (AEs), serious adverse events (SAEs), adverse events leading to treatment discontinuation (DAEs)

### 1.1.4 Exploratory objectives

<b>Exploratory Objective:</b>	<b>Outcome Measure:</b>
To assess health status for COPD patients treated with GFF	Change from baseline in EuroQol 5 Dimensions 5 Levels Questionnaire (EQ-5D-5L) over 24 weeks
To assess overall and COPD-specific Healthcare Resource Utilisation and work productivity loss for GFF	Health Economics Questionnaire for healthcare resource utilisation. Time to the first moderate or severe COPD exacerbation. Work Productivity and Activity Impairment - General Health (WPAI-GH)

## 1.2 Study design

This is a phase IIIb randomised, double-blind, double-dummy, multicentre, parallel group, 24 week study to assess the efficacy and safety of GFF 7.2/4.8 µg fixed-dose combination 2 inhalations twice daily compared to UV 62.5/25 µg fixed-dose combination 1 inhalation once daily in patients 40-95 years of age with moderate to very severe COPD.

Approximately 1400 patients will be screened, with 1000 patients expected to be randomised at approximately 150 sites in a 1:1 scheme to receive GFF or UV treatment. All patients will be required to be symptomatic (CAT  $\geq$ 10) at randomisation (Visit 3). Patients will be stratified based on incoming COPD treatment at screening: rescue only/maintenance monotherapy (long-acting muscarinic antagonist [LAMA] or selective long-acting  $\beta_2$ -agonist [LABA] or inhaled corticosteroids [ICS] only) vs. double maintenance therapy (LAMA/LABA or ICS/LABA). Patients on triple maintenance therapy will not be eligible.

Each patient will undergo a screening period of 1 to 4 weeks (-28 to -9 days). Post-bronchodilator FEV<sub>1</sub> will be tested to determine eligibility and COPD severity. Any treatment with ICS, LAMA, LABA and existing rescue medications will be stopped at the screening visit. During the screening period, patients will be treated with ipratropium bromide MDI 2 inhalations 4 times daily and albuterol/salbutamol MDI as needed. Albuterol/salbutamol MDI will also be used as rescue medication throughout the 24-week treatment period.

The first dose of investigational product (IP) will be taken at Visit 3 (Day 1). Since there are 2 different IP inhalers, a double-dummy technique will be applied to ensure the double-blinding of the study. Matched placebo to GFF and UV will be present with the same external appearance. During the 24-week double-blind treatment period, patients will take 2 inhalations

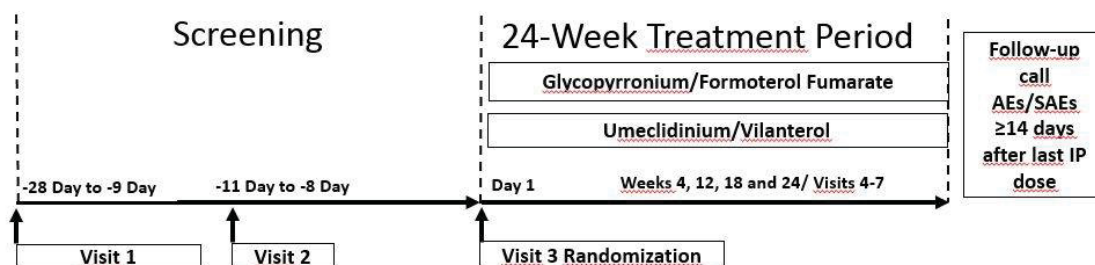
in the morning and in the evening from the GFF inhaler (active or placebo) and 1 inhalation in the morning from the UV inhaler (active or placebo).

After Visit 3, Patients will be evaluated at Visit 4 (Week 4), Visit 5 (Week 12), Visit 6 (Week 18) and Visit 7 (Week 24) using spirometry and patient reported outcomes. From Visit 2 onwards, patients will also enter data daily in an electronic Patient Reported Outcomes device (ePRO). For patients who remain on treatment throughout the study (i.e., complete Visit 7), a follow-up telephone call will be performed approximately 14 days after the last IP dose, in order to assess new or ongoing AE (as well as any concomitant medication administered to treat the mentioned AE) and COPD exacerbations.

Patients who discontinue study treatment prior to Week 24 (Visit 7) will complete a Treatment Discontinuation/Withdrawal Visit and be withdrawn from the study. A follow-up telephone call will be performed approximately 14 days after the last IP dose. No further data will be collected. In the event the Treatment Discontinuation/Withdrawal Visit is performed approximately 14 days post last IP dosing, a follow-up telephone call will not be required.

The study flow chart is shown in [Figure 1](#) (see below). The study plans are presented in Tables 1 and 2 of the clinical study protocol (CSP).

**Figure 1 Study flow chart**



### 1.3 Number of subjects

1000 patients (500 patients each in the GFF and UV treatment arms) are to be randomised in this study. Based upon Phase III studies of GFF, it is estimated that 12% of the patients randomised will be excluded from the Per Protocol analysis set, leaving around 440 patients in the co-primary non-inferiority analyses. This sample size will provide around 94% power to demonstrate that the difference between GFF vs UV in change from baseline in morning pre-dose trough FEV<sub>1</sub> over 24 weeks is greater than the non-inferiority margin of -50 mL. This calculation assumes a 1-sided significance level of 2.5% and a true difference of -10 mL. Should testing be conducted at a 1-sided significance level of 1.25% then the power would be 90%. Assumptions regarding variability for the co-primary endpoints are based on experience in previous GFF Phase III clinical studies. An effective standard deviation (SD) of 167 mL for

the change from baseline in morning pre-dose trough FEV<sub>1</sub> over 24 weeks has been assumed. This value is based on a per-visit SD of 200 mL and a within-patient (between-visit) correlation of 60%. Under these assumptions, for non-inferiority to be achieved the point estimate of the difference in pre-dose trough FEV<sub>1</sub> over 24 weeks must be above -28 mL when testing at a 2.5% significance level, or above -24.5 mL when testing at a 1.25% significance level.

For the co-primary endpoint of peak FEV<sub>1</sub> over 24 weeks and secondary lung function endpoint of peak IC over 24 weeks, the power to show non-inferiority is expected to be at least that of the co-primary endpoint of trough FEV<sub>1</sub>. Assuming an effective SD of 188 mL for peak FEV<sub>1</sub>, and a true difference of 50 mL then 500 patients would provide at least 97% power to demonstrate superiority of GFF to UV in peak FEV<sub>1</sub> with a 1-sided significance level of 1.25%.

For the secondary endpoint of the TDI focal score over 24 weeks a non-inferiority margin of -1 unit is defined. However, note that the power to show non-inferiority would remain at least 90% with a margin of -0.75 units (three quarters of the minimum clinically important difference [MCID]), assuming no difference between treatments and a standard deviation of 3 units.

## **2. ANALYSIS SETS**

### **2.1 Definition of analysis sets**

The co-primary non-inferiority analyses and all other non-inferiority analyses in this study will primarily be conducted using the Per Protocol analysis set. This will be used to compare GFF and UV when the treatments are used as planned at each visit, in the intended patients and without important protocol deviations which may affect efficacy. This will provide an estimate of the true biological efficacy of the treatments and counters the tendency for intention-to-treat analyses to favour a conclusion of equivalence. Non-inferiority comparisons will also be conducted in the full analysis set of all patients receiving randomised IP, allowing for variation in clinical application whilst on treatment, as a comparison of their effectiveness when used in practice. For selected endpoints this full analysis set will be used to provide an assessment of superiority, including for the co-primary analysis of the superiority of GFF relative to UV in peak FEV<sub>1</sub>.

For consistency, demographic and disease-related baseline characteristics will be presented using both Per Protocol analysis set and full analysis set.

#### ***All Patients Analysis Set***

This analysis set comprises all patients screened for the study and may be used for the reporting of disposition and screening failures.

#### **2.1.1 Efficacy analysis sets**

##### ***Full Analysis Set***

The full analysis set will be defined as all patients randomised who received at least 1 inhalation of IP from the GFF or UV inhaler (active or placebo). Patients will be analysed according to the treatment they were assigned to at randomisation, regardless of the treatment actually received.

This analysis set will be used to demonstrate consistency in non-inferiority hypothesis tests and as the primary means of assessing superiority hypotheses.

### ***Per Protocol Analysis Set***

The Per Protocol analysis set will be used as the primary means of assessing non-inferiority hypotheses.

The Per Protocol analysis set is the subset of the full analysis set containing patients with post-randomisation data obtained prior to important protocol deviations which may affect efficacy. Data obtained after such important protocol deviations will be excluded from analyses using this set. A detailed definition of the important protocol deviations and data points to be excluded are given in Section 2.2. Since receiving the wrong treatment will be a major protocol deviation, patients in the Per Protocol analysis set will be analysed as randomised (which for this population is identical to analysis by the actual treatment received).

#### **2.1.2 Safety analysis sets**

##### ***Safety Analysis Set***

All patients who received at least 1 inhalation of the randomised active IP that they were assigned to will be included in the safety analysis set. Patients will be classified according to the treatment they actually received. If a patient received more than 1 randomised treatment, they will be analysed and included in summaries according to the treatment they received the most. Any major deviations from the randomised treatment assignment will be listed and considered when interpreting the safety data. All safety summaries will be based on this analysis set.

#### **2.1.3 Other analysis sets**

##### ***Rescue Medication User analysis set***

Regional differences in rescue albuterol/salbutamol MDI usage are expected with some regions using virtually no rescue albuterol/salbutamol MDI at study entry. Therefore, the Rescue Medication User analysis set is defined as all patients in the full analysis set with average baseline rescue albuterol/salbutamol MDI use of  $\geq 1$  inhalation/day.

## **2.2 Violations and deviations**

### **2.2.1 Important protocol deviations**

The following general categories will be considered important protocol deviations and will be programmatically derived from the data.

- Patients without informed consent provided (Deviation 1).

- Patients with a severe COPD exacerbation (requiring hospitalisation) or more than 1 moderate exacerbation (requiring antibiotics or  $\geq 3$  days of systemic steroids) who did not discontinue IP by the time of next clinical visit or 7 days after the investigator was made aware of SAE following a severe exacerbation (Deviation 2).
- Patients requiring discontinuation from the study due to safety-related reasons who were not withdrawn (Deviation 3).
- Patients who deviate from key entry criteria relating to clinical diagnosis and severity of COPD (Deviation 4). These are inclusion criteria 3, 4, 5, 6, 7 and exclusion criteria 1 in the CSP version 2.
- Patients with treatment compliance in complete study period  $<70\%$  or  $>130\%$  for either the GFF or UV inhaler (Deviation 5).
- Patients who do not have the required washout of prohibited medications prior to baseline (Deviation 6). Please refer to CSP Section 7.7 for minimum washout periods. Exclusion criteria, the anatomical therapeutic chemical (ATC) codes and additional physician reviews will be used to identify the medications, where possible. Specifically:
  - Insufficient washout of prohibited COPD medications in CSP Table 8 except for short acting and ephedrine-containing medications. Theophylline or other xanthines are not considered an important protocol deviation if taken as a stable low dose prior to and throughout the study as defined in CSP Table 6.
  - Insufficient washout of any other investigational drug, any monoclonal or polyclonal antibody or strong CYP3A4 inhibitors with washout as defined in CSP Table 9.
  - Insufficient washout of systemic corticosteroids, other than a stable dose of an equivalent of 5 mg prednisone per day or 10 mg every other day, as per CSP Table 6.
- Patients who use certain prohibited or restricted medications outside of the limitations (Deviation 7). Please refer to CSP Section 7.7 for the permitted usage and restrictions of the medications during the study. Medications will be identified using ATC codes and additional physician reviews. Specifically:
  - The initiation of prohibited COPD medications in CSP Table 8 on or after Visit 2 except for short acting and ephedrine-containing medications. The initiation of theophylline or other xanthines is not considered an important protocol deviation if taken as a stable low dose prior to and throughout the study as defined in CSP Table 6 or if used to treat a COPD exacerbation for  $\leq 14$  days as per CSP Table 10. The ICS use to treat a COPD exacerbation for  $\leq 14$  days is allowed.



- The initiation of any other investigational drug or any monoclonal or polyclonal antibody after Visit 1, or of strong CYP3A4 inhibitors used for >4 weeks after Visit 3, as included in CSP Table 9.
  - The initiation of systemic corticosteroids and phosphodiesterase-4 inhibitors throughout the study, unless at a stable dose prior to and throughout the study as defined in CSP Table 6. The use of systemic corticosteroids to treat a COPD exacerbation for  $\leq 14$  days is allowed.
  - The use of antibiotics to treat a COPD exacerbation for >14 days, as per CSP Table 10.
- Patients who received IP other than that to which they were randomised to (Deviation 8).

The important protocol deviations will be listed and summarised by randomised treatment group. Some of these deviations that may affect efficacy will lead to exclusion of all data from a particular patient from the Per Protocol analysis set (Deviations 4, 5, 6), or require exclusion of data from the time point of deviations onwards from analyses using the Per Protocol analysis set (Deviations 7, 8). Deviations 1, 2, 3 above will not lead to any exclusion from the Per Protocol analysis set. None of the deviations will lead to patients or data points being excluded from other analysis sets described in Section 2.1.

In the case of severe site misconduct or duplicated patients (those included in the study more than once) it may be appropriate to remove sites, patients, or repeated patient information from the reporting of the trial, however this must be assessed dependent on the individual circumstance, and would be detailed prior to unblinding in a formal file note and in the documentation detailing the finalization of important protocol deviations as approved by AstraZeneca.

### 2.2.2 Other deviations

In addition to the important protocol deviations, patients with the following deviations will have their data from the affected visits removed from Per Protocol analyses of pulmonary function outcomes:

- Patients who do not take the correct morning and evening doses of blinded study inhalers (both GFF and UV, active or placebo) on the day preceding a study visit.
- Patients who do not comply with required restrictions prior to spirometry. The restrictions are: patient was not to smoke for at least 1 hour prior to study visit; patient was not to have rescue albuterol/salbutamol MDI or their screening period ipratropium bromide MDI (please refer to CSP Table 7) for at least 6 hours prior to study visit; patient was not to take IP in the morning prior to study visit during the treatment period.

The spirometry assessments at each post-baseline visit with the deviations below will be excluded from the Per Protocol analyses of pre-dose trough change from baseline endpoints in Section 3.2.1. Only assessments with non-missing time from IP dosing the previous day will be considered for exclusion.

- Pre-dose 60 min assessments (i.e.  $\geq 45$  min prior to IP dosing at the visit) that are not within  $11 \pm 1.5$  hours of GFF (active or placebo) dosing in the evening of the day prior to visit or not within  $23 \pm 1.5$  hours of UV (active or placebo) dosing in the morning of the day prior to visit.
- Pre-dose 30 min assessments (i.e.  $< 45$  min and  $\geq 0$  min prior to IP dosing at the visit) that are not within  $11.5 \pm 1.5$  hours of GFF (active or placebo) dosing in the evening of the day prior to visit or not within  $23.5 \pm 1.5$  hours of UV (active or placebo) dosing in the morning of the day prior to visit.

The deviations in sections above will be programmatically determined. In addition, monitoring notes or data listings will be reviewed to ensure that no other important protocol deviations need to be derived, and to check that those identified via programming are correctly classified. The final classification of important protocol deviations and other deviations relating to single visits and assessments will be made prior to database lock. Decisions made will be documented and approved by AstraZeneca prior to unblinding of the data.

### **3. PRIMARY AND SECONDARY VARIABLES**

#### **3.1 General Definitions**

##### **3.1.1 Definition of baseline**

All efficacy assessments are relative to pre-dose baseline obtained at randomisation at Visit 3. If there is no pre-dose value at Visit 3, then the baseline value will not be imputed and will be set to missing. No data known to be collected post first dose will be used in determining the baseline value.

For spirometry endpoints the mean of all available evaluable (meeting ATS/ERS criteria for acceptability) -60 and -30 minute pre-dose spirometry assessments conducted at Day 1 (Visit 3) will be used to establish baseline for all FEV<sub>1</sub>, forced vital capacity (FVC), and IC parameters. If patients missing either of these pre-dose assessments, the baseline value will be calculated from the single measurement. In patients missing both pre-dose values, the spirometry assessments from the screening period will not be used and the baseline will be set to missing.

For the diary symptom score parameters and rescue medication usage, baseline will be the average of the non-missing values from the ePRO data collected in the last 7 days before randomisation (i.e. excluding data recorded on the day of Visit 3). If at least half of the daily measures/scores within that period are missing, then baseline will be set to missing.

For TDI/Baseline Dyspnea Index (BDI), CAT, EQ-5D-5L and WPAI-GH, the latest pre-dose assessments at randomisation (Visit 3) or during the 7 days prior to randomisation will be used as the respective baselines.

### 3.1.2 Visit windows

For endpoints that present visit-based data, variables will be summarised according to the protocol-scheduled week for that visit. Data collected in spirometry and ePRO systems will be allocated to visit weeks based upon assigned windows so as to allow rescheduled assessments to be included.

The windows for the visits following baseline will be constructed in such a way that the upper limit of the interval falls half way between the 2 visits (the lower limit of the Week 4 visit will be Day 2). If an even number of days exists between 2 consecutive visits then the upper limit will be taken as the midpoint value minus 1 day. Visit windows are constructed so that every observation collected can be allocated to a particular visit. No visit windows will be defined for screening visits. Listings will display all values contributing to a visit week for a patient.

The visit windows for assessments conducted at every visit of Visit 4-Visit 7 are summarised in [Table 1](#).

**Table 1 Visit windows for assessments conducted at every visit of Visit 4-Visit 7**

Protocol-Scheduled Visit Week	Target Study Day	Visit Windows
Week 0 / Baseline	1	Study Day = 1
Week 4	29	$2 \leq \text{Study Days} \leq 56$
Week 12	85	$57 \leq \text{Study Days} \leq 105$
Week 18	127	$106 \leq \text{Study Days} \leq 147$
Week 24	169	$148 \leq \text{Study Days}$

For assignment of data to visit windows, study day will be defined as follows:

$$(\text{Date of assessment} - \text{date of randomisation}) + 1$$

By this definition, the day of randomisation will be study day 1 and the planned date of Visit 4 (Week 4) will be study day 29 (=28+1), for example.

For visit-based patient reported outcomes, if multiple assessments are recorded within a single visit window, please refer to the rules below.

- If there are 2 or more observations within the same visit window, then the non-missing observation closest to the scheduled visit will be used in the analysis.
- If 2 observations are equidistant from the scheduled visit, then the non-missing observation with the later collection date will be used in the analysis.

- If 2 observations are collected on the same day then the non-missing observation with the later collection time will be included in the analysis.

For spirometry assessments, multiple assessments pre and post dosing will be conducted at each scheduled visit. The non-missing values that have at least 1 effort that meets ATS/ERS criteria for acceptability will contribute to the analysis. The ATS/ERS criteria (as described in Section 3.2) will be applied in order to select the representative value from a series of repeat efforts before assigning the windows.

If multiple assessments with acceptable quality after applying the ATS/ERS criteria but with different study days are recorded within a single visit window, please refer to the rules below.

- If there are 2 or more study days within the same visit window, then the study day closest to the scheduled visit will be chosen and observations with acceptable quality on that study day will be used in the analysis.
- If 2 study days are equidistant from the scheduled visit, then the study day with the later collection date will be chosen and observations with acceptable quality on that study day will be used in the analysis.

Additionally, the spirometry assessments at each visit will be allocated to the derived nominal collection time windows using the time intervals specified in Table 2 and Table 3.

**Table 2 Time windows for FEV<sub>1</sub> and FVC assessments**

Protocol-Scheduled Collection Time	Time Windows of Minutes from IP dosing
Pre-dose 60 min	Time ≤ - 45
Pre-dose 30 min	-45 < Time ≤ 0
Post-dose 5 min	0 < Time < 10
Post-dose 15 min	10 ≤ Time < 23
Post-dose 30 min	23 ≤ Time < 45
Post-dose 1 hour	45 ≤ Time < 90
Post-dose 2 hours	90 ≤ Time < 180

**Table 3 Time windows for IC assessments**

Protocol-Scheduled Collection Time	Time Windows of Minutes from IP dosing
Pre-dose 60 min	Time ≤ - 45
Pre-dose 30 min	-45 < Time ≤ 0
Post-dose 1 hour	0 < Time < 90
Post-dose 2 hours	90 ≤ Time < 180

If multiple spirometry values with acceptable quality for the same parameter are recorded within a single time window, then the closest value (in minute and second) to the scheduled

time point will be used in the analysis, with the exception of peak change from baseline endpoints in Section 3.2.2 which are derived using all spirometry values with acceptable quality from the time of IP dosing to the end of the post-dose 2 hour time window.

If a visit window or a time window does not contain any observations, then the data will remain missing.

For Per Protocol analyses, data with important protocol deviations and other deviations relating to spirometry outcomes will not be considered when choosing observations in a visit or a time window for the parameters. These deviations and corresponding data exclusion are specified in Section 2.2. For spirometry outcomes, time windows will only be applied once both the ATS/ERS criteria have been applied and the data with deviations in Section 2.2 are excluded.

For symptom score parameters and rescue medication usage, the daily measurements will be aggregated based upon the time periods between clinic visits as specified in Table 4. Data from the day of the clinic visit will not be used. The definition of time period will depend on the actual days of the scheduled visits (as date of visit – date of randomisation + 1), instead of the scheduled days for those visits (i.e. allowing for the variation of  $\pm 3$  days for a visit schedule). Scheduled visits will be used to define the time periods, and if a scheduled visit is missing for a patient, the scheduled day for that visit (for example Day 29 for Week 4; refer to target day in 1) will be used for the lower or upper limit of the corresponding time period. If data is collected on the scheduled day of visit used to define the upper or lower limit of the time period, it will not contribute to the aggregate value for either time period.

If at least half of the daily measures/scores within a time period are missing, then the aggregated measures/scores for that period will be set to missing.

**Table 4 Time period for symptom scores and rescue medication usage**

Between Clinic Visits	Time Period
Baseline	$-7 \leq \text{Study Days} \leq -1$
Week 0 – Week 4	$2 \leq \text{Study Days} \leq (\text{Day of Visit 4}) - 1$
Week 4 – Week 12	$(\text{Day of Visit 4}) + 1 \leq \text{Study Days} \leq (\text{Day of Visit 5}) - 1$
Week 12 – Week 18	$(\text{Day of Visit 5}) + 1 \leq \text{Study Days} \leq (\text{Day of Visit 6}) - 1$
Week 18 – Week 24	$(\text{Day of Visit 6}) + 1 \leq \text{Study Days} \leq (\text{Day of Visit 7}) - 1$

### 3.2 Pulmonary Function Outcomes

All pulmonary function assessments that have at least 1 effort that meets ATS/ERS criteria for acceptability will be considered in the following steps to select a value at each time-point to be used in the analyses. The ATS/ERS criteria will be applied before assigning the visit or time windows.

- For FEV<sub>1</sub> and FVC, the largest acceptable and reproducible result from each set of repeat efforts at a protocol-scheduled time-point will be used in analyses. If reproducibility cannot be achieved then the largest result meeting acceptability criteria will be taken.
- For IC, after the removal of unacceptable efforts, if all remaining efforts are acceptable and reproducible results and there are at least 3 results, then the average of these will be used in analyses. If acceptability and reproducibility cannot be achieved for all remaining efforts, then the average of the results meeting acceptability criteria will be taken.
- If all the assessments at a specific time-point were deemed to be of unacceptable quality the pulmonary function assessments obtained at the time-point will not be included in any analyses and will be considered missing.

Further requirements specified in Section 2.2 will have to be met in addition to the ATS/ERS criteria in order for pulmonary function data to be used in the Per Protocol analyses.

### 3.2.1 Morning pre-dose trough FEV<sub>1</sub>

The co-primary endpoint of change from baseline in morning pre-dose trough FEV<sub>1</sub> is defined as the average of the -60 and -30 minute pre-dose values at each visit minus baseline. Baseline is defined as the average of the non-missing -60 minute and -30 minute values obtained prior to dosing at Visit 3.

In patients missing either of these pre-dose assessments, the morning pre-dose trough FEV<sub>1</sub> will be calculated from a single measurement. In patients missing both pre-dose values, the morning pre-dose trough FEV<sub>1</sub> at that visit will not be calculated.

The change from baseline in morning pre-dose trough FVC will be derived similarly.

### 3.2.2 Peak FEV<sub>1</sub> and Inspiratory Capacity post-dosing

The co-primary endpoint of peak change from baseline in FEV<sub>1</sub> is defined as the maximum of the FEV<sub>1</sub> assessments within the up to 2 hours post-dosing time windows (Table 2) at each visit minus baseline. Similarly, the secondary endpoint of peak change from baseline in IC is defined as the maximum of the IC assessments within the up to 2 hours post-dosing time windows (Table 3) at each visit minus baseline.

Further additional spirometry endpoints will also be derived. Peak change from baseline in FVC within 2 hours post-dose will be derived similarly to peak FEV<sub>1</sub>. The trapezoidal rule will be used to derive estimates of AUC<sub>0-2</sub> for FEV<sub>1</sub> and FVC at each visit, normalised by time from first to last non-missing value used in the derivation, as long as there are at least 2 non-missing time points during the first 2 hours post-dose time windows (Table 2). The derivation of AUC<sub>0-2</sub> endpoints will consider the spirometry time windows to use only those spirometry values selected within each time window. When applying the trapezoidal rule, change from baseline values at each time point will be used with their actual time of evaluation, and the

value at time 0 will be the average of -60 and -30 minute pre-dose trough values at that visit for the parameter.

### 3.2.3 Increase in FEV<sub>1</sub> at 5 minutes post-dosing

To assess the early onset of action, a secondary endpoint of this study will be the proportion of patients with increase of FEV<sub>1</sub> of  $\geq 100$  mL from baseline at 5 minutes on day 1. Only data assigned to the 5 minute window will be used to determine response. Patients with missing data will be considered to be non-responders for the analysis.

Additional categorical endpoints will be derived similarly as the proportion of patients with increase of FEV<sub>1</sub> of  $\geq 150$  mL from baseline at 5 minutes and the proportion of patients with increase of FEV<sub>1</sub> of  $\geq 12\%$  from baseline at 5 minutes. In addition the change from baseline in FEV<sub>1</sub> at each post-dose timepoint in [Table 2](#) on day 1 will also be derived.

## 3.3 Patient Reported Outcomes

### 3.3.1 Transition Dyspnea Index (TDI) focal score

As a secondary endpoint assessing dyspnea, the TDI focal score will be derived at post-randomisation visits relative to the baseline severity of dyspnea assessed using the BDI at Visit 3. The BDI/TDI is an instrument developed to provide a multidimensional measure of dyspnea in relation to activities of daily living. The BDI and TDI consist of 3 individual components: functional impairment, magnitude of task, and magnitude of effort. For the BDI, each of these 3 components are rated in 5 grades from 0 (i.e. very severe) to 4 (i.e. no impairment), and are summed to form a baseline total score from 0 to 12. For the TDI, changes in dyspnea are rated for each component by 7 grades from -3 (major deterioration) to +3 (major improvement), and are added to form a TDI focal score from -9 to +9. For the TDI, at each visit, if a response to any of the 3 questions is missing, then the focal score will also be considered missing. Otherwise, scoring and handling of missing items will be conducted in accordance with the user's guide for the TDI score.

The percentage of patients achieving the threshold of 1 unit or more in TDI focal score will be derived at each visit:

- TDI focal score  $\geq 1$  → Improvement
- $-1 < \text{TDI focal score} < 1$  → No change
- TDI focal score  $\leq -1$  → Deterioration

For the TDI responder analyses, responders will be defined as patients with TDI focal scores categorised as improvement. Patients with missing data will be considered to be non-responders for the analyses.

### **3.3.2 Early Morning Symptoms COPD Instrument (EMSCI) and Night Time Symptoms COPD Instrument (NiSCI)**

Patients will complete a daily ePRO questionnaire for their COPD symptoms with 2 parts covering symptoms “last night” and “this morning”.

The EMSCI collects data about the frequency and severity of early morning symptoms and the impact of COPD symptoms on early morning activity in patients with COPD.

The NiSCI collects data about the frequency and severity of night-time symptoms and the impact of COPD symptoms on night-time awakenings in patients with COPD.

The change from baseline in the 6-item EMSCI Symptom Severity Score is a secondary endpoint. The endpoint is derived by averaging the responses from a patient on the 6 item-level symptom scores (scored on a 4-point scale from 1 to 4); if the response to any score is missing, then the derived value will be considered missing.

The change from baseline in the 6-item NiSCI Symptom Severity Score will be derived similarly.

The change from baseline in the Overall COPD Symptom Severity scores (a single item asking about overall severity on a 5-point scale from 0 to 4) in the early morning and at night, Early Morning Activity score and Night-time Awakenings score (number of times the patient woke up during the night due to COPD symptoms) will also be derived.

Each daily measure will be aggregated based upon the time periods between clinic visits. Data from the day of the clinic visit will not be used in these aggregates, because IP is not taken in the morning on these days. The time periods are specified in Section 3.1.2. If the measure is available on more than half of the days in a given period then an aggregate score will be computed as the average of all present daily values. Otherwise the score will be considered missing for that period. ePRO data recorded during the last 7 days of the screening period prior to Visit 3 will be used to calculate the baseline.

### **3.3.3 COPD Assessment Test (CAT) score**

The COPD Assessment Test (CAT) is used to quantify the impact of COPD symptoms on health status. The CAT has a scoring range of 0-40, and it is calculated as the sum of the responses given for each of the 8 items (scored on a 6-point scale from 0 to 5), with higher scores indicating a higher impact of COPD symptoms on health status. If the response to 1 of the 8 items is missing, the missing item will be considered equal to the average of the 7 non-missing items for that patient. If more than 1 item is missing the score will be considered missing.

The change from baseline in CAT score will be derived at each visit relative to the baseline at Visit 3. The percentage of patients achieving a MCID threshold for CAT of 2 units or more will be derived at each visit:



- CAT score (change from baseline)  $\leq -2 \rightarrow$  Improvement
- $-2 <$  CAT score (change from baseline)  $< 2 \rightarrow$  No change
- CAT score (change from baseline)  $\geq 2 \rightarrow$  Deterioration

A CAT responder at the visit will be defined as a patient who had improvement based on the change from baseline in CAT score. Patients with missing data will be considered to be non-responders for the analyses.

### 3.3.4 Daily rescue medication use

The number of inhalations of rescue albuterol/salbutamol MDI will be recorded in the patient ePRO in the morning and evening. The mean daily number of inhalations of rescue albuterol/salbutamol MDI will be calculated overall and in the time periods between each clinic visit. These periods are specified in Section 3.1.2. ePRO data recorded during the last 7 days of the screening period prior to randomisation will be used to calculate the baseline. Change from baseline in mean daily rescue (albuterol/salbutamol MDI) use will then be derived. The denominator will be adjusted based on the number of days (or half days) with non-missing values (for example, a half day for daily number of inhalations if only daytime or night-time use is recorded on that day). The mean daytime and night-time rescue use and change in these parameters from baseline will be derived similarly.

A ‘day with no rescue use’ is defined from days where rescue albuterol/salbutamol MDI usage data is non-missing as any day where the patient reported no inhalations of rescue albuterol/salbutamol MDI in both the morning and evening. The percentage of days with no rescue use will be derived using a denominator of the number of full days in the period with no missing data.

### 3.3.5 EuroQol 5 Dimensions 5 Levels Questionnaire (EQ-5D-5L)

The EQ-5D-5L questionnaire assesses 5 dimensions which will be derived separately: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has 5 response options (no problems, slight problems, moderate problems, severe problems, and extreme problems) that reflect increasing levels of difficulty. A unique EQ-5D health state is referred to by a 5 digit code allowing for a total of 3125 health states. For example, state 11111 indicates no problems on any of the 5 dimensions. Utility values will be calculated by mapping the EQ-5D-5L descriptive system data onto the EQ-5D-3L valuation set. Mapping function developed by [Van Hout et al 2012](#) et al will be used for the calculation.

The patient will also be asked to rate current health status on a visual analogue scale from 0 to 100, with 0 being the worst imaginable health state. The change from baseline in visual analogue scale will be calculated.

### 3.3.6 Healthcare Resource Utilisation

Broad-based health care utilisation event information will be collected by the Investigator/authorised delegate via patient interview at each visit and recorded in the case report form (CRF) modules.

The number of days/times the following resources were utilised due to COPD-related and non-COPD-related events will be presented for each patient:

- Ambulance transport
- Hospitalisation, intensive care (days in intensive care)
- Hospitalisation, general care (days in general care)
- Emergency room visit
- Hospital admission or emergency department lasting over 24 hours
- Visit to specialist
- Visit to primary health care physician
- Other health care visit
- Home visit, physician
- Home visit, other health care
- Telephone call to physician
- Telephone call to nurse
- Telephone contact with other physician/health care provider
- Spirometry
- Advanced pulmonary function test
- Plain chest X-ray
- Computer tomography
- Oxygen treatment initiated

The number of moderate COPD exacerbations (defined as those requiring antibiotics or  $\geq 3$  days of systemic steroids) or severe COPD exacerbations (defined as those leading to in

patient hospitalisation, including >24 hours in emergency department/urgent care setting) will also be recorded.

The rate per year of COPD exacerbations in each treatment group will be calculated as (the total number of exacerbations from the date of randomisation to Visit 7 or Treatment Discontinuation/Withdrawal Visit, whichever occurs, \*365.25) divided by the total duration of exposure to randomised IP within the treatment group. Exacerbations will be considered as separate events provided that 7 or more days are between the recorded stop date of the earlier event and start date of the later event. Time during an exacerbation or in the 7 days following an exacerbation will not be included in the calculation of exposure for the denominator. The rate per year of moderate and severe COPD exacerbations will be calculated similarly.

Time to the first moderate or severe COPD exacerbation will be defined as the time from the date of randomisation until the start date of the first moderate or severe COPD exacerbation up to Visit 7 or Treatment Discontinuation/Withdrawal Visit, whichever occurs. Patients who have completed the treatment period and not experienced such an exacerbation will be censored at the date of Visit 7 (Week 24) or Day 183, whichever is earlier. Patients who withdrew from the study without experiencing a moderate or severe COPD exacerbation will be censored at the date of study withdrawal.

### **3.3.7 Work Productivity and Activity Impairment – General Health (WPAI-GH)**

The WPAI-GH will be used to measure self-reported productivity loss and consists of questions on COPD impact on patient's general health, ability to work and perform regular activities in daily living within 7 days.

There are a maximum of 6 questions that will be completed by patients.

Q1 = currently employed (yes/no)

Q2 = hours missed work due to health problems

Q3 = hours missed work due to other reasons

Q4 = hours actually worked

Q5 = degree health affected productivity while working (0-10 scale, with 0 meaning no effect)

Q6 = degree health affected regular activities (other than work) (0-10 scale, with 0 meaning no effect)

If the answer to question 1 is 'No, not currently employed', then the patient should skip to question 6.

Measures summarised will include the patient's work time missed (absenteeism), impairment at work or reduced on-the-job effectiveness (presenteeism), overall work impairment (absenteeism and presenteeism, ie, work productivity loss), and activity impairment outside

the work environment. The WPAI-GH outcomes are expressed as impairment percentages whereby higher scores indicate greater impairment and less productivity (i.e. worse outcomes).

The following calculations will be used to create the outcomes of interest:

- Number of work hours missed = Q2
- Absenteeism =  $Q2/(Q2+Q4)$
- Presenteeism =  $Q5/10$
- Work Productivity Loss =  $Q2/(Q2+Q4)+[(1-Q2/(Q2+Q4))x(Q5/10)]$
- Activity Impairment =  $Q6/10$

### **3.4 Safety and Tolerability**

Adverse events experienced by the patients will be collected throughout the entire study and will be coded by the AstraZeneca designee using the latest version of the Medical Dictionary for Regulatory Activities (MedDRA). A treatment emergent adverse event (TEAE) is an AE with an onset date on or after the first dose date of randomised treatment through the end of the study.

AEs that have missing causality will be assumed to be related to study drug. AEs with missing maximum intensity will be deemed as severe events. The handling of partial/missing dates for AEs is detailed in Appendix II.

## **4. ANALYSIS METHODS**

### **4.1 General principles**

Summary data will be presented in tabular format by treatment group. Categorical data will be summarised by the number and percentage of patients in each category. Continuous variables for parametric data will be summarised by descriptive statistics including N, mean, SD, median, and range. Data listings will be sorted by treatment group and patient number.

Minimum and maximum values will be reported to the same degree of precision as the raw data unless otherwise stated. Mean, median, SD and confidence intervals (CIs) will be reported to 1 further degree of precision. For categorical data, percentages will be rounded to 1 decimal place. All p-values will be nominal (i.e., not multiplicity adjusted) and will be rounded to 4 decimal places.

The data analyses will be conducted using the SAS<sup>®</sup> System (SAS Institute Inc., Cary, NC).

The change from baseline is computed as  $(\text{post-baseline value} - \text{baseline value})$ . Percent change from baseline is computed as  $((\text{post-baseline value} - \text{baseline value})/\text{baseline value}) \times 100\%$ . If either a post-baseline value or the baseline value is missing, the change from baseline value and the percent change from baseline will also be set to missing. Additional details regarding the definition of baseline are provided in Section 3.1.1.

#### 4.1.1 Statistical considerations

The primary objective of this study is to assess the effects of GFF relative to UV on lung function. This will be addressed by the following co-primary hypothesis tests:

- non-inferiority of GFF relative to UV in the change from baseline in morning pre-dose trough FEV<sub>1</sub> over 24 weeks;
- non-inferiority of GFF relative to UV in the peak change from baseline in FEV<sub>1</sub> within 2 hours post-dosing over 24 weeks;
- superiority of GFF relative to UV in the peak change from baseline in FEV<sub>1</sub> within 2 hours post-dosing over 24 weeks.

The non-inferiority null (H<sub>0</sub>) and alternative (H<sub>1</sub>) hypotheses, with  $\mu$  representing the mean change from baseline in each treatment group, correspond to a null hypothesis of inferiority and an alternative hypothesis of non-inferiority, as follows:

$$\begin{aligned} H_0: \mu_{\text{GFF}} - \mu_{\text{UV}} &\leq -\Delta && \text{(GFF is inferior to UV by } \Delta \text{ or more)} \\ H_1: \mu_{\text{GFF}} - \mu_{\text{UV}} &> -\Delta && \text{(GFF is inferior to UV by less than } \Delta) \end{aligned}$$

The non-inferiority analysis for each endpoint is assessed within a Per Protocol analysis set and if non-inferiority is demonstrated the consistency of this conclusion will be assessed within the full analysis set.

If non-inferiority is demonstrated, then for selected endpoints superiority hypotheses will also be addressed within the full analysis set. Here the null (H<sub>0</sub>) and alternative (H<sub>1</sub>) hypotheses are as follows:

$$\begin{aligned} H_0: \mu_{\text{GFF}} - \mu_{\text{UV}} &\leq 0 && \text{(GFF is not superior to UV)} \\ H_1: \mu_{\text{GFF}} - \mu_{\text{UV}} &> 0 && \text{(GFF is superior to UV)} \end{aligned}$$

The exception to this sequential testing of non-inferiority followed by superiority is the analysis of the proportion of patients with increase in FEV<sub>1</sub> of  $\geq 100$  mL from baseline at 5 minutes. For this endpoint testing will immediately be in terms of superiority hypotheses.

P-values will be reported as 1-sided for both the non-inferiority and superiority hypothesis testing. This will equivalently be presented by comparing the lower bound of CIs to the pre-specified margin. Further details of the hierarchical testing strategy are given in Sections 4.2.6.3 and 4.2.7.1.

The stratification factor for prior treatment in the statistical modelling will be based on the values in Interactive Web/Voice Response System (IWRS/IVRS) at randomisation, even if it is subsequently discovered that these values were incorrect.

## **4.2 Analysis methods**

### **4.2.1 Patient disposition**

Patient disposition will be summarised using the all patients analysis set. The total number of patients will be summarised for the following groups: those who enrolled and those who were not randomised (and reason). The number and percentage of patients within each treatment group will be presented by the following categories; randomised, received treatment with study drug, did not receive treatment with study drug (and reason), completed treatment with study drug, discontinued treatment with study drug (and reason), completed study, and withdrawn from study (and reason).

The number of randomised patients who were included in each analysis set will be displayed. The reason for exclusion from the Per Protocol analysis set of a patient due to important protocol deviations will be listed and tabulated, in addition to any patients excluded from other analysis sets.

The time to withdrawal from the study will be presented using the Kaplan Meier plot for the full analysis set and the Per Protocol analysis set.

The number of patients randomised by region, country and centre will be summarised by treatment group in the full analysis set.

### **4.2.2 Demography data and patient characteristics**

Demography data such as age, gender, race, and ethnicity will be summarised by treatment group for all patients in the full analysis set and the Per Protocol analysis set. Age will be derived from the date of informed consent minus date of birth, rounded down to the nearest integer. For patients in a country where date of birth is not recorded the age as recorded in the CRF will be used.

Various baseline characteristics will also be summarised by treatment for the full analysis set and the Per Protocol analysis set. These include patient characteristics (weight, height and BMI), pre and post bronchodilator lung function data at screening visit 2 and/or at baseline (i.e. FEV<sub>1</sub>, FVC, FEV<sub>1</sub>/FVC, bronchodilator responsiveness), and COPD-related disease characteristics such as smoking status, CAT score, COPD severity and duration, age at onset of COPD, and COPD exacerbation history.

Medical and surgical histories will be summarised by MedDRA Preferred Term (PT) within MedDRA System Organ Class (SOC) for the full analysis set.

Summaries of stratification factors as per IWRS/IVRS will be provided for the full analysis set and the Per Protocol analysis set.

### 4.2.3 Prior and concomitant medications

Medications recorded during the study will be classified into the following periods:

- Prior: Medications with stop date on or before the date of randomisation
- Concomitant: Medications with start date < the last day of IP and stop date > the date of randomisation
- Post-treatment: Medications with start date  $\geq$  the last day of IP

The handling of partial or missing dates for medications is detailed in Appendix II.

The number and percentage of patients who received each concomitant medications (by Anatomical Therapeutic Chemical [ATC] classification system codes and generic name) will be presented by treatment group for the full analysis set. Concomitant medications will be tabulated separately for those which are potentially prohibited/restricted and those which are allowed medications. Any prior (with the exception of medications at entry) or post-treatment medication will be included in the medication listing, but will not be included in the summary tables.

Prohibited/restricted medications will be defined following a physician review (prior to database lock) of the unique combinations of ATC code classifications and generic terms captured. Concomitant medication tables will not consider usage or washout criteria of prohibited/restricted medications, as defined in CSP Tables 6, 8, 9 and 10, only the identified medication or class. Randomised treatment will not be included in the medication tables.

The number and percentage of patients who received COPD and non-COPD treatment at entry will be summarised by treatment group for the full analysis set separately. These include medications with start date < the date of Visit 1 and stop date  $\geq$  (the date of Visit 1 - 7). COPD treatment at entry will additionally be tabulated based on the Per Protocol analysis set. Non-COPD treatment at entry will be tabulated by ATC codes only. Medications at entry will be flagged in the medication listing.

Medications will be classified according to the March 2017 version of the World Health Organisation Drug Dictionary (WHO-DD). Percentages will be calculated relative to the number of patients in the corresponding analysis set.

### 4.2.4 Extent of exposure

Extent of exposure summaries will be provided for the safety analysis set. The number of days on randomised treatment (duration of exposure = date of last dose of IP from active inhaler minus date of first dose of IP from active inhaler +1) will be summarised for each treatment group. In addition the following exposure categories will be derived:  $\leq 28$  days, 29-84 days, 85-126 days, 127-165 days, 166-172 days, 173-182 days and  $\geq 183$  days. The numbers and percentages of patients falling into each category will be presented for each treatment group.

#### 4.2.5 Compliance

Study treatment compliance and ePRO use compliance will be summarised descriptively for the full analysis set and the Per Protocol analysis set. The treatment period is defined as any day between the date of randomisation (Visit 3) to the date of last dose day inclusively.

Treatment compliance will be derived for the GFF inhaler (active or placebo) and UV inhaler (active or placebo) separately. This will be based on:

- Compliance in complete study period is defined as: total number of inhalations during the treatment period / (total number of inhalations expected in the morning and in the evening on a study day that is not the last day of treatment \* [days during the treatment period - 1] + total number of inhalations expected in the morning and evening on the last day of treatment) \*100
- Overall compliance in reported ePRO data is defined as: total number of inhalations during the treatment period / (number of inhalations expected in the morning on a study day \* days with ePRO data entered in the morning during the treatment period + number of inhalations expected in the evening on a study day \* days with ePRO data entered in the evening during the treatment period) \*100
- Morning compliance in reported ePRO data is defined as: total number of inhalations in the morning during the treatment period / (number of inhalations expected in the morning on a study day \* days with ePRO data entered in the morning during the treatment period) \*100
- Evening compliance in reported ePRO data is defined as: total number of inhalations in the evening during the treatment period / (number of inhalations expected in the evening on a study day \* days with ePRO data entered in the evening during the treatment period) \*100

If the last day of treatment is a visit day, the expected number of evening inhalations on the last day of treatment is 0 from the GFF inhaler. Otherwise, the expected number of inhalations for each study day is 2 in the morning and 2 in the evening from the GFF inhaler and 1 in the morning from the UV inhaler. The UV inhaler recorded as taken in the evening will be considered as 1 inhalation in the calculation of compliance. Patients who received no randomised treatment will have zero compliance.

ePRO use compliance will be derived based on the actual number of morning and evening ePRO entries during the treatment period compared to the expected number of entries during that period. 0 evening ePRO entries are expected on the last day of treatment if it is also a visit day. Otherwise, 1 morning and 1 evening ePRO entry is expected on a study day.

Overall, morning and evening ePRO use compliance during the treatment period will also be calculated:



- Overall ePRO use compliance is defined as: total number of morning and evening ePRO entries during the treatment period / ([days during the treatment period - 1] \* 2 + expected number of morning and evening ePRO entries on the last day of treatment) \* 100
- Morning ePRO use compliance is defined as: total number of morning ePRO entries during the treatment period / days during the treatment period \* 100
- Evening ePRO use compliance is defined as: total number of evening ePRO entries during the treatment period / (days during the treatment period - 1 + expected number of evening ePRO entries on the last day of treatment) \* 100

Treatment compliance will be summarised for GFF inhaler and UV inhaler separately in each treatment group. In addition the following compliance categories will be derived: >130%, >110% - ≤130%, >90% - ≤110%, ≥70% - ≤90%, <70%. The numbers and percentages of patients falling into each category will be presented for GFF inhaler and UV inhaler separately in each treatment group. ePRO compliance will be summarised descriptively and using the compliance categories by treatment group.

#### **4.2.6 Analysis of the primary variables**

##### **4.2.6.1 Trough FEV<sub>1</sub>**

The co-primary analysis of trough FEV<sub>1</sub> will be conducted using the Per Protocol analysis set. The change from baseline in morning pre-dose trough FEV<sub>1</sub> will be analysed using a linear mixed effects model repeated measures (MMRM) approach. The model will include baseline FEV<sub>1</sub> and bronchodilator responsiveness to albuterol/salbutamol MDI (%) as continuous covariates and stratification factor (prior treatment - rescue/maintenance monotherapy vs double maintenance therapy), region (North America vs Europe), visit, treatment, and treatment by visit, as categorical covariates (fixed effects). Baseline is defined as the average of the non-missing -60 min and -30 min values obtained prior to dosing at Visit 3. The Kenward-Roger approximation will be used to estimate the degrees of freedom.

An unstructured matrix will be used to model the variance-covariance structure within patient. If this model fit fails to converge, more parsimonious covariance structures will be considered to model the correlation between time-points from the same patient. The following covariance structures will be tried in order until convergence is reached: Toeplitz with heterogeneity, autoregressive with heterogeneity, Toeplitz, autoregressive and compound symmetry. If either the autoregressive with heterogeneity or autoregressive structure is used, patient will be included as a random effect.

A contrast will be used to obtain an estimate of the treatment difference over the entire 24-week treatment period (i.e. average treatment difference over visits giving each visit equal weight) with 2-sided 95% CI and 97.5% CI. Non-inferiority will be concluded using a 1-sided hypothesis test according to a significance level of either 1.25% or 2.5% based upon the hierarchical testing and recycling procedure described in Section 4.2.6.3. This is equivalent to

the lower bound of the 2-sided 95% or 97.5% CI being greater than the pre-specified margin of -50 mL. The 1-sided p-values will be produced accordingly.

From the same model contrasts will also be used to produce estimates of treatment difference and the corresponding 95% CI and 97.5% CI at each individual time-point (for example at Week 24). In addition, least squares means and corresponding 95% CIs will be provided for each treatment group at individual time-points and over 24 weeks, and presented on a line plot.

A similar analysis will be conducted in the full analysis set using all data obtained while on treatment.

Summary statistics for absolute values and change from baseline in morning pre-dose trough FEV<sub>1</sub> will be produced by treatment group and visit for both Per Protocol analysis set and full analysis set.

The assumption of normality in the data will be checked by visually inspecting the distribution of the residuals. Also, the model fit and the assumption of homogeneity of variance will be verified by inspecting scatter plots of residuals vs predicted values and residuals vs continuous covariates, and by inspecting box plots of residuals for categorical covariates (prior treatment, region, visit, treatment). A sensitivity analysis, if appropriate, will be performed using the MMRM approach allowing for heterogeneous variances in covariates that a lack of homogeneity of variance is evident.

The MMRM approach used in primary analysis relies on the assumption that data is missing at random (MAR) given the covariates included in the model. Supportive analyses to assess the robustness of conclusions to this assumption, including under Missing Not at Random (MNAR) scenarios, are outlined in Appendix I. In this analysis, data that are missing due to treatment discontinuation where the reason is reasonably attributable to treatment (for example tolerability or lack of efficacy) and data following initiation of prohibited medication (as described in important protocol deviation 7, Section 2.2.1) will be imputed under an assumption that such patients would have had a poorer outcome had they been followed for 24 weeks and not been able to escalate treatment. This outcome will use imputation from a random intercept model, with imputation centred at the model based estimate of the 5% percentile for the endpoint at the corresponding visit (conditional on covariates). A (multiply) imputed change from baseline value will be drawn from a normal distribution with mean based on the 5<sup>th</sup> percentile of this distribution at the corresponding visit, as described in Appendix I. Other missing data that could not be reasonably attributed to treatment are to be imputed using their conditional distributions as implied by MAR. Treatment discontinuations reasonably attributable to tolerability or lack of efficacy will be identified and documented prior to unblinding. Principles are outlined in Appendix I.

Subgroup analyses will be conducted comparing the co-primary endpoint trough FEV<sub>1</sub> between the treatments in the following subgroups of the Per Protocol analysis set:

- Baseline CAT score (10-14, 15-19,  $\geq 20$ )

- Prior treatment at entry (rescue therapy only, single maintenance therapy, LAMA/LABA maintenance therapy, LABA/ICS maintenance therapy)
- Region (North America, Europe)
- Bronchodilator responsiveness at Visit 2 ( $\geq 12\%$  and  $\geq 200$  mL,  $< 12\%$  or  $< 200$  mL)
- COPD severity at entry, as defined by Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria (moderate, severe, very severe)
- Smoking status (current smoker, former smoker)

For each subgroup, a MMRM model will be fitted using the same model terms as used for the primary analysis for each level of the subgroup separately. The prior treatment, region and bronchodilator responsiveness will not be included in the model when these are the subgroups being analysed. Similar output will be presented for the subgroups as for the primary analysis. Note that the study has not been designed or powered to assess efficacy within any pre-defined subgroups, and as such these analyses are considered exploratory.

#### **4.2.6.2 Peak FEV<sub>1</sub>**

The peak change from baseline in FEV<sub>1</sub> within 2 hours post-dosing will be analysed using a similar MMRM model to trough FEV<sub>1</sub> in Section 4.2.6.1. Testing of non-inferiority between GFF and UV will first be conducted in the Per Protocol analysis set. The estimates of treatment difference over 24 weeks and associated 2-sided 95% and 97.5% CIs will be referred to a non-inferiority margin of -50 mL.

The full analysis set will be used to assess the consistency of the non-inferiority conclusion using all evaluable data and a similar repeated measures model. This model will then be used for the formal assessment of superiority hypotheses. Testing of non-inferiority and superiority will be at a significance level of either 1.25% or 2.5% based upon the hierarchical testing and recycling procedure described in Section 4.2.6.3. The 1-sided p-values will be produced accordingly for both non-inferiority analyses and the superiority analysis.

Contrasts will be used to produce estimates of treatment difference at individual visits. Least squares means and corresponding 95% CIs will be provided for each treatment group at individual visits and over 24 weeks, and presented on a line plot. The peak change from baseline values will also be summarised descriptively by visit for each treatment group.

Subgroup analyses will be generated comparing the peak change from baseline in FEV<sub>1</sub> between treatment groups in the same way as previously specified for trough FEV<sub>1</sub> in Section 4.2.6.1 based on the full analysis set. Supportive analyses will be conducted on the full analysis set to assess the robustness of conclusions from the MMRM approach to the MAR assumption as described in Section 4.2.6.1 and Appendix I.

#### 4.2.6.3 Hierarchical testing strategy – primary endpoints

In order to strongly control the 1-sided type I error rate at 2.5% across all co-primary endpoint hypothesis tests and the non-inferiority hypothesis tests of the secondary endpoints a hierarchical testing procedure with recycling will be defined.

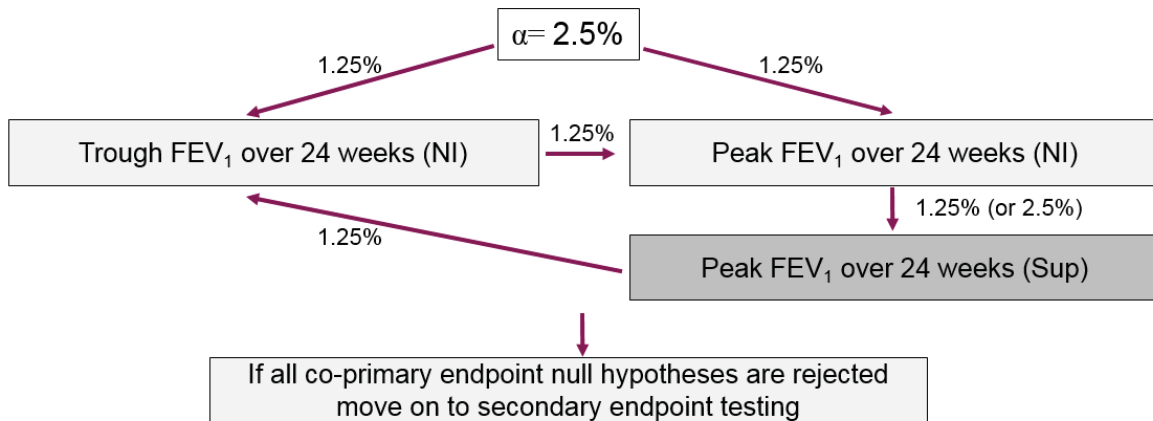
To account for the multiplicity of the two co-primary endpoints, the test mass (1-sided alpha of 2.5%) will initially be split equally between the tests of the non-inferiority null hypothesis for each of trough FEV<sub>1</sub> and peak FEV<sub>1</sub> (1-sided alpha of 1.25% each).

Should either of these null hypotheses be rejected, the test mass will be recycled (Burman et al 2009), as described in Figure 2. That is, after respective null hypotheses are rejected there will be recycling from each test as follows:

- from trough FEV<sub>1</sub> non-inferiority testing to peak FEV<sub>1</sub> non-inferiority testing
- from peak FEV<sub>1</sub> non-inferiority testing to peak FEV<sub>1</sub> superiority testing
- from peak FEV<sub>1</sub> superiority testing to trough FEV<sub>1</sub> non-inferiority testing

As such, respective hypothesis tests will be conducted at either a 1-sided 1.25% or 2.5% significance level depending upon this recycling.

**Figure 2 Testing hierarchy for primary endpoints**



NI = Non-inferiority, Sup = Superiority

Should all 3 null hypotheses relating to the co-primary endpoints be rejected then testing will proceed to the secondary endpoints using a 1-sided 2.5% significance level.

#### 4.2.7 Analysis of the secondary variables

Table 5 details the analyses to be performed for the endpoints, including the pre-planned subgroup and supportive analyses making clear which analysis and population is regarded as primary for that endpoint.

**Table 5 Statistical analyses and analysis populations**

Variable	Per Protocol analysis set	Full analysis set	Rescue Medication User analysis set
Pre-dose trough FEV <sub>1</sub> **	MMRM, for NI Subgroup analysis	MMRM MAR/MNAR analysis	-
Peak post-dose FEV <sub>1</sub> **	MMRM, for NI	MMRM, for Sup Subgroup analysis MAR/MNAR analysis	-
Peak post-dose IC*	MMRM, for NI	MMRM, for Sup Subgroup analysis MAR/MNAR analysis	-
Pre-dose trough FVC	MMRM	MMRM	-
Peak post-dose FVC	MMRM	MMRM	-
AUC <sub>0-2</sub> for FEV <sub>1</sub>	-	MMRM	-
AUC <sub>0-2</sub> for FVC	-	MMRM	-
Δ FEV <sub>1</sub> ≥100 mL at 5min, Day 1*	-	Logistic regression, for Sup	-
Δ FEV <sub>1</sub> ≥150 mL at 5min, Day 1	-	Logistic regression	-
Δ FEV <sub>1</sub> ≥ 12% at 5 min, Day 1	-	Logistic regression	-
FEV <sub>1</sub> post-dose timepoints, Day 1	-	MMRM	-
TDI focal score*	MMRM, for NI Subgroup analysis Responder analysis	MMRM Responder analysis MAR/MNAR analysis	-
TDI individual components	MMRM	MMRM	-
EMSCI symptom severity score *	MMRM, for NI	MMRM, for Sup Subgroup analysis	-
NiSCI symptom severity score	MMRM	MMRM Subgroup analysis	-
Overall COPD severity scores	Descriptive only	Descriptive only	-
Early morning activity score	Descriptive only	Descriptive only	-
Night-time awakening score	Descriptive only	Descriptive only	-
CAT score	MMRM Responder analysis	MMRM Subgroup analysis Responder analysis	-
Daily rescue use	-	-	MMRM
Daytime rescue use	-	-	MMRM
Night-time rescue use	-	-	MMRM

Variable	Per Protocol analysis set	Full analysis set	Rescue Medication User analysis set
Day with no rescue use	-	-	Descriptive only

\*\* Co-primary endpoints; \* Secondary endpoints in hierarchical testing; NI = Non-inferiority, Sup = Superiority.  
 Note: For endpoints with NI concluded on Per Protocol analysis set, supportive analyses will be conducted on full analysis set to demonstrate the consistency of this conclusion. CAT score and NiSCI will have treatment differences interpreted in relation to NI margin but not in hierarchical testing.

#### 4.2.7.1 Hierarchical testing strategy – secondary endpoints

A sequential testing hierarchy will be defined so as to strongly control 1-sided type I error at 2.5% across all non-inferiority hypothesis tests of secondary endpoints. The secondary endpoints over 24 weeks will be tested against their respective non-inferiority margins in the following sequence using a 1-sided significance level of 2.5%:

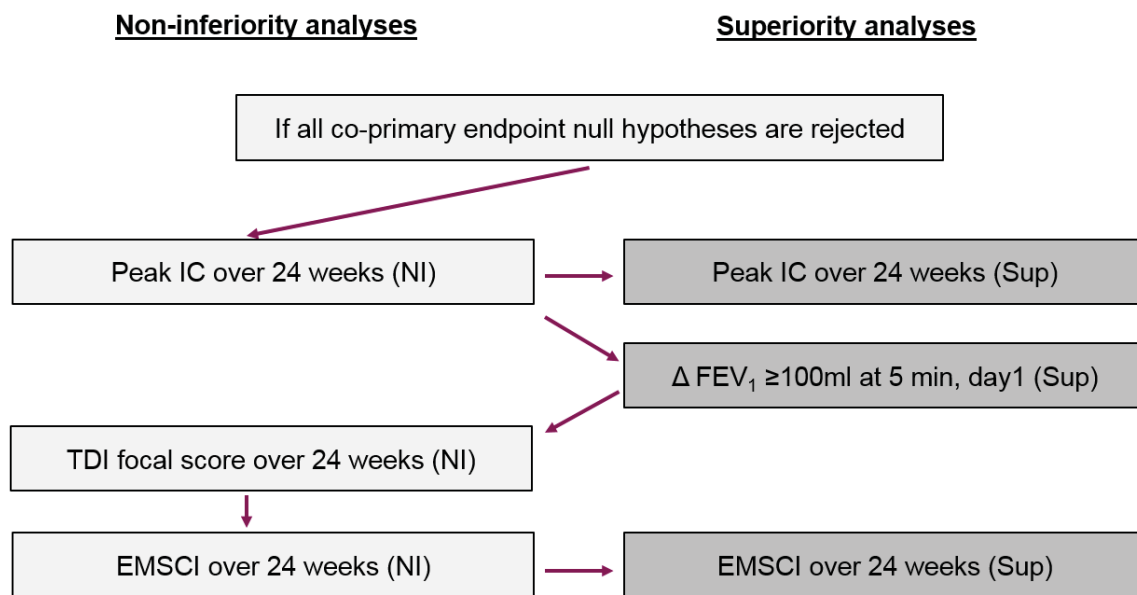
1. Peak change from baseline in IC within 2 hours post-dosing over 24 weeks (non-inferiority with option to proceed to superiority),
2. Proportion of patients with increase of FEV<sub>1</sub> of  $\geq 100$  mL from baseline at 5 minutes on Day 1 (superiority only),
3. Change from baseline in TDI focal score over 24 weeks (non-inferiority only)
4. Change from baseline in EMSCI Symptom Severity Score over 24 weeks (non-inferiority with option to proceed to superiority),

If non-inferiority is demonstrated for an endpoint, then testing may proceed to the next endpoint in the hierarchy, and a superiority test of the endpoint will also be conducted in the full analysis set for those endpoints indicated. Proceeding to the next endpoint is not contingent on demonstrating superiority.

An exception to this sequential non-inferiority/superiority testing procedure is the secondary endpoint Onset of action (proportion of patients with increase in FEV<sub>1</sub> of  $\geq 100$  mL from baseline at 5 minutes on Day 1). For this endpoint, superiority testing will immediately be carried out and superiority is required in order to move to the next step in the testing hierarchy.

This sequential testing hierarchy is illustrated in [Figure 3](#):

**Figure 3 Sequential testing hierarchy for secondary endpoints**



NI = Non-inferiority, Sup = Superiority

#### 4.2.7.2 Pulmonary Function over 24 weeks

The peak change from baseline in IC within 2 hours post-dosing will be analysed using a similar MMRM model to peak FEV<sub>1</sub> described in Section 4.2.6.2. The estimates of treatment difference over 24 weeks will also be referred to a non-inferiority margin of -50 mL.

If non-inferiority is demonstrated in the Per Protocol analysis then the full analysis set will be used to assess the consistency of the non-inferiority conclusion using all evaluable data and a similar repeated measures model. This model will be used for the formal assessment of superiority hypotheses and 1-sided p-values and 2-sided 95% CIs will be produced. The estimated treatment differences over 24 weeks and 95% CIs for peak change from baseline in IC will be presented graphically on a forest plot, along with the results of the co-primary variables.

Contrasts will be used to produce estimates of treatment difference at individual visits. Least squares means and corresponding 95% CIs will be provided for each treatment group at individual visits and over 24 weeks, and presented on a line plot. The peak change from baseline values will also be summarised descriptively by visit for each treatment group.

Subgroup analyses will be generated comparing peak change from baseline in IC between treatment groups in the same way as previously specified for morning pre-dose trough FEV<sub>1</sub> in Section 4.2.6.1 based on the full analysis set. Supportive analyses to assess the robustness of the repeated measures analyses to missing data as described in Section 4.2.6.1 are outlined in Appendix I.

Other continuous pulmonary function endpoints morning pre-dose trough and peak post-dose FVC will be analysed similarly using both the Per Protocol analysis set and full analysis set, and  $AUC_{0-2}$  for  $FEV_1$  and FVC will be analysed similarly using the full analysis set. The model term of baseline value will not be included (not applicable) when analysing  $AUC_{0-2}$  for  $FEV_1$  and FVC. Estimated treatment differences will be interpreted using 95% CIs, but will not be formally assessed for non-inferiority. Subgroup and supportive analyses above will not be performed for these endpoints.

#### **4.2.7.3 Pulmonary Function at Day 1**

On Day 1 at the assessment 5 minutes post-dosing, the proportion of patients achieving an improvement from baseline in  $FEV_1$  of  $\geq 100$  mL will be estimated for each treatment. Logistic regression will be used to compare the treatment groups with baseline and bronchodilator responsiveness to albuterol/salbutamol MDI as continuous covariates and treatment, region and stratification factor as categorical covariates. This analysis will use the full analysis set and proceed immediately to testing of the superiority hypothesis. One-sided p-values and odds ratios with 2-sided 95% CIs will be produced.

Similar analyses will be conducted at 5 minutes on Day 1 using alternative thresholds (increase in  $FEV_1$  of  $\geq 150$  mL from baseline and increase in  $FEV_1$  of  $\geq 12\%$  from baseline). A cumulative distribution function plot of the absolute increase in  $FEV_1$  at 5 minutes on Day 1 will also be presented by treatment group.

In addition, onset of action on day 1 will be assessed by analysing the change from baseline to each post-dose timepoint within the first 2 hours using the full analysis set. A similar MMRM model will be used to the co-primary analyses, but replacing visit with the post-dose timepoints in [Table 2](#). Differences will be interpreted using 95% CIs and least square means will be presented on a line plot illustrating the onset of action over the first 2 hours on day 1.

#### **4.2.7.4 Transition Dyspnea Index (TDI) Focal Score**

The difference between treatment groups in TDI over 24 weeks will be evaluated using a similar MMRM approach as for the co-primary endpoints described in [Section 4.2.6](#). The model will include BDI (in place of baseline) and bronchodilator responsiveness as continuous covariates. Stratification factor, region, visit, treatment, and treatment by visit will be included as categorical covariates.

One-sided p-values and point estimates with 2-sided 95% CIs will be produced for the treatment difference over 24 weeks. Non-inferiority will be evaluated on the Per Protocol analysis set, and concluded if the lower bound of this confidence interval is greater than the pre-specified margin of -1.0 units (the minimum clinically important difference). Estimates of treatment difference will also be produced at each individual time-point. Additionally, least squares means and corresponding 95% CIs will be provided for each treatment group at individual time-points and over 24 weeks, and presented on a line plot. A supportive analysis will also be produced using the full analysis set. Cumulative distribution function plots of TDI focal score at Week 24 will be presented by treatment group.



Subgroup analyses will be conducted comparing TDI focal score over 24 weeks between the treatment groups in the same subgroups as previously specified for morning pre-dose trough FEV<sub>1</sub> in Section 4.2.6.1 based on the Per Protocol analysis set.

For each subgroup, the same MMRM model defined above will be fitted separately for each level of the subgroup. The prior treatment, region and bronchodilator responsiveness will not be included in the model when these are the subgroups being analysed. Results from the model contrast will be presented in terms of the point estimate of treatment difference and corresponding 95% CI. The subgroup analyses will be considered exploratory and may only be supportive of the primary analysis of this endpoint.

Supportive analyses will be conducted on the full analysis set to assess the robustness of conclusions from the MMRM approach to the MAR assumption as described in Section 4.2.6.1 and Appendix I.

The individual components of the TDI: functional impairment, magnitude of task, and magnitude of effort will each be summarised relative to their corresponding BDI individual component score at baseline, using the same MMRM approach with estimated treatment differences over 24 weeks and at each visit, based on the Per Protocol analysis set and the full analysis set.

Furthermore as supportive analyses, responder analyses will be performed at each visit for the Per Protocol analysis set and the full analysis set, where responders are defined as an increase of 1.0 point or more in TDI focal score, and patients with missing data will be considered to be non-responders. Logistic regression will be used to compare the treatment groups with BDI as a continuous covariate and treatment group, region and stratification factor as categorical covariates. Odds ratios with 95% CIs will be produced. Summaries of the number and percentage of patients achieving an improvement, no change and deterioration will be provided by visit for each treatment group.

#### **4.2.7.5 Early Morning Symptoms COPD Instrument (EMSCI) and Night Time Symptoms COPD Instrument (NiSCI)**

The difference between treatment groups in the change from baseline in EMSCI Symptom Severity Score over 24 weeks will be evaluated using a similar MMRM approach as for the co-primary endpoints described in Section 4.2.6. Instead of visit, the relevant time interval will be used as a categorical covariate in the model. The model will include baseline and bronchodilator responsiveness as continuous covariates and stratification factor, region, time interval, treatment, and treatment by time interval, will be included as categorical covariates.

One-sided p-values and point estimates with 2-sided 95% CIs will be produced for the treatment difference over 24 weeks. Non-inferiority will be concluded if the lower bound of this confidence interval is greater than the pre-specified margin of -0.1 units. Estimates of treatment difference will also be produced at each individual time interval. Additionally, least squares means and corresponding 95% CIs will be provided for each treatment group at individual visits and over 24 weeks, and presented on a line plot.

If non-inferiority is demonstrated in the Per Protocol analysis then the full analysis set will be used to assess the consistency of the non-inferiority conclusion using all evaluable data and a similar repeated measures model. This model will be used for the formal assessment of superiority hypotheses and 1-sided p-values and 2-sided 95% CIs will be produced.

Estimates of least squares means from each treatment group and treatment differences will be produced similarly for the NiSCI Symptom Severity Score, based on the Per Protocol analysis set and the full analysis set. These will be interpreted using 95% CIs, in relation to a non-inferiority margin of -0.1.

Subgroup analyses, using the same approach and subgroup factor as specified in Section 4.2.6.1, will be performed for change from baseline in EMSCI and NiSCI Symptom Severity Scores respectively comparing the treatment difference over 24 weeks at each level of the subgroups, for patients in the full analysis set.

The change from baseline in Overall COPD Symptom Severity scores in the early morning and at night, Early Morning Activity score and Night-time Awakenings score, will be summarised descriptively by treatment group and time interval.

#### **4.2.7.6 COPD Assessment Test (CAT) score**

The difference between treatment groups in the change from baseline in CAT score over 24 weeks will be evaluated using a similar MMRM approach as for the co-primary endpoints described in Section 4.2.6, for the Per Protocol analysis set and the full analysis set. Treatment differences will be interpreted using 95% CIs in relation to non-inferiority margin of 2 units. Cumulative distribution function plot of the change from baseline values at Week 24 will be presented by treatment group for the full analysis set.

Subgroup analyses will be generated comparing the change from baseline in CAT score over 24 weeks between treatments in the same way as previously specified in Section 4.2.6.1, for the full analysis set. Baseline CAT score will not be included in the repeated measure models when it is the subgroup being analysed.

Responder analyses will be performed at each visit for the Per Protocol analysis set and the full analysis set, where responders are defined as a decrease of 2 units or more in the change from baseline value, and patients with missing data will be considered to be non-responders. Logistic regression will be used to compare the treatment groups with baseline CAT score as a continuous covariate and treatment group, region and stratification factor as categorical covariates. Odds ratios with 95% CI will be produced for each treatment comparison. Summaries of the number and percentage of patients achieving an improvement, no change and deterioration will be provided by visit for each treatment group.

#### **4.2.7.7 Daily Rescue Medication Use**

The difference between treatment groups in the change from baseline in mean daily rescue albuterol/salbutamol MDI use will be evaluated using a MMRM approach similar to the ones for co-primary endpoints described in Section 4.2.6, for patients in the Rescue Medication User analysis set. Instead of visit, the relevant time interval will be used as a categorical

covariate in the model. The model will include baseline and bronchodilator responsiveness as continuous covariates and stratification factor, region, time interval, treatment, and treatment by time interval, will be included as categorical covariates. Treatment differences will be interpreted using 95% CIs, but will not be formally assessed for non-inferiority.

The change from baseline in daytime and night-time mean albuterol/salbutamol MDI use will each be evaluated using the same MMRM approach as described above. Included in the models will be their corresponding baseline values. Summary statistics for these variables and percentage of days with no rescue use will be produced by treatment group and time period.

The overall daily, daytime and night-time rescue albuterol/salbutamol MDI use during the treatment period will also be summarised descriptively for each treatment group.

#### **4.2.8 Safety and tolerability**

Adverse event summaries, unless stated otherwise, will be based on TEAEs, as defined in Section 3.4. All AEs will be listed for each patient, regardless of whether or not they are treatment-emergent. All summaries will use the safety analysis set, and be presented by treatment group according to treatment received. No hypothesis tests will be performed.

An overall summary table will be produced showing the number and percentage of patients with at least 1 AE in any of the following categories: AEs, SAEs, AEs with outcome of death, AEs leading to discontinuation of IP (DAEs), and AEs deemed causally related to IP by the Investigator. The total number of AEs in the different AE categories in terms of AE counts will also be presented (i.e., accounting for multiple occurrences of the same event in a patient).

Adverse events, AEs with outcome of death, SAEs, DAEs, and AEs causally related to IP by the Investigator will be summarised by SOC and PT assigned to the event by MedDRA. For each PT, the number and percentage of patients reporting at least 1 occurrence will be presented, i.e. for a patient multiple occurrences of an AE will only be counted once. Serious adverse events deemed causally related to IP by the Investigator and/or AstraZeneca will also be summarised.

Each AE event rate (per 1000 patient years) will also be summarised by PT within SOC. For each PT, the event rate is defined as the total number of AEs divided by the total drug exposure of patients and then multiplied by  $365.25 \times 1000$  to present in terms of per 1000 patient years.

Summaries of the most common (frequency of >1% in any treatment group) AEs and non-serious AEs will be presented by PT. Adverse events will be summarised by preferred term and maximum intensity. If a patient reports multiple occurrences of the same AE, the maximum intensity will be taken as the highest recorded maximum intensity (the order being mild, moderate, and severe).

Separate listings of patients with AEs, AEs with outcome of death, SAEs, or DAEs will be presented.

#### **4.2.9 Exploratory analysis**

Exploratory variables as defined in Sections 3.3.5, 3.3.6 and 3.3.7 will be reported using the full analysis set.

The EQ-5D-5L responses from each dimension, the index score and the visual analogue scale will be summarised by treatment group. Shift tables will be produced for each dimension, and the change from baseline in index score and visual analogue scale will be summarised with descriptive statistics by visit.

The number and percentage of patients with COPD-related and non-COPD-related health resource utilisation will be presented by treatment group. The frequency and rate per year of COPD exacerbations, and moderate and severe COPD exacerbations will also be summarised.

Time to the first moderate or severe COPD exacerbation will be analysed using a Cox proportional hazards model. The model will include baseline FEV<sub>1</sub> (% PN), baseline CAT score, and bronchodilator responsiveness to albuterol/salbutamol MDI (%) as continuous covariates and stratification factor (prior treatment - rescue/maintenance monotherapy vs double maintenance therapy), COPD exacerbations in previous year (yes vs no), region (North America vs Europe) and treatment as categorical covariates. The estimated adjusted hazard ratio for GFF relative to UV will be displayed along with the associated Wald 2-sided 95% CI and 1-sided p-value for superiority assessment. Time to the first moderate or severe COPD exacerbation will be displayed graphically for each treatment group using a Kaplan-Meier curve.

WPAI-GH responses will be summarised using descriptive statistics by treatment group and visit.

## **5. INTERIM ANALYSES**

No interim analyses are planned for this study.

## **6. CHANGES OF ANALYSIS FROM PROTOCOL**

The CSP states that for the analysis of the co-primary endpoints and other similar analysis models patient will be considered a random effect. It is also stated that an unstructured covariance matrix will be used. These two statements are contradictory as if a random effect for patient is included there will be a diagonal residual covariance matrix. In the statistical analysis plan it is clarified that patient will only be included as a random effect if an autoregressive covariance structure is required, but that this will not be required if an unstructured or Toeplitz covariance structure can be used.

## **7. REFERENCES**

### **Burman et al 2009**

Burman C-F, Sonesson C, Guilbaud O. A recycling framework for the construction of Bonferroni-based multiple tests. *Statist. Med.* 2009; 28:739-761

### **Van Hout et al 2012**

Van Hout B, Janssen M, Feng Y, Kohlmann T, Busschbach J, Golicki D, Lloyd A, Scalone L, Kind P, Pickard AS. Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value in Health.* 2012; 15: 708-15