AstraZeneca

Statistical Analysis Plan

Study Code	PT010006
NCT#	NCT02497001
Date	09JANUARY2018

A Randomized, Double-Blind, Parallel-Group, 24-Week, Chronic-Dosing, Multi-Center Study to Assess the Efficacy and Safety of PT010, PT003, and PT009 Compared with Symbicort® Turbuhaler® as an Active Control in Subjects with Moderate to Very Severe Chronic Obstructive Pulmonary Disease

STATISTICAL ANALYSIS PLAN FOR STUDY PT010006 AND STUDY PT010007

Protocol Numbers:	PT010006 and PT010007
Investigational Drug and Drug Number:	BGF MDI; PT010 GFF MDI; PT003 BFF MDI; PT009 Symbicort® Turbuhaler®
Indication:	COPD
Dosage Form/Dose:	• BGF MDI 320/14.4/9.6 µg ex-actuator BID • GFF MDI 14.4/9.6 µg ex-actuator BID • BFF MDI 320/9.6 µg ex-actuator BID • Symbicort® Turbuhaler® 400/12 µg BID

PT010006 Protocol Title: A Randomized, Double-Blind, Parallel-Group, 24-Week, Chronic-Dosing, Multi-Center Study to Assess the Efficacy and Safety of PT010, PT003, and PT009 Compared with Symbicort® Turbuhaler® as an Active Control in Subjects with Moderate to Very Severe Chronic Obstructive Pulmonary Disease

PT010007 Protocol Title: A Randomized, Double-Blind, Parallel-Group, 28-Week, Chronic-Dosing, Multi-Center, Extension Study to Assess the Safety and Efficacy of PT010, PT003, and PT009 in Japanese Subjects with Moderate to Very Severe Chronic Obstructive Pulmonary Disease (COPD) compared with Symbicort® Turbuhaler® as an Active Control

Date of Issue: 09 Jan 2018

Version: Version 2.0

Signed Agreement on Statistical Analysis Plan

FINAL SIGN-OFF SIGNATURES

Primary Biostatistician and Peer Reviewer: [redacted]

Author: [redacted]

Approved by: [redacted]

Approved by: [redacted]

Approved by: [redacted]

Change Log			
Version No.	**Effective Date**	**Reason for the Change / Revision**	**Supersedes**
2.0	09 Jan 2018	• Non-inferiority margins for are provided for the following secondary endpoints: TDI, MCID in SGRQ, peak change from baseline in FEV_1, and E-RS Total Score. • The definition of ICS use at screening has been clarified. • Have clarified the definitions of the on-treatment and post-treatment periods. • Have clarified that subjects with premature treatment discontinuation are to be considered to be non-responders for the responder analyses. • Have specified how to estimate the attributable estimand for the SGRQ responder analysis. • Have provided more detail about the statistical analysis of the EuroQOL EQ-5D questionnaire. • Have clarified the baseline Exacerbation History categories to be using for sub-group analyses. • Have specified correlation analyses for subgroups. • Have clarified that the Attributable Estimand is considered to be secondary. • Minor edits have been made for clarity and grammatical correctness.	1.0

TABLE OF CONTENTS

List of Tables

List of Figures

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

AE	Adverse event
AESI	Adverse events of special interest
ALT	Alanine aminotransferase
ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
AR(1)	Autoregressive order 1
AST	Aspartate aminotransferase
ATS	American Thoracic Society
AUC	Area under the curve
BDI	Baseline Dyspnea Index
BFF MDI	Budesonide and Formoterol Fumarate Metered Dose Inhaler
BGF MDI	Budesonide, Glycopyrronium, and Formoterol Fumarate Metered Dose Inhaler
BID	Bis in die, twice daily
bpm	Beats per minute
BMI	Body mass index
CAT	Chronic Obstructive Pulmonary Disease Assessment Test
CCU	Coronary care unit
CCV	Cardio- and cerebrovascular
CD	Compact disc
CI	Confidence interval
CID	Clinically important deterioration
CKD-EPI	Chronic Kidney Disease Epidemiology Collaboration
cm	Centimeter
C_{min}	The minimum observed plasma concentration, expressed in concentration units
C_{max}	The maximum observed plasma concentration, expressed in concentration units

C_{avg}	Average concentration during a dosing interval
COPD	Chronic obstructive pulmonary disease
CTCAE	Common Terminology Criteria for Adverse Events
δ	Non-inferiority margin
DMC	Data monitoring committee
E-RS	Evaluating Respiratory Symptoms
ECG	Electrocardiogram
eCRF	Electronic case report form
eDiary	Electronic Diary
e.g.	Exempli gratia; for example
eGFR	Estimated glomerular filtration rate
EQ-5D	EuroQol 5 Dimensions Questionnaire
EQ-5D-5L	EuroQol 5 Dimensions Questionnaire 5-level
ER	Emergency room
EU	European Union
ex-actuator	Dose delivered from the actuator (i.e., mouthpiece) of the MDI
EXACT	Exacerbations of Chronic Pulmonary Disease Tool – Patient Reported Outcomes
FEV_1	Forced expiratory volume in the first second
FEF_{25-75}	Forced expiratory flow between 25% and 75% of FVC
FVC	Forced vital capacity
GFF MDI	Glycopyrronium and Formoterol Fumarate Metered Dose Inhaler
H_0	Null Hypothesis
H_1	Alternative hypothesis
hCG	Human chorionic gonadotropin
HCRU	Health Care Resource Utilization
HFA	Hydrofluoroalkane
HLGT	High-Level Group Term

HLT	High-Level Term
HPA	Hypothalamic-pituitary-adrenal
ICF	Informed consent form
ICS	Inhaled corticosteroid
ICU	Intensive care unit
i.e.	Id est; that is
ITT	Intent-to-treat
IWRS	Interactive web response system
JRS	Japanese Respiratory Society
λ_z	The terminal elimination rate constant, calculated from the slope of the terminal portion of the ln(drug concentration) versus time curve
L	Liter
LABA	Long-acting β_2 agonist
LAMA	Long-acting muscarinic antagonist
LLQ	lower limit of quantification
MACE	Major adverse cardiovascular event
MAR	Missing at random
MCAR	Missing completely at random
MCMC	Markov chain Monte Carlo
MCID	Minimal clinically important difference
MDI	Metered dose inhaler
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple imputation
MNAR	Missing not at random
µg	Microgram
mITT	Modified intent-to-treat
mL	Milliliter
mm	Millimeter

mmHg	Millimeter of mercury
msec (ms)	Millisecond
NHANES	National Health and Nutrition Examination Survey
OTC	Over-the-counter
PCS	Potentially clinically significant
PEFR	Peak expiratory flow rate
PFT	Pulmonary function test
PK	Pharmacokinetic
PMM	Pattern mixture model
PP	Per-protocol
PT	Preferred Term
PT003	Glycopyrronium and Formoterol Fumarate Inhalation Aerosol
PT009	Budesonide and Formoterol Fumarate Inhalation Aerosol
PT010	Budesonide, Glycopyrronium, and Formoterol Fumarate Inhalation Aerosol
QoL	Quality of life
QTcF	QT corrected using Fridericia's formula
RM	Repeated measures
ROM	Read-only memory
SABA	Short-acting β_2-agonist
SAE	Serious adverse event
SAMA	Short-acting muscarinic antagonist
SAP	Statistical analysis plan
SC	Serum cortisol
SD	Standard deviation
SGRQ	St. George's Respiratory Questionnaire
SMQ	Standard MedDRA Query
SOC	System Organ Class

$t_{1/2}$	The apparent terminal elimination half-life, expressed in hours.
TBH	Turbuhaler
TDI	Transition Dyspnea Index
TEAE	Treatment-emergent adverse event
TLFs	Tables, listings, and figures
t_{max}	The time to reach maximum observed plasma concentration (C_{max}), expressed in hours
ULN	Upper limit of normal
US	United States
VAS	Visual analog score
WAVE	Weighted average
WHO-DD	World Health Organization Drug Dictionary

Trademark Information

Trademarks Not Owned By Pearl

Atrovent

SAS

Ventolin

1. INTRODUCTION

This Statistical Analysis Plan (SAP) outlines the statistical methods for the display, summary and analysis of data to be performed at the end of Pearl Therapeutics, Inc. (Pearl) Study PT010006, and its extension sub-study, PT010007, in Japanese subjects. Pearl Therapeutics is a member of the AstraZeneca group of companies. Sponsor means either Pearl or AstraZeneca. The SAP should be read in conjunction with the study protocol. This version of the SAP has been developed using the PT010006-01 Amended Protocol (Version 3.0 dated 25 August 2017) and the PT010006 CRF (Version 02 dated 19 February 2016) as well as the PT010007 Protocol (Version 2.0 dated 17 March 2017) and the PT010007 CRF (Version Revision 01 dated 25 May 2017).

2. STUDY OBJECTIVES AND ENDPOINTS

2.1 Study Objectives

The overall objective of PT010006 is to assess the efficacy and safety of treatment with BGF MDI 320/14.4/9.6 µg (micrograms) (budesonide, glycopyrronium, and formoterol fumarate metered dose inhaler), GFF MDI 14.4/9.6 µg (glycopyrronium and formoterol fumarate metered dose inhaler), BFF MDI 320/9.6 µg (budesonide and formoterol fumarate metered dose inhaler), and Symbicort® Tubuhaler® (TBH) 400/12 µg over 24 weeks in subjects with moderate to very severe chronic obstructive pulmonary disease (COPD).

The objective of the Japanese extension sub-study PT010007 is to assess the long term safety and tolerability of treatment with BGF MDI 320/14.4/9.6 µg, GFF MDI 14.4/9.6 µg, BFF MDI 320/9.6 µg, and Symbicort® Turbuhaler® 400/12 µg over 52 weeks in Japanese subjects with moderate to very severe COPD.

2.2 Objectives for PT010006:

2.2.1 Primary Objective

- To assess the effects of BGF MDI, GFF MDI, BFF MDI, and Symbicort TBH on lung function.

2.2.2 Secondary Objectives

- To assess the effects of BGF MDI, GFF MDI, BFF MDI, and Symbicort TBH on dyspnea.
- To assess the effects of BGF MDI, GFF MDI, BFF MDI, and Symbicort TBH on quality of life (QoL).
- To assess the effects of BGF MDI, GFF MDI, BFF MDI, and Symbicort TBH on symptoms of COPD.
- To assess the effects of BGF MDI, GFF MDI, BFF MDI, and Symbicort TBH on COPD exacerbations.

- To determine the time to onset of action of BGF MDI, GFF MDI, BFF MDI, and Symbicort TBH.

2.2.3 Safety Objectives

- To assess the safety of BGF MDI, GFF MDI, BFF, and Symbicort TBH.

2.2.4 Healthcare Resource Utilization Objective

- To assess overall and COPD-specific Healthcare Resource Utilization (HCRU) of BGF MDI, GFF MDI, BFF MDI, and Symbicort TBH.

2.3 Objectives for Sub-studies in PT010006

2.3.1 12-Hour Pulmonary Function Test (PFT) Sub-study Objective

- To assess the effect of BGF MDI, GFF MDI, BFF MDI, and Symbicort TBH on PFT parameters over 12 hours.

2.3.2 Pharmacokinetic Sub-study Objective

- To characterize the steady state pharmacokinetics of budesonide, glycopyrronium, and formoterol based on pharmacokinetic (PK) assessments.

2.3.3 HPA Axis Sub-study Objective

To assess the effect of BGF MDI, GFF MDI, BFF MDI, and Symbicort TBH on hypothalamic-pituitary-adrenal (HPA) axis function.

2.4 Objectives for PT010007:

2.4.1 Primary Objective

- To evaluate the long-term safety and tolerability of BGF MDI, GFF MDI, BFF MDI, and Symbicort TBH in Japanese subjects with moderate to very severe COPD.

2.4.2 Other Objectives

- To assess the effect of BGF MDI, GFF MDI, BFF MDI, and Symbicort TBH on lung function.
- To assess the effect of BGF MDI, GFF MDI, BFF MDI, and Symbicort TBH on COPD exacerbations.
- To assess the effect of BGF MDI, GFF MDI, BFF MDI, and Symbicort TBH on symptoms using the change in rescue medication use as an indirect measure of symptom control.

2.5 Study Endpoints

2.5.1 Efficacy Endpoints for PT010006

The primary endpoints, treatment comparisons of interest, and analysis timeframes may differ by country or region due to local regulatory agency requirements. The 3 different registration approaches will be called: (1) Japan/China, (2) Europe (EU)/Canada, and (3) United States (US). Countries not specifically mentioned will be decided by regulatory requirements and included in one of the three defined registration approaches. The delineation of multiplicity controls for the primary and secondary measures will be separated by approach.

2.5.1.1 Primary Efficacy Endpoints

Primary Endpoint (Japan/China Approach)

- Change from baseline in morning pre-dose trough FEV_1 (forced expiratory volume in the first second) over Weeks 12 to 24 (BGF MDI versus BFF MDI, BGF MDI versus GFF MDI, and BFF MDI versus Symbicort TBH)

Primary Endpoints (European Union [EU] and Canada Approaches)

- FEV_1 area under the curve from 0 to 4 hours (AUC_{0-4}) over 24 weeks (BGF MDI versus BFF MDI and BGF MDI versus Symbicort TBH)
- Change from baseline in morning pre-dose trough FEV_1 over 24 weeks (BGF MDI versus GFF MDI and BFF MDI versus Symbicort TBH [non-inferiority])

Primary Endpoints (United States [US] Approach)

- FEV_1 AUC_{0-4} at Week 24 for the comparison of BGF MDI to BFF MDI
- Change from baseline in morning pre-dose trough FEV_1 at Week 24 for the comparison of BGF MDI to GFF MDI

2.5.1.2 Secondary Efficacy Endpoints

Endpoints that are not considered primary for a specific approach or region have been included under secondary endpoints for that region; as a result, secondary endpoints may differ between approaches.

Secondary Endpoints (Japan/China Approach):

- Change from baseline in morning pre-dose trough FEV_1 over 24 weeks
- FEV_1 AUC_{0-4} over Weeks 12 to 24
- Change from baseline in St. George Respiratory Questionnaire (SGRQ) total score over Weeks 12 to 24
- Transition dyspnea index (TDI) focal score over Weeks 12 to 24

- Change from baseline in average daily rescue Ventolin Hydrofluoroalkane (HFA) use over 24 weeks
- Peak change from baseline in FEV_1 within 4 hours post-dosing over Weeks 12 to 24
- Time to clinically important deterioration (CID)
- Time to onset of action on Day 1

Secondary Endpoints (Europe [EU] and Canada Approach):

- Change from baseline in morning pre-dose trough FEV_1 over 24 weeks (BGF MDI versus BFF MDI)
- TDI focal score over 24 weeks (EU only)
- Change from baseline in SGRQ total score over 24 weeks
- Change from baseline in average daily rescue Ventolin HFA use over 24 weeks
- Peak change from baseline in FEV_1 within 4 hours post-dosing over 24 weeks
- Rate of Moderate or Severe COPD Exacerbations
- Change from baseline in the Evaluating Respiratory Symptoms in COPD (E-RS: COPD) total score (RS-Total Score) over 24 weeks (EU only)
- Time to CID
- Time to onset of action on Day 1

Secondary Endpoints (US Approach):

- Change from baseline in morning pre-dose trough FEV_1 over 24 weeks
- Percentage of subjects achieving an minimal clinically important difference (MCID) of 4 units or more in SGRQ total score (SGRQ responders) at Week 24
- Change from baseline in average daily rescue Ventolin HFA use over 24 weeks.
- Peak change from baseline in FEV_1 within 4 hours post-dosing at Week 24
- Rate of Moderate or Severe COPD Exacerbations
- Time to onset of action on Day 1

2.5.1.3 **Other Efficacy Endpoints**

Wherever stated, analyses of an endpoint at each post-randomization visit will be performed only at time points where the endpoint will be assessed per the schedule of assessments.

Day 1 Assessments:

- Change from baseline at each post-dose time point in FEV_1, forced vital capacity (FVC), peak expiratory flow rate (PEFR), and forced expiratory flow between 25% to 75% (FEF_{25-75})
- Proportion of subjects achieving an improvement from baseline in FEV_1 using different thresholds (e.g., ≥10%, ≥12%, ≥100 mL [milliliter], ≥200 mL; and ≥12% and ≥200 mL)

Assessments Over 24 Weeks (Unless Otherwise Stated):

- Rate of moderate or severe COPD exacerbations
- Rate of COPD exacerbations of any severity
- Rate of severe COPD exacerbations
- Time to treatment failure (treatment discontinuation for any cause, moderate or severe exacerbation, or death)
- Time to first moderate or severe COPD exacerbation
- Time to first COPD exacerbation of any severity
- Time to first severe COPD exacerbation
- Time to CID
- Time to sustained CID
- Additional spirometry assessments over 24 weeks, over Weeks 12 to 24, and at each post-randomization visit:
  - Change from baseline in morning pre-dose trough for FEV_1, FVC, PEFR, and FEF_{25-75}
  - Peak change from baseline within 4 hours in FEV_1, FVC, PEFR, and FEF_{25-75}
  - FEV_1 AUC_{0-4}, FVC AUC_{0-4}, PEFR AUC_{0-4}, and FEF_{25-75} AUC_{0-4}
- Change from baseline in: the EXACT total score, E-RS: COPD total score (RS-Total Score), as well as 3 subscale scores symptom (RS-Breathlessness, RS-Cough and Sputum, and RS-Chest Symptoms over each 4-week interval of the 24-week Treatment Period
- TDI focal score at each post-randomization visit
- Individual components of the TDI: functional impairment, magnitude of task, and magnitude of effort over 24 weeks, over Weeks 12 to 24, and at each post-randomization visit
- Percentage of subjects achieving an MCID threshold of 1 unit or more on average in TDI focal score over 24 weeks and separately over Weeks 12 to 24
- Changes from baseline at each post-randomization visit for SGRQ total score
- Change in individual domain scores of SGRQ: Symptoms, Activity, and Impacts over 24 weeks, over Weeks 12 to 24, and at each post-randomization visit
- Percentage of subjects achieving an MCID of 4 units or more in SGRQ total score at Week 24, over 24 weeks, and separately over Weeks 12 to 24
- Quality-of-Life Endpoints: European Quality-of-Life-5 Dimensions (EQ-5D-5L) scored at randomization and each post-randomization visit

2.5.1.4 **Efficacy Endpoints for PT010007:**

All efficacy endpoints for PT010007 are exploratory. The data from this 28-week study will be combined with the 24 weeks of data obtained from Study PT010006 to provide safety and efficacy data over 52 weeks of treatment.

- Change from baseline in morning pre-dose trough FEV_1 over 52 weeks and at each post-randomization visit

- Change from baseline in average daily rescue Ventolin HFA (albuterol sulfate) use
- Percentage of days with no rescue Ventolin HFA use
- Rate of moderate or severe COPD exacerbations
- Rate of COPD exacerbations of any severity
- Change from baseline in morning pre-dose trough over 52 weeks, and at each post-randomization visit for:
    - FVC
    - PEFR
    - FEF_{25-75}
- Change from baseline over 52 weeks and over each 4-week interval of the 52-week Treatment Period in:
    - the EXACT total score
    - the RS-Total Score
    - 3 subscale scores symptom (RS-Breathlessness, RS-Cough and Sputum, and RS-Chest Symptoms)

2.5.2 Safety Endpoints

The safety endpoints for both PT010006 and PT010007 include:

- Adverse events (AEs), Treatment-emergent AEs, serious adverse events (SAEs), adverse events of special interest (AESIs)
- Major Adverse Cardiovascular Events (MACE)
- Confirmed cases of pneumonia
- 12-lead electrocardiograms (ECGs): Change from baseline in heart rate, PR interval, QRS axis, QRS interval, QT interval, and QTcF (Fridericia Corrected QT) interval
- Clinical laboratory testing
- Vital signs measurements

Safety endpoints for PT010006 are observed over 24 weeks, and those for PT010007 over 52 weeks.

2.5.3 Sub-Study Endpoints (for PT010006)

2.5.3.1 12-hour PFT Sub-Study

The primary 12-hour PFT endpoint is:

- FEV_1 AUC_{0-12} at Week 24

Additional assessments at Week 24:

- FEV_1 at each time point
- Serial spirometry parameters including FEV_1 AUC_{0-6}, FEV_1 $AUC_{6-12,}$ and time to peak FEV_1
- FVC, PEFR, and FEF_{25-75} will be evaluated using AUC_{0-12}

2.5.3.2 Pharmacokinetic Sub-study Endpoints:

The PK endpoints at Week 24 (i.e. steady-state, where "i.e." denotes "id est; that is") include:

- Area under the plasma concentration-time curve from time 0 to 12 hours post dose (AUC_{0-12})
- Time to reach maximum observed plasma concentration (t_{max})
- Maximum observed plasma concentration (C_{max})
- Terminal elimination rate constant (λ_z)
- Terminal elimination half-life ($t_{1/2}$)
- Minimum observed plasma concentration (Cmin)
- Time-average concentration during a dosing interval (C_{avg})
- %Fluctuation
- %Swing

2.5.3.3 HPA Axis Sub-study Endpoints:

Primary Endpoint:

- Ratio to Baseline of the 0- to 24-hour weighted mean serum cortisol (SC) concentration curve at Visit 10a (Week 24)

2.5.4 Health Care Resource Utilization Endpoints (for Study PT010006)

- The number of days missed from work due to COPD
- The number of days that primary caregivers of subjects missed from work as a result of the subject's COPD
- The percentage of subjects with telephone calls to health-care providers
  - Call to any health-care provider (physician or other)
  - Calls to physician
  - Calls to other healthcare provider
- The mean number of telephone calls to health-care providers
  - Call to any health-care provider (physician or other)
  - Calls to physician
  - Calls to other healthcare provider
- The percentage of subjects with visits to health-care providers
  - Visits to any health-care provider (general practitioner [GP], specialist, or other)

  - Visits to GP
  - Visits to specialist
  - Visits to other health-care provider
- The mean number of visits to health-care providers
  - Visits to any health-care provider (GP, specialist, or other)
  - Visits to GP
  - Visits to specialist
  - Visits to other health-care provider
- The percentage of subjects with Emergency Room (ER) visits
- The mean number of visits to ERs
- The percentage of subjects hospitalized
- The mean number of subject hospitalizations
- The mean number of days in the hospital
- The mean number of hospitalizations in which subject spent some time in the Intensive Care Unit (ICU) or the Coronary Care Unit (CCU)
- The percentage of subjects hospitalized with some time spent in the ICU or CCU
- The mean number of days in the hospital with some time spent in the ICU or CCU
- The mean number of hospitalizations in which subject spent No time in the ICU or the CCU
- The percentage of subjects hospitalized with No time in the ICU or CCU
- The mean number of days in the hospital with No time spent in the ICU or CCU
- The mean number of days in ICU
- The percentage of subjects in the ICU
- The mean number of days in CCU
- The percentage of subjects in the CCU
- The percentage of subjects who required ambulance transport
- The mean number of times ambulance transport was required

3. STUDY DESIGN AND ANALYTICAL CONSIDERATIONS

3.1 Study Designs

3.1.1 Overall Study Designs and Plan

PT010006 study is a multi-center, randomized, double-blind, parallel-group, chronic-dosing (24 weeks), active-controlled study to assess the efficacy and safety of BGF MDI, GFF MDI, BFF MDI, and open-label Symbicort TBH as an active control in subjects with moderate to very severe COPD that remain symptomatic (COPD Assessment Test [CAT]≥10) on two or more inhaled maintenance treatments.

This study will be conducted at approximately 160 sites, contributing approximately 10 to 20 subjects per site. Subject participation in the PT010006 study and all sub-studies will be determined at Screening, prior to any study procedures. Approximately 1800 subjects with moderate to very severe COPD will be randomized in a 2:2:1:1 scheme into the study to provide

approximately 1600 subjects to complete the study. Approximately 600 subjects each will be randomized to the BGF MDI and GFF MDI treatment groups, and 300 subjects each will be randomized to the BFF MDI and Symbicort TBH treatment groups. Randomization will be stratified by reversibility to Ventolin HFA, country, and disease severity.

All Japanese sites that participated in Study PT010006 will be eligible to contribute subjects to extension study PT010007. It is planned that approximately 324 Japanese subjects with moderate to very severe COPD will continue into study PT010007 for an additional 28 weeks to provide approximately 300 subjects to complete the study. Based on the randomization ratio from the Study PT010006, PT010007 will evaluate approximately 100 completed Japanese subjects in the BGF MDI and GFF MDI arms, and approximately 50 completed Japanese subjects in the BFF MDI and Symbicort TBH arms.

In PT010006, subjects who discontinue study treatment prior to Week 24 (Visit 10a) will be encouraged to remain in the study to complete all remaining study visits during the 24 week treatment period. Subjects who agree to continue to be followed post treatment discontinuation will sign an informed consent form (ICF) addendum. All subjects who agree to continue study participation beyond treatment discontinuation will complete a Treatment Discontinuation/Withdrawal Visit prior to transitioning back to regularly scheduled study visits. Subjects participating in a sub-study who choose to discontinue from treatment will only complete regularly scheduled visits and not complete any remaining sub-study assessments. Treatment discontinuation subjects will return to appropriate maintenance COPD medications, per the investigators discretion.

If a subject chooses not to continue with study assessments, at a minimum the subject will complete the Treatment Discontinuation/Withdrawal Visit (refer to the Schedule of Events in the Study Protocol). These subjects will return to appropriate maintenance COPD medications, per the investigators discretion. A follow-up telephone call will be performed at least 14 days after the last study drug dose. In the event the Treatment Discontinuation/Withdrawal Visit is performed >14 days post last study drug dosing, a follow-up TC will not be required. These subjects will be followed for vital status at 24 weeks post randomization in accordance with the informed consent.

Sub-Studies:

The PT010006 study will include the following 3 sub-studies:

12-Hour PFT Sub-study (US sub-study only): Serial PFTs will be conducted over 12 hours in a subset of approximately 600 randomized subjects (200 subjects from each of BGF MDI and GFF MDI treatment groups, and 100 subjects from each of the BFF MDI and Symbicort TBH treatment groups) at Visit 10a (Week 24). On the test day, additional serial spirometry will be obtained at 6, 8, 10, 11.5 and 12 hours post-dose.

Pharmacokinetic Sub-study: PK assessments will be performed in a subset of subjects who participate in the PFT sub-study. Approximately 240 randomized subjects (80 subjects from each

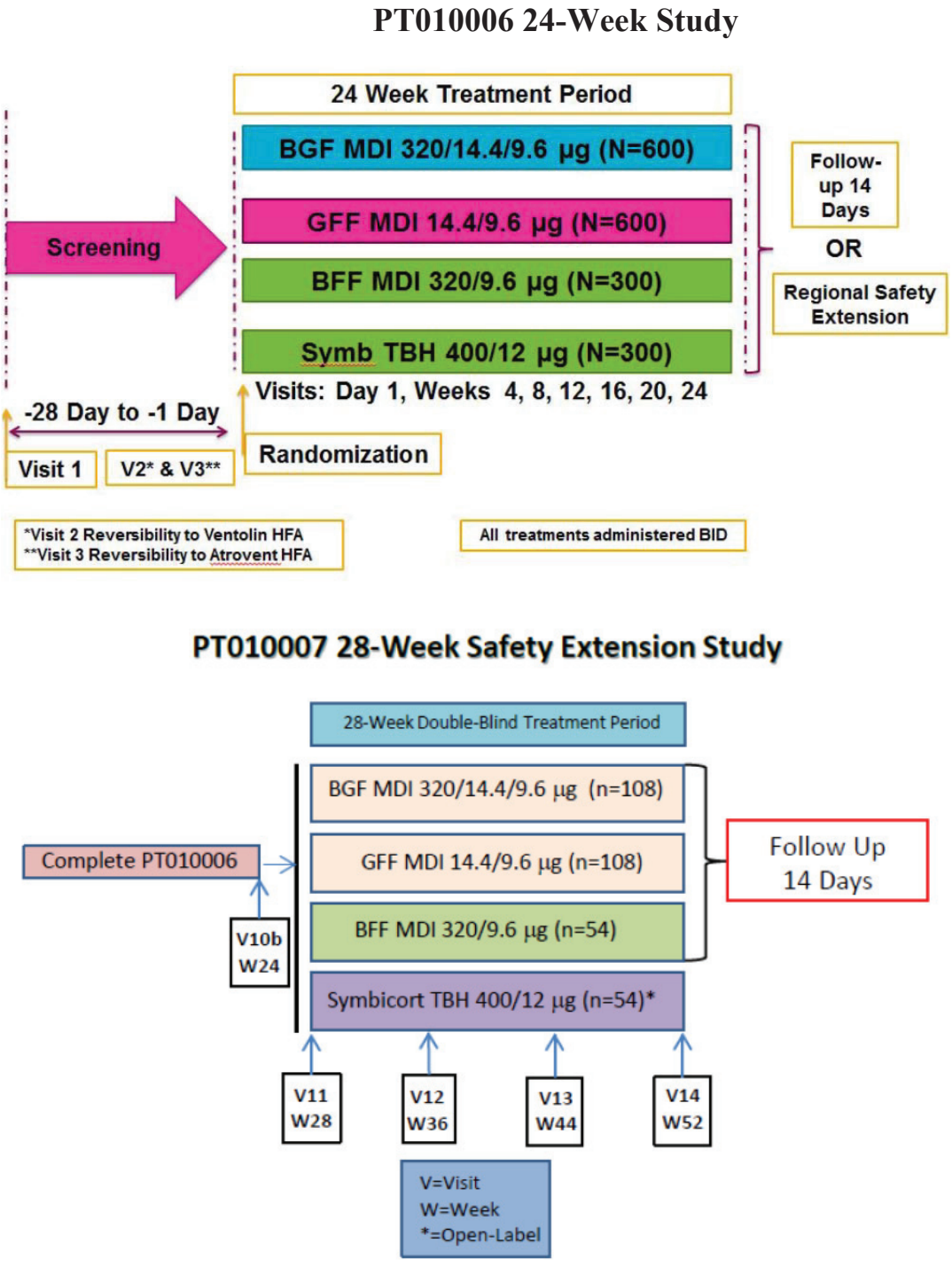
of the BGF MDI and GFF MDI treatment groups, and 40 subjects from each of the BFF MDI and Symbicort TBH treatment groups) will be assessed at Visit 10a (Week 24).

HPA Axis Sub-study: Adrenocorticosteroid activity will be assessed in a subset of subjects in the PK sub-study. SC will be measured in approximately 108 randomized subjects (36 subjects from each of the BGF MDI and GFF MDI treatment groups, and 18 subjects from each of the BFF MDI and Symbicort TBH treatment groups) over 24 hours, between Visits 3 and 4 prior to dosing at Randomization and Visit 10a (Week 24).

The Schedules of Events and Timed Assessments are in the study protocols.

The overall study design is summarized and illustrated in Figure 1.

Figure 1 Study Design



Abbreviations: BFF = Budesonide and Formoterol Fumarate; BGF = Budesonide, Glycopyrronium, and Formoterol Fumarate; GFF = Glycopyrronium and Formoterol Fumarate; MDI = metered dose inhaler; TBH = Tubuhaler; V = visit.

3.1.2 Prior, Concomitant, Post-Treatment, Prohibited Medications, and Other Restrictions (if applicable)

All prescription and over-the-counter (OTC) medications taken by the subject within 30 days before Visit 1 (Screening) will be recorded on the prior/concomitant medications electronic case report form (eCRF). All concomitant medications taken during the study will be recorded on the Concomitant Medications eCRF page with indication, total daily dose, dose regimen, and dates of drug administration.

3.2 Hypothesis Testing

For the primary comparisons, the null hypothesis for each pair-wise comparison will be that the mean treatment difference is zero (mean treatment effects are equal). The alternative hypothesis is that the mean treatment difference is greater than (less than) zero (mean treatment effects are not equal). All comparisons will be for superiority except that the comparison of BFF MDI to Symbicort TBH will be for non-inferiority and will use margins (δs) of -50 mL for the lower bound of a 2-sided 95% CI for the treatment difference for morning pre-dose trough FEV_1 and -75 mL for FEV_1 AUC_{0-4}. Margins for secondary endpoints are provided with their descriptions below in Section 6.4. P-values will be reported as 2-sided.

The primary null (H_0) and alternative (H_1) hypotheses with μ representing the mean are:

- H_0: $\mu_{BGF} = \mu_{GFF}$
  H_1: $\mu_{BGF} \neq \mu_{GFF}$

- H_0: $\mu_{BGF} = \mu_{BFF}$
  H_1: $\mu_{BGF} \neq \mu_{BFF}$

- H_0: $\mu_{BFF} \leq \mu_{Symbicort} - 50$ mL for morning pre-dose trough FEV_1
  H_1: $\mu_{BFF} > \mu_{Symbicort} - 50$ mL for morning pre-dose trough FEV_1

Secondary and other efficacy analyses will involve the above hypotheses applied to secondary efficacy endpoints. The directionality – being "< "or ">" -- of H_1 will depend on the endpoint.

3.3 Interim Analysis

No interim efficacy analyses are planned for this study.

A Data Monitoring Committee (DMC) will review safety data approximately every 6 months. Further detail is given in the DMC Charter.

3.4 Sample Size

For Study PT010006, it is estimated that a sample size of 1800 subjects (600 per arm in the BGF MDI and GFF MDI groups and 300 per arm in the BFF MDI and Symbicort TBH groups) will provide the following power estimates, all assuming Type I error control at a 2-sided alpha level

of 0.05 unless specified otherwise: 99% power to detect a difference of 75 mL between BGF MDI and BFF MDI in FEV_1 AUC_{0-4} over 24 weeks; 96% power to detect a difference of 35 mL between BGF MDI and GFF MDI in morning pre-dose trough FEV_1 over 24 weeks and approximately 92% power over Weeks 12 to 24; 97% power to detect a difference of 50 mL between BGF MDI and BFF MDI in morning pre-dose trough FEV_1 over Weeks 12 to 24; and 96% power to demonstrate non-inferiority of BFF MDI to Symbicort TBH in morning pre-dose trough FEV_1 over 24 weeks and approximately 92% power over Weeks 12 to 24 based on a margin of 50 mL (one-sided, alpha=0.025) assuming no true difference.

Assumptions regarding variability for the primary endpoint are based on Pearl's experience with Phase IIb and III clinical studies. A composite value standard deviation (SD) of 200 mL for the change from baseline at each visit has been assumed for trough FEV_1 and 220 mL for FEV_1 AUC_{0-4}. Dropout is anticipated to be approximately 12% by the end of the study. Based on the repeated measures (RM) analysis, an effective SD for the change over 24 weeks of 157 mL and 173 mL for trough FEV_1 and FEV_1 AUC_{0-4}, respectively, is assumed. For Weeks 12 to 24, an effective SD for trough FEV_1 of 171 mL is assumed.

For Study PT010007, the sample size of 324 (108 per arm in the BGF MDI and GFF MDI groups and 54 per arm in the BFF MDI and Symbicort TBH groups) includes all Japanese subjects who were enrolled in Study PT010006. The sample size was not calculated to achieve statistical power but was selected to provide approximately 100 completing subjects in the BGF MDI and GFF MDI arms.

4. DATA AND ANALYTICAL QUALITY ASSURANCE

The overall quality assurance procedures for the study data, statistical programming and analyses are described in Standard Operating Procedures (SOPs) of Everest Clinical Research. Detailed data management procedures are documented in the study Data Management Plan, Data Validation Check Specifications, and Integrated Safety Data Review Plan. Detailed statistical and programming quality control and quality assurance procedures are documented in the Statistical Analysis and Programming QC/QA Plan.

Transfer of PFT data from the central PFT laboratory (iCardiac PFT Global) to Everest Clinical Research will be defined in the iCardiac DMP (Data Management Plan), and data handling rules related to this data are included in Appendix 1 of this SAP. The quality of all PFTs obtained at each time point will be graded independently at iCardiac by qualified personnel. Quality grading assessments will be based on ATS (American Thoracic Society)/ERS criteria and will be included in data transfers.

5. ANALYSIS POPULATIONS

5.1 Population Definitions

5.1.1 Intent-to-Treat (ITT) Population (PT010006 only)

The **ITT Population** is defined as all subjects who are randomized to treatment and receive any amount of the study treatment in Study PT010006. Subjects will be analyzed according to randomized treatment group. Data obtained after discontinuation of treatment, but prior to withdrawal from the study, will be included. The ITT population will be used for sensitivity analyses.

5.1.2 Modified Intent-to-Treat (mITT) Population (PT010006 only)

The **mITT Population** is a subset of the ITT Population, defined as all subjects with post-randomization data obtained prior to discontinuation from treatment in Study PT010006. Any data collected after completion of or discontinuation from randomized study medication will be excluded. Subjects will be analyzed according to randomized treatment group. (Note that a subject who used a study treatment, but took less than one full dose of treatment will qualify for this population). The mITT Population will be the primary population for all efficacy analyses except for the non-inferiority analyses. Note: The knowledge that a subject did not have a COPD exacerbation constitutes an efficacy assessment.

5.1.3 Japanese Modified Intent-to-Treat (mITT) Population

The **Japanese mITT Population** is defined as the subgroup of the mITT Population who are Japanese subjects (enrolled at sites in Japan) from Study PT010006, regardless of participation in Study PT010007. Subjects will be analyzed according to the active treatment they were assigned to at randomization in Study PT010006. Data from both Study PT010006 and Study PT010007 will be included.

5.1.4 Rescue Ventolin User Population (RVU) (PT010006 Only)

Regional differences in rescue Ventolin HFA usage are expected with subjects in some countries using virtually no rescue medication at study entry. Therefore, the **RVU Population** is defined as all subjects in the ITT Population with mean baseline Rescue Ventolin use of ≥ 1.0 puff/day.

5.1.5 Per-Protocol (PP) Population (PT010006 only)

The **PP Population** is a subset of the ITT Population, defined as all subjects with post-randomization data obtained prior to any major protocol deviations in Study PT010006. Data obtained after any major protocol deviation or discontinuation from treatment will be excluded. Since receiving the wrong treatment is a major protocol deviation, data after the deviation from such subjects will be excluded from the PP Population. If the first treatment received is the wrong treatment then the subject will be excluded entirely from the PP population. Any evaluability criteria with a potential impact on efficacy results will be identified during Blinded Data Review Meeting (BDRM) prior to database lock. Major protocol deviations, therefore, can result in exclusion of all data from a particular subject from the PP Population or require

exclusion of data from a specific time point and/or subsequent time points for an endpoint. The PP Population will be the main population for all non-inferiority analyses.

Protocol deviations and criteria for exclusion from the PP Population will be established at a BDRM prior to database lock. Reasons for exclusion from the PP Population will include, but are not limited to, the following:

1. An incorrect diagnosis of COPD.
2. Subjects who do not have an established clinical history of COPD and severity where an established clinical history of COPD and severity is to be identified at the BDRM.
3. For those subjects who use rescue Ventolin HFA less than 6 hours before study visit, all data post-Ventolin HFA administration will be considered missing for that day.
4. For subjects who take any protocol-prohibited medication that would affect spirometry assessments on the date of an assessment, spirometry measurements taken at that assessment will be excluded.
5. For spirometry endpoints, the subject will not be eligible for the endpoint-specific, visit-specific PP Population if the subject did not take study medication in the evening prior to the visit day.
6. Subjects are also excluded from the PP Population for PFT endpoints if the following condition is true:
   - Subjects who cannot meet protocol-specified baseline stability criteria. FEV_1 baseline stability is defined as the mean of the -60 minute and -30 minute pre-dose FEV_1 assessments at Visit 4 being within ±20% or 200 mL of the mean of the pre-bronchodilator FEV_1 assessments obtained at the 2 preceding visits (average of pre-dose FEV_1 assessments obtained at Visit 2 and Visit 3).

Subjects who fail to meet any of the three restriction criteria prior to spirometry will be handled as follows. The 3 restrictions are (1) subject was not to smoke for at least 4 hours prior to study visit and throughout the duration of each study visit, (2) subject was not to use xanthine-containing products (i.e., coffee, tea, cola and chocolate) for at least 6 hours prior to study visit and for the duration of each study visit, and (3) subject was not to have COPD bronchodilator medications for at least 6 hours prior to study visit. Spirometry data for these subjects at the affected visits will be removed from the PP Population for PFT endpoints. Such restrictions will be applied only if data pertaining to the meeting of the criteria (including the timing) were collected.

5.1.6 Safety Population (PT010006 only)

The **Safety Population** is defined as all subjects who are randomized to treatment and receive at least one dose of the study treatment in Study PT010006. However, subjects will be analyzed according to treatment received rather than randomized. If a subject received more than one randomized treatment, they will be analyzed and included in summaries according to the treatment they received the most. Subjects receiving no study treatment will be excluded, as will

subjects who have no post-dose safety assessments. Note: The statement that a subject had no AEs also constitutes a safety assessment.

5.1.7 Japanese Safety Population

The **Japanese Safety Population** is a subset of the Safety Population that is defined as all Japanese subjects (enrolled at sites in Japan) who received any amount of study medication in Study PT010006, regardless of participation in Study PT010007. Subjects will be analyzed according to the actual treatment they received. Data from both Study PT010006 and Study PT010007 will be included.

5.1.8 PT010007 Safety Population

The **PT010007 Safety Population** is defined as all subjects who received any amount of study medication in Study PT010007. Subjects will be analyzed according to the actual treatment they received. Data from both Study PT010006 and Study PT010007 will be included.

5.1.9 PK Population

The **PK Population** is defined as all randomized and treated subjects who have sufficient data to reliably calculate at least one PK parameter in Study PT010006. Subjects will be analyzed according to treatment received rather than randomized.

The PK Population will be determined after review of the clinical study data (e.g., concomitant medications, dosing information from the subject diary, and adverse events). Prior to the final PK analysis, subject data as well as protocol deviations will be reviewed in a blinded manner by Everest and Pearl at the BDRM for inclusion/exclusion into the PK Population.

5.1.10 HPA Axis Population

The HPA Axis Population is defined as all subjects who participated in the HPA axis sub-study. The HPA Axis Population is a subset of the Safety Population without protocol deviations which could affect SC endpoint and whose serum samples did not have confounding factors that would affect the interpretation of the results. Exclusion from the HPA Axis Population will be established at a blinded data review meeting prior to database lock. Subjects will be analyzed according to treatment received rather than randomized.

5.2 Populations for Primary and Sensitivity Analyses

PT010006:

Demographics will be summarized for the mITT, PP, Safety, and Non-randomized Populations as well as for subjects participating in the 12-hr PFT, PK, and HPA Axis sub-studies.

Extent of exposure will be summarized for the Safety Population. The Safety Population will be used to summarize safety.

Efficacy Analyses will be performed for the ITT, mITT, and PP Populations. The mITT Population will be used for the primary efficacy analyses, with the ITT and PP populations being considered supportive with one exception. The PP Population will be primary for the non-inferiority comparisons of BFF MDI vs. Symbicort TBH.

PK will be summarized for the PK Population. HPA Axis results will be summarized for the HPA Axis Population.

Selected analyses will be performed for specified sub-populations. See the Subgroup Analyses section (Section 6.4.9).

PT010007:

When the final database for PT010007 becomes available, selected safety and efficacy results will be generated using the Japanese Safety and Japanese mITT populations respectively. Selected AE summaries will be repeated on the PT010007 Safety Population.

6. STATISTICAL ANALYSIS

Study PT010006 analyses will be performed when the PT010006 final database is available, with PT010007 analyses to be performed later once the PT010007 final database is available.

All data collected contributing to the analysis will be provided in listings. Data for all subjects who are randomized will be included in the subject data listings. Data for non-randomized subjects will be listed where available.

All safety and efficacy parameters will be summarized by treatment unless specified otherwise.

Continuous variables will be summarized with descriptive statistics (the number of non-missing values, mean, standard deviation, median, minimum, and maximum). Additionally, the 25^{th} and 75^{th} percentiles will be presented when appropriate based on historical knowledge of the normality or non-normality of the distribution of underlying data.

Categorical variables will be summarized with frequency counts and percentages (where appropriate).

6.1 Data Handling Rules and Definitions, Including Handling of Missing Data

Missing data will be maintained as missing in the analysis datasets, unless specified otherwise. For variables where missing data are imputed, the analysis dataset will contain a new variable with the imputed value and the original variable value will be maintained as missing.

Data Imputation for Adverse Events Summaries by Severity and Relationship to Study Drug

For the AE summaries by severity (mild, moderate, or severe), an AE with missing severity will be deemed as severe. For the AE summaries by relationship to study drug, an AE with a missing relationship to study drug will be deemed as definitely related. Imputed values will not be listed in data listings.

Data Imputation for Laboratory, Vital Sign, and ECG Summaries (Continuous Parameters)

Data from unscheduled visits will not be used for by-visit summaries. Data from both scheduled and unscheduled visits will be used for the end-of-treatment summary, for shift tables and for determining incidence of clinically significant values.

Data Imputation (All Laboratory Summaries)

Laboratory values of '>=x' or '<=x' will be taken as the value of x in the analyses. If a laboratory value is prefixed with'>': the available original value +0.001 will be used for table summaries; if a laboratory value is prefixed with '<', then the original value –0.001 will be used in table summaries.

Study Dates and Day of Assessment or Event

Study Day and Day of Assessment or Event definitions are provided in Appendix 1, Data Handling Rules.

On-treatment COPD exacerbations

An exacerbation will be considered "on-treatment" if its start date is before or on the last treatment date. For treatment discontinuations, this definition is extended to include exacerbations starting one day after the last treatment date. (If it is decided during a clinic visit to discontinue study drug and to switch to a treatment for the ongoing exacerbation symptoms, the subject typically would not take the morning dose of study drug at that visit, and their exacerbation start date will be one day after the last treatment date. Such exacerbations will still be considered "on-treatment").

6.2 Subject Disposition and Analysis Populations

A disposition table for PT010006 for all subjects randomized will be provided *(Table 1.1.1.1)*. This tabulation will include the number of subjects in each randomized treatment who were not treated, who received the study treatment, who discontinued treatment prematurely, who discontinued treatment prematurely but completed the study, who discontinued treatment prematurely and withdrew from the study, who withdrew from the study prematurely, and who completed the study. The number and percentage of randomized subjects included in the mITT, Japanese mITT, ITT, PP, Safety, Japanese Safety, 12-hr PFT Sub-study, PK, and HPA Axis Populations will also be tabulated *(Table 1.1.1.1)*. A disposition table for subjects in PT010006 and PT010007 will also be provided (*Table 1.1.1.2*). Informed consent is listed in *Listing 9.7*.

The numbers of subjects randomized and in the analysis populations will be provided by country, center, and treatment in *Table 1.1.2*. The number of subjects randomized by stratification factor and cross-classification of reversibility to Ventolin HFA and disease severity, using interactive web response system (IWRS) data and Clinical data for the stratification-factor levels, will be tabulated in *Tables 1.1.5.1, 1.1.5.2, 1.1.6.1,* and *1.1.6.2*, respectively. If there are any subjects who took study treatment other than what was randomized during the study, both the treatment assigned at randomization and actual treatment(s) received during the Treatment Period will be listed *(Listing 1.3)*. The duration of actual treatment will also be listed (*Listing 1.3)*. A list of the discrepancies between IWRS-based and clinical-data-based stratification factors will be provided (Listing 1.6).

A summary of reasons subjects were not randomized will be provided for all subjects not randomized (*Table 1.1.3*). A listing of reasons subjects were not randomized will also be provided *(Listing 1.4)*. Subjects excluded from the ITT, mITT, PP, Safety, PK and HPA Axis analysis populations will be summarized (*Table 1.1.4)* for all subjects randomized. Reasons for premature discontinuation from study treatment will be summarized for the Safety Population and for the Japanese Safety Population *(Table 1.2.1* and *Table 1.2.2* respectively).

The number and percentage of subjects in the ITT Population who withdrew from the study will be tabulated by reason for withdrawal *(Table 1.2.3)*.

The reason for exclusion from the PP Population will be tabulated by study treatment for all mITT subjects *(Table 1.3.1)*. The reason for exclusion of a subject from the ITT, mITT, PP, Safety, PK and HPA Axis Populations or exclusion of partial data (at some but not all time points) for a subject will be listed for all randomized subjects *(Table 1.1.4)*. A listing of subjects who did not comply with restrictions on smoking, use of rescue medication, and xanthine-containing products (protocol deviations requiring removal of data from the PP Population analysis) just prior to spirometry will be provided in *Listing 6.1.1*. Use of rescue medication at pre-dose or during the post-dose assessments on each specific test day (yes/no), will be tabulated in *Listing 6.1.3*. In addition, the eligibility information (inclusion/exclusion criteria with any waivers granted) of all subjects who are randomized will be listed *(Listing 2.1).*

The number and percentage of subjects with changes in smoking status after the start of study treatment will be tabulated by randomized treatment, by visit and overall during the study, in *Table 1.13* (Safety Population) and *Table 1.14* (Japanese Safety Population) *(Listing 1.5)*.

6.3 Demographic and Baseline Characteristics and Extent of Exposure

The definitions for the derived demographic or baseline characteristic variables can be found in Appendix 1.

6.3.1 Demography, Physical Characteristics, CAT

Subject demographics, total CAT score, use of inhaled corticosteroids (ICS) at screening, and smoking status/history will be summarized for the mITT, PP, and Safety Populations and for Non-Randomized subjects *(Tables 1.4.1.1* through *1.4.5.1,* respectively, and *Listing 1.2)*. The

ITT population does not need to be tabulated because it is the same as the mITT population for demographics. If the Safety Population has the same treatment assignment as the mITT, then it will be identical as well and hence not produced. Use of ICS (yes/no) will be summarized for all populations except for the Non-Randomized subjects. Demographics will also be summarized for the subjects in the 12-hr PFT sub-study *(Tables 1.4.6.1* through *1.4.6.4)*, HPA Axis Population (*Tables 1.4.6.5* through *1.4.6.8*), and PK Population (*Tables 1.4.6.9* through *1.4.6.12*). Demographics and baseline characteristics will also be summarized for subjects in the Japanese mITT, Japanese Safety, and PT010007 Safety Populations (*Tables 1.4.7.1, 1.4.7.2, 1.4.7.3).*

Demographic and baseline characteristic variables summarized will include the following:

- Age
- Age Group
- Gender
- Race
- Ethnicity (Hispanic or Non-Hispanic)
- The CAT total score and total score category (<10, ≥10, <15, ≥15, <20, ≥20, Missing)
- Used ICSs at Screening (all populations except for Non-Randomized subjects)
- Baseline eosinophil count (<150 cells per mm^3 vs. ≥150 cells per mm^3)
- Baseline exacerbation history (0, 1, ≥2)
- Smoking status (current vs. former smoker)
- Number of years smoked
- Average number of cigarettes smoked per day
- Number of pack years smoked, calculated as (number of cigarettes per day/20) x number of years smoked
- Weight
- Height
- Body mass index (BMI)

Screening and pre-treatment CAT data will be listed *(Listing 4.2)*.

6.3.2 COPD History, Screening/Baseline Spirometry, and Reversibility

Duration of COPD and the number of years prior to the start of study medication that COPD was first diagnosed (calculated as [Date of First Dose of Study treatment in the study – Date COPD First Diagnosed] /365.25) will be summarized by treatment and for all subjects for the mITT, Safety, Japanese mITT, and Japanese Safety Populations and listed *(Tables 1.5.1, 1.5.4, 1.5.5, and 1.5.8* and *Listing 4.1)*. Severity of COPD at Screening Visit 2 post-Ventolin HFA will also be included in these summaries. History of moderate or severe COPD exacerbations within the past 12 months will be summarized and listed for subjects in the Safety, Japanese Safety, PT010007, mITT and Japanese mITT Populations *(Table 1.9.4, Table 1.9.5, Table 1.9.6, Table*

1.9.7 and *Listing 4.3)*. Severity of COPD is defined in terms of baseline FEV_1, as found in the Data Handling Rules (Appendix 1).

Descriptive statistics will be provided for screening period pre-bronchodilator and post-bronchodilator and baseline spirometry parameters *(Tables 1.6.1, 1.6.2, 1.6.3, 1.6.4,* and *1.6.5,* for the mITT, ITT, PP, Japanese mITT Populations, and 12-hr PFT Sub-study, respectively, and *Listings 2.2 and 2.3*).

Characterization of Reversibility:

Reversibility to Ventolin HFA (short-acting β_2-agonist, SABA) will be evaluated at Visit 2. Reversibility to Atrovent HFA (short-acting anticholinergic) will be evaluated at Visit 3. Reversibility to Ventolin HFA (obtained at Visit 2) will be used as a stratification variable at randomization to ensure an even distribution of reversibility across the treatment arms. Reversibility to Atrovent HFA will be used to characterize the population.

Reversibility (%) to a bronchodilator is defined as 100 x (the change from pre-bronchodilator HFA to post-bronchodilator HFA for FEV_1)/pre-bronchodilator HFA FEV_1. Reversible (Yes/No) is defined as improvement in FEV_1 post-bronchodilator HFA administration compared to pre-bronchodilator HFA of >=12% and >=200mL.

Reversibility to Ventolin HFA at Screening Visit 2 and reversibility to Atrovent HFA at Screening Visit 3 will be summarized for the mITT, 12-hr PFT Sub-study, and Japanese mITT Populations and listed *(Tables 1.7.1 to 1.7.5,* and *Tables 1.8.1* to *1.8.5* for Ventolin HFA and Atrovent HFA reversibility, respectively, and *Listings 2.2* and *2.3* for Ventolin HFA reversibility and Atrovent HFA reversibility respectively, and *Listing 5.2* for Atrovent HFA and Ventolin HFA dispensing). The number and percentage of subjects reversible will be included in these summaries. A summary of the change in FEV_1 from pre-dose FEV_1 to post-bronchodilator assessment will also be included. If multiple time points are available post-bronchodilator, then the one with the highest FEV_1 will be used.

Additionally, the number and percentage of subjects meeting each of the following response criteria will be summarized for both Ventolin HFA and Atrovent bronchodilators:

- ≥12% improvement post-bronchodilator in FEV_1 from pre-bronchodilator
- ≥150 mL improvement post-bronchodilator in FEV_1 from pre-bronchodilator
- ≥200 mL improvement post-bronchodilator in FEV_1 from pre-bronchodilator

6.3.3 Medical and Surgical History at Screening, Reproductive Status and Pregnancy Testing

Medical and Surgical History at Screening will be summarized for the Safety Population and Japanese Safety Population and listed for all randomized subjects *(Table 1.9.1.1, Table 1.9.1.2,* and *Listing 4.4)*. Cardiovascular medical history of interest at Screening will be summarized for

the Safety Population and Japanese Safety Population and listed for all randomized subjects *(Table 1.9*.2.1 and Table 1.9.2.2 and *Listing 4.5)*.

Screening Reproductive Status and Pregnancy Testing Results will be listed *(Listing 4.6)*.

Subjects' history of moderate or severe COPD exacerbations within the past 12 months will be summarized for the Safety, Japanese Safety, PT010007 Safety, mITT, and Japanese mITT Populations (*Tables 1.9.3* through *1.9.7*).

6.3.4 Prior, Concomitant, and Post-Treatment Medications/Treatments

All prescription and OTC medications taken by the subject during 30 days before Screening will be recorded on the Concomitant Medications case report form (CRF) page.

Coding: Verbatim medication/treatment terms will be coded by Everest Clinical Research and will be assigned a preferred term (PT) and an ATC (anatomic therapeutic class) term using the latest version of the World Health Organization Drug Dictionary (WHO-DD) available (version: 3Q2016 or later).

Multiple ATC assignments: If there are multiple ATC codes assigned to the same concomitant medication, the "primary" one based on a Pearl medical evaluation will be used. All prior medication taken by the subject within 30 days of Screening for the study and all concomitant therapy taken by the subject while on study will be recorded in the eCRF.

Prior medication/treatment is any medication/treatment taken prior to study treatment, even if this medication continued to be taken on the day of the start of study treatment in the study or afterward *(Appendix 1)*.

Concomitant medication/treatment is any medication/treatment reported as being taken after the start of the randomized study treatment in the study and being taken on or before the date prior to the last dose of study treatment for the subject. A medication with an onset date on or after the date of discontinuation from or completion of randomized study treatment for the subject will not be considered concomitant, but will be considered a **Post-Treatment medication/treatment**.

Any medication/treatment which cannot be identified as Prior, Concomitant, or Post-Treatment will be considered as being in each of the categories that are possible from the available information.

Concomitant COPD-related, COPD exacerbation-related, and Non-COPD related medications/treatments will be summarized by preferred term and actual treatment received for the Safety Population *(Tables 1.11.1, 1.11.1b, 1.11.2, 1.11.5.1 to 1.11.8)* and Japanese Safety Population *(Tables 1.11.3, 1.11.4)*. COPD-related summaries will not include the COPD-exacerbation medications. Prior, concomitant/post-treatment COPD, COPD exacerbation, and Non-COPD medications will be displayed in separate listings *(Listings 4.7* and *4.8*, respectively).

Reported prior medications for COPD and non-COPD-related medications will be tabulated for the Safety Population *(Tables 1.10.1.1* and *1.10.2)* and the Japanese Safety Population *(Tables 1.10.3* and *1.10.4)* and listed separately *(Listings 4.7* and *4.8*, respectively).

Prior COPD Medications will be tabulated (for the Safety population) for subjects having received any one, two, all three, or none of the following treatments whether in fixed combination products or separate are used at the time of screening and separately for post-treatment (for any duration): (1) a muscarinic antagonist, (2) a β2 agonist, and (3) an ICS *(Table 1.10.1.2)*. For this purpose, scheduled SAMA (Short-acting muscarinic antagonist) or SABA treatments are included. In addition, tabulations for long-acting muscarinic antagonists (LAMA) and long-acting β2 agonists (LABA) will also be included.

Post-treatment medications will be tabulated for subjects having received any one, two, all three, or none of the following treatments: (1) a muscarinic antagonist, (2) a β2 agonist, and (3) an ICS *(Table 1.11.5.2)*.

6.3.5 Extent of Exposure to Study Medication and Compliance

Subject's exposure to a study treatment will be determined by the duration of time (days) for which the doses were administered, defined as "([End date of treatment – Date of first dose of treatment] + 1)". Percent compliance is defined as (total number of puffs of study treatment taken on a study day/total expected puffs taken on a study day) averaged across all days of a subject's dosing between start of study treatment and last day on study treatment x 100. The expected number of puffs for a test day which is the last date of treatment will be 2, and the expected number of puffs for the last date of treatment which is not a test day will be 4 when a PM dose is taken but will be 2 otherwise; the expected number of puffs on dates prior to the last date of treatment will be 4.

The number of days of exposure to study treatment will be summarized for each treatment for the Safety Population and Japanese Safety Population. The total person-years of exposure for a treatment group, defined as the total exposure in the study across all subjects in the treatment, will also be provided by treatment *(Table 1.12.1* for the Safety Population, *Table 1.12.2* for the Japanese Safety Population, *Table 1.12.3* for PT010007 Safety Population). In addition, treatment compliance will be provided in this summary. The treatment compliance will be categorized into 7 different groups depending on the degree of compliance: 0 – <20%, ≥20 – <40%, ≥40 – <60%, ≥60 – <80%, ≥80 – ≤100%, >100 – ≤120%, and >120%. Descriptive statistics (n, mean, standard deviation, median, minimum and maximum) for percent compliance will also be provided by treatment. Treatment compliance will be reported in *Listing 5.3*. A listing of treatment dosing and dispensing information will be provided in *Listing 5.1.* Any comments related to study medication or any other additional study comments will be listed *(Listing 9.6)*.

6.4 Efficacy Analyses

For study PT010006, there are four pairwise comparisons of treatments of interest, namely,

- BGF MDI vs. BFF MDI,
- BGF MDI vs. GFF MDI,
- BGF MDI vs. Symbicort TBH, and
- BFF MDI vs. Symbicort TBH.

However, summary statistics will be provided by randomized treatment and for each treatment difference for all comparisons. All comparisons will be performed for testing superiority except that the comparison of BFF MDI to Symbicort TBH will be for non-inferiority. Superiority comparisons will chiefly use the mITT population. Non-inferiority analyses will chiefly use the PP population, unless specifically stated otherwise. In order for non-inferiority to be met, the confidence bound (on the side of lower efficacy for BFF) from a 95% two-sided confidence interval (for the treatment difference or treatment ratio [as applicable]) must be less extreme than the non-inferiority margin.

6.4.1 Estimands

The primary estimand of interest is called the efficacy estimand and is the effect of the randomized treatments in all subjects assuming continuation of randomized treatments for the duration of the study regardless of actual compliance. There are three additional estimands of interest. One is called the attributable estimand and is the effect of treatment in subjects attributable to the randomized treatment. For this estimand, discontinuation of randomized medication for reasons such as tolerability or lack of efficacy are considered unfavorable outcomes. Another estimand of interest is called the treatment policy estimand. This estimand is the effect of randomized treatment over the study period regardless of whether randomized treatment is continued. The final estimand of interest is called the per protocol estimand. This estimand is the effect of treatment on subjects who are compliant with the protocol (i.e. no major protocol deviations), including the use of randomized medication.

The primary analysis for the efficacy estimand will be conducted using the mITT Population where only data obtained prior to subjects discontinuing from randomized treatment will be utilized. This assumes that efficacy observed on treatment is reflective of what would have occurred after discontinuation of randomized treatment had they remained on treatment.

The second estimand of interest is the attributable estimand. Analyses of the attributable estimand will be conducted in the mITT Population. Data that are missing due to treatment discontinuation will be imputed based on the 5^{th} percentile of the reference arms' distribution if the reason is reasonably attributable to tolerability or lack of efficacy. The 5^{th} percentile applies to an endpoint for which a higher value is a better outcome; however the 95^{th} percentile applies to an endpoint for which a lower value is a better outcome. Other missing data are to be imputed using the observed data model, i.e. assumed to be missing at random (MAR). The number of imputations used for the derivation of the attributable estimand will be between 100 and 1000.

More detail about the computation of the attributable estimand will be provided in subsequent sections (especially 6.4.4.1) and in the Details Appendix to this SAP.

Treatment discontinuations reasonably attributable to tolerability or lack of efficacy will be identified during the BDRM and documented in the minutes prior to unblinding. Discontinuations will be attributed to tolerability if the subject had an adverse event determined by the investigator to be possibly, probably, or definitely related to study drug, and for which study drug was permanently discontinued. Discontinuations will be attributed to lack of efficacy if 'lack of efficacy' is indicated to be the primary reason for discontinuation from study drug. For the remaining discontinuation categories, where specific reasons or criteria frequently need to be considered, decisions will be made and documented at the BDRM. Once these subjects are identified, post-treatment discontinuation FEV_1 values for each patient will be imputed based on the 5^{th} percentile of the reference arms' distribution.

The third estimand of interest is the treatment policy estimand. Analyses of the treatment policy estimand will be conducted in the ITT Population, in which all observed data will be utilized regardless of whether subjects remain on randomized treatment.

Finally, the last estimand of interest is the per protocol estimand. Analysis of this estimand will use the PP Population.

6.4.2 Baselines and Baseline Covariates for Analysis

The mean of all evaluable 60- and 30-minute pre-dose spirometry assessments conducted at Day 1 (Visit 4) will be used to establish baseline for all FEV_1, FVC, $FEF_{25-75,}$ and PEFR parameters.

For the diary symptom score parameters and rescue medication usage, baseline will be the mean of the non-missing values from the diary data collected in the last seven days of the Screening Period.

For the SGRQ scores, baseline will be the value of the score calculated using the Day 1 questionnaire data collected prior to the start of randomized study treatment.

Baseline COPD exacerbation history is set to 0, 1, or ≥ 2 moderate or severe exacerbations in the last 12 months (from the Visit 1 CRF page).

Baseline percent predicted FEV_1 is the mean of the 30 minute and 60 minute values of FEV_1 prior to dosing on Day 1 (Visit 4).

ICS use at screening (Yes or No) is to be defined as follows. A subject was considered to have had "ICS Use at Screening" if:

- the subject was taking a medication that contained a component (active ingredient) that is in the WHODRUG SDG (standardized drug grouping) of "CORTICOSTEROIDS", and

- the route of administration was “INHALED”, and the medication was used at any time during the screening period (or in the 30 days prior to the screening period). Medications may be but are not necessarily ongoing medications.

Baseline blood eosinophil count is the mean of non-missing blood eosinophil count values prior to the first dose of study medication.

Baseline age is the age in years at the time of Informed Consent.

Baseline post-bronchodilator FEV_1 is the highest value of FEV_1 obtained approximately 30 minutes after dosing with Ventolin (or later if there are repeated assessments) at Visit 2 unless it is missing, in which case the post-Atrovent value at Visit 3 will be used.

The baseline post-bronchodilator percent predicted FEV1 is the baseline post-bronchodilator FEV_1 divided by the predicted FEV_1 and multiplied by 100.

Baseline percent reversibility to Ventolin is 100 x (POST-PRE)/PRE, where POST is the highest value of FEV_1 obtained 30 minutes (or later if there are repeated assessments) after dosing with Ventolin at Visit 2 and PRE is the mean of the 30 minute and 60 minute values of FEV_1 prior to dosing with Ventolin at Visit 2.

6.4.3 Visits and Time Windows for Visit-Based Efficacy Assessments

Efficacy data obtained during unscheduled visits will not be used for any of the pre-defined efficacy analyses. Efficacy from scheduled and unscheduled visits will be listed.

For efficacy analysis or derivation of AUC based on time points, the change from baseline in PFT assessments will be allocated to derived nominal collection time windows using the time intervals specified for each below.

Table 1 Analysis Study Time Window for Spirometry Assessments

Calculated Study Time Window	Time Interval for the Study Time Window
Pre-dose 60 min.	≥45 minutes prior to dose
Pre-dose 30 min.	≥0 to <45 minutes prior to dose
Post-dose 5 min.	>0 to 9 min. post-dose
Post-dose 15 min.	10 to 22 min. post-dose
Post-dose 30 min.	23 to 44 min. post-dose
Post-dose 1 hr.	45 to 89 min. post-dose
Post-dose 2 hrs.	90 to 179 min. post-dose
Post-dose 4 hrs.	3 to <5 hrs. post-dose
Post-dose 6 hrs.	5 to <7.5 hrs. post-dose
Post-dose 8 hrs.	7.5 to <9 hrs. post-dose
Post-dose 10 hrs.	9 to <10.75 hrs. post-dose
Post-dose 11.5 hrs.	10.75 to <11.75 hrs. post-dose
Post-dose 12 hrs.	11.75 to < 14 hrs. post-dose, but must be prior to any subsequent dose of study medication or maintenance medication.

Note: The minutes are rounded to the nearest whole number before applying time windows.

If there are multiple spirometry values for the same parameter within the same post-baseline study time window on the same day, the last value will be chosen for analysis.

Analyses for the efficacy endpoints are presented in sections 6.4.4 to 6.4.7.

6.4.4 Primary Efficacy Analyses

Analyses for the primary endpoint are presented in this section along with analyses for any of the secondary or other efficacy endpoints related to the primary endpoint. Calculation of FEV_1 AUC_{0-4} for the efficacy, treatment policy, and per protocol estimands will require at least one non-missing post-dose value.

6.4.4.1 Change from Baseline in Morning Pre-Dose Trough FEV_1

PT010006:

Change from baseline in morning pre-dose trough FEV_1 at each visit is defined as the mean of the 60 and 30 minute pre-dose values minus baseline. In subjects missing either of these pre-dose assessments, the value will be calculated from the single measurement. In subjects missing both pre-dose values, morning pre-dose trough FEV_1 at that visit will not be calculated.

The change from baseline in morning pre-dose trough FEV_1 will be analyzed using an RM linear mixed model. The model will include treatment, visit, treatment by visit interaction, and ICS use at Screening as categorical covariates and baseline FEV_1, baseline eosinophil count, and percent reversibility to Ventolin HFA as continuous covariates. Baseline FEV_1 is defined as the mean of the non-missing -60 minute and -30 minute values obtained prior to dosing at Visit 4, and baseline eosinophil count is defined as the mean of non-missing eosinophil counts prior to the first dose of study treatment. An unstructured correlation (UN) matrix will be used to model correlation within a subject. If the UN model fails to converge, then a first-order autoregressive (AR(1)) structure will be used instead. In the AR(1) model, subject will be included as a random effect.

Contrasts will be used to obtain estimates of the treatment differences over Weeks 12 to 24 (primary for Japan/China), over 24 weeks (i.e. Weeks 4 to 24) (primary for EU and Canada; secondary for Japan/China; secondary for US), at Week 24 (primary for US), and at each post-randomization visit ("other" endpoint). Subject data from the first 24 weeks post-randomization will be included in the analyses. Two-sided p-values and point estimates with two-sided 95% confidence intervals (CIs) will be produced for each treatment difference.

All comparisons will be for superiority except that the comparison of BFF MDI to Symbicort TBH will be for non-inferiority and will use a margin of -50 mL for the lower bound of a 2-sided 95% CI for the treatment difference. The primary analysis will be conducted using the efficacy estimand except for the non-inferiority comparisons of BFF MDI vs. Symbicort TBH, for which the per protocol estimand will be primary treatment *(Tables* and *Figures 2.1.1, 2.1.3, 2.1.5,* and *Figures 2.1.6.1,* to *2.1.6.10* for the efficacy estimand, the attributable estimand, the treatment policy estimand, and the per protocol estimand, respectively). Efficacy data from unscheduled visits will not be used for this analysis. Additional figures will display the change from baseline in morning pre-dose trough FEV_1 by completion vs. discontinuation patterns for the efficacy estimand, the attributable estimand, and the treatment policy estimand (*Figures 2.1.6.11* to *2.1.6.14*).

The attributable estimand will use the mITT Population but then impute missing post-treatment discontinuation data based on the 5^{th} percentile of the reference arms' (BFF's and GFF's combined) distribution if the reason for discontinuation is attributable to lack of efficacy or/tolerability. All other missing data will be imputed using the observed data model. The variance used for the multiple imputation is described in the Details Appendix to this SAP (Appendix 6). The number of imputations used for the derivation of the attributable estimand will be between 100 and 1000. Work by Seaman, White and Leacy (2014) and Cro (2017) show that Rubin's rules can be validly used in conjunction with so called control-based multiple imputation methods, of which our attributable analysis is one type. Given these results we believe the attributable analysis to be conservative from a Type I error control perspective.

Sensitivity analyses will be conducted to assess robustness of the analyses to missing data (See Section 6.4.4.4).

PT010007:

The change from baseline in morning pre-dose trough FEV_1 over 52 weeks (Weeks 4 to 52) and at each post-baseline visit will be analyzed for the efficacy estimand with the Japanese mITT Population using an RM linear mixed model. The model will include treatment, visit, and treatment by visit interaction, and ICS use at screening as categorical covariates and baseline FEV_1, baseline eosinophil count, and percent reversibility to Ventolin HFA as continuous covariates. Baseline is defined as the mean of the non-missing -60 minute and -30 minute values obtained prior to dosing at Visit 4. An unstructured correlation matrix will be used to model variability across time within each subject. If the UN model fails to converge, then an AR(1) structure will be used instead. In an AR(1) model, subject will be included as a random effect. Two-sided p-values and point estimates with two-sided 95% CIs will be produced for each treatment difference. The analysis will assess the treatment effects over the entire 52 weeks of treatment, i.e. subject data from the first 24 weeks of Study PT010006 will be included in the analyses *(Table* and *Figure 2.1.2* and *Figures 2.1.2.1* and *2.1.2.2* for the Japanese mITT Population).

6.4.4.2 **FEV_1 AUC_{0-4}**

FEV_1 AUC_{0-4} will be analyzed for PT010006.

FEV_1 AUC_{0-4} is the area under the curve for the change from baseline for FEV_1 and is calculated using the trapezoidal rule. The AUC will be normalized by converting it into a weighted average, which will be accomplished through dividing by the time in hours from dosing to the last measurement included (typically 4 hours). Only one non-missing post-dose value is required for the calculation of AUC. Actual time from dosing will be used if available; otherwise scheduled time will be used.

The differences between treatment groups in FEV_1 AUC_{0-4} over 24 weeks (i.e. Weeks 4 to 24) (primary for EU/ Canada), over Weeks 12 to 24 (secondary for Japan/China), at Week 24 (primary for US), and at Day 1 and each post-randomization visit ("other" endpoint), will be evaluated using an RM linear mixed model with baseline FEV_1, percent reversibility to Ventolin HFA, and baseline eosinophil count as continuous covariates and treatment, visit, treatment by visit interaction, and ICS use at Screening as categorical covariates. Two-sided p-values and point estimates with two-sided 95% CIs will be produced for each treatment difference of interest (*Tables and Figures 2.1.8 to 2.1.11,* and *Figures 2.1.12.1*to *2.1.12.8* for the efficacy, attributable, treatment policy, and per protocol estimands, respectively). Additional figures will display FEV_1 AUC_{0-4} by completion vs. discontinuation patterns for the efficacy estimand, the attributable estimand, and the treatment policy estimand (*Figures 2.1.12.9* to *2.1.12.12*).

All comparisons will be for superiority except that the comparison of BFF MDI to Symbicort TBH will be for non-inferiority and will use a margin of -75 mL for the lower bound of the 95% CI for the treatment difference. The primary analysis will be conducted using the efficacy estimand except for the non-inferiority comparisons of BFF MDI vs. Symbicort TBH, for which the per protocol estimand will be the primary population.

For the attributable estimand of FEV_1 AUC_{0-4} (for the analysis at Week 24, the analysis over Weeks 12-24, and the analysis over 24 weeks), data that are missing due to treatment discontinuation will be imputed in a similar manner as in the attributable estimand for morning pre-dose trough FEV_1 in Section 6.4.4.1 above.

Sensitivity analyses will be conducted to assess robustness of the analyses to missing data (See Section 6.4.4.4).

Study PT010007 did not have post-dose spirometry and therefore the analysis of FEV_1 AUC_{0-4} is not possible.

6.4.4.3 **Assumptions Checks and Removal of Outliers in Sensitivity Analyses**

In general, the distribution of spirometry measures is well-approximated by a normal distribution. Under some circumstances (for example during a COPD exacerbation unrelated to treatment), atypical values can arise. Such values may disproportionately affect model-based estimates of the fixed effect and variance parameters. Prior to database lock and unblinding, the change from baseline values for efficacy endpoints will be examined as part of data quality management. This may include production of normal probability plots, kernel density estimates, and normal order outlier statistics. Based on this blinded evaluation, if atypical values are identified, nonparametric methods or data transformations (e.g. logarithmic or normal rank transformation) will be considered. If erroneous values are detected, every effort will be made to correct them prior to database lock. If these values cannot be corrected, they will be considered for removal from analysis. These analyses will be conducted if warranted to demonstrate the robustness of the primary and secondary results and reported in the Statistical Methods Appendix to the clinical study report (CSR).

The assumption of normality in the change from baseline in the morning trough FEV_1 data will be checked by visually inspecting the distribution of the residuals. Also, model fit and the assumption of homogeneity of variance will be verified by inspection of scatter plots of predicted vs. residuals, residuals vs. treatment, residuals vs. ICS use (yes/no), and by box plots of residuals for model variables with a potential effect on variance (treatment, visit, and ICS use). Plots for scaled (marginal) residuals will be prepared (option=VCIRY on the model statement and ODS graphics option allows the production of plots using these residuals). As a sensitivity analysis, if appropriate, the linear RM model analysis will be conducted by allowing for heterogeneity of variance between treatments, visits (if unstructured correlation matrix fail to converge), and/or ICS use categories (yes/no). Note that the unstructured correlation matrix structure allows for heterogeneity among the visits.

Some further assumptions checks are mentioned in the Details Appendix to this SAP.

6.4.4.4 **Sensitivity Analyses for Missing Data**

Sensitivity analyses will be conducted for FEV_1 and AUC_{0-4} to evaluate the robustness of the primary analysis findings to missing data.

Robustness of results to missing data will be explored using tipping point analyses (Ratitch 2013).

For the attributable estimand for morning pre-dose trough FEV_1 and FEV_1 AUC_{0-4}, data that are missing due to treatment discontinuation will be imputed based on the 5^{th} percentile of the GFF and BFF reference arms' distribution if the reason is reasonably attributable to tolerability or lack of efficacy (see Sections 6.4.4.1 and 6.4.4.2). Other missing data are to be imputed using the observed data model.

Analyses of the treatment policy estimand will be conducted in the ITT Population where all observed data will be utilized regardless of whether subjects remain on randomized treatment. A sensitivity analysis will be conducted where missing data in the treatment arm are imputed with the benefit of treatment decremented by up to 500 mL until the p-value ≥ 0.05.

The following table summarizes the multiple imputation-based sensitivity analyses under the PMM (pattern mixture model) framework that will be undertaken,

Table 2 Sensitivity Analyses for Morning Pre-dose Trough FEV_1 and FEV_1 AUC_{0-4}

Efficacy Estimand		Attributable Estimand	Treatment Policy Estimand
mITT Population		mITT Population	ITT Population
Tipping point analysis #1: All missing data are imputed using the observed data model except that for subjects in the treatment arm (not the comparator arm) values that are considered MNAR are imputed with the change from baseline in the treatment arm decremented by up to 500 mL until the p-value ≥ 0.05.	Tipping point analysis #2: All missing data are imputed using the observed data model except that for subjects in the treatment arm (not the comparator arm) values are imputed with the change from baseline in the treatment arm decremented by up to 500 mL until the p-value ≥ 0.05.	Tipping point analysis: MI based on the 5th percentile of the reference arms' distribution if treatment discontinuation is due to tolerability or lack of efficacy of study drug (as in the primary analysis of this estimand). Otherwise, all missing data are imputed using the observed data model except that for subjects in the treatment arm (not the comparator arm), values are imputed with the change from baseline in the treatment arm decremented by up to 500 mL until the p-value ≥ 0.05.	Tipping point analysis: All missing data are imputed using the observed data model except that for subjects in the treatment arm (not the comparator arm) values are imputed with the change from baseline in the treatment arm decremented by up to 500 mL until the p-value ≥ 0.05.

MNAR = Missing not at random. Missingness determined to be potentially MNAR will be defined and documented in the BDRM minutes prior to unblinding. The tipping point will be shown to at least a precision of 10 mL. Imputed values may not be impossible values – i.e. changes from baseline that would imply a negative FEV1 value. Thus the values will be imputed from a truncated distribution.

The primary analysis is for the efficacy estimand, which includes data collected up until the time of discontinuation of treatment. The efficacy estimand quantifies the difference in outcomes for all patients as if they continued on their initially randomized treatment. The primary analysis uses a linear mixed model and assumes that all missing data are MAR or MCAR (missing completely at random).

Although the analysis for the attributable estimand starts with the same amount of missingness, less remains after imputation for missingness deemed attributable to the treatments is performed. These remaining missing data are imputed using the observed data model in the main analysis under the assumption of MAR. More detail about the computation of the attributable estimand is provided in subsequent sections and in the Details Appendix to this SAP.

Tipping-point analyses will be conducted to examine the impact of varying the treatment mean for missing data in subjects who discontinue BGF MDI. Multiple imputation (MI) techniques

will be used to impute the missing data for these patients by varying the mean in the treatment arm. The change from baseline in the treatment arm will be decremented by up to 500 mL until the p-value for the comparison of treatment to comparator becomes ≥ 0.05. A total of 10 imputations will be used for each set of tipping point analyses. This imputation technique will be applied in sensitivity analyses as described below.

Tipping Point Analyses of the Primary Estimand:

Tipping Point #1: this first set of analyses will impute diminished effects only for subjects on BGF MDI whose missing data are determined to be MNAR.

Tipping Point #2: this analysis will impute diminished effects for all missing data in the BGF MDI arm.

Note that for both tipping point analyses, all other missing data will be imputed using the observed data model.

Tipping Point Analysis of the Attributable Estimand:

For the attributable estimand, by definition, missing data in all arms due to tolerability and lack of efficacy are already imputed based on the 5th percentile of the reference arms' distribution, therefore the remaining missing data imputed using the observed data model in the main analysis are likely MAR or MCAR. Hence, there is no need to conduct a tipping analysis like #1 planned for the efficacy estimand. A tipping point analysis like #2 will be conducted where the non-attributable missing data will be imputed using progressively diminished effects.

Tipping Point Analysis of the Treatment Policy Estimand:

For the treatment policy estimand, a tipping point analysis like #2 will be conducted where missing data in the treatment arm will be imputed using progressively diminished effects.

In all of these analyses, the imputed values that would have been seen are then combined with the observed values to provide a complete dataset. These data are then analyzed using the same linear mixed model used for the primary analysis. This analysis is repeated multiple times and the results are combined using Rubin's formulae [Rubin, 1987].

For the tipping point analyses, tables giving results for each progressively diminished effect will be produced. Figures of delta (decrement in treatment effect) versus p-values will also be produced. Details of the sensitivity analyses are discussed in the Statistical Methods Appendix to the CSR.

Cumulative Responder Analysis

Additional sensitivity analyses will be implemented based on a cumulative responder approach (Farrar et al., 2006) for the change from baseline in morning pre-dose trough FEV_1 over Weeks

12-24, over 24 weeks, and at Week 24 *(Tables 2.1.14.1 to 2.1.14.3* for the efficacy estimand, *Tables 2.1.14.1a* to *2.1.14.3a* for treatment policy estimand*)*, and change from baseline in FEV_1 AUC_{0-4} over 24 weeks and at Week 24 (*Tables 2.1.14.4* and *2.1.14.5* for the efficacy estimand and *Tables 2.1.14.4b* and *2.1.14.5b* for the treatment policy estimand). A cumulative distribution plot by treatment arm will also be produced. The observed change from baseline in morning pre-dose trough FEV_1 over Weeks 12-24 (over 24 weeks and at Week 24) will be plotted on the X axis, while the proportion of responders (subjects that equal or exceed that level of change) will be plotted on the Y axis *(Figures 2.1.14.1* to *2.1.14.3* for the efficacy estimand, *Figures 2.1.14.1a* and *2.1.14.3a* for the treatment policy estimand*),* and a cumulative proportion of responders plot for change from baseline in FEV_1 AUC_{0-4} over 24 weeks and at Week 24 (*Figures 2.1.14.4* and *2.1.14.5* for the efficacy estimand and *Figures 2.1.14.4a* and *2.1.14.5a* for the treatment policy estimand). Subjects without post-baseline data or who discontinue treatment for any reason will be considered non-responders in the analysis. For display purposes only, the range of the X axis will be from -1 to +1 liters [L] by increments of 0.01 liters in order to avoid the undue influence of outlying values. The cumulative responder curves for each treatment will then be compared pairwise using Kolmogorov-Smirnov tests. Cumulative responder analyses for the attributable estimand will not be performed because they are not well defined: methodology to apply Rubin's rules for combining multiply imputed data for such an analysis is not readily available.

6.4.5 Analysis of Secondary Efficacy Variables

The following variables appear as secondary for 1 or more registration approaches: morning pre-dose trough FEV_1 over 24 Weeks, FEV_1 AUC_{0-4} over Weeks 12-24, peak FEV_1, TDI focal score (at Week 24, over Weeks 12-24, and over 24 weeks), rescue Ventolin HFA usage over 24 weeks, SGRQ total score over Weeks 12-24 and over 24 weeks, percentage of subjects achieving an MCID of 4 units or more in SGRQ total score at Week 24, RS-Total score over 24 weeks, time to CID, the rate of moderate or severe COPD exacerbations, and time to onset. Multiplicity will be controlled for the secondary variables as described in Section 6.4.11. The main analysis for secondary endpoints will use the efficacy estimand. The treatment policy estimand will include all post-treatment discontinuation data and will serve as a sensitivity analysis. The attributable estimand will also be estimated as a sensitivity analysis for the secondary endpoints with the exception of time of onset on Day 1.

For the attributable estimand for continuous variables, namely, rescue Ventolin use, the SGRQ total score, and the RS-Total Score, missing data are imputed based on the 95^{th} percentile of the reference arms' distribution if the reason for discontinuation is attributable to lack of efficacy or/tolerability, while for TDI and Peak FEV_1, missing data are imputed based on the 5^{th} percentile of the reference arms' distribution if the reason for discontinuation is attributable to lack of efficacy or/tolerability. All other missing data will be imputed using the observed data model. The variance used for the multiple imputation is described in the Details Appendix to this SAP. Time to onset on Day 1 is excluded since there is anticipated to be very little missing data at this early time point. The imputation for the rate of moderate or severe COPD exacerbations and time to CID is described in the respective sections below.

The non-inferiority margins will be set to 0.75 for TDI, 10% for achievement of MCID in SGRQ, 75 mL for peak change from baseline in FEV_1, and -1.5 for E-RS total score.

6.4.5.1 **Transition Dyspnea Index**

The TDI is collected in Study PT010006 only.

Assessments of dyspnea will be obtained using the BDI/TDI (where BDI is the Baseline Dyspnea Index). The BDI/TDI questionnaire can be found in Protocol Appendix 10.

At Randomization (Visit 4), the severity of dyspnea at baseline will be assessed using the BDI. BDI components are functional impairment, magnitude of task, and magnitude of effort (*Listing 6.1.6)*. The possible range of values for each BDI component score is 0 (very severe impairment) to 4 (no impairment). The BDI component scores are summed to determine the BDI focal score (0 to 12) (i.e., the lower the score, the worse the severity of dyspnea).

At subsequent visits (as per Schedule of Events: see of the Schedule of Events in the PT010006 protocol), change from baseline will be assessed using the TDI. TDI components include: Change in Functional Impairment, Change in Magnitude of Task, and Change in Magnitude of Effort (*Listing 6.1.6)*. The TDI component score ranges from -3 (major deterioration) to +3 (major improvement). The sum of all component scores yields the TDI focal score (-9 to +9) (i.e., the lower the score, the more deterioration from baseline).

The difference between treatment groups in TDI focal score over Weeks 12-24 and over 24 weeks and at each post-randomization visit will be analyzed using a similar RM approach as for the primary endpoint, but using BDI instead of baseline FEV_1 in the model, and adding baseline post-bronchodilator percent predicted FEV_1 as a continuous covariate. Thus, the model will include treatment, visit, treatment by visit interaction, and ICS use at Screening as categorical covariates and BDI, baseline eosinophil count, percent reversibility to Ventolin HFA and baseline post-bronchodilator percent predicted FEV_1 as continuous covariates. The treatment comparisons at each visit are considered to be “other” endpoints. Data from all study treatments will be included in the modeling. Scoring and handling of missing items will be conducted in accordance with the user’s guide for the TDI score. Two-sided p-values and point estimates with 2-sided 95% CIs will be produced for each treatment difference (*Tables* and *Figures 2.2.1* for the efficacy estimand).

Analyses in the attributable estimand, the treatment policy estimand, and the per protocol estimand over Weeks 12-24 and over 24 Weeks will be conducted as supportive (*Tables* and *Figures 2.2.2, 2.2.3* and *2.2.5,* respectively).

The attributable estimand will be computed in a similar manner as the attributable estimand is computed for change from baseline in morning pre-dose trough FEV_1 at Week 24 as described in Section 6.4.4.1.

All comparisons will be for superiority except that the comparison of BFF MDI to Symbicort TBH will be for non-inferiority rather than superiority and will use a margin of -0.75 for the lower bound of the two sided 95% CI for the treatment difference. The main analysis will be conducted using the efficacy and attributable estimands except for the non-inferiority comparisons of BFF MDI vs. Symbicort TBH, for which the per protocol estimand analysis will be main analysis.

In addition, the difference between treatments for the individual components of the TDI (as "other" endpoints): functional impairment, magnitude of task, and magnitude of effort will each be analyzed over 24 weeks, over Weeks 12-24, and at each post-baseline visit using the same modeling approach as for the TDI total score *(Table 2.2.7* and *Figures 2.2.7.1-2.2.7.3* for the functional impairment, magnitude of task, and magnitude of effort for the efficacy estimand).

Furthermore, as supportive analyses, responder analyses will be performed (for the TDI focal score) where responders are defined as subjects with a response of 1.0 point or more (corresponding to at least a minor improvement) on average over 24 weeks and on average over Weeks 12-24. Logistic regression with SAS PROC GENMOD will be used to compare the treatment groups with BDI, baseline eosinophil count, post-bronchodilator percent predicted FEV_1, and percent reversibility to Ventolin HFA as continuous covariates and treatment, and ICS use at Screening as categorical covariates. P-values and odds ratios with 95% CIs will be produced for each of the four treatment comparisons *(Table 2.2.13* for the efficacy estimand).

For the TDI, at each visit, if a response to any of the three questions is missing, then the focal score will also be considered missing. For the TDI responder analyses, subjects without post-baseline data or who discontinue treatment for any reason will be considered to be non-responders.

TDI and BDI data will be listed in *Listing 6.1.6*.

6.4.5.2 **Peak FEV_1**

Peak FEV_1 will be included in the analyses of the efficacy, attributable, treatment policy, and per protocol estimands as long as there is at least one non-missing post-dose value.

Peak FEV_1 will be analyzed for Study PT010006.

The peak change from baseline in FEV_1 within 4 hours post-dosing (assessed within a visit) over Weeks 12-24 (Japanese/China Approach), over 24 weeks (EU and Canada approaches), and at Week 24 (US Approach) will be analyzed and summarized similarly to morning pre-dose trough FEV_1. The peak change from baseline on Day 1 and at each post randomization visit will also be analyzed. Peak FEV_1 will be included in efficacy, attributable, and treatment policy estimand analyses as long as there is at least one non-missing post-dose value during the first 4 hours post-dose. For the per protocol estimand analysis, Peak FEV_1 will be calculated if there are at least two non-missing FEV_1 data-points during the first 4 hours post-dose (*Tables* and *Figures 2.3.1, 2.3.2, 2.3.3,* and *2.3.5* for the efficacy, attributable, treatment policy, and per protocol estimands, respectively).

The attributable estimand will be computed in a similar manner as the attributable estimand is computed for change from baseline in morning pre-dose trough FEV_1 at Week 24 as described in Section 6.4.4.1.

Additional sensitivity analyses will be implemented based on a cumulative responder approach as described in Farrar 2006 for the peak change from baseline in FEV_1 within 4 hours post-dose over 24 Weeks (*Table 2.3.4* for the efficacy estimand*)*. A cumulative distribution plot by treatment arm (Farrar et al., 2006) will also be produced. The observed peak change from baseline in FEV_1 will be plotted on the X axis, while the proportion of responders (subjects that equal or exceed that level of change) will be plotted on the Y axis (*Figure 2.3.4* for the efficacy estimand). Subjects without post-baseline data or who discontinue treatment for any reason will be classified as non-responders. For display purposes only, the range of the X axis will be from -1 to +1 liters [L] by increments of 0.01 liters in order to avoid the undue influence of outlying values. The cumulative responder curves for each treatment will then be compared pairwise using Kolmogorov-Smirnov tests. Cumulative responder analyses for the attributable estimand will not be performed because they are not well defined: methodology to apply Rubin's rules for combining multiply imputed data for such an analysis is not readily available.

Tipping Point Analyses for Peak Change from Baseline in FEV_1 within 4 hours Post-Dose at Week 24

Robustness of results to missing data will be explored using tipping point analyses (Ratitch 2013). A brief overview of the approach is summarized in the table below. Details of the methods are similar to sensitivity analyses of FEV_1 and AUC_{0-4} (found in Sections 6.4.4.4 and in the Details Appendix to this SAP), but with the model for peak change from baseline in FEV_1 described above. Multiple-imputation results will be combined using Rubin's formulae [Rubin, 1987]. Details of the methods may be found in the Details Appendix to this SAP.

Table 3 Sensitivity Analyses for Peak Change from Baseline in FEV_1

Efficacy Estimand		Attributable Estimand	Treatment Policy Estimand
mITT Population		mITT Population	ITT Population
Tipping point analysis #1: All missing data are imputed using the observed data model except that for subjects in the treatment arm (not the comparator arm) values that are considered MNAR are imputed with the change from baseline in the treatment arm decremented by up to 500 mL until the p-value ≥ 0.05.	Tipping point analysis #2: All missing data are imputed using the observed data model except that for subjects in the treatment arm (not the comparator arm) values are imputed with the change from baseline in the treatment arm decremented by up to 500 mL until the p-value ≥ 0.05.	Tipping point analysis: MI based on the 5th percentile of the reference arms' distribution if treatment discontinuation is due to tolerability or lack of efficacy of study drug (as in the primary analysis of this estimand). Otherwise all missing data are imputed using the observed data model except that for subjects in the treatment arm (not the comparator arm) values are imputed with the change from baseline in the treatment arm decremented by up to 500 mL until the p-value ≥ 0.05.	Tipping point analysis: All missing data are imputed using the observed data model except that for subjects in the treatment arm (not the comparator arm) values are imputed with the change from baseline in the treatment arm decremented by up to 500 mL until the p-value ≥ 0.05.

MNAR = Missing not at random. The tipping point will be shown to at least a precision of 10 mL. Imputed values may not be impossible values – i.e. changes from baseline that would imply a negative FEV_1 value. Thus the values will be imputed from a truncated distribution.

6.4.5.3 St. George's Respiratory Questionnaire

The SGRQ is collected in Study PT010006.

The SGRQ will be used to provide the health status/health-related QoL measurements in this study (see PT010006 Protocol Appendix 10). The SGRQ contains 50 rated items divided into three domains: "Symptoms" concerned with respiratory symptoms, their frequency, and severity; "Activity" concerned with activities that cause or are limited by breathlessness; and "Impacts" which covers a range of aspects concerned with social functioning and psychological disturbances resulting from airway disease. Individual items of SGRQ data will be listed (*Listings 6.1.7* for All Subjects Randomized).

A score will be calculated for each component and a "Total" score will also be calculated (*Listings 6.1.9* for All Subjects Randomized). In each case, the lowest possible value is zero and the highest is 100. Higher values correspond to greater impairment of QoL.

The difference between treatment groups in the change from baseline in SGRQ over 24 weeks and over Weeks 12 to 24 will be evaluated using a similar RM approach as for TDI focal score, but adding country as a covariate, and with baseline SGRQ score replacing BDI in the model. Thus the model will include treatment, visit, treatment by visit interaction, ICS use at Screening, and country as categorical covariates and baseline SGRQ score, baseline eosinophil count, percent reversibility to Ventolin HFA and baseline post-bronchodilator percent predicted FEV_1 as continuous covariates. Scoring and handling of missing items will be conducted in accordance with the user's guide for the SGRQ. Each response is to be given a unique empirically derived weight between 0 and 100, the weights of all responses are then summed up and divided by the maximum possible score and expressed as a percentage. Missing SGRQ total scores will not be imputed. Two-sided p-values and point estimates with 2-sided 95% CIs will be produced for each treatment difference. The comparison of BFF MDI to Symbicort TBH will be for non-inferiority rather than superiority and will use a margin of 3 units for the upper bound of the two sided 95% CI for the treatment difference.

The main analysis of the SGRQ will be for the efficacy estimand. Supportive analyses will use the attributable estimand, the treatment policy estimand, and per protocol estimand over Weeks 12-24 and over 24 Weeks *(Tables* and *Figures 2.4.1, 2.4.3, 2.4.4,* and *2.4.5* for the efficacy, attributable, treatment policy, and per protocol estimands, respectively).

The attributable estimand (for the analysis of the change from baseline in total SGRQ score) will be computed in a similar manner as the attributable estimand is computed for change from baseline in morning pre-dose trough FEV_1 at Week 24 as described in Section 6.4.4.1 except that the 95^{th} percentile will be used instead of the 5^{th} percentile.

As additional supportive analyses (i.e. as other efficacy endpoints), the difference between treatments at each of the individual visits will be evaluated and summarized *(Tables* and *Figures 2.4.1, 2.4.4,* and *2.4.5)*. Individual domains of the SGRQ (as "other" endpoints) will also be analyzed in a similar fashion as the total score *(Tables 2.4.7* for the efficacy estimand).

Responder analyses will be performed where responders are defined as subjects with an improvement of (i.e. a decrease in the total SGRQ score of) $\geq$4.0 points at Week 24 (secondary endpoint for US Approach), and on average over 24 weeks (EU and Canada Approaches) and Weeks 12-24 (Japan/China Approach). For the SGRQ responder analyses, subjects without post-baseline data or who discontinue treatment for any reason will be considered non-responders. Logistic regression will be used to compare the treatment groups with baseline SGRQ Score, baseline eosinophil count, baseline post-bronchodilator percent predicted FEV_1, and percent reversibility to Ventolin HFA as continuous covariates, and treatment, country, and ICS use at Screening as categorical covariates. P-values and odds ratios with 95% CIs will be produced for each treatment comparison (*Table 2.4.13* for the efficacy estimand and *Tables 2.4.14, 2.4.15,* and

2.4.16 for the attributable estimand, the treatment policy estimand, and the per protocol estimand, respectively).

The attributable estimand (for responder analysis of SGRQ) will be computed as follows. First, multiple imputations will be performed on the continuous total SGRQ scores in a similar manner as for the attributable estimand that is computed for change from baseline in morning pre-dose trough FEV_1 (Section 6.4.4.1), except that the 95th percentile will be used instead of the 5th percentile. After that, it will be determined whether the subject has attained the MCID. The analysis will proceed using logistic regression as described above, followed by combining of results across the multiple imputations using the formulae of Rubin [Rubin, 1987].

Additional sensitivity analyses will be implemented based on a cumulative responder approach as described in Farrar 2006 for the change from baseline in SGRQ score over 24 weeks (*Table 2.4.2* for the efficacy estimand). A cumulative distribution plot by treatment arm (Farrar et al., 2006) will also be produced. The observed change from baseline in SGRQ over 24 weeks will be plotted on the X axis, while the proportion of responders (subjects that equal or exceed that level of change) will be plotted on the Y axis (*Figure 2.4.2* for the efficacy estimand). Subjects without post-baseline data or who discontinue treatment for any reason will not be considered non-responders at all values. For display purposes only, the range of the X axis will be from -8 to +8 by increments of 0.1 in order to avoid the undue influence of outlying values. The cumulative responder curves for each treatment will then be compared pairwise using Kolmogorov-Smirnov tests. Cumulative responder analyses for the attributable estimand will not be performed because it is not well defined.

Tipping Point Analyses for Percentage of Subjects achieving an MCID of 4 Units or More in SGRQ Total Score at Week 24

Robustness of results to missing data will be explored using tipping point analyses (Ratitch 2013). A brief overview of the approach is summarized in the table below. Details of the methods are similar to sensitivity analyses of FEV_1 and $AUC_{0\text{-}4}$ (found in Section 6.4.4.4 and in the Details Appendix to this SAP), but with the model for SGRQ total score described above. Multiple-imputation results will be combined using Rubin's formulae [Rubin, 1987]. Details of the methods may be found in the Details Appendix to this SAP.

Table 4 Sensitivity Analyses for Percentage of Subjects Achieving an MCID of 4 Units or More in SGRQ Total Score at Week 24

Efficacy Estimand		Attributable Estimand	Treatment Policy Estimand
mITT Population		mITT Population	ITT Population
Tipping point analysis #1: All missing data are imputed using the observed data model except that for subjects in the treatment arm (not the comparator arm) values that are considered MNAR are imputed with the SGRQ total score in the treatment arm increased by a maximum of 16 units until the p-value ≥ 0.05.	Tipping point analysis #2: All missing data are imputed using the observed data model except that for subjects in the treatment arm (not the comparator arm) values are imputed with the SGRQ total score in the treatment arm increased by a maximum of 16 units until the p-value ≥ 0.05.	Tipping point analysis: MI based on the 95th percentile of the reference arms' distribution if treatment discontinuation is due to tolerability or lack of efficacy of study drug (as in the primary analysis of this estimand). Otherwise all missing data are imputed using the observed data model except that for subjects in the treatment arm (not the comparator arm) values are imputed with the SGRQ total score in the treatment arm increased by a maximum of 16 units until the p-value ≥ 0.05.	Tipping point analysis: All missing data are imputed using the observed data model except that for subjects in the treatment arm (not the comparator arm) values are imputed with the SGRQ total score in the treatment arm increased by a maximum of 16 units until the p-value ≥ 0.05.

MNAR = Missing not at random. The tipping point will be shown to at least a precision of 0.01 units. Imputed values may not be impossible values – i.e. a negative SGRQ total score. Thus the values will be imputed from a truncated distribution.

The multiple imputation will be applied to the continuous total SGRQ scores within the already-stated repeated-measures analysis framework for total SGRQ score. Missing values will first be imputed for the missing total SGRQ scores prior to the computation of whether the subject has attained the MCID (for the sensitivity analysis). The analysis using the imputed data will proceed using logistic regression as described above, followed by a combining of results across the multiple imputations using the formulae of Rubin [Rubin, 1987].

6.4.5.4 **Rescue Ventolin HFA Use**

PT010006:

The number of puffs of rescue Ventolin HFA taken in the previous 12 hours since the previous (AM or PM) dose will be recorded in the subject diary in the morning and evening. The mean daily number of puffs of rescue Ventolin HFA used by subjects during the study will be calculated overall and for each of the 4-week intervals during the treatment period and provided in a diary data listing *(Listing 6.1.3* for All Subjects Randomized). Diary data recorded during the last 7 days of the 10- to 14-day Screening Period will be used to calculate the baseline. For every interval of time over which the mean number of puffs of rescue will be calculated, records with missing values will be ignored in both the numerator and denominator. As such, the denominator will be adjusted based on the number of days (including half days) with non-missing values.

That is, the mean daily number of puffs of daytime rescue use (M_DT) will be set to the total number of daytime puffs divided by the number of half-days when daytime rescue use was recorded. The mean daily number of puffs of nighttime rescue use (M_DN) will be set to the total number of nighttime puffs divided by the number of half-days when the nighttime rescue use was recorded. The mean daily rescue use (puffs) is then two multiplied by the mean of M_DT and M_DN.

The difference between treatment groups in the change from baseline in rescue Ventolin HFA usage over 24 weeks will be evaluated using a linear RM analysis of covariance (ANCOVA) model which will include treatment, 4-week time interval (Interval 1 – Interval 6), treatment by time-interval interaction, and ICS use at Screening as categorical covariates, and baseline rescue Ventolin HFA use, baseline eosinophil count, percent reversibility to Ventolin HFA, and baseline post-bronchodilator percent predicted FEV_1 as continuous covariates. A UN matrix will be used to model additional autocorrelation within subject. If the UN model fails to converge, an AR(1) structure will be used instead. In the AR(1) model, subject will be included as a random effect. Contrasts will be used to obtain estimates of the treatment differences over the entire 24 weeks. Two-sided p-values and point estimates with two-sided 95% CIs will be produced for each of the four treatment differences. The comparison of BFF MDI to Symbicort TBH will be for non-inferiority rather than superiority and will use a margin of 0.75 puffs/day for the upper bound of the two sided 95% CI for the treatment difference. The main analysis will be conducted using the efficacy estimand (*Table* and *Figure 2.5.1*) restricted to the Rescue Ventolin User (RVU) Population. Supportive analyses will use the attributable, the treatment policy estimand, and the per protocol estimand (*Tables* and *Figures 2.5.3 and 2.5.5*) restricted to the RVU Population. Other supportive analyses will use the efficacy estimand not restricted to the RVU Population and the treatment policy estimand not restricted to the RVU Population (*Tables* and *Figures 2.5.6* and *2.5.7*).

The attributable estimand (for the analysis of average daily rescue Ventolin HFA use) will be computed in a similar manner as the attributable estimand is computed for change from baseline

in morning pre-dose trough FEV_1 at Week 24 as described in Section 6.4.4.1 except that the 95th percentile will be used instead of the 5th percentile.

As supportive analyses, the treatment difference for each 4-week interval and over Weeks 12 to 24 will be evaluated and summarized. Additionally, as supportive analyses, daytime rescue Ventolin® HFA use and night-time rescue Ventolin HFA use will be evaluated and summarized in a similar fashion. Two-sided p-values and point estimates with 2-sided 95% CIs will be produced for each treatment difference (*Tables* and *Figures 2.5.9* and *2.5.13* for the efficacy estimand restricted to the RVU Population).

Additional sensitivity analyses will be implemented based on a cumulative responder approach as described in Farrar 2006 for the change from baseline in average daily rescue medication over 24 weeks (*Table 2.5.1a* for the efficacy estimand). A cumulative distribution plot by treatment arm (Farrar et al., 2006) will also be produced. The observed change from baseline in average daily rescue medication over 24 weeks will be plotted on the X axis, while the proportion of responders (subjects that equal or exceed that level of change) will be plotted on the Y axis (*Figure 2.5.1a* for the efficacy estimand. Subjects without post-baseline data or who discontinue treatment for any reason will be considered non-responders for all values. For display purposes only, the range of the X axis will be from the 1st percentile to the 99th percentile irrespective of treatment in order to avoid the undue influence of outlying values. The cumulative responder curves for each treatment will then be compared pairwise using Kolmogorov-Smirnov tests. Cumulative responder analyses for the attributable estimand will not be performed because they are not well defined.

Tipping Point Analyses for Rescue Ventolin HFA Use Over 24 Weeks

Robustness of results to missing data will be explored using tipping point analyses (Ratitch 2013). A brief overview of the approach is summarized in the table below. Details of the methods are similar to sensitivity analyses of FEV_1 and AUC_{0-4} (found in Sections 6.4.4.4 and in the Statistical Methods Appendix to this SAP), but with the model for rescue Ventolin HFA use described above. Multiple-imputation results will be combined using Rubin's formulae [Rubin, 1987]. Details of the methods may be found in the Details Appendix to this SAP (Appendix 6).

Table 5 Sensitivity Analyses for Rescue Ventolin HFA Use

Efficacy Estimand		Attributable Estimand	Treatment Policy Estimand
mITT Population		mITT Population	ITT Population
Tipping point analysis #1: All missing data are imputed using the observed data model except that for subjects in the treatment arm (not the comparator arm) values that are considered MNAR are imputed with the number of puffs in the treatment arm increased by up to 4 puffs per day until the p-value ≥ 0.05.	Tipping point analysis #2: All missing data are imputed using the observed data model except that for subjects in the treatment arm (not the comparator arm) values are imputed with the number of puffs in the treatment arm increased by up to 4 puffs per day until the p-value ≥ 0.05.	Tipping point analysis: MI based on the 95th percentile of the reference arms' distribution if treatment discontinuation is due to tolerability or lack of efficacy of study drug (as in the primary analysis of this estimand). Otherwise all missing data are imputed using the observed data model except that for subjects in the treatment arm (not the comparator arm) values are imputed with the number of puffs in the treatment arm increased by up to 4 puffs per day until the p-value ≥ 0.05.	Tipping point analysis: All missing data are imputed using the observed data model except that for subjects in the treatment arm (not the comparator arm) values are imputed with the number of puffs in the treatment arm increased by up to 4 puffs per day until the p-value ≥ 0.05.

MNAR = Missing not at random. The tipping point will be shown to at least a precision of 0.02 puffs/day. Imputed values may not be impossible values – i.e. a negative number of puffs of rescue Ventolin HFA. Thus the values will be imputed from a truncated distribution.

PT010007:

The mean change from baseline in rescue use will be summarized by treatment group over 52 weeks and over each post-randomization 4-week interval (Interval 1 to Interval 13) (*Table* and *Figure 2.5.2* for the efficacy estimand with the Japanese mITT Population). Listing 6.1.3 in PT010006 will be generated for Japanese Randomized subjects and updated with information through 52 weeks.

6.4.5.5 **RS-Total Score**

PT010006:

The EXACT is a 14-item patient reported outcome (PRO) instrument from the daily diary which will be used to measure the effect of treatment on exacerbations, and on the severity of respiratory symptoms. Mean change from baseline in the daily EXACT Total Score, the 11-item RS-Total Score, as well as 3 subscale scores, RS-Breathlessness, RS-Cough and Sputum, and RS-Chest Symptoms, will be calculated over each post-randomization 4-week interval of the 24-week Treatment Period. Higher scores indicate a more severe condition. The last 7 days of the 10 to 14 day Screening Period will be used to calculate the baseline. The mean change from baseline in the EXACT Total Score, RS-Total Score, RS-Breathlessness, RS-Cough and Sputum, and RS-Chest Symptoms over each 4-week interval will be analyzed using a similar RM model as for TDI to estimate treatment effects over 24 weeks and over Weeks 12 to 24, but using the corresponding baseline mean score instead of the BDI as a covariate (*Tables and Figures 2.6.1, 2.6.3, 2.6.4,* and *2.6.5* for the efficacy estimand, the attributable estimand, and the treatment policy estimand*,* and *Tables 2.6.13* and *2.6.14* for the efficacy estimand and the efficacy estimand in the Japanese mITT Population). The analyses on the attributable estimand, treatment policy estimand, and per protocol estimand will pertain to the RS-Total Score only. Instead of visit, the number of the relevant respective 4-week interval (Interval 1 to Interval 6) will be used as a categorical covariate in the model. Thus the model will include treatment, time interval, treatment by time-interval interaction, and ICS use at Screening as categorical covariates and baseline score, baseline eosinophil count, percent reversibility to Ventolin HFA and baseline post-bronchodilator percent predicted FEV_1 as continuous covariates. An unstructured correlation matrix will be used to model additional autocorrelation within subject. If the UN model fails to converge, then an AR(1) structure will be used instead. In the AR(1) model, subject will be included as a random effect. The RS-Total score over 24 weeks is a secondary efficacy endpoint. The EXACT Total score, RS-Total score and the RS subscale scores are "other" endpoints. EXACT data will be listed in *Listing 6.1.5*. The analysis of RS-Total score will be secondary for the EU only.

The attributable estimand (for the analysis of RS-Total score) will be computed in a similar manner as the attributable estimand is computed for change from baseline in morning pre-dose trough FEV_1 at Week 24 as described in Section 6.4.4.1 except that the 95^{th} percentile will be used instead of the 5^{th} percentile.

Additional sensitivity analyses will be implemented based on a cumulative responder approach as described in Farrar 2006 for the change from baseline in daily RS-Total Score over 24 weeks (*Table 2.6.1a* for the efficacy estimand). A cumulative distribution plot by treatment arm (Farrar et al., 2006) will also be produced. The observed change from baseline in mean daily RS-Total score over 24 weeks will be plotted on the X axis, while the proportion of responders (subjects that equal or exceed that level of change) will be plotted on the Y axis (*Figure 2.6.1a* for the efficacy estimand). Subjects without post-baseline data or who discontinue treatment for any reason will be considered non-responders for all values. For display purposes only, the range of the X axis will be from the 1st percentile to the 99th percentile irrespective of treatment in order

to avoid the undue influence of outlying values. The cumulative responder curves for each treatment will then be compared pairwise using Kolmogorov-Smirnov tests. Cumulative responder analyses for the attributable estimand will not be performed because they are not well defined.

PT010007:

The mean change from baseline in RS-Total Score, EXACT Total Score, RS-Breathlessness, RS-Cough and Sputum, and RS-Chest Symptoms will be summarized by treatment group over 52 weeks and over each post-randomization 4-week interval (Interval 1 to Interval 13) (*Tables* and *Figures 2.6.2* and *2.6.8* and *Table 2.6.14*)

6.4.5.6 **Time to Onset of Action Assessed Using FEV_1 on Day 1**

This analysis will be conducted for PT010006.

The onset of action will be determined for each treatment using the post-dosing FEV_1 assessments from Day 1. The onset of action for each product (BGF MDI, GFF MDI, and BFF MDI) will be defined as the first time point where the mean change from baseline exceeds 100 mL. Supportive analyses may be conducted using alternatively definitions of onset of action. The resulting tables and figures are *Table* and *Figure 2.7.1* for the efficacy estimand.

6.4.5.7 **Other Spirometry Endpoints**

Analyses of other endpoints will be performed for the efficacy estimand only.

PT010006:

The analysis for between-treatment comparisons of changes from baseline in morning pre-dose trough FEV_1 over 24 weeks, over Weeks 12-24 and at each post-randomization visit through Week 24 has already been described in Section 6.4.4.1 (*Tables* and *Figures 2.1.1, and 2.1.3*). Analyses for FEV_1 AUC_{0-4} over 24 weeks, over Weeks 12-24, and at each post-randomization visit have been described in a similar manner in Section 6.4.4.2 (*Table* and *Figure 2.1.9* for efficacy estimand). Peak change from baseline within 4 hours in FEV_1 over 24 weeks, over Weeks 12 to 24, and at Day 1 and at each post-randomization visit where measured through Week 24 will be estimated and compared between treatment groups using a linear mixed RM model with the same model as pre-dose trough FEV_1 (*Table* and *Figure 2.3.1* for the efficacy).

Similar analyses will be conducted for FVC, PEFR, and FEF_{25-75} over 24 weeks, over Weeks 12 to 24, and at each post-randomization visit where measured for the mITT and ITT Populations, respectively. The baseline covariate for each model will be endpoint-specific (*Tables* and *Figures 2.8.1* to *2.8.14* for change from baseline in morning pre-dose trough FVC, PEFR, and FEF_{25-75}; *Tables* and *Figures 2.9.1* to *2.9.14* for peak change from baseline within 4 hours for FVC, PEFR, and FEF_{25-75}; *Tables* and *Figures 2.10.1* to *2.10.14* for FVC AUC_{0-4}, PEFR AUC_{0-4}, and FEF_{25-75} AUC_{0-4}). The analyses for change from baseline at each post-dose time point in FEV_1, FVC, PEFR, and FEF_{25-75} on Day 1 will be provided (*Tables* and *Figures 2.7.1* and *2.7.3*

for FEV_1 and *2.11.7* to *2.11.19* for FVC, PEFR, and FEF_{25-75}, respectively; *Listings 6.3.1* to *6.3.7* for the efficacy estimand).

On Day 1 during the first four hours post-dosing and by time point, the proportion of subjects achieving an improvement from baseline in FEV_1 using different thresholds (i.e., ≥10%, ≥12%, ≥100 mL, ≥200 mL, and "≥12% and ≥200 mL") will be estimated for each treatment. Subjects without post-baseline data will not be considered in the analysis. Logistic regression, performed with SAS PROC GENMOD, will be used to compare the treatments, adjusting for baseline FEV_1, reversibility to Ventolin HFA, and baseline eosinophil count as continuous covariates, and ICS use at Screening as a categorical covariate. The odds ratio for treatment will be determined, along with the Wald two-sided 95% CI. The Wald chi-square test will be used to calculate p-values for comparisons between treatments (*Tables* and *Figures 2.12.1, 2.12.3, 2.12.4, 2.12.5,* and *2.12.7* for the efficacy estimand).

PT010007:

The analysis for changes from baseline in morning pre-dose trough FEV_1 over 52 weeks (Weeks 4 to 52) and at each post-randomization visit has been described in Section 6.4.4.1 (*Table 2.1.2* and *Figures 2.1.2, 2.1.2.1,* and *2.1.2.2* for the Japanese mITT Population).

Similar analyses will be performed on Change from baseline in morning pre-dose trough for FVC, PEFR, and FEF_{25-75} over 52 weeks (Weeks 4 to 52), and at each post randomization visit measured for the Japanese mITT Populations (*Tables* and *Figures 2.8.2, 2.8.8,* and *2.8.14 for FVC, PEFR, and FEF_{25-75}).*

6.4.5.8 **Rate of COPD Exacerbations**

COPD Exacerbations

A **COPD exacerbation** will be defined as a change in the subject's usual COPD symptoms that lasts 2 or more days, is beyond normal day-to-day variation, is acute in onset, and may warrant a change in regular medication. The change in symptoms must include at least one major COPD symptom and at least one other major (dyspnea, sputum volume, and sputum color) or minor symptom (cough, wheeze, sore throat, cold symptoms, and fever without other cause).

Exacerbations will be considered **moderate** if they result in:

- Use of systemic corticosteroids and/or antibiotics for at least 3 days; a single depot injectable dose of corticosteroids will be considered equivalent to a 3-day course of systemic corticosteroids.

Exacerbations will be considered **severe** if they result in:

- An inpatient COPD-related hospitalization (documentation stating that the subject was hospitalized for the COPD exacerbation or a record of the subject being admitted for ≥24

hours to an observation area, the emergency department, or other equivalent healthcare facility depending on the country and healthcare system).

- COPD-related death.

Moderate-or-severe COPD exacerbations will be entered in the eCRF.

Additionally, the investigator may identify certain events (recorded on the same CRF page) which don't entirely meet the criteria above as exacerbations; the justifications supporting the investigator's judgment will be recorded on a separate page on the eCRF.

COPD exacerbations not meeting the criteria for moderate or severe COPD exacerbations will be considered to be mild COPD exacerbations. For more detail about moderate-or-severe, severe, and any-severity COPD exacerbation events (and their start and end dates) and how they are operationally defined, see the subsections titled "Duration of COPD Exacerbation", "Moderate-or-Severe Exacerbation and Severe Exacerbation: Operational Definitions:", and "Exacerbation of any Severity: Operational Definition:".

Tipping Point Analyses for Rate of Moderate or Severe COPD Exacerbations

Robustness of results to missing data will be explored using tipping point analyses (Ratitch 2013). A brief overview of the approach is summarized in the table below. Details of the methods are similar to sensitivity analyses of FEV_1 and AUC_{0-4} (found in Sections 6.4.4.4 and in the Details Appendix to this SAP), but with the negative binomial model for rate of moderate or severe COPD exacerbations described below.

Table 6 Sensitivity Analyses for Rate of Moderate or Severe COPD Exacerbations

Efficacy Estimand		Attributable Estimand	Treatment Policy Estimand
mITT Population		mITT Population	ITT Population
Tipping point analysis #1: All missing data are imputed using the observed data model except that for subjects in the treatment arm (not the comparator arm) values that are considered MNAR are imputed with rate in the treatment arm increased by up to 1.5 exacerbations/year until the p-value ≥ 0.05.	Tipping point analysis #2: All missing data are imputed using the observed data model except that for subjects in the treatment arm (not the comparator arm) values are imputed with the rate in the treatment arm increased by up to 1.5 exacerbations/year until the p-value ≥ 0.05.	Tipping point analysis: MI based on the 95th percentile of the reference arms' distribution if treatment discontinuation is due to tolerability or lack of efficacy of study drug (as in the primary analysis of this estimand). Otherwise all missing data are imputed using the observed data model except that for subjects in the treatment arm (not the comparator arm) values are imputed with rate in the treatment arm increased by up to 1.5 exacerbations/year until the p-value ≥ 0.05.	Tipping point analysis: All missing data are imputed using the observed data model except that for subjects in the treatment arm (not the comparator arm) values are imputed with the rate in the treatment arm increased by up to 1.5 exacerbations/year until the p-value ≥ 0.05.

MNAR = Missing not at random. The tipping point will be shown to at least a precision of 0.02 exacerbations/year.

Duration of COPD Exacerbation PT010006:

- The rate of moderate or severe COPD exacerbations will be analyzed using negative binomial regression as implemented in SAS PROC GENMOD. Treatments will be compared adjusting for baseline post-bronchodilator percent predicted FEV_1 and baseline eosinophil count as continuous covariates and baseline COPD exacerbation history (0, 1, >=2), country, and ICS use at Screening (yes/no) as categorical covariates. COPD exacerbations will be considered separate events provided that there are more than 7 days between the recorded stop date of the earlier event and the start date of the later event. Time at risk of experiencing an exacerbation will be used as an offset variable in the model.

- For the efficacy estimand, the time at risk is defined as time of exposure to randomized treatment – not during or within 7 days after an exacerbation (of equal or greater severity)

– until the last dosing date. More precisely, this is the amount of time between the date of first dose of study medication and the date of premature discontinuation from study medication plus + 1 day or the date of completion of study medication minus the number of days while the subject was experiencing any exacerbation and minus the seven days subsequent to any exacerbation. Any days subsequent to the date of premature discontinuation from study medication + 1 day or the date of completion of study medication are not subtracted. For the treatment policy estimand, time at risk is defined as follow-up time (equivalently, the total exposure to randomized medication and post-treatment follow-up for efficacy) – not during or within 7 days after an exacerbation (of equal or greater severity) – up to the last recorded date (of any assessment or contact) for the subject (including telephone contact) (*Tables 2.13.1.1* and *2.13.4.1* for the efficacy estimand, and the treatment policy estimand, respectively). The per protocol estimand is not applicable for this endpoint.

- Any imputed COPD exacerbation events for the attributable estimand or for sensitivity analyses will have an assumed duration of 10 days.

- For moderate-or-severe COPD exacerbations that were identified apart from an electronic diary (eDiary) alert, the symptom information is listed in *Listing 6.1.2.2*.

- The rate (and number of) exacerbations and the percentage of subjects who experience exacerbations will be summarized for moderate-or-severe exacerbations (*Tables 2.13.1, 2.13.3, and 2.13.4* for the efficacy, attributable and treatment policy estimands, respectively).

- The rate of moderate or severe COPD exacerbations for subjects with ≥1 exacerbation in the previous year will be summarized for the efficacy estimand (*Tables 2.13.6*).

-

- The rate of moderate or severe COPD exacerbations for completers vs. discontinuations (*Table 2.13.1b*).

PT010007:

- The rate of moderate-or-severe COPD exacerbations and the percentage of subjects who experience exacerbations will be summarized by treatment group (*Table 2.13.2* for the efficacy estimand with the Japanese mITT Population).

Duration of COPD Exacerbation

For moderate or severe exacerbations, the duration is defined by the length of prescribed treatment (using the eCRF COPD exacerbation page), whereas for mild exacerbations, the duration is defined by the length of symptoms.

For moderate or severe COPD exacerbations, the start date will be defined as the start date of prescribed treatment with a systemic corticosteroid or systemic antibiotic and the stop date will be defined as the latter of the last day of prescribed treatment with a systemic corticosteroid or systemic antibiotic (if applicable). In order to ensure that the same event is not counted twice, consecutive or concurrent moderate or severe COPD exacerbations with equal to or fewer than 7 days between the recorded stop date of the earlier event and start date of the later event will be considered the same event and assigned the maximum severity between the two.

For mild COPD exacerbations, start date will be defined as the onset of worsened symptoms as recorded by the subject in the eDiary, and the stop date will be defined as the last day of worsened symptoms. In order to ensure that the same event is not counted twice, consecutive or concurrent mild COPD exacerbations with equal to or fewer than 7 days between the recorded stop date of the earlier event and start date of the later event will be considered the same event.

In addition, in order to not double-count exacerbations, eDiary data from dates within 7 days prior to or after a moderate or severe exacerbation will not be counted as an additional mild COPD exacerbations. This implies that continuing worsened symptoms that meet the definition of a mild exacerbation would need to be present at least 2 days prior to the 7-day period immediately preceding the start date of a moderate or severe COPD exacerbation in order to be considered a separate event. Similarly, worsened symptoms would need to be present for at least 2 days after the 7-day period immediately following a moderate or severe COPD exacerbation to be considered a separate event.

Analyses of each severity of exacerbation will account for the time that subjects are at risk of having an exacerbation of that severity or greater. Time during or immediately following – i.e. within 7 days of – an exacerbation will not be considered as part of the time that the subject was at risk. However, time during or immediately following an exacerbation of lower severity will be included since, for example, a subject experiencing a mild exacerbation is still at risk of the event increasing in severity and becoming a moderate exacerbation. Moderate and severe COPD exacerbations occurring within 7 days of one another will be coalesced into a single COPD exacerbation event with the severity of "severe" (see details below).

Moderate-or-Severe Exacerbation and Severe Exacerbation: Operational Definitions:

- Moderate exacerbations and severe exacerbations will be defined based on information from the COPD Exacerbation eCRF page. A time interval from a single COPD exacerbation eCRF page will be designated as being during an event of a moderate-or-

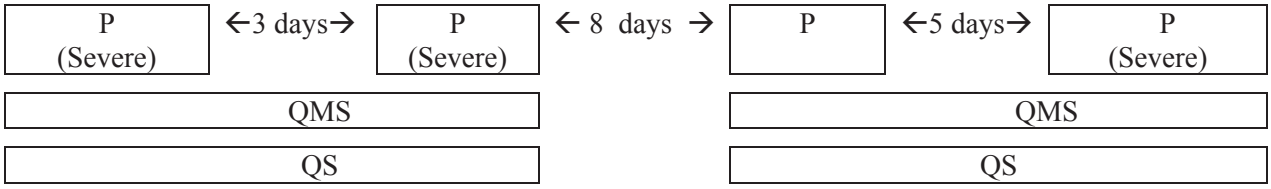
severe COPD exacerbation if for that interval, either antibiotics or oral corticosteroids were administered for the exacerbation.

Call this time interval a "P-Interval". The start date of the P-Interval is the earliest start date of the above, and the stop date will be defined as the last stop day of the above. If the subject was hospitalized due to the exacerbation or if the exacerbation led to a COPD-related death, then the severity of "severe" will be assigned to this P-interval; otherwise the severity of "moderate" will be assigned. The later of the stop date of the treatment with a systemic corticosteroid and the stop date of the treatment with an antibiotic will be the end date of the COPD exacerbation (i.e. the end of the P-Interval).

An overarching interval of (any number of) such P-Intervals – including any P-Intervals with an end date not more than 7 days prior to the start date of some other P-Interval or with a start date not more than 7 days after the end date of some other P-Interval – and including the days in any gaps between them – will be called an "QMS-Interval". This QMS-interval will represent the consolidated duration of several exacerbations recorded on different CRF pages. This QMS-Interval will be considered to be a single event of a moderate-or-severe COPD exacerbation. See Figure 2.

- A P-interval of severe COPD exacerbation is called a "severe" P-Interval. Any QMS interval that contains at least one "severe" P-Interval will also be called a "QS-Interval". This QS-Interval will be considered to be a single event of a severe COPD exacerbation. See Figure 2.

Figure 2 Overarching Intervals of Moderate-or-Severe (QMS) and Severe (QS) COPD Exacerbations



A P-interval is a moderate-or-severe COPD exacerbation instance from a single CRF page.
In a "Severe" P-Interval [denoted in the figure as "P (severe)"], the maximum severity of the COPD exacerbation is "severe".
A QMS interval is an overarching moderate-or-severe COPD exacerbation event encompassing multiple CRF pages.
A QS interval is an overarching severe COPD exacerbation event encompassing multiple CRF pages.

Exacerbation of any Severity: Operational Definition:

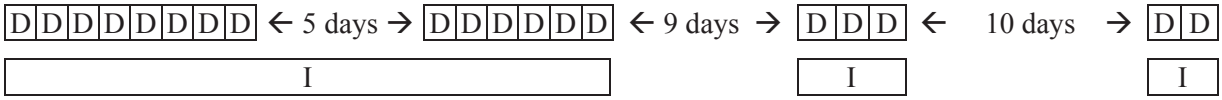
- Using eDiary data, a day will be designated as being during an event of a COPD exacerbation of some severity if (1) there was at least one major symptom and there was at least one other major or minor symptom and if (2) on an adjacent day there was at least

one major symptom and there was at least one other major or minor symptom. Denote such a day as a "Category-D" day.

- An interval of (any number of) such Category-D days – including any Category-D days not more than 7 days apart from some other Category-D day – and including the days in any gaps between them – will be called an "I-Interval". See Figure 3.

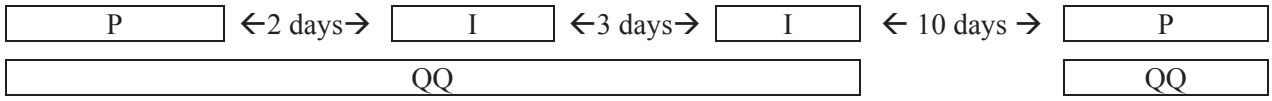
- An overarching interval coalescing (any number of) P-Intervals and I-Intervals – including any such P-or-I-intervals with an end date not more than 7 days prior to the start date of some other P-or-I-Interval or with a start date not more than 7 days after the end date of some other P-or-I-interval – and including the days in any gaps between them – will be called a "QQ-Interval". This QQ-interval will represent the consolidated duration of several exacerbations recorded on different CRF pages or identified from subject diary data. This QQ-Interval will be considered to be a single event of an any-severity COPD exacerbation. See Figure 4.

Figure 3 Overarching Intervals (I) of Mild-Moderate-or-Severe COPD Exacerbation Events Based on eDiary Symptom Data



A Category-D day is a day with mild-moderate-or-severe COPD exacerbation based on e-diary symptom data.
An I-Interval is an overarching mild-moderate-or-severe COPD exacerbation event encompassing multiple clusters of e-diary symptom days.

Figure 4 Overarching Intervals (QQ) of Mild-Moderate-or-Severe COPD Exacerbation Events Incorporating Both CRF Data and eDiary Symptom Data



A P-Interval is a moderate-or-severe COPD exacerbation instance from a single CRF page.
An I-Interval is an overarching mild-moderate-or-severe COPD exacerbation event based on e-diary symptom data.
A QQ-Interval is an overarching mild-moderate-or-severe COPD exacerbation event – encompassing multiple P-Intervals and I-Intervals – incorporating both CRF data and e-diary symptom data.

In summary, we combine CRF-based moderate-or-severe COPD exacerbation events if they are close enough together in time (Figure 2). We also combine severe COPD exacerbation events with other severe COPD exacerbation events or with moderate-or-severe COPD exacerbation events if they are close enough together in time (Figure 2) – thus forming a single severe COPD exacerbation event. We also combine mild-moderate-or-severe COPD exacerbations if they are close enough together in time; this coalescing is done first within-data-source (CRF [Figure 2] or diary [Figure 3]) and then between the two sources (Figure 4).

Time-at-Risk for COPD Exacerbations of Various Severities: Operational Definition

- During a time when a subject is not experiencing a severe COPD exacerbation (i.e. QS interval) – and is not in the seven days following a severe COPD exacerbation – a subject is considered to be at risk of having a severe exacerbation. During a time when a subject is not experiencing a moderate-or-severe COPD exacerbation (i.e. QMS interval) – and is not in the seven days following a moderate-or-severe COPD exacerbation – a subject is considered to be at risk of having a moderate-or-severe exacerbation. During a time when a subject is not experiencing an any-severity COPD exacerbation (i.e. QQ interval) – and is not in the seven days following an any-severity COPD exacerbation – a subject is considered to be at risk of having an any-severity exacerbation.

Overarching coalesced intervals (i.e. events) of COPD exacerbation will be listed for severe exacerbations, moderate -to-severe exacerbations, and any-severity exacerbations (*Listing 6.1.2.3*). A severe COPD exacerbation event must be classified also as a moderate-or-severe event and also as an any-severity event. A moderate-or-severe COPD exacerbation event must be classified also as an any-severity event.

PT010006:

- Rate of COPD exacerbations of any severity will be analyzed in a manner similar to the rate of moderate or severe COPD exacerbations *(Table 2.13.11.1* for the efficacy estimand).

- Rate of severe COPD exacerbations will be analyzed in a manner similar to the rate of moderate or severe COPD exacerbations *(Table 2.13.12.1* for the efficacy estimand).

PT010007:

- Rate of COPD exacerbations of any severity will be analyzed in a manner similar to the rate of moderate or severe COPD exacerbations *(Table 2.13.11.2* for the efficacy estimand with the Japanese mITT Population).

- Rate of severe COPD exacerbations will be analyzed in a manner similar to the rate of moderate or severe COPD exacerbations *(Table 2.13.12.2* for the efficacy estimand with the Japanese mITT Population).

The count of COPD exacerbations of any severity is the number of QQ-Intervals (for a subject) as defined previously. Time at risk of experiencing an exacerbation will be used as an offset variable in the model. Time during an exacerbation (of any severity) or in the 7 days following an exacerbation (of any severity) will not be included in the calculation of exposure (i.e. time at risk). Data related to COPD exacerbations of any severity are listed in *Listings 6.1.2.1, 6.1.2.2, 6.1.2.3,* and *6.1.4*. For moderate-or-severe COPD exacerbations that were identified apart from an eDiary alert, the symptom information is listed in *Listing 6.1.2.2*.

6.4.6 Time to Clinically Important Deterioration

Time to CID will be analyzed for PT010006.

Two definitions of CID will be used:

- CID: is ≥ 100 mL decrease (from baseline) in trough FEV_1, or ≥ 4 points increase (from baseline) in SGRQ total score, or a TDI focal score of -1 point or less, or treatment-emergent moderate-or-severe COPD exacerbation occurring up to Week 24.

- Sustained CID: is ≥ 100 mL decrease (from baseline) in trough FEV_1, or ≥ 4 points increase (from baseline) in SGRQ total score, or a TDI focal score of -1 point or less, any of which is occurring on two consecutive analysis visits or for ≥ 50% of all available subsequent analysis visits, or a treatment emergent moderate-or-severe COPD exacerbation occurring up to Week 24

Time to CID analysis will be performed for each CID definition using the Cox regression model. The model will include baseline post-bronchodilator percent predicted FEV_1, baseline eosinophil count, baseline COPD exacerbation history (0, 1, >=2), country, and ICS use at Screening (yes/no)**.** Time to a CID event will be based on the component event which occurs first. Subjects who do not have a CID event will be censored at the earliest day among the component censoring times. COPD exacerbations happening after Week 24 will not be counted as CID events. Estimated adjusted hazard ratios will be displayed along with associated 95% CI and p-values (*Tables 2.60.1,* to *2.60.5*). Time to CID will be displayed for each treatment group using a Kaplan-Meier curve (*Figures 2.60.1* to *2.60.5*). Time to CID will be analyzed for the efficacy estimand, the attributable estimand, the treatment policy estimand, and the per protocol estimand. The comparison of BFF MDI to Symbicort TBH will be for non-inferiority with the per protocol estimand rather than superiority and will use a margin of 1.1 for the upper bound of the two sided 95% CI for the hazard ratio.

For the analysis of the attributable estimand, missing data that have been reasonably attributed to tolerability or lack of efficacy will be imputed based on either the 5^{th} or the 95^{th} percentile (see below) of the reference arms' distribution. The attributable estimand for time to CID will employ multiple by-visit imputation of pre-dose trough FEV_1 and the SGRQ total score (at visits for which they are missing) (as described in Sections 6.4.4.1 and 6.4.5.3, respectively) and multiple imputation of time to first moderate-or-severe COPD exacerbation. A complete dataset for the COPD exacerbation count will be created analogously to that described in the analysis of Rate of Moderate or Severe Exacerbations in Section 6.4.5.8; however, the timing of imputed events is also needed. These will be obtained for each imputed event by randomly drawing a value from the uniform distribution over the interval that starts with time of treatment discontinuation + 1 day (in study days) and ends at 24 weeks. The attributable estimand for time to CID will also employ multiple by-visit imputation of the TDI focal score in a manner similar to that for SGRQ. The application of the percentile penalty to the attributable estimand will be carried out for the four component variables simultaneously.

6.4.7 Analysis of Other Endpoints

6.4.7.1 Percentage of Days with "No Rescue Ventolin HFA Use" Over the Treatment Period

PT010006:

As a supportive analysis, percentage of days with "no rescue Ventolin HFA use" over 24 weeks will be analyzed. A "day with no rescue use" is defined as any day where the subject reported zero puffs of rescue Ventolin HFA. The rescue Ventolin HFA usage diary data from days where rescue Ventolin HFA usage data is non-missing will be used to ascertain the days with "no rescue Ventolin HFA use". The percentage of days with no rescue use will be calculated as 100 x (number of days no rescue Ventolin use over the entire treatment period / number of days with non-missing rescue Ventolin use over the entire treatment period). The percentage of days with "no rescue use" will be summarized by treatment and analyzed using ANCOVA with baseline average daily rescue Ventolin HFA use, percent reversibility to Ventolin HFA, baseline post-bronchodilator percent predicted FEV_1, baseline blood eosinophil count as continuous covariates and ICS use at Screening as a categorical covariate. Two-sided p-values and point estimates with 2-sided 95% CIs will be produced for each treatment difference (*Tables 2.15.1* for the efficacy estimand).

For the efficacy estimand, the analyses of Percentage of Days with "No Rescue Ventolin HFA Use" will be restricted to the Rescue Ventolin User Population.

PT010007:

Percentage of days with "no rescue Ventolin HFA use" over 52 weeks (PT010007) will be summarized descriptively (Table 2.15.2 for the efficacy estimand with the Japanese mITT Population).

6.4.7.2 Time to First COPD Exacerbation

Time to First COPD Exacerbation will be analyzed for PT010006.

Time to first COPD exacerbation of any severity is the time from first dose of study medication (or from randomization for any subjects randomized but not treated) to the time of onset of the first COPD exacerbation (mild, moderate, or severe). Time to first severe COPD exacerbation is the time from first dose of study medication (or from randomization for any subjects randomized but not treated) to the time of onset of the first COPD exacerbation (severe). Time to first moderate or severe COPD exacerbation is the time from first dose of study medication (or from randomization for any subjects randomized but not treated) to the time of onset of the first COPD exacerbation (moderate or severe).

The time to first COPD exacerbation of any severity will be analyzed up through Week 24 using a Cox regression model. Treatment comparisons will be performed using the model, adjusting for baseline percent predicted FEV_1 and baseline eosinophil count as continuous covariates and

baseline COPD exacerbation history (0, 1, ≥2), country, and ICS use at Screening (Yes/No) as categorical covariates. Estimated adjusted hazard ratios relative to the comparator for each treatment comparison will be displayed along with the associated Wald two-sided 95% CIs and p-values (*Table 2.13.13* for the efficacy estimand).

Time to first COPD exacerbation of any severity will be displayed graphically for each treatment using a Kaplan-Meier curve (*Figure 2.13.13* for the efficacy estimand). Subjects who do not experience a COPD exacerbation over treatment period will be censored at the Week 24 visit. Subjects who withdraw from the study without experiencing a COPD exacerbation will be censored at the date of withdrawal or the last date of treatment, whichever is later. Time to first moderate or severe COPD exacerbation will be analyzed and displayed similarly to the time to first COPD exacerbation of any severity (*Table* and *Figure 2.13.19* for the efficacy estimand and Listing 6.1.2.3).

Time to first severe COPD exacerbation will be analyzed similarly to the analysis for time to first COPD exacerbation of any severity (Table 2.13.20 and Figure 2.13.19).

6.4.7.3 **Time to Treatment Failure**

Time to Treatment Failure will be analyzed for PT010006.

Treatment failure is defined as a moderate or severe COPD exacerbation or premature discontinuation from treatment for any reason or death. Time to treatment failure will be displayed graphically for each treatment group using a Kaplan-Meier curve and analyzed using a log-rank test to compare the curves between the treatments (*Figure 2.14.1*). Subjects who do not experience a treatment failure will be censored at their Week 24 Visit. The time to treatment failure will be analyzed using the efficacy estimand. The model will include treatment, baseline post-bronchodilator percent predicted FEV_1, baseline eosinophil count, baseline COPD exacerbation history (0, 1, >=2), country, and ICS use at Screening (yes/no). Estimated adjusted hazard ratios will be displayed along with associated 95% CI and p-values (*Table 2.14.1*).

6.4.7.4 **European Quality-of-Life-5 Dimension-5 Level Questionnaire**

Data from the EQ-5D-5L will be analyzed for PT010006.

The data will be weighted to calculate an index score based upon subjects' responses to the 5 dimensions. The visual analogue scale (VAS) will be scored from 0 (worst imaginable health state) through 100 (best imaginable health state) to represent the subject's self-report concerning how bad or how good their health was during that day.

EQ-5D will be presented in three different ways:
1. Presenting results from the EQ-5D-5L descriptive system as a health profile at baseline, at all visits, and at EoT (%, n) by domain
2. Presenting results of the VAS as a measure of overall self-rated health status - baseline scores, scores at each visit, changes from baseline at each visit, and mean VAS score over the treatment period

3. Presenting results from the EQ-5D-5L index score (using UK value set) baseline, each visit, changes from baseline to each visit, and the mean index score over the treatment period.

The percentage of subjects' categorical responses to each of the 5-dimensions will be summarized (*Table 2.16.1* for the efficacy estimand). Descriptive statistics for the index score (*Table 2.16.7* for the efficacy estimand) and VAS (*Table 2.16.13* for the efficacy estimand) will be presented by treatment group. VAS scores over 24 weeks may be analyzed using a similar RM model as is used for the TDI, but using baseline EQ-5D VAS score as a covariate instead of BDI (*Table* and *Figure 2.16.7* for the index score and *Table 2.16.13* for VAS for the efficacy estimand). EQ-5D data are listed in *Listing 6.1.11*.

For calculations of index score, the method recommended by the national institute for health and care excellence (NICE) August 2017 will be applied. Cross-walk between EQ-5D-3L value set and EQ-5D-5L descriptive system have been developed by Van Hout et al 2012 (Van Hout *et al.* 2012) and this cross-walk value set for EQ-5D-5L will be used to calculate the index score (Van Reenan 2015). Appendix 10 contains the SAS/SPSS codes for crosswalk between 5L and 3L for calculation of index score.

No imputation will be made for missing data in either the EQ-5D-5L or VAS responses.

The compliance of completing the EQ-5D-5L questionnaires is a critical issue in the QoL and health-state evaluation, and will be described by post-randomization visit, by displaying the number and percentage of subjects who were assessed (per subject, at least 1 question answered) at each visit *(Table 2.16.19* for the efficacy estimand).

6.4.8 12-Hour Pulmonary Function Tests

12-Hour PFTs will be analyzed for PT010006.

FEV_1 AUC_{0-12} will be measured in a subset of approximately 600 randomized subjects at Day 1 (Visit 4) and Week 24 (Visit 10a). Area under the curve at Week 24 will be calculated using the trapezoidal rule, after first subtracting the baseline FEV_1 value, and transformed into a weighted average by dividing by the time in hours from dosing of the last measurement included (typically 12 hours). Spirometry data are listed in *Listings 6.2, 6.3.9* for FEV_1, *6.3.12* for FVC, *6.3.15* for PEFR, and *6.3.18* for FEF_{25-75} for the efficacy estimand). Participation in the 12-hr PFT Sub-study is listed in *Listing 9.7*.

At least one non-missing post-dose value is required for the calculation of AUC. Actual time from dosing will be used if available; otherwise scheduled time will be used. The differences between treatments in FEV_1 AUC_{0-12} at Day 1 and Week 24 will be evaluated using an ANCOVA with baseline FEV_1, baseline eosinophil count, and percent reversibility to Ventolin HFA as continuous covariates and treatment and ICS use at Screening as categorical covariates. Two-sided p-values and point estimates with two-sided 95% CIs will be produced for each

treatment difference (*Table 2.17.1* to *2.17.*4 for the efficacy, attributable, treatment policy, and per protocol estimands for FEV_1 AUC_{0-12}, and *Table 2.17.7* and *Figure 2.17.7* for FEV_1 over 12 hours post-dose for the efficacy estimand). The comparison of BFF MDI to Symbicort TBH will be for non-inferiority rather than superiority and will use a margin of -75 mL for the lower bound of the two sided 95% CI for the treatment difference; for that comparison, the analysis on the per protocol estimand will be the main analysis. As additional supportive analyses, FEV_1 AUC_{0-4}, FEV_1 AUC_{0-6}, and FEV_1 AUC_{6-12} will be calculated and analyzed in a similar fashion as FEV_1 AUC_{0-12} (*Tables 2.17.28 for* $AUC_{0-4,}$ *2.17.31* for $AUC_{0-6,}$ and *2.17.34* for $AUC_{6-12,}$ for the efficacy estimand). Similar analyses (as for AUC_{0-4} and AUC_{0-12}) will be performed for FVC, PEFR, and FEF_{25-75} *(Tables 2.17.37* for FVC $AUC_{0-4,}$ *2.17.40* for FVC $AUC_{0-12,}$ *2.17.43* for PEFR $AUC_{0-4,}$ *2.17.46* for PEFR $AUC_{0-12,}$ *2.17.49* for FEF_{25-75} $AUC_{0-4,}$ *2.17.52* for FEF_{25-75} $AUC_{0-12,}$ for efficacy estimand). The baseline covariate will be endpoint-specific.

Treatments will be compared using change from baseline at each post-dose time point over 12 hours at Week 24 (Visit 10a) for the following variables: FEV_1, FVC, PEFR, and FEF_{25-75}. The differences between treatments will be evaluated using an ANCOVA with baseline of the respective endpoint, baseline eosinophil count, and percent reversibility to Ventolin HFA as continuous covariates and treatment and ICS use at Screening as categorical covariates *(Tables* and *Figure 2.17.7* and *Listing 6.3.9* for FEV_1, *2.17.13* and *Listing 6.3.12* for FVC, *2.17.19* and *Listing 6.3.15* for PEFR, and *2.17.25* and *Listing 6.3.18* for FEF_{25-75} for the efficacy estimand). The 12-hour post-dose trough is defined as the average of the 11.5 and 12 hour post-dose values. In subjects missing either of these assessments, the value will be calculated from the single measurement. In subjects missing both values, this value will be missing.

The attributable estimand (for the analysis of FEV_1 AUC_{0-12}) will be computed in a similar manner as the attributable estimand is computed for change from baseline in morning pre-dose trough FEV_1 at Week 24 as described in Section 6.4.4.1.

6.4.9 Subgroup Analyses

Selected tabulations will be provided within the following subgroups for the primary endpoints:

- China
- Japan and Non-Japan
- Asia (Asia is defined by country rather than by race and includes China and Japan.)
- Each Individual Country (i.e. US, Canada, China, and Japan)
- Severity of COPD (based on FEV_1):
    - Moderate
    - Severe
    - Very Severe
- GOLD Categories (based on CAT, FEV_1, and Exacerbations per year):

  - B, which is defined as: (% predicted FEV_1 post-bronchodilator ≥ 50% AND no COPD hospitalizations AND at most one COPD exacerbation in the last 12 months) AND CAT ≥ 10
  - D, which is defined as: (% predicted FEV_1 post-bronchodilator < 50% OR at least one COPD hospitalizations OR at least two COPD exacerbations in the last 12 months) AND CAT ≥ 10
- Reversibility to Ventolin HFA:
  - No
  - Yes
- Baseline Eosinophil Count:
  - <150 cells per mm^3
  - ≥150 cells per mm^3
- Racial groups (for the primary efficacy and the below safety endpoints only):
  - Black
  - White
  - Native Hawaiian or Pacific Islander (for safety only)
  - American Indian or Alaska Native (for safety only)
  - Asian
  - Other (for safety only)
- Age groups (for the primary efficacy and the below safety endpoints only):
  - Age < 65 years
  - Age ≥ 65 years
- Sex groups (for the primary efficacy and the below safety endpoints only):
  - Male
  - Female
- ICS Use at Screening:
  - Yes
  - No
- Post-bronchodilator FEV_1 (<50%, ≥50% Predicted)
- Exacerbation history (0, ≥1 in the last year)

The distribution of baseline characteristics used in the primary model – ICS use at screening, baseline FEV_1, baseline eosinophil count, and percent reversibility to Ventolin HFA – will be summarized within subgroups, both overall and by treatment group.

The following tables and figures will be provided by subgroup:

- Change from Baseline in Trough FEV_1 (excluding the exacerbation history subgroups).

- FEV_1 area under the curve from 0 to 4 hours, except Japan (no post-dose spirometry assessments collected) and Asia and the exacerbation history subgroups.
- Rate of Moderate or Severe COPD exacerbations (for country, baseline eosinophil count, and exacerbation history [0, ≥1 in the last year] subgroups only).

In addition, the following subgroup analyses will be presented for the efficacy estimand for the China (*Tables* and *Figures 8.X.X.X.X*), Japan/non-Japan (*Tables* and *Figures* 6.*X.X.X.X*), and Asia (*Tables* and *Figures 9.X.X.X.X*) subgroups, unless indicated otherwise (see also Section 6.4.9.1 and 6.4.9.2):

- TDI Focal Score
- TDI Responder (China and Asia only)
- Change from Baseline in Average Daily Rescue Ventolin HFA Use
- Change from Baseline in Average Daily Daytime Rescue Ventolin HFA Use (China only)
- Change from Baseline in Average Daily Nighttime Rescue Ventolin HFA Use (China only)
- Change from Baseline in SGRQ Total Score
- SGRQ Responder (China and Asia only)
- Peak Change from Baseline in FEV_1 (L) within 4 hours post-dosing (China only)
- Change from Baseline in FEV_1 (L) at 5-Min Post-Dose on Day 1 (China only)
- Change from Baseline in FEV_1 (L) at 15-Min Post-Dose on Day 1 (China only)
- RS-Total Score
- EXACT Total Score (China and Asia only)
- Time to CID
- Change from Baseline in Trough FVC (China and Asia only)
- Percent responders analyses for Day 1 (China only)
- Rate of Moderate or Severe COPD Exacerbations
- Rate of Severe COPD exacerbations (China and Asia only)
- EuroQoL 5 Dimensions categorical responses (China and Asia only)

In the notation X.X.X.X used above, X can refer to any integer.

Additional analyses for China and Asia subgroups may be found in Appendix 8 and Appendix 9, respectively.

Each subgroup will be analyzed separately using the same model that was used for the overall (combined subgroups) analysis. Estimates for the treatment effect and for the treatment

differences will be displayed in the efficacy endpoint tables for each subgroup *(Tables 6.1.1* to *6.11.40)*.

For each subgroup analysis, a test for the treatment-by-subgroup interaction will be performed using the same model that was used for the overall (combined subgroups) analysis but with the addition of terms for subgroup and the treatment-by-subgroup interaction. A table will be provided with the p-value for the test of the treatment-by subgroup interaction *(Table 6.13* for the efficacy estimand*)*. Should any country/region effects be identified, shrinkage estimates may be generated in order to further understand the impact of these effects (Carroll and Fleming, 2013).

Subgroup analyses of the trough FEV_1 will be conducted in the baseline eosinophil count-high ($\geq$150 cells per mm^3) and the baseline eosinophil count-low (<150 cells per mm^3) subgroups. It is acknowledged 150 cells per mm^3 may not ultimately be the appropriate threshold for evaluation of treatment benefit. Thus, additional analyses will evaluate alternative thresholds, and the results from these analyses could then inform thresholds for future clinical studies. This exploration will include using additive mixed models that combine nonparametric regression for the relationship of eosinophil levels to trough FEV_1 as well as potentially using subgroups defined by different cut points.

6.4.9.1 **China and Asia Subgroups**

The China subgroup is defined as all subjects enrolled in sites in China. The Asia subgroup is defined as all subjects enrolled in sites located in Japan and China.

Corresponding ITT, mITT, RVU, PP and Safety analysis populations, as well as other relevant analysis sets, for the China and Asia subgroups are defined as described in Section 5.1 and restricted to the China/Asia subgroups. Populations described for primary analyses in Section 5.2 are applied to China/Asia subgroups.

6.4.9.2 **Subgroup Analyses China and Asia**

To support registration in China, a separate Clinical Study Report will be written to present the study results in the China and Asia subgroups with the objective of demonstrating consistency with the overall Study PT010006 population. Select subject disposition, demographic and baseline characteristics, extent of exposure, prior and concomitant medications, efficacy, and safety analyses will be repeated for the China/Asia subgroups. Appendix 8 and Appendix 9 contain the table of contents for Post-Text TFLs in the China and Asia subgroups, respectively.

Analyses in the China and Asia subgroups will proceed as described in Section 6.4.9. All analyses based on the China and Asia subgroups will be considered exploratory. No adjustment for multiplicity will be made; thus the hierarchical testing detailed in Section 6.4.11 to control the Type I error will not be employed directly to the China and Asia subgroups.

6.4.10 Correlations

Pearson correlation coefficients will be generated between the primary and secondary continuous endpoints (from Japan/China Approach, EU/Canada Approach, US Approach). SGRQ total score will be used in place of the SQRQ responder. The mITT population will be used.

Note that for morning trough FEV_1, AUC_{0-4} FEV_1, TDI, and change from baseline in SGRQ, the estimates over 24 weeks were obtained as LS means from MMRM analyses, and were not derived at the subject level. For the purpose of the correlation analysis, the endpoints over 24 weeks will be represented by simple averages of available data over 24 weeks.

The correlations will be organized in a matrix, with its upper triangle filled with pairwise Pearson correlation coefficients. All treatment groups will be pooled (*Table 2.61.1* to *2.61.3*).

6.4.11 Control of Type I Error

There are 3 separate plans for control of Type I error, corresponding to each of the 3 registration approaches for Japan/China, EU/Canada, and US. Each of the plans will test the primary endpoint using the efficacy estimand first for superiority comparisons, followed by the attributable estimand, which is considered a secondary endpoint. Additional secondary endpoints will only be tested using the efficacy estimand for Type I error control. Non-inferiority comparisons will be made using the per-protocol estimand. All secondary endpoints with the exception of time to onset are included in the Type I error control plans described below.

6.4.11.1 **Japan/China Approach:**

The comparisons of interest for the Japan/China approach are BGF MDI versus BFF MDI and BGF MDI versus GFF MDI, both for superiority, and BFF MDI versus Symbicort TBH for non-inferiority.

Strong control of the Type I error rate will be maintained at the 2-sided 0.05 level for the primary endpoint across key treatment comparisons using a sequential approach across comparisons, and then for the secondary measures, Type I error control will be maintained within comparison using a combination of sequential and simultaneous approaches as detailed below. Based on positive dependence of the test statistics (Sarkar, 2008; Sarkar and Chang 1997), simultaneous control of Type I error for the relevant secondary measures will be achieved using the Hochberg procedure (Hochberg, 1988).

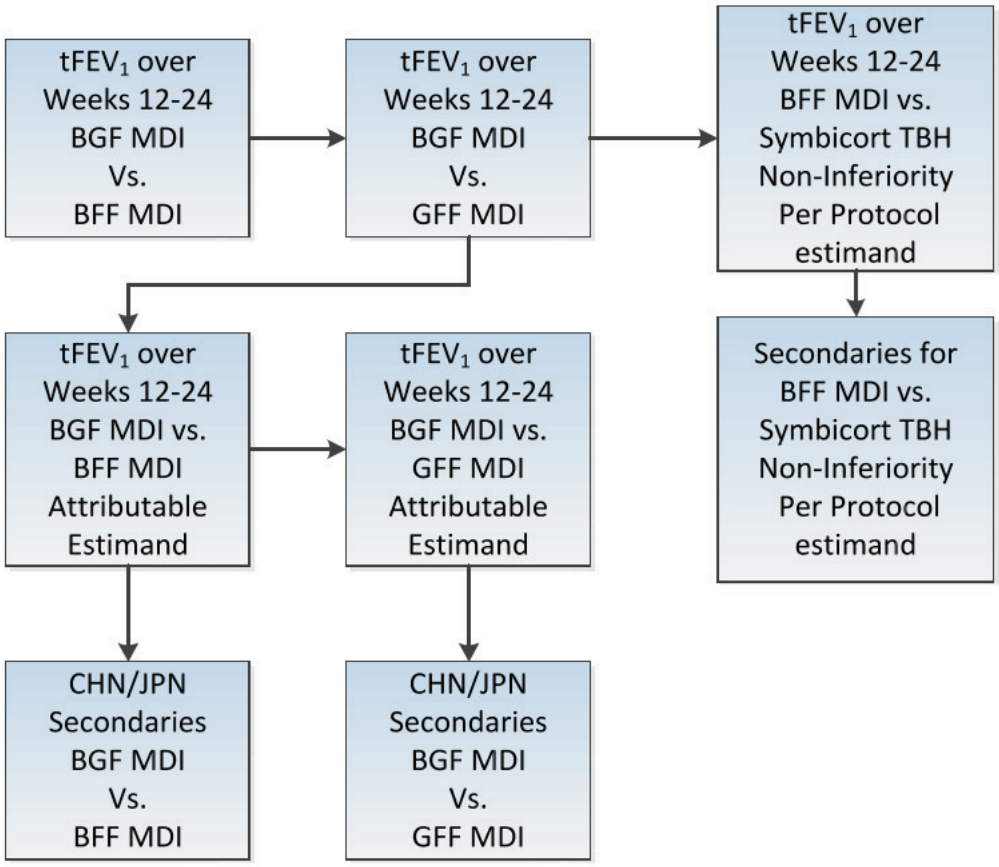
The following 3 between-treatment comparisons will be conducted, in the order stated, for morning pre-dose trough FEV_1 over Weeks 12 to 24: BGF MDI versus BFF MDI, BGF MDI versus GFF MDI, using the efficacy estimand, and BFF MDI versus Symbicort TBH (non-inferiority) using the per-protocol estimand. Each comparison will be made only if the preceding comparison in the sequence is statistically significant.

If the comparisons for BGF MDI versus BFF MDI and BGF MDI versus GFF MDI are both statistically significant at a 2-sided alpha level of 0.05, testing will proceed to comparison of the respective attributable estimands in the same order. If the comparisons of the attributable estimands are also significant, then the secondary measures within each comparison will be

tested using the efficacy estimand. Type I error will be controlled at 0.05 within each comparison (BGF MDI versus BFF MDI, BGF MDI versus GFF MDI) for the remaining secondary endpoints through simultaneous testing under the Hochberg procedure with a 2-sided alpha of 0.05.

If non-inferiority is established for the comparison of trough FEV_1 between BFF MDI and Symbicort TBH, then tests of the additional secondary measures for this comparison will be interpreted without any additional control of Type I error. Non-inferiority margins appear throughout the document in description and analyses of the endpoints as applicable.

Figure 5 Type I Error Control: Japan/China Approach



$tFEV_1$ = morning pre-dose trough FEV_1. Secondaries = secondary efficacy endpoints. Each subsequent hypothesis is tested only if statistical significance was attained in the precursor(s) hypotheses. Unless stated otherwise, all comparisons use the efficacy estimand.

6.4.11.2 EU/Canada Approach

The comparisons of interest for the EU/Canada approach are: BGF MDI versus GFF MDI, BGF MDI versus BFF MDI, and BGF MDI versus Symbicort TBH, all for superiority, and the comparison of BFF MDI vs. Symbicort for non-inferiority. All comparisons are evaluated over 24 weeks unless stated otherwise.

Strong control of the Type I error rate will be maintained at the 2-sided 0.05 level for the key comparisons using a sequential approach for the primary endpoints and then for the secondary measures Type I error control will be maintained within comparison using a combination of sequential and simultaneous approaches as detailed below.

The following 4 comparisons will be conducted first, in the order they appear below:

- FEV_1 AUC_{0-4} for BGF MDI versus BFF MDI using the efficacy estimand

- Trough FEV_1 for BGF MDI versus GFF MDI using the efficacy estimand
- FEV1 AUC_{0-4} for BGF MDI versus BFF MDI using the attributable estimand
- Trough FEV_1 for BGF MDI versus GFF MDI using the attributable estimand

All subsequent comparisons below will use only the efficacy estimand.

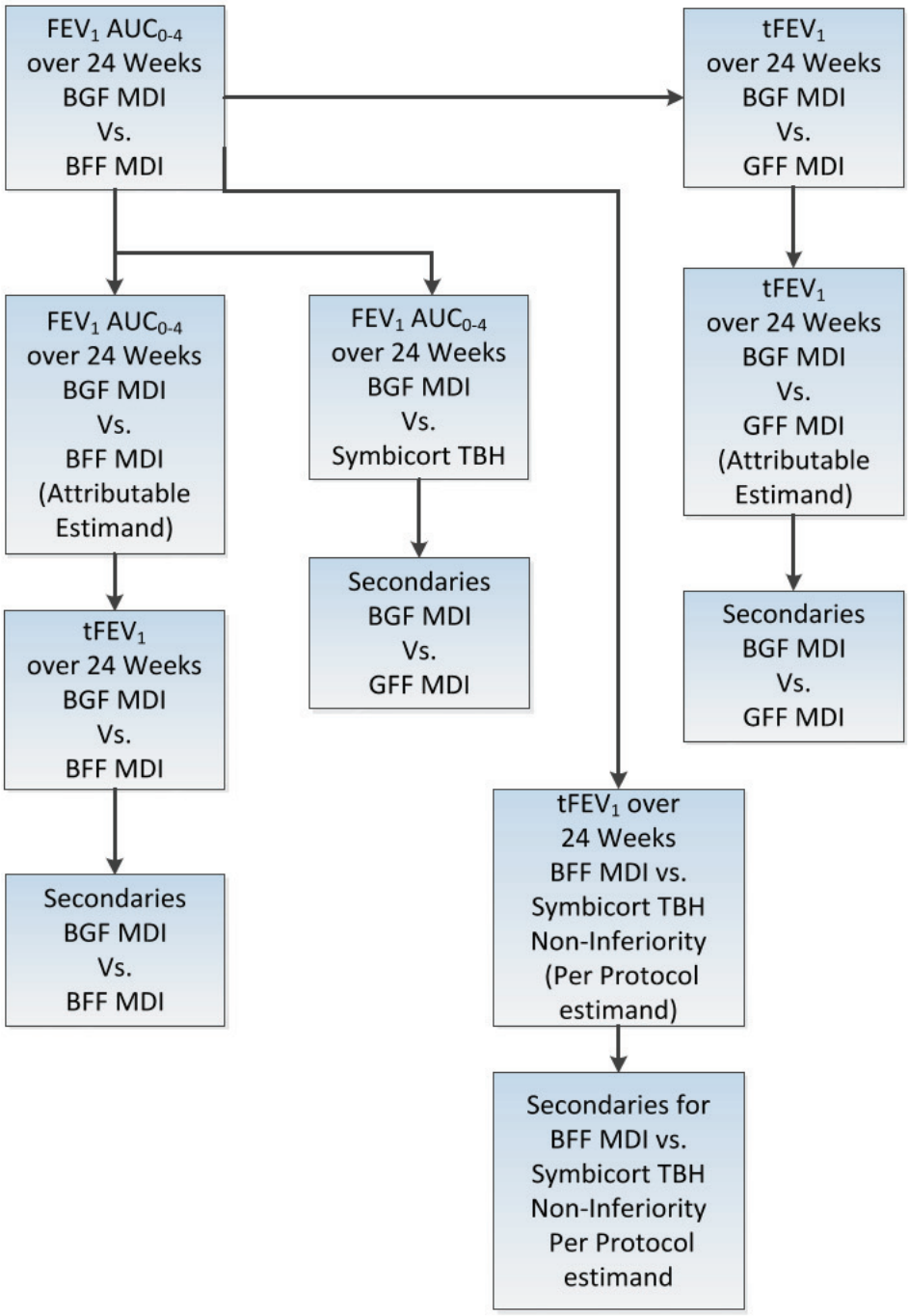
If the comparison of FEV_1 AUC_{0-4} between BGF MDI and BFF MDI using the attributable estimand above is statistically significant, testing will proceed to the secondary comparison of BGF MDI versus BFF MDI for change in morning pre-dose trough FEV_1 using a 2-sided 0.05 level test. If this test is also significant, testing will proceed to the remaining secondary endpoints. BGF MDI versus BFF MDI will be simultaneously compared among these secondary endpoints using the Hochberg procedure with a 2-sided alpha of 0.05.

If the comparison of BGF MDI versus GFF MDI for change in morning pre-dose trough FEV_1 using the attributable estimand is statistically significant, testing will proceed to the remaining secondary endpoints for BGF MDI vs. GFF MDI using the efficacy estimand. BGF MDI versus GFF MDI will be simultaneously compared among the secondary endpoints using the Hochberg procedure with a 2-sided alpha of 0.05.

If the comparison of FEV_1 AUC_{0-4} for BGF MDI versus BFF MDI is statistically significant using the attributable estimand, testing will also proceed to a comparison of BGF MDI versus Symbicort TBH for of FEV_1 AUC_{0-4} using the efficacy estimand. If statistically significant, the remaining secondary endpoints for BGF MDI versus Symbicort TBH will be simultaneously compared among the secondary endpoints using the Hochberg procedure with a 2-sided alpha of 0.05.

Finally, if the comparison of FEV_1 AUC_{0-4} over 24 weeks for BGF MDI versus BFF MDI is statistically significant, testing will proceed to the non-inferiority comparisons of BFF MDI versus Symbicort TBH. If non-inferiority is established, tests of the additional secondary measures for this comparison will be interpreted without any additional control of Type I error. Non-inferiority margins appear throughout the document in description and analyses of the endpoints as applicable.

Figure 6 Type I Error Control: EU Approach



$tFEV_1$ = morning pre-dose trough FEV_1. Secondaries = secondary efficacy endpoints. Each subsequent hypothesis is tested only if statistical significance was attained in the precursor(s) hypotheses. Unless stated otherwise, all comparisons use the efficacy estimand.

6.4.11.3 **US Approach**

The comparisons of interest for the US approach are: BGF MDI versus GFF MDI and BGF MDI versus BFF MDI, both for superiority, and at Week 24 unless stated otherwise. Statistical significance for nominal comparisons will be at alpha=0.05.

Strong control of the Type I error rate will be maintained at the 2-sided 0.05 level for the key comparisons using a sequential approach for the primary endpoints and then for the secondary measures Type I error control will be maintained within comparison using a combination of sequential and simultaneous approaches as detailed below.

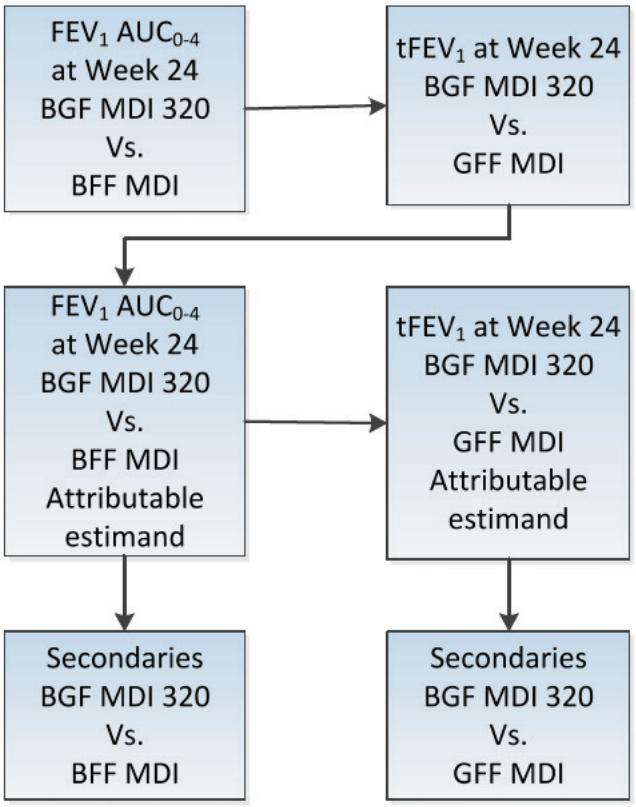
The following 4 comparisons will be conducted first, in the order they appear below:

- FEV_1 AUC_{0-4} for BGF MDI versus BFF MDI using the efficacy estimand
- Trough FEV_1 for BGF MDI versus GFF MDI using the efficacy estimand
- FEV_1 AUC_{0-4} for BGF MDI versus BFF MDI using the attributable estimand
- Trough FEV_1 for BGF MDI versus GFF MDI using the attributable estimand

If the comparison of FEV_1 AUC_{0-4} between BGF MDI versus BFF MDI using the attributable estimand is statistically significant, the remaining secondary endpoints for BGF MDI versus BFF MDI will be compared simultaneously using the efficacy estimand, and using the Hochberg procedure with a 2-sided alpha of 0.05.

If the comparison of trough FEV_1 AUC_{0-4} between BGF MDI versus GFF MDI for the attributable estimand is statistically significant, the secondary endpoints for BGF MDI versus GFF MDI will be simultaneously compared using the Hochberg procedure with a 2-sided alpha of 0.05.

Figure 7 Type I Error Control: US Approach



$tFEV_1$ = morning pre-dose trough FEV_1. Secondaries = secondary efficacy endpoints. Each subsequent hypothesis is tested only if statistical significance was attained in the precursor(s) hypotheses. Unless stated otherwise, all comparisons use the efficacy estimand.

6.5 Safety Analysis

All safety analyses for PT010006 are based on the Safety Population. Most safety analyses for PT010007 employ the Japanese Safety Population. A few analyses for PT010007 employ the PT010007 Safety Population. Hypothesis testing will not be performed for any safety analyses.

For PT010006, all AE data, clinically significant laboratory values, vital signs, and ECG values will be categorized according to their onset date into the following study periods:

- Events occurring during the Treatment Period are events with an onset date on or after the first date of dose and up to and including the date of completion of study treatment or the day after the date of premature discontinuation from study treatment. Events known to have occurred before the time of the first dose of study treatment are not included.
- Events occurring during the Post-treatment-discontinuation Follow-up are events with an onset date on [or after] the day after the dateof completion of study treatment or on or after the day after the day after (i.e. 2 days after) the date of premature discontinuation of study treatment. The exception is that deaths are still considered to be during the Treatment Period if any adverse event that led to that death is during the Treatment Period.

Any AEs, clinically significant laboratory values, vital signs, and ECG values during the Treatment Period will be tabulated and listed. Beginning on the day after the date of discontinuation from or completion of study medication, any new clinically significant ECGs, laboratory values, and vital signs will not be included in the tabulation or the computation of incidence rates, but will still be listed. Any new AEs, SAEs, and deaths during the Post-treatment-discontinuation Follow-up will be tabulated and listed. Tabulations of the incidence of deaths, AEs by system organ class (SOC), and SAEs by SOC will be provided using information collected after treatment discontinuation (as alternative tables).

6.5.1 Adverse Events

The version of the Medical Dictionary for Regulatory Activities (MedDRA) that is current at the time of database lock will be used to code verbatim terms for AEs for final analysis of the data. A glossary of MedDRA preferred terms used for adverse events reported in the study along with the associated Investigator's verbatim term will be provided in *Listing 7.2*.

An adverse event is considered on-treatment (i.e. treatment-emergent) if an event occurs after the first dose of study medication in the study, or if the AE worsened during the study after the first dose of study medication in the study (intensity and/or severity changed to a worsened grade) and the event onset is on or before the date of discontinuation from or completion of study medication. An adverse event that begins on the same date as the first dose of study medication is treatment-emergent if the AE begins after the time of first dose or if the time of AE onset is unknown. Adverse events with onset after the date of premature discontinuation from study treatment plus one day or after the date of completion of study treatment will not be considered

treatment-emergent, but will be tabulated separately (Table 3.2.1.18) and listed in adverse event data listings (Listing 7.1). . Adverse events that occur between the time the subject signs the informed consent form (ICF) for the study and the time when that subject is randomized are to be recorded as medical history unless the event met the definition of an SAE. Additionally, if an AE has an onset date on treatment and has an outcome of death, that death will be considered on treatment even if the date of death is after the last date of treatment+1.

The incidence of an AE will be defined as the number and percentage of subjects experiencing an event. Adverse events will be tabulated at the level of the MedDRA preferred term and the MedDRA SOC for the Safety Population and for the Japanese Safety Population. These will also be summarized by duration of exposure for the Japanese Safety Population for both Studies PT010006 and PT010007 (using the combined data from both studies). No hypothesis tests will be performed.

An overview table will be prepared for the Safety Population, for the Japanese Safety Population, and for the PT010007 Safety Population with the incidences of subjects with at least one TEAE, at least one serious TEAE, at least one TEAE related to study treatment, at least one serious TEAE related to study treatment, at least one TEAE leading to premature discontinuation, at least one serious TEAE leading to premature discontinuation, and a report of death (*Tables 3.1, 3.1.13,* and *3.1.14*). This overview table will also be presented by duration of exposure for the Japanese Safety Population (*Tables 3.1.15* to *3.1.18*).

The following will be done for events with irregular onset dates**.** All treatment-emergent adverse events (TEAEs) will be included in the data listings regardless of the completeness of the onset dates. Partial dates will be imputed in order to determine if an AE is treatment-emergent using the imputation rules in Appendix 1; however, imputed dates will not be provided in the data listings.

All adverse events, whether treatment-emergent or not, will be included in the listings. Reported adverse events by SOC, preferred term, treatment, country, center, subject and onset day will be provided *(Listing 7.1)*. Reported adverse events by treatment, country, center, subject, and onset date will be presented in *Listing 7.3*. SAE-specific report information will be listed in *Listing 7.7*.

The listing of adverse events will provide the severity, maximum severity, relationship to study drug, action taken and outcome for each adverse event. Adverse events leading to permanent discontinuation of study treatment will be listed for the Safety Population and for the Japanese Safety Population *(Table 3.6.1, Table 3.6.2)*. A listing of any reported deaths during the study (prior to randomization, during the Treatment Period, or during the Post- treatment-discontinuation Follow-up) will be provided (*Table 3.15.2.1*); study treatment taken prior to the death and the number of days since the last dose of this study treatment at the time of the death will be included in the listing.

Summary tabulations of the following will be prepared for all subjects in the Safety Population, for each treatment, for each primary SOC, and for each preferred term within an SOC:

1. The incidence of all TEAEs *(Tables 3.2.1.1, 3.2.1.2, Table 3.2.1.3)*
2. The incidence of all TEAEs by the duration of exposure for the Japanese Safety Population (*Tables 3.2.1.4* to *3.2.1.7*)
3. The incidence of subjects with adverse events by SOC during the Post-treatment-discontinuation Follow-up (*Tables 3.2.1.18, 3.2.1.19)*
4. The incidence of TEAEs occurring in SMQs (Standard MedDRA Queries)/groupings of interest *(Tables 3.2.3.1, 3.2.3.2)*
5. The incidence of non-serious TEAEs occurring in ≥5% of subjects in a treatment *(Tables 3.2.4.1, 3.2.4.2)*
6. The incidence of all treatment-related TEAEs *(Tables 3.4.1, 3.4.2)*
7. The incidence of discontinuation from study treatment due to a TEAEs *(Tables 3.5.1, 3.5.2)*
8. The incidence of treatment-emergent serious adverse events *(Tables 3.7.1.1.1, 3.7.1.1.2, 3.7.1.1.3)*
9. The incidence of subjects with SAEs by SOC during the Post-treatment-discontinuation Follow-up (*Tables 3.7.2.3, 3.7.2.4)*
10. The incidence of all treatment-related treatment-emergent serious adverse events *(Tables 3.9.1, 3.9.2)*
11. The incidence of all TEAEs by highest severity to treatment (*Tables 3.11.1.1* through *3.11.4.2* for the four treatments)
12. A summary tabulation will also be prepared for the incidence of TEAEs occurring in at least 2% of subjects in any treatment (*Tables 3.2.2.1, 3.2.2.2* sorted by descending frequency of events in a preferred term).
13. In addition, to control for possible differences in exposure between the treatments, the following AE and SAE summaries will be presented with the frequency and rate of occurrence (total number of events per 1000 person-years of exposure) by treatment, primary SOC, and preferred term:
    a) Frequency and rate of AEs *(Tables 3.3.1, 3.3.2)*
    b) Frequency and rate of SAEs *(Tables 3.8.3.1, 3.8.3.2)*
    c) Frequency and rate of neoplasms (*Tables 3.10.3.1 and 3.10.3.2 – All Cancer, 3.10.4.1 and 3.10.4.2* - Excluding Non-Melanoma Skin Cancer).

Safety summaries will be performed for AEs (overall summary of AEs and incidence of AEs by MedDRA SOC and preferred term in *Tables 7.1.1* to *7.2.1.17f*), SAEs (*Tables 7.3.1.2* to *7.3.1.13f*), and AEs of special interest (*Tables 7.7.1* to *7.7.3*) for China, Asia, and Japan subgroups.

In addition to AE, SAE, and AEs of special interest summaries for the Japan subgroup described above, treatment-related AEs and AEs leading to discontinuation will also be summarized for the Japan subgroup (*Tables 7.4* to *7.6*) for PT010006.

Safety summaries will be performed for AEs (overall summary of AEs and incidence of AEs by MedDRA SOC and preferred term, and SAEs) for age, sex, and race subgroups.

6.5.1.1 Adverse Events of Special Interest

AESIs have been defined based on known effects of s LAMAs, LABAs, and ICS. These include but are not limited to cardiovascular effects, ocular disorders, urinary retention, gastrointestinal disorders, and anticholinergic effects for LAMAs; cardiovascular, tremor effects, hyperglycemia, and hypokalemia for LABAs; and local (e.g., candidiasis and voice effects) and systemic (e.g., bone and skin effects, diabetes control, ocular and taste effects, adrenal suppression) steroid class effects and lung infection for ICS.

SMQs will be utilized when possible, and a selection of high-level group terms (HLGTs), high-level terms (HLTs), and preferred terms (PTs) will be utilized to represent other situations. The terms proposed to be used in the assessment of AESIs associated with ICS, LAMAs, and LABAs are listed in Table 7. Standardized MedDRA queries will be utilized when possible and a selection of preferred terms in other situations *(Appendix 5)*.

Table 7 Adverse Events of Special Interest

Medical Concept	Selection of MedDRA Terms
Adrenal suppression	Adrenal cortical hypofunctions HLT
Agitation or anxiety	Collection of PTs
Anticholinergic effects[a]	Anticholinergic syndrome SMQ Dry mouth PT
Bone fracture	Collection of HLGTs, HLTs, and PTs.
Candidiasis	Collection of PTs
Cardiovascular	Cardiac arrhythmias SMQ Cardiac failure SMQ Ischemic heart disease SMQ Torsades de Pointe/QT prolongation SMQ
Cardiovascular death	Collection of PTs
Cerebrovascular condition	CNS haemorrhages and cerebrovascular conditions SMQ
Diabetes mellitus	Hyperglycaemia/new onset diabetes mellitus SMQ
Dysgeusia or ageusia	Collection of PTs
Dysphonia or aphonia	Collection of PTs
Gastrointestinal	Gastrointestinal perforation, ulceration, haemorrhage or obstruction SMQ Gastrointestinal obstruction SMQ
Headache	Headache (PT)
Hypercortisolism	Collection of PTs
Hypertension	Blood pressure ambulatory increased (PT) Blood pressure increased (PT) Blood systolic increased (PT)

Medical Concept	Selection of MedDRA Terms
Hypokalemia	Collection of PTs
Lower respiratory tract infections other than pneumonia	Bronchitis (PT) Bronchitis viral (PT) Bronchitis bacterial (PT) Lower respiratory tract infection (PT) Lower respiratory tract infection viral (PT) Lower respiratory tract infection bacterial (PT) Infective exacerbation of chronic obstructive airway disease (PT)
Ocular effects	Visual disorders HLT Glaucoma SMQ increased intraocular pressure collection of PTs Cataract collection of PTs
Osteoporosis and osteopenia	Osteoporosis/osteopenia (SMQ)
Palpitation	Palpitations PT
Paradoxical bronchospasm	Collection of PTs
Pneumonia	Collection of PTs
Psychiatric effect	Collection of PTs
Skin effects	Skin atrophy (PT) Skin striae (PT) Acne (PT) Contusion (PT) Ecchymosis (PT) Increased tendency to bruise (PT) Petechiae (PT) Purpura (PT) Malassezia folliculitis (collection of PTs) Hypertrichosis (collection of PTs) Alopecia (collection of PTs)
Sleep effects	Initial insomnia (PT) Insomnia (PT) Sleep disorder (PT)
Sudden death	Collection of PTs
Throat irritation	Collection of PTs
Tremor	Tremor HLT
Urinary retention	Collection of PTs
Weight gain	Collection of PTs

Abbreviations: CNS=central nervous system.

[a] This medical concept is uniquely associated with LAMAs.

Appendix 5 (which will be based on the latest version of MedDRA available at the time of database lock) provides detail on selection of terms (narrow/wide designations for preferred terms are provided).

The incidence of adverse Events in MedDRA SMQs/Groupings of Interest by Term will be tabulated for the Safety Population, and the Japanese Safety Population *(Tables 3.2.3.1* and *3.2.3.2)*, as well as for the China, Asia, and Japan Subgroups of the Safety Population (*Tables 7.7.1* to *7.7.3*).

6.5.1.2 MACE Events Determined by Clinical Endpoint Committee

The clinical endpoint committee (CEC) will review and adjudicate serious cardio- and cerebrovascular (CCV) events as MACE. MACE events are defined as the following:

• Cardiovascular death

• Non-fatal myocardial infarction (MI)

• Non-fatal stroke

The CEC will review and assess these non-fatal serious CCV events and all deaths as to whether or not they fulfill criteria (based on CEC working practices) for MACE.

MACE events will be summarized by adjudicated CRF category and treatment group (*Tables 3.13.1.1, 3.13.1.2*). The assessment of MACE events will include the rate of confirmed MACE events (*Tables 3.13.2.1, 3.13.2.2*). Adjudicated MACE events will be listed in *Listing 7.4*.

The incidence of subjects with adjudicated MACE AEs by category will be summarized in *Table 3.13.3.1* and *Table 3.13.3.2*.

6.5.1.3 Pneumonia Events Determined by Adjudication Committees

All AEs/SAEs with preferred terms that could relate to pneumonia will be adjudicated to provide a more complete assessment of all physician-reported pneumonias. The incidence of confirmed pneumonia events will be tabulated (*Tables 3.13.3.1, 3.13.3.2*). The assessment of pneumonia events will include the rate of confirmed pneumonia events (*Tables 3.14.2.1, 3.14.2.2*). Adjudicated pneumonia events will be listed in *Listing 7.4*.

In order to account for specific patient risk factors, data permitting, time to first pneumonia will be compared between treatments using Cox proportional hazards (*Tables 3.14.3.1, 3.14.3.2*). Specific patient risk factors (baseline FEV_1, age, and medical history of pneumonia in the last 5 years [Yes or No]) will be evaluated for inclusion.

The incidence of subjects with adjudicated pneumonia AEs by category will be summarized in *Table 3.13.3.1* and *Table 3.13.3.2*.

6.5.1.4 Cause of Death Determined by Adjudication Committees

Causes of death will be listed by subject and summarized by treatment for (1) all-cause mortality, (2) mortality of probable cardiovascular cause, (3) mortality of probable respiratory cause and

(4) mortality of probable other causes using the Safety Population based on (A) cases reported during the active Treatment Period and (B) cases reported during the active Treatment Period plus the Post-treatment-discontinuation Follow-up *(Tables 3.15.2.1* and *3.15.2.2)*. The incidence of subjects with a death event will be tabulated by adjudicated CRF category and treatment during the Treatment Period *(Tables 3.15.1.1, 3.15.1.2)* and during the Post-treatment-discontinuation Follow-up (*Tables 3.15.1.3* and *3.15.1.4)*. To control for possible differences in exposure between treatments, the death will be summarized with frequency and rate of occurrence (total number of events per 1000 person-years of exposure) by treatment, primary SOC, and preferred term (*Tables 3.15.3.1, 3.15.3.2*). Adjudicated death events will be listed in *Listing 7.4*.

6.5.2 Clinical Laboratory Measurements

Lab parameters collected include the following:

Table 8 Lab Parameters

Hematology	
Hemoglobin	Mean corpuscular hemoglobin
Hematocrit	Mean corpuscular hemoglobin concentration
White blood cell count with differential	Mean corpuscular volume
Red blood cell count	Eosinophils
Platelet count	
Clinical Blood Chemistry	
Liver Enzyme and Other Liver Function Tests	**Other Clinical Blood Chemistry**
Alanine aminotransferase	Albumin
Aspartate aminotransferase	Blood Urea Nitrogen (BUN)
Alkaline phosphatase	Calcium[a]
Bilirubin, total	Chloride[a]
Gamma-glutamyl transferase	Cholesterol
	Bicarbonate
	Creatinine[a]
	Glucose[a]
	Magnesium
	Potassium[a]
	Phosphate
	Protein, total
	Sodium[a]
	Triglycerides
Urinalysis	
Macroscopic examination including specific gravity, pH, protein, glucose, ketones, blood, and urobilinogen.	
Other Tests:	
PT010006	
Pregnancy test (women of childbearing potential only): serum hCG at Visit 1 (Screening) and Final Visit in PT010006 (Visit 10a) or Treatment Discontinuation Visit; urine hCG at Visit 7 (Week 12)	
PT010007	
Pregnancy test (women of childbearing potential only): serum hCG at Visit 14 only and urine hCG at Visit 12 (Week 36) or Treatment Discontinuation Visit.	
Creatinine clearance will be estimated by the CKD-EPI formula [Levey, 2009].	

Abbreviations: CKD-EPI=Chronic Kidney Disease Epidemiology Collaboration; hCG=human chorionic gonadotropin

[a] Parameters included in the Basic Metabolic Panel.

Hematology

A Clinically Significant Laboratory Abnormality as identified by the investigator after the start of study treatment will be recorded as an Adverse Event and tabulated as an AE in the AE analysis. Abnormalities occurring prior to the start of treatment will be noted in medical history and presented in a data listing. Per protocol, the criteria for a "clinically significant" laboratory abnormality are:

a. A laboratory abnormality that leads to a dose-limiting toxicity (e.g., an abnormality that results in study drug dose reduction, suspension or discontinuation)
b. A laboratory abnormality that results in any therapeutic intervention (i.e., concomitant medication or therapy)
c. Other laboratory abnormality judged by the Investigator to be of any particular clinical concern (e.g., significant fall in hemoglobin not requiring transfusion)

All laboratory data will be stored in the database with the units in which they were originally reported. Laboratory data not reported in International System of Units (SI units; Système International d'Unités) will be converted to SI units before data analysis.

Individual clinical laboratory variables for hematology and clinical chemistry and kidney function, including creatinine clearance, will be provided in listings *(Listing 8.1* for hematology, *Listing 8.2* for blood chemistry and kidney function*, Listing 8.3* for urinalysis, and *Listing 4.6 for* pregnancy test results at screening and after the start of treatment). Data will be listed in SI units where available. Comments for laboratory testing will be listed *(Listing 8.4)*. For listings, laboratory values will be flagged as Low or High based on the reference ranges provided by the central laboratory, LabCorp Laboratories *(Appendix 4)*.

The baseline measurement for a laboratory parameter will be the last available measurement prior to the start of dosing.

Table 9 Analysis Study Time Window for Clinical Lab Assessments

Calculated Study Time Window	Time Interval for the Study Time Window
Pre-dose 1 hr.	≥0 min. prior to dose
Post-dose 30 min.	>0 to <75 min. post-dose
Post-dose 2 hrs.	≥75 min. to <4 hrs. post-dose

Note: The minutes are rounded to the nearest whole number before applying time windows.

The laboratory-value windows will be applied only for calcium, chloride, glucose, potassium, and sodium (i.e. the laboratory parameters that are sometimes assessed post-dose) and these windows will be applied only at the following visits: Visit 4 (Day 1), Visit 5 (Week 4), and Visit

10a (Week 24). For other laboratory parameters and other visits, windows will not be applied. The rationale is that for other laboratory parameters and other visits, post-dose assessments are not to be assessed.

If there are multiple laboratory values for the same parameter at pre-dose of a visit or within the same post-dose study time window (if applicable) at a visit, the last value will be chosen for analysis.

Summary statistics (n, mean, median, standard deviation, minimum, and maximum) for the baseline assessment and for the pre-dose value and change from baseline at each post-baseline visit and end of treatment for scheduled lab assessments of continuous laboratory variables including serum potassium and glucose will be tabulated. “End of Treatment” is defined as the last non-missing assessment during the treatment period. Data from unscheduled visits will not be used for the by-visit summaries but both scheduled-visit values and unscheduled-visit values are candidates for the end-of-treatment summary. Data from both scheduled and unscheduled visits will be listed. The summaries will be provided by treatment (*Tables 3.16.1.1* through *3.16.4.2,* for hematology, blood chemistry, kidney function, and urinalysis, respectively).

Data from unscheduled visits will not be used for the by-visit summaries but both scheduled-visit data and unscheduled-visit data are candidates for clinically significant values, for the end-of-treatment summary, and for shift tables. Shift tables will be produced using the categories defined by the Common Terminology Criteria for Adverse Events (CTCAE) Version 4.03 grades for the Safety Population (*Tables 3.16.5.1, 3.16.6.1, and Table 3.16.7.1*) for hematology, chemistry, kidney function and urinalysis, respectively) and the Japanese Safety Population (*Tables 3.16.5.2, Table 3.16.6.2, Table 3.16.7.2* and *Table 3.16.8.2* for hematology, chemistry, kidney function and urinalysis, respectively). For these shift tables, for each treatment, the subject’s pre-dose grade will be cross-tabulated by the subject’s maximum post-baseline grade during the treatment; also, the subject’s maximum post-baseline grade during treatment will be tabulated for all baseline grades combined. Percentages of subjects in each maximum post-baseline grade for a treatment will be calculated for each pre-dose grade for the treatment and also for all baseline grades combined. Laboratory abnormal values on-treatment will be flagged as High or Low values based on laboratory reference ranges provided by LabCorp Laboratories (found in Appendix 4) as per Pearl, Inc. These flags along with the reference ranges will be provided in the laboratory data listings.

Potentially Clinically Significant Laboratory Values Above/Below a Clinically Relevant Threshold on-treatment, based on CTCAE 4.03 and other criteria, will be identified based on the following thresholds:

Table 10 Potentially Clinically Significant (PCS) Laboratory Parameter Criteria

Parameter	Post-Baseline Criteria
Hematology	
Hemoglobin	<8.0 g/dL (<80 g/L)
	Increase of >40 g/L to a value above the ULN (upper limit of normal)
White Blood Cell Count	<2000/μL
	>35,000/μL
Platelet Count	<50,000/μL
	>999,000/μL
Chemistry	
eGFR-EPI (where eGFR denotes estimated glomerular filtration rate)	<30 mL/min/1.73 m^2
AST (aspartate aminotransferase)	>3 x ULN
ALT (alanine aminotransferase)	>3 x ULN
Alkaline Phosphatase	>5 x ULN
Total Bilirubin	>2 x ULN
Blood Glucose* (random values)	<2.2 mmol/L (<39.6 mg/dL)
	>13.9 mmol/L (>250 mg/dL) if baseline is below or equal to 10.0 mmol/L (180 mg/dL), >16.7 mmol/L (>300 mg/dL) if baseline is greater than 10.0 mmol/L (180 mg/dL).
Serum Potassium	<3.0 mmol/L
	>6.0 mmol/L

*CTCAE 4.03 criteria are based on fasting glucose values. However, subjects were not required to fast prior to obtaining blood glucose values.

Clinically significant laboratory values will be tabulated for the Safety Population (*Table 3.16.9*). Since a reduction in potassium and an increase in blood glucose are known class effects of beta-agonists, all potassium or glucose assessments for subjects who experienced newly occurring or worsening potentially clinically significant values after start of the study treatment will be provided in separate listings (*Tables 3.16.10.1, 3.16.10.2, 3.16.11.1,* and *3.16.11.2*). For all laboratory parameters other than glucose and potassium noted in Table 8, all laboratory data for the parameter identified as potentially clinically significant for a subject will be listed (*Table 3.16.12.1*- Safety Population and 3.16.12.2- Japanese Safety Population).

6.5.3 Vital Signs

Changes from Baseline in on-treatment supine or seated systolic blood pressure, supine or seated diastolic blood pressure, and heart rate will be evaluated, where baseline is defined as the mean of all available pre-dose measurements taken prior to the start of dosing at the Randomization Visit (Visit 4). If there are no Visit 4 pre-dose values, the baseline will be defined

as the mean of pre-bronchodilator values at Visit 2 and Visit 3. No Hypothesis testing will be performed.

A **Clinically Significant Abnormality** in vital signs identified by the investigator will be recorded as an Adverse Event if it occurs after the start of treatment. These adverse events will be included in the AE summaries; abnormalities prior to the start of treatment will be noted in medical history and listed.

Potentially clinically significant changes in systolic and diastolic blood pressure will be defined based on the following criteria provided by Pearl, Inc.:

Table 11 Potentially Clinically Significant Criteria for Systolic and Diastolic Blood Pressure Parameters

Parameter (mmHg)	Post-Baseline Criteria
Systolic Blood Pressure, increase	≥180 and increase from baseline ≥20
Systolic Blood Pressure, decrease	≤90 and decrease from baseline ≥20
Diastolic Blood Pressure, increase	≥105 and increase from baseline ≥15
Diastolic Blood Pressure, decrease	≤50 and decrease from baseline ≥15

mmHG = millimeter of mercury.

Potentially clinically significant changes in heart rate will be assessed as follows:

Table 12 Potentially Clinically Significant Criteria for Heart Rate Parameters

Parameter	Post-Baseline Criteria
Tachycardia Event	≥110 bpm and increase ≥15% from baseline
Bradycardia Event	≤50 bpm and decrease ≥15% from baseline

bpm = beats per minute.

Vital sign measurements (Heart rate, systolic blood pressure, diastolic blood pressure and body temperature, weight, height) during the study will be displayed in a vital signs listing *(Listing 9.1)*.

A summary of baseline weight, height, and BMI will be presented by treatment (*Tables 1.4.1.1, 1.4.3.1, 1.4.4.1 and 1.4.5.1* for the mITT, PP, and Safety Populations, and all subjects not randomized respectively). The ITT Population does not need to be summarized because it is the same as the mITT Population at baseline. The Safety Population may not be needed either.

Summary statistics (n, mean, median, standard deviation and range) of the absolute value and change from baseline for systolic blood pressure, diastolic blood pressure, and heart rate, will be tabulated by treatment, visit, and time point. Baseline will be defined as the mean of the values prior to dosing at Visit 4 (Day 1). These summaries (*Table 3.17.1.1, Table 3.17.1.2*) will be

prepared for baseline and each scheduled post-baseline nominal time point at each scheduled post-baseline visit and end of treatment. "End of Treatment" will be summarized for each scheduled post-baseline time point (pre-dose 1 hr, and post-dose 30 minutes and 2 hour). "End of Treatment" for each of these assessment points is defined as the last non-missing on-treatment assessment available for the time point. Data from unscheduled visits will not be used for the by-visit summaries, but both scheduled-visit data and unscheduled-visit data are candidates for the end-of-treatment summary. Data from both scheduled and unscheduled visits will be listed. Time windows will be derived for each post-baseline visit using the time intervals for the study time windows detailed in Table 13. No hypothesis tests will be performed.

Table 13 Analysis Study Time Windows for Vital Signs Assessments

Calculated Study Time Window	Time Interval for the Study Time Window
Pre-dose	≥0 min. prior to dose
Post-dose 30 min.	>0 to <75 min. post-dose
Post-dose 2 hr	≥ 75 min. to < 4 hrs. post-dose
Post-dose 12 hrs.	≥8 hrs. to <16 hrs. post-dose

Note that minutes are rounded to the nearest whole number before applying time windows.

If there are multiple vital sign values for the same parameter at pre-dose assessments after Visit 4 or within the same post-dose study time window at a visit, the last value will be chosen for analysis.

Data from unscheduled visits will not be used for the by-visit summaries but both scheduled-visit data and unscheduled-visit data are candidates for clinically significant values and for the end-of-treatment summary.

The percentage of subjects with potentially clinically significant values for vital signs at any time post-dose at a visit will be summarized by treatment based on the criteria in Table 11 and Table 12 *(Table 3.17.2.1- Safety Population, Table 3.17.2.2- Japanese Safety Population)*.All vital sign assessments for subjects with potentially clinically significant values will be listed *(Tables 3.17.3.1, 3.17.3.2, 3.17.4.1, Table 3.17.4.2)*.

6.5.4 12-Lead Electrocardiogram Measurements

Changes from baseline in Heart Rate, PR Interval, QRS Axis, QRS Interval, QT Interval and QTcF interval will be calculated where baseline is defined as the mean of the pre-dose measurements taken prior to the start of treatment at the randomization visit (Visit 4). If there are no Visit 4 pre-dose values, the baseline will be defined as the mean of the pre-bronchodilator values at Visit 2 and Visit 3. The QTcF (Fridericia Corrected QT) is defined as $[QT/(RR^{1/3})]$. Heart rate (bpm) is estimated as 60,000/RR, where RR is in units of ms. These assessments will be tabulated for each treatment and assessment time.

A **Clinically Significant Abnormality** for a 12-Lead ECG measurement identified by the investigator as a clinically significant abnormality will be recorded as an Adverse Event if it occurred after the start of study treatment. These adverse events will be included in the AE summaries.

All 12-Lead ECG measurements for the Safety Population will be listed *(Listing 9.2).* Summary statistics (mean, median, standard deviation and range) for raw values and change from baseline values in Heart Rate, PR Interval, QRS Axis, QRS Interval, QT Interval and QTcF interval will be calculated, where baseline is defined as the mean of the pre-dose measurements taken prior to the start of treatment at Visit 4 (Day 1). These assessments will be tabulated for each treatment and each scheduled nominal time point (derived using the time intervals for the study time windows detailed below in Table 14) at each visit and at end of treatment (*Table 3.18.1, Table 3.18.2*). "End of Treatment" will be summarized for each scheduled post-baseline time point (pre-dose 1 hour, post-dose 30 minutes, and post-dose 2 hours). End of Treatment for each of these assessment points is defined as the last non-missing on-treatment assessment available for the time point. Data from unscheduled visits will not be used for the by-visit summaries but both scheduled-visit data and unscheduled-visit data are candidates for the end-of-treatment summary. Data from both scheduled and unscheduled visits will be listed. Mean pre-dose change from baseline for heart rate and QTcF will be plotted across post-baseline visits by treatment (*Figure 3.18.1a* and *Figure 3.18.1e*). ECG data from subjects with pacemakers will not be included in analyses, but will be listed.

Table 14 Analysis Study Time Window for ECG Assessments

Calculated Study Time Window	Time Interval for the Study Time Window
Pre-dose 1 hr.	≥0 min. prior to dose
Post-dose 30 min.	>0 to <75 min. post-dose
Post-dose 2 hrs.	≥75 min. to <4 hrs. post-dose
Post-dose 12 hrs.	≥8 hrs. to <16 hrs. post-dose

Note: The minutes are rounded to the nearest whole number before applying time windows.

If there are multiple ECG values for the same parameter at pre-dose of a visit date (other than for Visit 4) or within the same post-dose study time window on a visit date, the last value will be chosen for analysis.

Data from unscheduled visits will not be used for the by-visit summaries but both scheduled-visit data and unscheduled-visit data are candidates for clinically significant values and for the end-of-treatment summary.

Table 15 Criteria for PCS ECG Values

Parameter	Post-Baseline Criteria
QTcF Prolongation	(1) ≥ 500 msec if < 500 msec at study baseline and ≥ 15 msec change from study baseline
	(2) ≥ 530 msec if ≥ 500 msec at study baseline and ≥ 15 msec change from study baseline
	(3) ≥ 500 msec and ≥ 15 msec change from study baseline
	(4) Change of ≥60 msec from study baseline regardless of initial value

msec = millisecond

Potentially clinically significant ECG parameter values will be identified based on criteria listed in Table 15. The number and percentage of subjects who had such values observed any time post-dose will be tabulated for each treatment (*Table 3.18.3, 3.18.4*) and listed (*Tables 3.18.5, 3.18.6* for QTcF prolongation). No hypothesis tests will be performed.

6.5.5 Healthcare Resource Utilization

Data on HCRU will be collected at all visits post-baseline and summarized by treatment group.

The following variables will be calculated unadjusted (per subject) over the entire Treatment Period and tabulated by actual treatment received for those subjects for whom they or one or more of their family members missed work:

- The number of days missed work due to COPD.
- The number of days that caregivers of subjects missed from work as a result of the subject's COPD.

The following variables will be tabulated by actual treatment received and relationship to COPD (COPD-related, not COPD-related, and combined). The mean and the mean per person-year will be calculated across all subjects in a treatment.

- The percentage of subjects with telephone calls to health-care providers:
    - Calls to any health-care provider (physician or other)
    - Calls to physician
    - Calls to other healthcare provider
- The mean number of telephone calls to health-care providers:
    - Calls to any health-care provider (physician or other)
    - Calls to physician
    - Calls to other healthcare provider

- The percentage of subjects with visits to health-care providers:
    - Visits to any health-care provider (GP, specialist, or other)
    - Visits to GP
    - Visits to specialist
    - Visits to other health-care provider
- The mean number of visits to health-care providers:
    - Visits to any health-care provider (GP, specialist, or other)
    - Visits to GP
    - Visits to specialist
    - Visits to other health-care provider
- Ambulance Transport
    - The percentage of subjects who required ambulance transport
    - The mean number of times ambulance transport was required
- ER Visits
    - The percentage of subjects with ER visits
    - The mean number of visits to ERs
- Hospitalizations
    - The percentage of subjects hospitalized
    - The mean number of subject hospitalizations
    - The mean number of days in the hospital
- Hospitalizations with some time spent in the ICU or CCU
    - The percentage of subjects hospitalized with some time spent in the ICU or CCU
    - The mean number of subject hospitalizations with some time spent in the ICU or CCU
    - The mean number of days in the hospital with some time spent in the ICU or CCU
- Hospitalizations with No time spent in the ICU or CCU
    - The percentage of subjects hospitalized with No time spent in the ICU or CCU
    - The mean number of subject hospitalizations with No time spent in the ICU or CCU
    - The mean number of days in the hospital with No time spent in the ICU or CCU
- ICU
    - The percentage of subjects in the ICU
    - The mean number of days in ICUs
- CCU
    - The percentage of subjects in the CCU
    - The percentage of subjects who required ambulance transport

Analyses will be performed using the mITT Population.

Descriptive statistics (n, mean, standard deviation, median, minimum and maximum) will be provided by actual treatment received for the number of days missed from work, the number of days that family members of subjects missed from work per year overall during the study *(Table 3.20.1* and *Listing 9.4)*.

Also, descriptive statistics will be provided by actual treatment received and relationship to COPD (related, not-related, and total) overall during the entire Treatment Period for the following variables: the number of telephone calls to health-care providers, the number of visits to health-care providers, the number of ER visits, the number of number of times ambulance transport was required, the number of subject hospitalizations, the number of days in the hospital, the number of days in the ICU, and the number of days in the CCU (*Table 3.20.2* and *Listings 9.4* and *9.5*).

6.5.6 Pharmacokinetic Analysis

Blood samples for PK assessments will be taken prior to administration of study drug at Week 24 and at 2, 5, and 20 minutes, and at 1, 2, 4, 8, 10, and 12 hours post-dose. Actual sampling time points relative to dosing will be used for PK assessments and analysis where available. If the actual sampling time is unknown, the scheduled protocol time may be used for the calculation of derived PK parameters.

The concentration-time data reported by the bioanalytical laboratory will be evaluated for inclusion in the PK analysis dataset.

The PK analysis will be performed for subjects in the PK Population.

PK parameters will be estimated by non-compartmental analysis (NCA) using the software Phoenix® WinNonlin® (Pharsight Corporation, US). From the plasma budesonide, glycopyrronium and formoterol concentration-time data, the following PK parameters will be estimated for each subject where possible:

AUC_{0-12}	The area under the plasma concentration-time curve from time 0 to 12 hours post dose
C_{max}	The maximum observed plasma concentration, expressed in concentration units
t_{max}	The time to reach C_{max}, expressed in hours
λ_z	The terminal elimination rate constant, calculated from the slope of the terminal portion of the ln(drug concentration) versus time curve
$t_{1/2}$	The apparent terminal elimination half-life, expressed in hours, calculated as $\ln 2/ \lambda_z$
C_{min}	The minimum observed plasma concentration, expressed in concentration units
C_{avg}	Average concentration during a dosing interval
Fluctuation	Degree of fluctuation $[(C_{max}-C_{min})/C_{avg}]$
Swing	$[(C_{max}-C_{min})/C_{min}]$

AUC_{0-12} will be calculated using the linear-log trapezoidal method.

The PK parameters C_{max}, C_{min} and time to C_{max} (t_{max}) will be obtained from the observed values.

A 12-hour post-dose sampling schedule may not permit adequate estimation of λ_z considering the known PK of the products. However, λ_z will be estimated for each subject where feasible by linear regression analysis, calculated from the slope of the terminal portion of the ln(drug concentration) versus time curve. Selection of data points to include in the estimation of λ_z for each subject for each treatment for each analyte will be based on the following criteria:

- All samples used should preferably fall in the log-linear elimination phase.
- At least 3 samples above the lower limit of quantification (LLQ) should be used in the estimation.
- C_{max} must not be used in the estimation.

In order for the selection to take place the adjusted r^2 value reported in Phoenix® WinNonlin® must be above 0.7.

Fluctuation, swing and $t_{1/2}$ will be calculated using the formulas listed above.

For the purposes of parameter estimation, plasma concentration values below the LLQ will be set to missing in the NCA. Missing values (e.g., no blood sample collected, no value obtained at analysis) will be treated as missing and excluded from the NCA. If there are ≥2 consecutive missing concentration values, the estimation of PK parameters will be evaluated on a case-by-case basis.

All concentration-time data reported by the bioanalytical laboratory, for each analyte, for each treatment, will be listed for subjects participating in the PK sub-study. Actual sample collection times will be detailed in the listing along with the scheduled nominal sample collection times. In

addition, all calculated PK parameters for each treatment for each analyte will be listed for subjects participating in the PK sub-study.

Descriptive statistics for plasma concentrations of budesonide, formoterol, and glycopyrronium, by treatment, visit and time point will be summarized. Descriptive statistics will include the number of observations (n), mean (CV%), SD, standard error (SE), median, minimum (min), maximum (max), geometric mean, and geometric coefficient of variation (*Tables 4.1 – 4.3*).

Descriptive statistics for PK parameters of budesonide, glycopyrronium, and formoterol, will be summarized by treatment and visit. Descriptive statistics will include the number of observations (n), mean (CV %), SD, median, minimum, maximum, geometric mean, and geometric coefficient of variation. For the PK parameter t_{max}, only the number of observations (n), mean, median, minimum (min), and maximum (max) will be presented (*Tables 4.4-4.6*).

The plasma concentration-time profiles for individual and mean plasma concentrations of budesonide, formoterol, and glycopyrronium, will be presented for each treatment on the linear/linear scale and on the linear/log-linear scale. Mean and individual plots will be separate for each analyte. Nominal sampling time points relative to dosing will be used for all mean plots. Actual sampling time points will be used for all individual plots (*Figures 4.1.1* through *4.6.2*).

Non-compartmental parameter estimates for budesonide, formoterol, and glycopyrronium AUC_{0-12} and C_{max} will be natural-log transformed and analyzed using mixed effect models (*Tables and Figures 4.7 – 4.9*).

The relative bioavailability of budesonide, formoterol, and glycopyrronium delivered via BGF MDI to delivery via GFF MDI, BFF MDI, or Symbicort TBH will be evaluated and summarized. Separate analysis of variance (ANOVA) models with fixed effects for treatment, will be fit for each analyte (*Tables 4.7 – 4.9*).

6.5.7 HPA Axis Analysis

Twenty-four-hour SC levels will be obtained in a subset of PK sub-study subjects over 24 hours between Visits 3 and 4 prior to dosing at Randomization and Visit 10a (Week 24).

Cortisol sampling will occur 30 minutes pre-dose, and 1, 2, 4, 8, 10, 12, 14, 16, 20, 22, and 24 hours post-dose. All SC concentrations reported will be listed.

The data reported by the bioanalytical laboratory will be evaluated for inclusion in the HPA Axis analysis dataset.

The HPA Axis analysis will be performed for subjects in the HPA Axis Population.

The following parameters will be estimated for each subject where possible:

- 0-24 hour weighted mean SC concentration

The 24-hour SC weighted mean concentration (AUC/t) will be obtained by dividing the AUC by the sample collection time interval within 24 hours; t is the difference in hours between last non-missing time minus first non-missing time included in the AUC. The AUC will be calculated using the trapezoidal rule from the first non-missing time point to the last non-missing time point.

Baseline is defined as the weighted mean 0-24 h from Day -1/1. If only one time point is available at either baseline or at Week 24, time-weighted average will use the single concentration value at that time point; in this case, baseline time-weighted average is just the single concentration value collected at baseline.

Concentrations below the assay's LLQ will be imputed with a value of LLQ/2 when deriving the weighted mean and AUC. If an observation is missing between 2 non-missing observations, the AUC will be calculated using the reported values at the adjacent time points.

The primary analysis will be performed on the log-transformed ratio from baseline of the weighted mean SC for subjects in the HPA Axis Population. An ANCOVA model with baseline (log-transformed) value, gender, age, and treatment group as covariates will be used to draw comparisons between BGF MDI versus GFF MDI, BFF MDI versus GFF MDI, and Symbicort TBH versus GFF MDI. Estimated ratios and their 90% confidence intervals will be presented (*Table 5.1*).

SC concentration values will be listed (*Listing 11.1*). Baseline and Week 24 weighted mean SC concentration will be listed for each subject (*Listing 11.2*).

The ratio from baseline to Week 24 (Visit 10a) in the 0-24 hour weighted mean SC concentration at will be summarized by treatment group (*Table 5.1*).

Descriptive statistics (n, mean, SD, median, minimum, maximum, geometric mean, and geometric coefficient of variation) for SC concentration by time point (for both Screening and Week 24) will be tabulated (*Table 5.2*). SC concentration profiles over 24 hours will be graphically presented by treatment group using box plots over time, at baseline and at Week 24 (*Figures 5.2.1 – 5.2.4).*

The percentage of subjects observing an at least 30% increase from baseline in 0- to 24-hour weighted mean SC concentration will be summarized by treatment group. The percentage of subjects observing an at least 30% decrease from baseline in 0- to 24-hour weighted mean SC concentration will also be summarized by treatment group.

6.5.8 Physical Examination

Any physical examination abnormality reported after the start of treatment for a subject is to be reported as an adverse event. Thus, these will be included in listings of adverse events and summarized in adverse event summaries. Abnormalities seen at the Screening physical examinations will be recorded as Medical History and listed.

7. CHANGES FROM METHODS PLANNED IN THE PROTOCOL

Any further refinements to methods planned in this SAP will be identified in the BDRM minutes.

8. STATISTICAL SOFTWARE

Data processing, statistical screening, descriptive reporting and analysis of the efficacy and safety data will be performed using SAS (Version 9.2 or higher). Graphs may also be produced using R (R Development Core Team, 2003).

9. REFERENCES

Anzures-Cabrera J, Higgens PT. Graphical displays for meta-analysis: An overview with suggestions for practice. Research Synthesis Methods, Published online on 23 Mar 2010 (Wiley Interscience).

Carroll KJ, Fleming TR. Statistical Evaluation and Analysis of Regional Interactions: The PLATO Trial Case Study. *Statistics in Biopharmaceutical Research*, 2013*;* 5:2, 91-101. DOI: 10.1080/19466315.2013.783878
COPD Assessment Test (CAT) (http://www.catestonline.org/)

Cro, S (2017) Relevant Accessible Sensitivity Analysis for Clinical Trials with Missing Data. PhD thesis, London School of Hygiene & Tropical Medicine. DOI: 10.17037/PUBS.03817571 (http://researchonline.lshtm.ac.uk/3817571/1/2017_EPH_PhD_CRO_S.pdf).

Devlin N, Shah K, Feng Y, Mulhern B, van Hout B. (2016), Valuing Health-Related Quality of Life: An EQ-5D-5L Value Set for England. Office of Health Economics, Research Paper 16/01.

EuroQol Group. EQ-5D. Accessed at www.euroqol.org/about-eq-5d/how-to-use-eq-5d.html (November, 2014).

Farrar JT, Dworkin RH, Max MB. Use of the cumulative proportion of responders analysis graph to present pain data over a range of cut-off points: Making clinical trial data more understandable. J Pain Symptom Manage 2006; 31:369-377

Hochberg Y, A sharper Bonferroni procedure for multiple tests of significance. Biometrika 1988; 75, 4, pp. 800-2.

Japanese Respiratory Society (JRS). Guidelines for the Diagnosis and Treatment of COPD, 4th edition, 2013.

Jennison C, Turnbull BW. Group Sequential Methods: Applications to Clinical Trials. Chapman and Hall/CRC: Florida, 2000.

Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. Ann Intern Med. 2009;150(9):604-612.

Little RJA, Rubin DB. Statistical analysis with missing data, (2nd edn.). John Wiley and Sons, Inc.: New York, 2002.

Molenberghs G, Kenward MG. Missing data in clinical studies. John Wiley and Sons, Inc.: New York, 2007.

National Center for Health Statistics. 1996. NHANES III Reference Manuals and Reports. Data Dissemination Branch, Hyattsville, MD. CD-ROM No. 6-0178 (1096).

National Research Council. Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. The prevention and treatment of missing data in clinical trials. The National Academies Press: Washington, DC, 2010.

R Development Core Team. R: A language and environment for statistical computing, 2003. R Foundation for Statistical Computing, Vienna, Austria.

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

Rubin DB. M. Multiple imputation for nonresponse in surveys. John Wiley and Sons, Inc.: New York, 1987.

SAS Institute Inc. 2008. SAS/STAT 9.2 User's Guide. Cary, NC: SAS Institute Inc.

Seaman SR, White IR, Leacy FP. Comment on "Analysis of Longitudinal Trials With Protocol Deviations: A Frameword for Relevant, Accessible Assumptions, and Inference via Multiple Imputation," by Carpenter, Roger, and Kenward. Journal of Biopharmaceutical Statistics, 2014; 24(6):1358-1362.

Van Hout B, Janssen MF, Feng YS, 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 15, 2012 (708-715).

van Reenan M, Janssen B. EuroQol Group EQ-5D-5L User Guide: Basic information on how to use the EQ-5D-5L instrument. https://euroqol.org/wp-content/uploads/2016/09/EQ-5D-5L_UserGuide_2015.pdf (April 2015)

White IR, Royston P, and Wood AM. Multiple imputation using changed equations: issues and guidance for practice. Statistics in Medicine, 2011; 30: 377-399.