AstraZeneca

Statistical Analysis Plan

Study Code	PT009002
NCT #	NCT02766608
EudraCT	2016-000154-34

A Randomized, Double-Blind, Parallel Group, Multi-Center Study to Assess the Efficacy and Safety of PT009 Compared to PT005, PT008, and Open-label Symbicort® Turbuhaler®, as an Active Control, on Lung Function over a 24-Week Treatment Period in Subjects With Moderate to Very Severe COPD

STATISTICAL ANALYSIS PLAN FOR STUDY PT009002

Protocol Number:	PT009002-02 (Version 3.0)
Investigational Drug and Drug Number:	BFF MDI; PT009 FF MDI; PT005 BD MDI; PT008 Symbicort® Turbuhaler®
Indication:	COPD
Dosage Form/Dose:	• BFF MDI 320/9.6 µg BID • BFF MDI 160/9.6 µg BID • FF MDI 9.6 µg BID • BD MDI 320 µg BID • Symbicort® Turbuhaler® 400/12 µg BID

PT009002 Protocol Title: A Randomized, Double-Blind, Parallel Group, Multi-Center Study to Assess the Efficacy and Safety of PT009 Compared to PT005, PT008, and Open-label Symbicort® Turbuhaler®, as an Active Control, on Lung Function over a 24-Week Treatment Period in Subjects with Moderate to Very Severe COPD

Date of Issue: [redacted]

Version: Version 1.1

Signed Agreement on Statistical Analysis Plan

FINAL SIGN-OFF SIGNATURES

Primary Biostatistician and Peer Reviewer: [redacted]

Signature Date

Author: [redacted]

Approved by: [redacted]

Signature Date

Approved by: [redacted]

Signature Date

Approved by: [redacted]

Signature Date

Approved by: [redacted]

Signature

Change Log			
Version No.	**Effective Date**	**Reason for the Change / Revision**	**Supersedes**
1.1	[redacted]	Figures 7 and 8 (Type I error control charts for EU approach) were fixed to reflect that the non-inferiority testing is performed using the Per Protocol Estimand.	1.0

TABLE OF CONTENTS

List of Tables

List of Figures

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

AE	Adverse event
AESI	Adverse event of special interest
ALT	Alanine aminotransferase
ANCOVA	Analysis of covariance
AR(1)	Autoregressive order 1
AST	Aspartate aminotransferase
ATS	American Thoracic Society
AUC	Area under the curve
AUC_{0-4}	Area Under the Curve From 0 to 4 Hours
AUC_{0-12}	Area Under the Curve From 0 to 12 Hours
BD	Budesonide
BDI	Baseline Dyspnea Index
BFF MDI	Budesonide 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 Equation (according to National Kidney Disease Education Program)
cm	Centimeter
COPD	Chronic obstructive pulmonary disease
CRF	Case Report Form

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 1 second
FEF_{25-75}	Forced expiratory flow between 25% and 75% of FVC
FF	Formoterol Fumarate
FVC	Forced vital capacity
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
ICF	Informed consent form

ICS	Inhaled corticosteroid
ICU	Intensive care unit
ID	Identification
i.e.	Id est; that is
ITT	Intent-to-treat
IWRS	Interactive web response system
L	Liter
LABA	Long-acting β_2 agonist
LAMA	Long-acting muscarinic antagonist
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
PMM	Pattern mixture model
PP	Per-protocol
PT	Preferred Term
PT005	Formoterol Fumarate Inhalation Aerosol
PT008	Budesonide Inhalation Aerosol
PT009	Budesonide 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
SAP	Statistical analysis plan
SD	Standard deviation
SGRQ	St. George's Respiratory Questionnaire
SMQ	Standard MedDRA Query
TBH	Turbuhaler
TC	Telephone call
TDI	Transition Dyspnea Index
TEAE	Treatment-emergent adverse event
ULN	Upper limit of normal
US	United States
VAS	Visual analog scale

Trademark Information

Trademarks Not Owned By Pearl

KoKo Spirometer®

SAS® Software

Symbicort®

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 PT009002. The SAP should be read in conjunction with the study protocol. This version of the SAP has been developed using the PT009002-02 Amended Protocol (Version 3.0 [redacted]) and the PT009002 case report form (CRF) (Version 01[redacted]).

2. STUDY OBJECTIVES AND ENDPOINTS

2.1 Study Objectives

The overall objective is to assess the efficacy and safety of treatment with BFF MDI 320/9.6 µg (budesonide and formoterol fumarate metered dose inhaler), BFF MDI 160/9.6 µg, FF MDI 9.6 µg (formoterol fumarate metered dose inhaler), BD MDI 320 µg (budesonide metered dose inhaler), and Symbicort® Turbuhaler® (TBH) 400/12 µg over 24 weeks in subjects with moderate to very severe chronic obstructive pulmonary disease (COPD).

2.1.1 Primary Objective

- To assess the effects of BFF MDI relative to FF MDI and BD MDI on lung function

2.1.2 Secondary Objectives

- To assess the effects of BFF MDI relative to FF MDI and Symbicort TBH on COPD exacerbations
- To assess the effects of BFF MDI relative to FF MDI, BD MDI, and Symbicort® TBH on symptoms of COPD
- To assess the effects of BFF MDI relative to FF MDI, BD MID, and Symbicort® TBH on quality of life (QoL)
- To determine the time to onset of action on Day 1

2.1.3 Safety Objectives

- To assess the safety of BFF MDI, FF MDI, BD MDI, and Symbicort® TBH

2.1.4 Healthcare Resource Utilization (HCRU) Objective

- To assess overall and COPD-specific HCRU of BFF MDI relative to FF MDI, BD MDI, and Symbicort® TBH

2.1.5 Pulmonary Lung Function (PFT) Sub-Study Objectives

- To characterize forced expiratory volume in 1 second (FEV_1) over 12 hours at Week 12

2.2 Study Endpoints

The primary endpoints, treatment comparisons of interest, and analysis timeframes may differ by country or region due to local regulatory agency requirements. The Sponsor has defined two different registration approaches in this study. The registration approaches will be called: (1) United States (US) and (2) European Union (EU). The US approach is for countries or regions where the primary endpoints are generally evaluated at a point in time. The EU approach is for registration purposes in countries or regions where the primary endpoints are generally evaluated over a period of time. The multiplicity controls for the primary analyses are delineated by approach in section 6.4.7.

2.2.1 Primary Efficacy Endpoints

US APPROACH

- Change from baseline in morning pre-dose trough FEV_1 at Week 24 (BFF MDI versus FF MDI)
- Change from baseline in FEV_1 area under the curve from 0 to 4 hours (AUC_{0-4}) at Week 24 (BFF MDI vs BD MDI)

EU APPROACH

- Change from baseline in morning pre-dose trough FEV_1 over 24 weeks (BFF MDI vs FF MDI; BFF MDI 320/9.6 µg vs Symbicort® TBH, non-inferiority)
- Change from baseline in FEV_1 AUC_{0-4} over 24 weeks (BFF MDI vs BD MDI; BFF MDI 320/9.6 µg vs Symbicort® TBH, non-inferiority)

2.2.2 Secondary Efficacy Endpoints

US APPROACH

- Percentage of subjects achieving an MCID of 4 units or more in Saint George's Respiratory Questionnaire (SGRQ) total score at Week 24 (BFF MDI vs FF MDI; BFF MDI vs BD MDI)
- Change from baseline in morning pre-dose trough FEV_1 at Week 24 (BFF MDI vs BD MDI)
- Peak change from baseline in FEV_1 at Week 24 (BFF MDI vs BD MDI)
- Change from baseline in average daily rescue Ventolin Hydrofluoroalkane (HFA) use over 24 weeks (BFF MDI vs BD MDI)
- Time to onset of action on Day 1 (BFF MDI vs BD MDI)
- Time to first moderate or severe COPD exacerbation (BFF MDI vs FF MDI)

EU APPROACH

- Change from baseline in morning pre-dose trough FEV_1 over 24 weeks (BFF MDI versus BD MDI)
- Transition Dyspnea Index (TDI) focal score over 24 weeks (BFF MDI vs FF MDI; BFF MDI vs BD MDI; BFF MDI 320/9.6 µg vs Symbicort® TBH, non-inferiority)

- Percentage of subjects achieving an MCID of 4 units or more in SGRQ total score over 24 weeks (BFF MDI vs FF MDI; BFF vs BD MDI; BFF MDI 320/9.6 µg vs Symbicort® TBH, non-inferiority)
- Change from baseline in average daily rescue Ventolin HFA use over 24 weeks (BFF MDI vs BD MDI)
- Peak change from baseline in FEV_1 over 24 weeks (BFF MDI vs BD MDI)
- Change from baseline in the Evaluating Respiratory Symptoms in COPD (E-RS: COPD) total score (RS-Total Score) over 24 weeks (BFF MDI vs FF MDI; BFF MDI vs BD MDI; BFF MDI 320/9.6 µg vs Symbicort® TBH, non-inferiority)
- Time to onset of action on Day 1 (BFF MDI vs BD MDI)
- Time to first moderate or severe COPD exacerbation (BFF MDI vs FF MDI)
- Time to clinically important deterioration (CID) (BFF MDI vs FF MDI)

2.2.3 Other Efficacy Endpoints

Unless already categorized as a secondary endpoint in one of the regulatory approaches above, all of the following endpoints will be categorized as "Other efficacy endpoints", with the treatment comparisons of interest being BFF MDI vs FF MDI, BFF MDI vs BD MDI, and BFF 320/9.6 µg MDI vs Symbicort® TBH (non-inferiority).

DAY 1 ENDPOINTS

- Change from baseline at each post-dose time point and in AUC_{0-4} for FEV_1, forced vital capacity (FVC), peak expiratory flow rate (PEFR), and forced expiratory flow from 25% to 75% (FEF_{25-75})
- Proportion of subjects achieving an improvement from baseline in FEV_1 using different thresholds (e.g., ≥10%, ≥12%, ≥15%, ≥100 mL, ≥200 mL; and ≥12% and ≥200 mL)

ENDPOINTS OVER 24 WEEKS (Unless Otherwise Stated):

- Rate of moderate or severe COPD exacerbations
- Rate of COPD exacerbations of any severity
- Time to treatment discontinuation for any cause
- Time to treatment failure (treatment discontinuation for any cause, moderate or severe exacerbation, or death)
- Time to CID
- Time to sustained CID
- Time to Death
- Additional spirometry assessments over Weeks 12 to 24, over 24 weeks, 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, the Evaluating Respiratory Symptoms in COPD (E-RS: COPD) Total Score (RS-Total Score), as well as 3 subscale scores (RS-Breathlessness, RS-Cough and Sputum, and RS-Chest Symptoms) over 24 weeks and over each 4-week interval of the 24-week Treatment Period
- TDI focal score over Weeks 12 to 24, over 24 weeks, and at each post-randomization visit
- Individual components of the TDI: functional impairment, magnitude of task, and magnitude of effort over Weeks 12 to 24, over 24 weeks, and at each post-randomization visit
- Percentage of subjects achieving a minimal clinically important difference (MCID) threshold of 1 unit or more on average in TDI focal score over Weeks 12 to 24 and over 24 weeks
- Changes from baseline over 24 weeks and at each post-randomization visit for SGRQ total score
- Change in individual domain scores of SGRQ: Symptoms, Activity, and Impacts over Weeks 12 to 24, over 24 weeks, 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 Weeks 12 to 24, and over 24 weeks
- Quality-of-Life Endpoints: European Quality-of-Life-5 Dimensions (EQ-5D-5L) Questionnaire scored at each post-randomization visit

2.2.4 Safety Endpoints

The safety endpoints for this study include:

- Adverse events (AEs), Treatment-emergent AEs, Serious AEs (SAEs), AEs of special interest (AESIs)
- 12-lead electrocardiograms (ECGs)
- Clinical laboratory values (hematology and clinical chemistry)
- Vital signs measurements (blood pressure and heart rate)

2.2.5 Health Care Resource Utilization Endpoints

- 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
  - 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 (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 subjects 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 subjects spent no time in the ICU or CCU
- The percentage of subjects hospitalized 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
- 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

2.2.6 12-hour Pulmonary Function Test Sub-Study Endpoints

The assessments will provide PFT profiles over 12-hours post-dose at Week 12.

Primary Endpoint:

- FEV_1 AUC_{0-12}

Other Endpoints

- Peak FEV_1
- FVC, PEFR, and FEF_{25-75} will be evaluated using AUC_{0-12} and at peak

3. STUDY DESIGN AND ANALYTICAL CONSIDERATIONS

3.1 Study Design

3.1.1 Overall Study Design and Plan

This is a Phase III, multi-center, randomized, double-blind, parallel-group, chronic-dosing (24 weeks), lung function study to assess the efficacy and safety of BFF MDI (320/9.6 µg and 160/9.6 µg) compared to FF MDI 9.6 µg, BD MDI 320 µg and Symbicort® TBH as an open-label active control, administered twice daily (BID), in subjects with moderate to very severe COPD.

This study will be conducted at approximately 200 sites, contributing approximately 10 to 14 subjects per site. Subjects meeting all inclusion criteria and no exclusion criteria will be randomized into this study. A total of 2420 subjects with moderate to very severe COPD will be randomized in a 3:3:3:1:1 scheme to one of the five treatment groups:

- BFF MDI 320/9.6 µg BID (660 subjects)
- BFF MDI 160/9.6 µg BID (660 subjects)
- FF MDI 9.6 µg BID (660 subjects)
- BD MDI 320 µg BID (220 subjects)
- Symbicort® TBH 400/12 µg BID (open-label, 220 subjects)

Approximately 1900 subjects are expected to complete the study. Randomization will be stratified by reversibility to Ventolin [(≥ 12% and 200 mL) or (< 12% or < 200 mL improvement over pre-bronchodilator FEV_1), measured at Visit 2], post-bronchodilator FEV_1 (< 50% or 50% to < 80% predicted, measured at Visit 2), blood eosinophil count (< 150 or ≥ 150 cells per mm^3), and country. Enrollment will be targeted to achieve a 2:1 ratio for blood eosinophil strata with twice as many randomized subjects in the ≥ 150 cells per mm^3 category.

Table 1 Planned Sample Size

	BFF MDI 320/9.6 µg	BFF MDI 160/9.6 µg	FF MDI 9.6 µg	BD MDI 320 µg	Symbicort® TBH 400/12 µg	Overall
Allocation Ratio	3	3	3	1	1	
Subjects Enrolled	660	660	660	220	220	2420

Subjects Completing the Study	518	518	518	173	173	1900

Subjects who discontinue randomized treatment prior to Week 24 (Visit 9) 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 the sub-study who discontinue randomized treatment will only complete regularly scheduled visits and not complete any remaining sub-study assessments. Subjects who discontinue randomized treatment will be returned to appropriate maintenance COPD medications, per the investigator's 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 investigator's 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 telephone call (TC) will not be required. These subjects will be followed for vital status at 24 weeks post randomization in accordance with the informed consent.

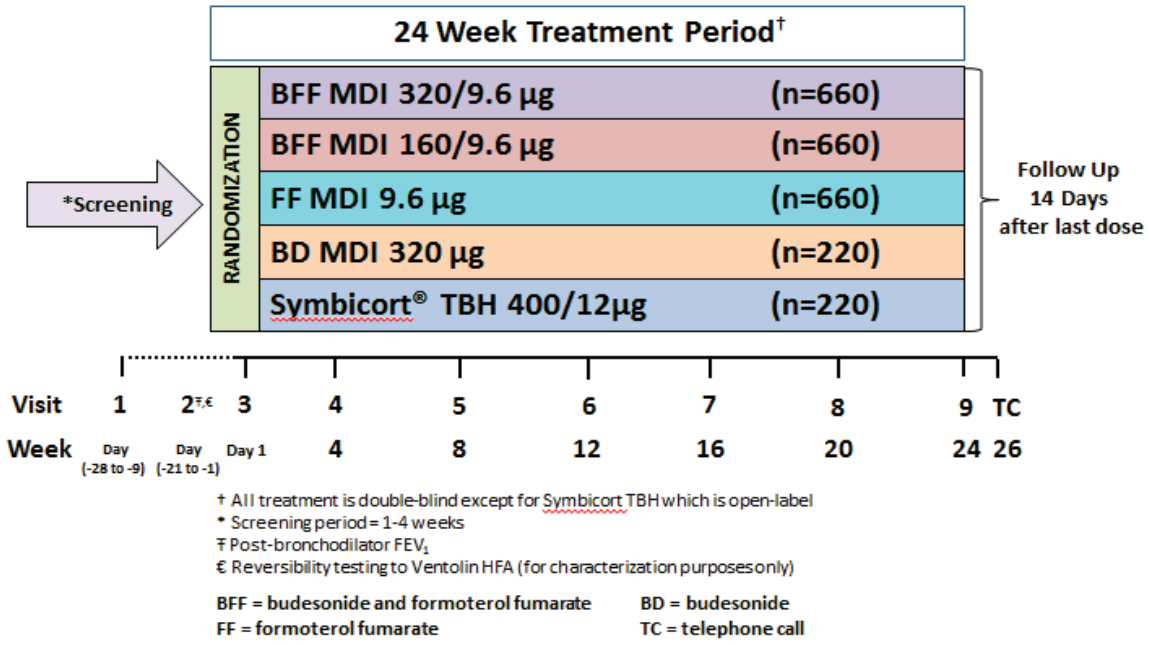
Sub-Study

12-Hour Pulmonary Function Test Sub-study: Serial PFTs will be conducted over 12 hours in a subset of approximately 510 randomized subjects (153 subjects from each of BFF MDI arms and the FF MDI arm, and 51 subjects from the BD MDI arm) at Week 12 (Visit 6) only. On the test day, additional serial spirometry will be obtained at 6, 8, 10, 11.5 and 12 hours post-dose.

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

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

Figure 1 Study Design



3.1.2 Prior, Concomitant, Post-Treatment, Prohibited Medications, and Other Restrictions

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 CRF (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. Refer to the Protocol for information about prohibited medications.

3.2 Hypothesis Testing

For the primary comparisons, the general null hypothesis for each pair-wise comparison to a BFF MDI mono-component will be that the mean treatment difference is zero (mean treatment effects are equal). The alternative two-sided hypothesis is that the mean treatment difference is greater or less than zero (mean treatment effects are not equal). The primary comparisons of BFF MDI 320/9.6 µg to Symbicort® TBH will be for non-inferiority and will use a margin (δ) of 50 mL for morning pre-dose trough FEV_1 and a margin of 75 ml for FEV_1 AUC_{0-4}. P-values will be reported as 2-sided.

The primary null (H_0) and alternative (H_1) hypotheses, with µ representing the mean, are presented below. The superiority hypotheses will take the following forms depending on the control group:

- $H_0: \mu_{BFF\ xxx/9.6} = \mu_{FF\ 9.6}$
  $H_1: \mu_{BFF\ xxx/9.6} \neq \mu_{FF\ 9.6}$

- H_0: $\mu_{BFF\ xxx/9.6} = \mu_{BD\ 320}$
  H_1: $\mu_{BFF\ xxx/9.6} \neq \mu_{BD\ 320}$

The non-inferiority hypotheses will take the following forms:

- H0: $\mu_{Symbicort} - \mu_{BFF\ 320/9.6} \geq \delta$
  H1: $\mu_{Symbicort} - \mu_{BFF\ 320/9.6} < \delta$

Where δ will be 50 or 75 mL. Two-sided 95% confidence interval (CI) for $\mu_{Symbicort} - \mu_{BFF}$ will be computed. If the upper bound is observed to be less than δ, non-inferiority of BFF MDI 320/9.6 µg relative to Symbicort® TBH will be declared.

Secondary and other efficacy analyses will involve the above hypotheses applied to secondary efficacy endpoints. The directionality of the non-inferiority hypotheses will be reversed for E-RS total score. 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.

3.3 Interim Analysis

No interim efficacy analyses are planned for this study.

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

3.4 Sample Size

It is estimated that a sample size of 2420 subjects (660 per arm in the BFF MDI and FF MDI groups and 220 per arm in the BD MDI and Symbicort® TBH groups) will provide power estimates as summarized in Table 2. All calculations assume Type I error control at a 2-sided alpha level of 0.05 and 20% dropout rate.

Table 2 Power Estimates

Endpoint	Assumed Difference	At Week 24 (US)	Over 24 Weeks (EU)
Trough FEV_1	BFF - FF = 40 mL	90%	99%
	Symbicort® - BFF = 0 mL *		96%
AUC_{0-4} FEV_1	BFF - BD = 100 mL	99%	99%
	Symbicort® - BFF = 0 mL *		99%

* Non-inferiority comparisons using the margins of δ = 50 mL for the pre-dose trough FEV_1 and δ = 75 mL for AUC_{0-4} FEV_1.

Assumptions regarding variability for the primary endpoints are based on Pearl's experience with Phase IIb and III clinical studies. A 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}. Based on the repeated measures analysis and anticipated dropout of approximately 20%, an effective SD for the change over 24 weeks of 157 mL and 200 mL for trough FEV_1 and FEV_1 AUC_{0-4}, respectively, is assumed. For Weeks 12 to 24, an effective SD for the change in trough FEV_1 of 175 mL is assumed.

The non-inferiority margin of 50 mL for the evaluation of pre-dose trough FEV_1 represents the approximate anticipated treatment effect in this endpoint. The non-inferiority margin of 75 mL for the evaluation of FEV_1 AUC_{0-4} represents a value less than the anticipated treatment effect in this endpoint.

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 Technologies) 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 PFT's obtained at each time point will be graded independently at iCardiac by qualified personnel. Quality grading assessments will be based on American Thoracic Society (ATS)/ERS criteria and will be included in data transfers.

5. ANALYSIS POPULATIONS

5.1 Population Definitions

5.1.1 Intent-to-Treat (ITT) Population

The **Intent-To-Treat Population** is defined as all subjects who are randomized to treatment and receive any amount of the study treatment. Subjects will be analyzed according to the treatment they were assigned at randomization. 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

The **Modified Intent-to-Treat Population** is a subset of the ITT Population, defined as all subjects with post-randomization data obtained prior to discontinuation from treatment. Any data collected after completion of or discontinuation from randomized study medication will be excluded from the mITT analysis but will still be included in the ITT analysis. Subjects will be

analyzed according to randomized treatment group. (Note that a subject who used a study treatment, but took less than one full daily 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 Rescue Ventolin User Population

Differences in rescue Ventolin HFA usage are expected across the study with some subjects using virtually no rescue medication at study entry. In order to represent the population of patients who may benefit from study treatment and reduce their use of rescue medication, the **Rescue Ventolin User (RVU) Population** is defined as all subjects in the ITT Population with average baseline Rescue Ventolin use of ≥ 1 puff/day.

5.1.4 Per-Protocol (PP) Population

The **Per-Protocol Population** is a subset of the ITT Population, defined as all subjects with post-randomization data obtained prior to any major protocol deviations. Data obtained after any major protocol deviation or discontinuation from treatment will be excluded. Since receiving the wrong treatment is a major protocol deviation, subjects in the PP population will be analyzed as randomized (which for this population is identical to analysis by actual treatment received). Any evaluability criteria with a potential impact on the primary efficacy results will be identified in a blinded fashion from review of data listings prior to unblinding. Major protocol deviations (protocol violations), 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 the blinded data review meeting (BDRM) prior to database lock. Reasons for exclusion from the PP Population will include, but are not limited to, the following:

- An incorrect diagnosis of COPD.
- 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.
- CAT < 10 at Screening (Visit 1).
- Subjects who do not meet protocol-specified FEV_1 baseline stability criteria. FEV_1 baseline stability is defined as the average of the -60 minute and -30 minute pre-dose FEV_1 assessments at Visit 3 being within ±20% or 200 mL of the pre-bronchodilator FEV_1 assessments obtained at Visit 2.

5.1.4.1 Record-level Exclusion of PFT Data from the PP Population

PFT records are excluded from the PP population under the following conditions:

- For subjects who take any protocol-prohibited medication that would affect PFT assessments on the date of an assessment, PFT measurements taken at that assessment will be excluded from the PP population.

- For those subjects who require rescue Ventolin HFA less than 6 hours before study visit, or who take any dose of rescue Ventolin HFA during the visit, all spirometry data post-Ventolin HFA administration will be considered missing for that day for the PP population.

- Subjects have to meet three restriction criteria prior to spirometry: (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 and for the duration of each in-clinic study visit, and (3) subject was not to have COPD bronchodilator medications for at least 6 hours prior to study visit. For subjects who fail to meet any of the restriction criteria, PFT data at the affected visits will be removed from the PP population. Such restrictions will be applied only if data pertaining to these criteria (including the timing) were collected.

- Subjects who did not take the evening dose of study medication on the day prior to a visit will have their PFT data excluded from the PP population for that visit.

5.1.5 Safety Population

The **Safety Population** is defined as all subjects who are randomized to treatment and receive any amount of the study treatment. 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. A subject who used a study treatment, but took less than 1 full daily dose of treatment will qualify for this population. Note: the statement that a subject had no AEs also constitutes a safety assessment.

5.2 Populations for Primary and Sensitivity Analyses

Demographics will be summarized for the mITT, PP, RVU, Safety, and Non-Randomized Populations, as well as for subjects participating in the 12-hr PFT sub-study.

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

Efficacy analyses will be performed for the ITT, mITT, and PP Populations. The mITT Population will be considered the primary population for the efficacy analyses, with the ITT and PP populations being considered supportive with several exceptions. The ITT Population will be used as primary for some efficacy analyses (e.g., time to death), and the PP Population will be

considered primary for the non-inferiority comparisons of BFF MDI 320/9.6 µg vs. Symbicort® TBH. Rescue medication endpoints will be analyzed with mITT, ITT, PP, and RVU populations.

6. STATISTICAL ANALYSIS

Analyses will be performed when the 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 25th and 75th percentiles will be presented when appropriate based on historical knowledge of the normality or non-normality 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 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 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.

Pre-dose spirometry values will use the average of the non-missing -60 minutes and -30 minutes values. Weekly averages for eDiary-based parameters (where eDiary means electronic diary) will use all non-missing values.

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

Disposition for all randomized subjects will be tabulated *(Table 1.1.1)* and listed *(Listing 1.2).* The 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 withdrew from the study prematurely, and who completed the study. The number and percentage of randomized subjects included in the mITT, PP, RVU, Safety, and 12-hr PFT Sub-study Populations will also be tabulated *(Table 1.1.1)*. Informed consent is listed in *Listing 9.6*.

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 will be tabulated in *Table 1.1.4.* 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 subjects with discrepant IWRS-based and actual stratification factors will also be provided (*Listing 1.7*).

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, RVU, and Safety analysis populations will be listed (*Listing 1.6)* for all subjects randomized. Reasons for premature discontinuation from study treatment will be summarized for the Safety Population *(Table 1.2.1).* Similarly, reasons for subjects' withdrawal from the study will be summarized for the ITT Population (*Table 1.2.2*).

Time to discontinuation of treatment and withdrawal from the study will be presented graphically by means of the Kaplan-Meier plots (*Tables and Figures 1.2.3 and 1.2.4*).

The reason for exclusion from the PP Population will be tabulated by study treatment for all mITT subjects *(Table 1.3.1)*. 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) prior to spirometry (Section 5.1.4.1) 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 listed *(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 (ICSs) at screening, and smoking status/history will be summarized for the mITT, RVU, PP, and Safety Populations and for Non-Randomized subjects *(Tables 1.4.1 through 1.4.5, 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 and baseline characteristics. If the Safety Population has the same treatment assignment as the mITT, then these summaries will be identical as well and hence not produced. Inhaled corticosteroid use (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 (the 12-hr PFT mITT Population) *(Table 1.4.6)*.

Demographic and baseline characteristic variables summarized will include the following:

- Age
- Age Group
- Age of onset of COPD
- Gender
- Race
- Ethnicity (Hispanic or Non-Hispanic)
- COPD Assessment Test (CAT) total score and total score category (<10, ≥10, <15, ≥15, <20, ≥20, Missing)
- Used inhaled corticosteroids 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 (Group as 0, 1, and ≥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, 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 all subjects for the mITT and Safety Populations and listed *(Tables 1.5.1, 1.5.2* and *Listing 4.1)*. A summary for the Safety Population will only be performed if the Safety Population is different from the mITT/ITT Population. 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 and mITT Populations *(Tables 1.9.1, 1.9.2,* and *Listing 4.3)*.

Descriptive statistics will be provided for screening period pre-bronchodilator and post-bronchodilator and baseline spirometry parameters *(Tables 1.6.1* to *1.6.4* for the mITT, PP, RVU, and 12-hr PFT Sub-study Populations, respectively, and *Listing 2.2)*.

Characterization of Reversibility:

Reversibility to Ventolin HFA will be evaluated at Visit 2 and used as a stratification variable at randomization to ensure an even distribution of reversibility across the treatment arms. A subject is considered reversible if the improvement in FEV1 at 30 minutes post-Ventolin is ≥ 12% and ≥ 200 mL.

Reversibility to Ventolin HFA at Screening Visit 2 will be summarized for the mITT and 12-hr PFT Sub-study Populations and listed *(Tables 1.7.1, 1.7.2, Listing 2.2* and *Listing 5.2* for Ventolin HFA dispensing). The number and percentage of subjects reversible will be included in these summaries. Also included will be a summary of the change in FEV_1 from pre-dose FEV_1 to post-bronchodilator assessment. 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 Ventolin HFA bronchodilator:

- ≥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 listed for all randomized subjects *(Table 1.8.1.1 and Listing 4.4)*. Cardiovascular medical history of interest at Screening will be summarized for the Safety Population and listed for all randomized subjects *(Table 1.8.1.2 and Listing 4.5)*.

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

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

All prescription and OTC medications taken by the subject during 30 days before Screening and all concomitant therapy taken by the subject while on study will be recorded on the Prior and 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 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.

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 to 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, COPD-Exacerbation, and Non-COPD related medications/treatments will be summarized by preferred term and actual treatment received for the Safety Population (*Tables 1.11.1 to 1.11.3*). 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 to 4.9,* respectively).

Reported prior medications for COPD, COPD-Exacerbation, and non-COPD-related medications will be tabulated for the Safety Population (*Tables 1.10.1 to 1.10.4*) and listed separately *(Listings 4.7* to *4.9*, 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 separately): (1) a muscarinic antagonist, (2) a β2 agonist, and (3) an inhaled corticosteroid *(Table 1.10.2)*. 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 1.11.2)*.

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. 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)*. 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%. Also provided in this summary will be descriptive statistics (n, mean, standard deviation, median, minimum and maximum) for percent compliance 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

6.4.1 Estimands

Four estimands of interest are defined for this study: efficacy estimand, attributable estimand, treatment policy estimand, and per protocol estimand.

The primary estimand of interest is the efficacy estimand and is the effect of the randomized treatment in all subjects assuming continuation of randomized treatments for the duration of the study regardless of actual compliance. 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 attributable estimand 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 is considered a bad outcome. Analyses of the attributable estimand will be conducted using the mITT Population. Data that are missing due to treatment discontinuation will be imputed based on the 5^{th} or 95^{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 higher value is a worse outcome. For this purpose, FF MDI and BD MDI are considered the reference arms. The estimated average of the two reference means will be used. 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 6.4.3.1) and in the Details Appendix of this SAP.

Treatment discontinuations reasonably attributable to tolerability or lack of efficacy will be identified during the BDRM and documented in the BDRM minutes prior to unblinding. Discontinuations will be attributed to tolerability if the subject had an adverse event determined by the investigator to be 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.

The treatment policy estimand is the effect of randomized treatment over the study period regardless of whether randomized treatment is continued. Analyses of the treatment policy estimand will be conducted using the ITT Population, in which all observed data will be utilized regardless of whether subjects remain on randomized treatment.

The per protocol 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.

Analysis of this estimand will use the PP Population. Exclusions from the PP Population will be finalized at the BDRM prior to database lock and recorded in the meeting minutes.

There are five pairwise comparisons of treatments of interest, namely, BFF MDI (2 doses: 320/9.6 µg and 160/9.6 µg) vs. FF MDI, BFF MDI (2 doses: 320/9.6 µg and 160/9.6 µg) vs. BD MDI, and BFF MDI 320/9.6 µg vs. Symbicort® TBH. Estimation results will be provided by randomized treatment and for each treatment difference for all comparisons, in each estimand.

All comparisons will be performed for testing superiority except that the comparison of BFF MDI 320/9.6 µg to Symbicort® TBH will be for non-inferiority. Non-inferiority analyses will use the per protocol estimand, unless specifically stated otherwise.

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 3) 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 average 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.

For COPD exacerbation history, baseline is whether the subject had a moderate-or-severe COPD exacerbation (Yes/No) in the last 12 months prior to Visit 1 (from the Visit 1 CRF page).

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

- the subject was taking a medication that contained a glucocorticoid component (active ingredient) that is listed 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).

Baseline blood eosinophil count is the average 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 available value of FEV_1 obtained after dosing with Ventolin at Visit 2.

Baseline percent reversibility to Ventolin is 100 x (POST-PRE)/PRE, where PRE is the mean of the available 30 minute and 60 minute values of FEV_1 prior to dosing with Ventolin at Visit 2, and POST is the post-bronchodilator FEV_1 value defined above.

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 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 below.

Table 3 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.

6.4.3 Primary Efficacy Analysis

Analyses will be conducted on the efficacy estimand, on the attributable estimand, on the treatment policy estimand, and on the per protocol estimand. The primary analyses of the efficacy, treatment policy, and per protocol estimands use only observed data. The attributable

estimand will use the mITT Population but then impute missing data as described in section 6.4.1.

6.4.3.1 Change from Baseline in Morning Pre-Dose Trough FEV_1

Change from baseline in morning pre-dose trough FEV_1 is the primary endpoint in the US approach (BFF vs. FF at Week 24) and the EU approach (BFF vs. FF and non-inferiority of BFF 320/9.6 µg vs. Symbicort® over 24 weeks). It is also a secondary endpoint in the US approach (BFF vs. BD at Week 24) and the EU approach (BFF vs. BD over 24 weeks).

Change from baseline in morning pre-dose trough FEV_1 at each visit is defined as the average 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. Spirometry data from unscheduled visits will not be used for this analysis. Assessments obtained during early termination visits will be used if their timing is consistent with the next scheduled collection of spirometry data.

The change from baseline in morning pre-dose trough FEV_1 will be analyzed using a repeated measures (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 blood eosinophil count, and percent reversibility to Ventolin HFA as continuous covariates. An unstructured covariance (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 at Week 24, over weeks 12 to 24, over 24 weeks, and at each post-randomization visit. Two-sided p-values and point estimates with two-sided 95% CIs will be produced for each treatment difference.

The primary analysis will be conducted for the efficacy estimand. The analysis of the primary endpoints with the attributable estimand is considered secondary. Analyses will be also conducted for the treatment policy and per protocol estimands as supportive except for the non-inferiority comparisons of BFF MDI 320/9.6 µg vs. Symbicort® TBH, for which the per protocol estimand will be primary (*Table 2.1.1* and *Figure 2.1.1* for the efficacy estimand, *Table 2.1.2* and *Figure 2.1.2* for the attributable estimand, *Table 2.1.3* and *Figure 2.1.3* for the treatment policy estimand*, Table 2.1.4 and Figure 2.1.4* for the per protocol estimand). All comparisons will be for superiority except that the comparison of BFF MDI 320/9.6 µg to Symbicort® TBH will be for non-inferiority and will use a margin of 50 mL.

For the attributable estimand (for the analysis at Week 24 and the analysis over 24 weeks), multiple imputation for missing values for morning pre-dose trough FEV_1 will use mean changes from baseline based on the 5th percentile of the reference arms' distribution when missingness is reasonably attributable to tolerability or lack of efficacy (see Section 6.4.1). Other missing data are to be imputed using the observed data model. The variance used for the multiple imputation

are described in the Details Appendix to this SAP. 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.

Exploration of the robustness of findings to missing data is discussed in Section 6.4.3.4.

6.4.3.2 FEV_1 AUC_{0-4}

FEV_1 AUC_{0-4} is the primary endpoint in the US approach (BFF vs. BD at Week 24), and the EU approach (BFF vs. BD and non-inferiority of BFF 320/9.6 µg vs. Symbicort® over 24 weeks).

FEV_1 AUC_{0-4} is the area under the curve for FEV_1 calculated using the trapezoidal rule, after the subtraction of the baseline FEV_1 value, and the AUC will be transformed into a weighted average by dividing by the time (in hours) from dosing to the last measurement included (typically 4 hours). For all estimands, 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} at Week 24, over 24 weeks, over weeks 12 to 24, and at Day 1 and each post-randomization visit, 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. Covariance structure will be chosen by the same approach as for the morning pre-dose trough FEV_1 model. Two-sided p-values and point estimates with two-sided 95% CIs will be produced for each treatment difference of interest (*Tables 2.1.5 to 2.1.8* for the efficacy estimand, attributable estimand, treatment policy estimand, and per protocol estimand, respectively).

All comparisons will be for superiority except that the comparison of BFF MDI 320/9.6 µg to Symbicort® TBH will be for non-inferiority and will use a margin of 75 mL. The primary analysis will be conducted using the efficacy estimand except for the non-inferiority comparison of BFF MDI 320/9.6 µg vs. Symbicort® TBH, for which the per protocol estimand will be primary.

For the attributable estimand of FEV_1 AUC_{0-4} (for the analysis at Week 24 and over 24 weeks), data that are missing due to treatment discontinuation will be imputed in a similar manner to that of missing data for the attributable estimand for morning pre-dose trough FEV_1 in Section 6.4.3.1 above. The analysis of this endpoint with the attributable estimand is considered secondary.

6.4.3.3 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, 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.

The assumption of normality for the change from baseline in the morning trough FEV_1 and $AUC_{0\text{-}4}$ 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 covariance model fail to converge), and/or ICS use categories (yes/no). Note that the unstructured covariance structure allows for heterogeneity among the visits.

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

6.4.3.4 Sensitivity Analyses

Sensitivity analyses will be conducted for the change from baseline in morning pre-dose trough FEV_1 and FEV_1 $AUC_{0\text{-}4}$ 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 efficacy, attributable and treatment policy estimands. The following table summarizes the multiple imputation-based sensitivity analyses under the PMM (pattern mixture model) framework that will be undertaken.

Table 4 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. 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 that 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.

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 will be 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 BFF 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 Analysis of the Primary Estimand:

- Tipping Point #1: this first set of analyses will impute diminished effects only for subjects on BFF MDI whose missing data are determined to be MNAR.
- Tipping Point #2: this analysis will impute diminished effects for all missing data in the BFF 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 (*Tables 2.1.10.1 to 2.1.10.8* for the change from baseline in morning pre-dose trough FEV_1, *Tables 2.1.11.1 to 2.1.11.8* for FEV_1 AUC_{0-4}). Figures of delta (decrement in treatment effect) versus p-values will also be produced (*Figures 2.1.10.1a* to *2.1.10.8h* for the change from baseline in morning pre-dose trough FEV_1, *2.1.11.1a to 2.1.11.8h* for FEV_1 AUC_{0-4}). Details of the sensitivity analyses will be 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 at Week 24 and over 24 weeks for the efficacy estimand estimand *(Tables 2.1.12.1* to *2.1.12.3)* and treatment policy estimand (*Tables 2.1.12.2 to 2.1.12.4*), and change from baseline in FEV_1 AUC_{0-4} at Week 24 and over 24 weeks (*Tables 2.1.13.1 and 2.1.13.3* for the efficacy estimand, *Tables 2.1.13.2* and *2.1.13.4* for treatment policy estimand).

Cumulative distribution plots by treatment arm will also be produced. The observed change from baseline in morning pre-dose trough FEV_1 at Week 24 and 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 *(Figures 2.1.12.1* and *2.1.12.3* for the efficacy estimand, *Figures 2.1.12.2* and *2.1.12.4* for treatment policy estimand). Similarly, the cumulative proportion of responders plots for FEV_1 AUC_{0-4} will be generated (*Figures 2.1.13.1* and *2.1.13.3* for the efficacy estimand, *Figures 2.1.13.2* and *2.1.13.4* for treatment policy estimand). Subjects without post-baseline data 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.

6.4.4 Secondary Efficacy Analyses

Secondary efficacy analyses will be conducted for the efficacy estimand, the attributable estimand, the treatment policy estimand, and the per protocol estimand. The efficacy estimand will be considered primary for superiority comparisons, and the per protocol estimand will be considered primary for the non-inferiority comparisons of BFF 320/9.6 µg vs. Symbicort®. The analysis of the primary endpoints with the attributable estimand is considered to be a secondary efficacy analysis; otherwise, the attributable estimand will be supportive. The other estimands will be supportive.

6.4.4.1 Transition Dyspnea Index

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 7.
TDI focal score is a secondary efficacy endpoint in the EU approach (BFF vs. FF, BFF vs. BD, and non-inferiority of BFF 320/9.6 µg vs. Symbicort® over 24 weeks).

At Randomization (Visit 3), 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 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. 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). The efficacy estimand using on-treatment data will be considered the main analysis. Analyses for the attributable estimand, treatment policy estimand, and the analysis of on-treatment data for the per protocol estimand will be conducted as supportive (*Tables* and *Figures 2.2.2* to *2.2.4*, 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 6.4.1.

All comparisons will be for superiority except that the comparison of BFF MDI 320/9.6 µg to Symbicort® TBH will be for non-inferiority rather than superiority, and will use a margin of 0.75.

The main analysis will be conducted using the efficacy estimand (with analyses on the attributable, treatment policy, and per protocol estimands as supportive) except for the non-inferiority comparisons of BFF MDI 320/9.6 µg vs. Symbicort® TBH, for which the per protocol estimand analysis will be the main analysis.

In addition, the difference between treatments for the individual components of the TDI: 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 focal score *(Tables 2.2.5* and *Figures 2.2.5.1-2.2.5.3* for the functional impairment, magnitude of task, and magnitude of effort for the efficacy estimand).

Furthermore, responder analyses will be performed for the TDI focal score where responders are defined as subjects with a response of 1.0 points or more (corresponding to at least a minor improvement) on average over 24 weeks and on average over weeks 12-24. Logistic regression will be used to compare the treatment groups with BDI, baseline eosinophil count, 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 treatment comparison *(Table 2.2.6* 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 will be considered to be non-responders in the analysis.

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

6.4.4.2 Peak Change from Baseline in 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.

The peak change from baseline in FEV_1 within 4 hours post-dosing is a secondary endpoint in the US approach (BFF vs. BD at Week 24) and the EU approach (BFF vs. BD over 24 weeks).

The peak change from baseline in FEV_1 within 4 hours post-dosing at Week 24, over weeks 12-24 and over 24 weeks will be evaluated as a secondary efficacy analysis. This analysis will use the same modeling as with morning pre-dose trough FEV_1. The peak change from baseline on Day 1 and at each post randomization visit will also be analyzed (*Tables* and *Figures 2.3.1* to *2.3.4* for the efficacy estimand, attributable estimand, treatment policy estimand, and per protocol estimand, 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 6.4.1.

Tipping Point Analyses for Peak Change from Baseline in FEV_1 at Week 24

Robustness of results to missing data will be explored using tipping point analyses (Ratitch 2013). Details of the methods are similar to sensitivity analyses of FEV_1 and AUC_{0-4} (found in Section 6.4.3.4 and in the Details Appendix to this SAP, and using the maximum value of δ = 500 mL), 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].

Additional sensitivity analyses will be implemented based on a cumulative responder approach as described in section 6.4.3.4 for the peak change from baseline in FEV_1 within 4 hours post-dose at Week 24 (*Tables 2.3.5.1 and 2.3.5.2* for the efficacy and treatment policy estimand, respectively). A cumulative distribution plot by treatment arm will also be produced (*Figure 2.3.5.1 and 2.3.5.2* for the efficacy and treatment policy estimand, respectively).

6.4.4.3 St. George's Respiratory Questionnaire

Percentage of subjects achieving an MCID of ≥4 in SGRQ total score is a secondary endpoint in the US approach (BFF vs. FF and BFF vs. BD at Week 24) and in the EU approach (BFF vs. FF, BFF vs. BD and non-inferiority of BFF 320/9.6 µg vs. Symbicort® over 24 weeks). Change in

SGRQ total score and change in individual domain score from baseline over 24 weeks, over weeks 12 to 24, and at each post-baseline visit are 'other' endpoints.

The SGRQ will be used to provide the health status/health-related QoL measurements in this study (see Protocol Appendix 8). 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.

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.

Responder analyses of SGRQ as a secondary efficacy endpoint will be performed, where responders are defined as subjects with an improvement of ≥4.0 points at Week 24, on average over 24 weeks, and over weeks 12-24. For the SGRQ responder analyses, subjects who discontinue treatment for any reason or are missing all post-baseline data will be considered to be non-responders for the analysis using the efficacy estimand or per protocol estimand. For the treatment policy estimand, subjects who withdraw from the study 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 and ICS use at Screening as categorical covariates. P-values and odds ratios with 95% CIs will be produced for each treatment comparison (*Tables 2.4.3 to 2.4.6* for the efficacy, attributable, treatment policy and per protocol estimands).

All comparisons will be for superiority except that the comparison of BFF MDI 320/9.6 µg to Symbicort® TBH will be for non-inferiority rather than superiority, and will use a margin of 10%.

The main analysis will be conducted using the efficacy estimand (with analyses on the attributable, treatment policy, and per protocol estimands as supportive) except for the non-inferiority comparisons of BFF MDI 320/9.6 µg vs. Symbicort® TBH, for which the per protocol estimand will be the main analysis.

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 6.4.1), 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].

The difference between treatment groups in the change from baseline in SGRQ total score over 24 weeks, over Weeks 12 to 24, and each post-baseline visit will be evaluated using a similar RM approach as for TDI focal score, but using baseline SGRQ score replacing BDI in the model. Thus the model will include treatment, visit, and treatment by visit interaction, ICS use at Screening 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. 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 *(Table* and *Figure 2.4.1* for the efficacy estimand).

Individual domains of the SGRQ will also be analyzed in a similar fashion as the total score *(Table 2.4.2* for the efficacy estimand).

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). Details of the methods are similar to sensitivity analyses of FEV_1 and AUC_{0-4} (found in Section 6.4.3.1 and in the Details Appendix to this SAP, and using the maximum value of $\delta = 16$ units), but with the model for SGRQ total score described above.

6.4.4.4 Rescue Ventolin HFA Use

Change from baseline in average daily rescue Ventolin HFA use is a secondary endpoint in the US approach (BFF vs. BD over 24 weeks) and the EU approach (BFF vs. BD over 24 weeks).

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 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, baseline FEV_1, and percent reversibility to Ventolin HFA as continuous covariates. A UN matrix will be used to model additional autocorrelation within subject. If this 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 treatment difference. The main analysis will be conducted using the efficacy estimand (*Table* and *Figure 2.5.1*). Supportive analyses will use the attributable estimand, treatment policy estimand, the per protocol estimand, and the RVU Population (*Tables* and *Figures 2.5.1* to *2.5.4, 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 6.4.3.1 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 2.5.5, 2.5.6, 2.5.8 and 2.5.9* for the efficacy estimand and the RVU Population).

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). Details of the methods are similar to sensitivity analyses of FEV_1 and AUC_{0-4} (found in Sections 6.4.3.4 and in the Details Appendix to this SAP, and using the maximum value of $\delta = 4$ puffs), but with the model for rescue Ventolin HFA use described above. Multiple-imputation results will be combined using Rubin's formulae [Rubin, 1987].

Additional sensitivity analyses will be implemented based on a cumulative responder approach as described in section 6.4.3.4 for the average daily rescue Ventolin HFA use over 24 weeks (*Tables 2.5.10 and 2.5.11* for the efficacy estimand and treatment policy estimand). A cumulative distribution plot by treatment arm will also be produced (*Figures 2.5.10 and 2.5.11* for the efficacy estimand and treatment policy estimand). Cumulative responder analysis for the attributable estimand will not be performed as methodology to apply Rubin's rules for combining multiply imputed data for such an analysis is not readily available.

6.4.4.5 E-RS Total Score

Change from baseline in E-RS Total score is a secondary efficacy endpoint in the EU approach (BFF vs. FF, BFF vs. BD and non-inferiority of BFF 320/9.6 µg vs. Symbicort® over 24 Weeks).

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 E-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. The last 7 days of the Screening Period will be used to calculate the baseline.

The mean change from baseline in E-RS Total Score, RS-Breathlessness, RS-Cough and Sputum, RS-Chest Symptoms and the EXACT Total Score over each 4-week interval will be analyzed using a similar RM model as for TDI to estimate treatment effects over 24 weeks, but using the corresponding baseline mean score instead of the BDI as a covariate. 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, and 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. A UN correlation matrix will be used to model additional autocorrelation within subject. If this 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.

The E-RS Total score over 24 weeks is a secondary efficacy endpoint. The RS subscale scores and the EXACT Total score are "other" endpoints. EXACT data will be listed in *Listing 6.1.5*. The analysis of E-RS Total score will be secondary for the EU only. Two-sided p-values and point estimates with 2-sided 95% CIs will be produced for each treatment difference (*Tables* and *Figures 2.6.1 to 2.6.4* for the efficacy estimand, attributable estimand, treatment policy estimand, and per protocol estimand, respectively; *Tables 2.6.5 and 2.6.6* for EXACT total and domain scores). Analyses on the attributable estimand, treatment policy estimand, and per protocol estimand will pertain to the E-RS Total score only.

The attributable estimand (for the analysis of E-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 6.4.3.1 except that the 95th percentile will be used instead of the 5th percentile.

6.4.4.6 Time to Onset of Action Assessed Using FEV_1 on Day 1

Time to onset of action on Day 1 is a secondary efficacy endpoint in the US and the EU approaches (BFF vs. BD).

The onset of action for BFF MDI will be evaluated on Day 1 by comparing BFF MDI vs BD

MDI in the mean change from baseline in FEV_1 at the 5 minute post-dose timepoint. Comparisons will also be made at additional timepoints but will not be controlled for multiplicity. Baseline is defined as the average of the non-missing -60 minute and -30 minute values obtained prior to dosing at Day 1 (Visit 3).

To perform this analysis, a linear model with repeated measures will be fit to the Day 1 post-baseline data. This model will include baseline FEV_1, percent reversibility to Ventolin HFA, and baseline eosinophil count as continuous covariates and time point, treatment, the treatment by time point interaction, and ICS use at screening as categorical covariates. An unstructured correlation matrix will be used to model additional autocorrelation within subject. If this 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. The resulting tables and figures are *Tables* and *Figures 2.7.1 to 2.7.2* for the efficacy and per protocol estimands, respectively.

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%, ≥15%, ≥100 mL, ≥200 mL, and ≥12% and 200 mL) will be estimated for each treatment. Subjects without post-baseline data will be considered to be non-responders in the analysis.

The percentage of responders at each post-dose time point will be summarized by treatment.

Logistic regression 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. Rate differences will also be estimated, using non-linear modeling to construct the CI *(Tables 2.7.3 to 2.7.8* for the efficacy estimand).

The median time to response will also be presented by treatment. The median will be defined as the first time point on Day 1 such that at least 50% of the subjects of the given treatment group have achieved the response at or before that time. If the 4 Hour time point does not meet this condition, the median is not defined. Since the number of time points of FEV_1 collection is limited, constructing a CI or performing treatment comparison is not feasible for the median time to response. These estimates will be incorporated in *Table 2.7.3*.

6.4.4.7 Time to First Moderate or Severe COPD Exacerbation

Time to first moderate or severe COPD exacerbation is a secondary endpoint in the US and the EU approaches (BFF vs. FF). The non-inferiority comparison of BFF 320/9.6 µg to Symbicort® is not part of the type I error control strategy.

Time to first moderate or severe COPD exacerbation is the time from first dose of study medication to the time of onset of the first moderate or severe COPD exacerbation.

Only on-treatment exacerbations will be included for calculating the time to first moderate or severe COPD exacerbation for the efficacy estimand and per-protocol estimand (see section 6.1).

Exacerbations occurring after the premature discontinuation of treatment will be considered for the treatment policy estimand.

For the attributable estimand, missing data that have been reasonably attributed to tolerability or lack of efficacy will be imputed based on the 95th percentile of the reference arms' distribution. The imputed value will be drawn from a negative binomial distribution with mean exacerbation rate (and variance) based on the 95th percentile of the reference arms' distribution, with estimates set to the average of estimates for the two reference treatments from the analysis of the exacerbation rates. Other missing data are to be imputed using the observed data model, i.e. assumed to be missing at random (MAR) or missing completely at random (MCAR). Further information about the computation of the attributable estimand is described in the Details Appendix to this SAP (Appendix 6); 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 (in study days) and ends at 24 weeks.

The time to first moderate or severe COPD exacerbation will be analyzed up through the Week 24 Visit using a Cox regression model. Treatment comparisons will be performed using the model, adjusting for percent predicted post-bronchodilator FEV_1, baseline eosinophil count, baseline COPD exacerbation history (Yes/No), country, and ICS use at Screening (Yes/No). SAS PROC PHREG will be used. 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 (*Tables 2.12.1 to 2.12.4* for the efficacy estimand, attributable estimand, treatment policy estimand, and per-protocol estimand).

Time to first moderate or severe COPD exacerbation will be analyzed and displayed graphically for each treatment using a Kaplan-Meier curve (*Tables* and *Figures 2.12.1 to 2.12.4* for the efficacy estimand, attributable estimand, treatment policy estimand, per protocol estimand, and *Listing 6.1.2.3*). Subjects who complete the study (and the study treatment) and do not experience a COPD exacerbation over treatment period will be censored at the Week 24 visit. For the efficacy and per-protocol estimands, subjects who do not experience a COPD exacerbation and discontinue treatment early will be censored at the date of treatment discontinuation. For the treatment policy estimand, subjects who do not experience a COPD and withdraw from the study will be censored at the last date of last assessment or contact (including telephone contact), and subject who do not experience a COPD exacerbation and complete the trial will be censored at the date of the Week 24 Visit. For the non-inferiority comparison of BFF MDI 320/9.6 µg MDI to Symbicort®, a non-inferiority margin for the hazard ratio of 1.1 will be employed.

Kaplan-Meier curves for time to moderate or severe COPD exacerbations will be constructed for completers vs. discontinuations (*Figure 2.12.1.1*), and, in separate plots, for MAR/MCAR vs. MNAR, and attributable vs. non-attributable discontinuation types (*Figures 2.12.1.2 and 2.12.1.3*).

<u>Tipping Point Analyses for Time to First Moderate or Severe COPD Exacerbation</u>

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.

Table 5 Sensitivity Analyses for Time to First 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 the 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 using the 95th percentile of the reference arms' distribution (for the rate of moderate or severe COPD exacerbation) 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 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 multiple imputation will be applied to the moderate or severe COPD exacerbation events within the negative-binomial analysis framework for the rate of moderate or severe COPD exacerbations, using values of δ that increase the rate in the treatment arm by up to 1.5 exacerbations/year. For this method, an underlying negative binomial stochastic process for the rate of exacerbations is assumed and post-treatment-discontinuation counts will be imputed conditional upon the reason for treatment discontinuation (see Appendix 6 for details).

A dataset with event counts through Week 24 will be created; 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 (in study days) and ends at Week 24. Missing values will first be imputed for the missing COPD exacerbation events prior to the computation of the time to the first moderate or severe COPD exacerbation (for the sensitivity analysis). After imputation, the analysis will proceed to use Cox regression (as described above) and subsequently the multiple-imputation results will be combined using Rubin's formulae [Rubin, 1987].

6.4.4.8 Time to Clinically Important Deterioration

Clinically important deterioration (CID) is defined as ≥ 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. Time to CID is a secondary efficacy endpoint in the EU approach (BFF vs. FF). The non-inferiority comparison of BFF 320/9.6 µg to Symbicort® is not part of the type I error control strategy.

Time to CID analysis will be performed using the Cox regression model. The model will include treatment, baseline post-bronchodilator percent predicted FEV_1, baseline eosinophil count, baseline COPD exacerbation history (Yes/No), 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 2.15.1 to 2.15.4* for the efficacy estimand, attributable estimand, treatment policy estimand, and per-protocol estimand). Time to CID will be displayed for each treatment group using a Kaplan-Meier curve (*Figures 2.15.1 to 2.15.4* for the efficacy estimand, attributable estimand, treatment policy estimand, and per protocol estimand).

For the attributable estimand, missing data that have been reasonably attributed to tolerability or lack of efficacy will be imputed based on either the 5th or the 95th percentile of the reference arms' distribution. The attributable estimand for time to CID will be computed by applying the percentile penalty to each of four component variables simultaneously; it uses multiple by-visit imputation of pre-dose trough FEV_1 (at visits for which they are missing) (as described in Section 6.4.3.1), of the SGRQ total score (as described in Section 6.4.4.3), of the TDI focal score in a manner similar to that for SGRQ, and of time to first moderate-or-severe COPD exacerbation (as described in Section 6.4.4.7).

6.4.5 Other Efficacy Analyses

6.4.5.1 Other Spirometry Endpoints

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 6.4.3.1 (*Tables* and *Figures 2.1.1* to *2.1.6*).

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 the 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 estimand).

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 efficacy estimand and treatment policy estimand, respectively. The baseline covariate for each model will be endpoint-specific (*Tables* and *Figures 2.8.1* to *2.8.3* for change from baseline in morning pre-dose trough FVC, PEFR, and FEF_{25-75}; *Tables* and *Figures 2.9.1* to *2.9.3* for peak change from baseline within 4 hours for FVC, PEFR, and FEF_{25-75}; *Tables* and *Figures 2.10.1* to *2.10.3* 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* for FEV_1 and *2.11.1* to *2.11.3* for FVC, PEFR, and FEF_{25-75}, respectively, for the efficacy estimand; *Listings 6.3.1* to *6.3.7*).

6.4.5.2 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 and 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".

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 (Yes, No), 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 discontinuation from or 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 discontinuation from or completion of study medication are not subtracted.

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*.

The number of exacerbations and the percentage of subjects who experience exacerbations, exacerbation rates, and rate ratios comparing treatments will be summarized for the efficacy estimand for moderate-or-severe exacerbations (*Table 2.12.7*), for any severity of exacerbations (*Table 2.12.8).*

Follow-up time at risk will be summarized and displayed graphically for the efficacy estimand for moderate or severe COPD exacerbations (*Table and Figure 2.12.7a*) and for any severity of exacerbations (*Table and Figure 2.12.8a).*

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). If the subject dies before being treated or hospitalized, then the start and stop date will be the date of death. 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 that are moderate or severe, eDiary data from dates within 7 days of a moderate or severe exacerbation will not be included as 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. Thus, for example, in the analysis of severe COPD exacerbations, subjects will still be considered to be at risk of an exacerbation even during or within 7 days after a mild or moderate exacerbation. Likewise, in the analysis of moderate-or-severe COPD exacerbations, subjects will still be considered to be at risk of an exacerbation even during or within 7 days after a mild exacerbation.

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 either antibiotics or oral corticosteroids were administered for the exacerbation.

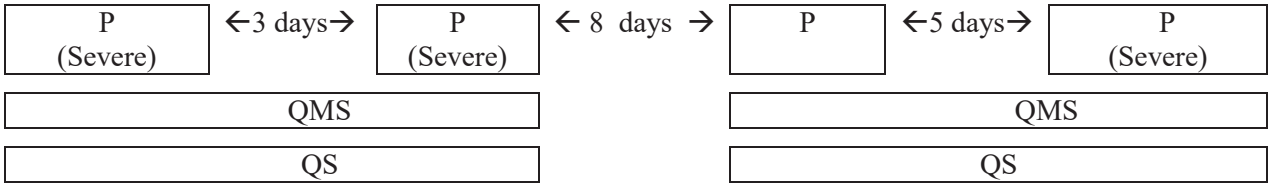
Call this time interval a “P-Interval”. The start date of the P-Interval is the earliest start date of the above, and the stop date will be defined as the last stop day of the above. If the subject was hospitalized due to the exacerbation or if the exacerbation led to a COPD-related death, then the

severity of "severe" will be assigned to this P-interval; otherwise the severity of "moderate" will be assigned. The later among 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, 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

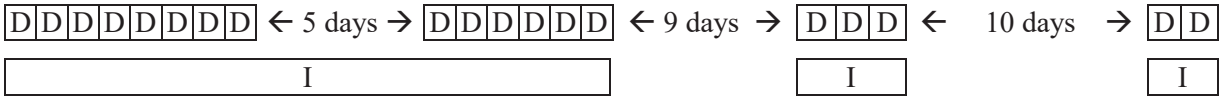
Using eDiary data, a day will be designated as being during an event of a COPD exacerbation of some severity if (1) there was at least one major symptom and there was at least one other major or minor symptom and if (2) on an adjacent day there was at least one major symptom and there was at least one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

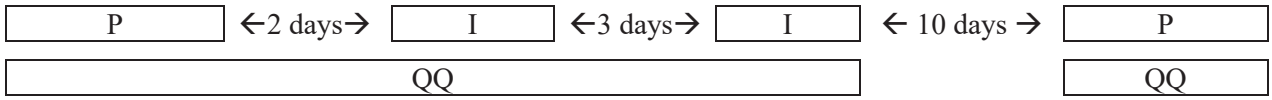
This QQ-interval will represent the consolidated duration of several exacerbations recorded on different CRF pages or identified from subject diary data. This QQ-Interval will be considered to be a single event of an any-severity COPD exacerbation. See Figure 4.

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 (or severe COPD exacerbation events) if they are close enough together in time (Figure 2). 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.

Rate of COPD exacerbations of any severity will be analyzed in a manner similar to the rate of moderate or severe COPD exacerbations *(Table 2.12.8* for the efficacy estimand).

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*.

6.4.5.3 Time to Treatment Failure

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 2.13.1*). Subjects who do not experience a treatment failure will be censored at their Week 24 Visit or Day 168, whichever comes first. The time to treatment failure will be analyzed using the Cox model for the efficacy estimand. The model will include treatment, baseline post-bronchodilator percent predicted FEV_1, baseline eosinophil count, baseline COPD exacerbation history (Yes/No), 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*).

6.4.5.4 Time to Treatment Discontinuation for Any Cause

Time to discontinuation from treatment for any cause will be summarized and graphically displayed using a Kaplan-Meier plot (*Table and Figure1.2.3*). Statistical comparisons will not be performed. Time to study discontinuation will be summarized similarly, but using the treatment policy estimand (*Table and Figure 1.2.4*).

6.4.5.5 Time to Death

For time to death (all causes), subjects will be censored at the date of last contact. The primary analysis will use the treatment policy estimand. The Cox regression model will include treatment, baseline post-bronchodilator percent predicted FEV_1, and baseline age. Estimated adjusted hazard ratios will be displayed along with associated 95% CI and p-values (*Table* and *Figure 2.14.1* for the treatment policy estimand). The analysis of time to death will be conducted contingent upon having at least 30 deaths in the study. Otherwise, the analysis will be limited to counts and listings.

6.4.5.6 Time to Sustained Clinically Important Deterioration

Sustained CID is defined as ≥ 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 sustained CID analysis will be performed using the Cox regression model. The model will include baseline post-bronchodilator percent predicted FEV_1, baseline eosinophil count, baseline COPD exacerbation history (Yes/No), country, and ICS use at Screening (yes/no)**.** Time to a sustained CID event will be based on the component event which occurs first. Subjects who do not have a sustained 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 sustained CID events. Estimated adjusted hazard ratios will be displayed along with associated 95% CI and p-values (*Table 2.15.5*). Time to sustained CID will be displayed for each treatment group using a Kaplan-Meier curve (*Figure 2.15.5*).

6.4.5.7 European Quality-of-Life-5 Dimension-5 Level Questionnaire

The European Quality-of-Life-5 Dimension-5 Level Questionnaire (EQ-5D-5L) 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-5L 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 (*Tables 2.16.1* for the efficacy estimand). Descriptive statistics for the index score (*Table 2.16.2* for the efficacy estimand) and VAS (*Table 2.16.3* for the efficacy estimand) will be presented by treatment group. VAS scores over 24 weeks will be analyzed using a similar RM ANCOVA model which is used for primary efficacy analysis. The model will include treatment, visit, treatment-by-visit interaction, country, and gender as categorical covariates and baseline score and age as continuous covariates (*Tables* and *Figures 2.16.2* for the index score and *Table 2.16.3* 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 8 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 that visit *(Table 2.16.4* for the efficacy estimand).

6.4.5.8 12-Hour Pulmonary Function Tests

12-hour lung function data will be collected in a subset of approximately 510 randomized subjects at Week 12 (Visit 6). 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*.

Descriptive summaries for change from baseline in FEV_1, FVC, PEFR, and FEF_{25-75} will be presented at each post-dose time point over 12 hours and at 12-hour post-dose trough at Week 12 (Visit 6) *(Tables* and *Figures 2.17.2, Listing 6.3.9* for FEV_1; *Tables* and *Figures 2.17.5,* and *Listing 6.3.12* for FVC; *Tables* and *Figures 2.17.8,* and *Listing 6.3.15* for PEFR; *Tables* and *Figures 2.17.11,* 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.

Area under the curve AUC_{0-12} will be calculated using the trapezoidal rule, after first having subtracted the baseline FEV_1 value, and transformed into a weighted average by dividing by the time in hours between the first and the last measurement included. For both the efficacy and per protocol estimand analysis, 1 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 Week 12 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 for the efficacy estimand (*Tables* and *Figures 2.17.1* for AUC_{0-12}). In addition, the peak change from baseline in FEV_1 over 12 hours at Week 12 will be estimated and compared between treatments using an ANCOVA with the same model as described for AUC FEV_1 (*Table 2.17.3* for the efficacy estimand).

Similar analyses will be performed for AUC_{0-12} of FVC, PEFR, and FEF_{25-75} *(Tables 2.17.4* for FVC $AUC_{0-12,}$ *2.17.7* for PEFR $AUC_{0-12,}$ *2.17.10* for FEF_{25-75} AUC_{0-12} for the efficacy estimand). The baseline covariate will be endpoint-specific.

6.4.6 Subgroup Analyses

Subgroup analyses will be performed for change from baseline in morning pre-dose trough FEV_1, AUC_{0-4} FEV_1, and rate of moderate or severe COPD exacerbations (efficacy estimand only). The following subgroups will be considered:

- History of Moderate or Severe COPD Exacerbation in the last 12 Months:
  - Yes
  - No
- Baseline Eosinophil Count:
  - <150 cells per mm^3
  - ≥150 cells per mm^3
- Country

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 4.1.1* to *4.3.3).*

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 4.4* for the efficacy estimand*)*.

Eosinophil Cut Point Exploration

Subgroup analyses of trough FEV_1 will be conducted in the baseline eosinophil count-high (≥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 (*Figures 4.5.1, 4.5.1.1 and 4.5.1.2*) as well as potentially using subgroups defined by different cut points.

6.4.7 Control of Type I Error

The same general Type I error control strategy will be employed for the US and EU regulatory approaches. In each approach, Type I error will be controlled for the following analyses:

- The primary analyses under the efficacy estimand
- The secondary analysis of the primary endpoints under the attributable estimand
- The secondary analyses

6.4.7.1 US Approach

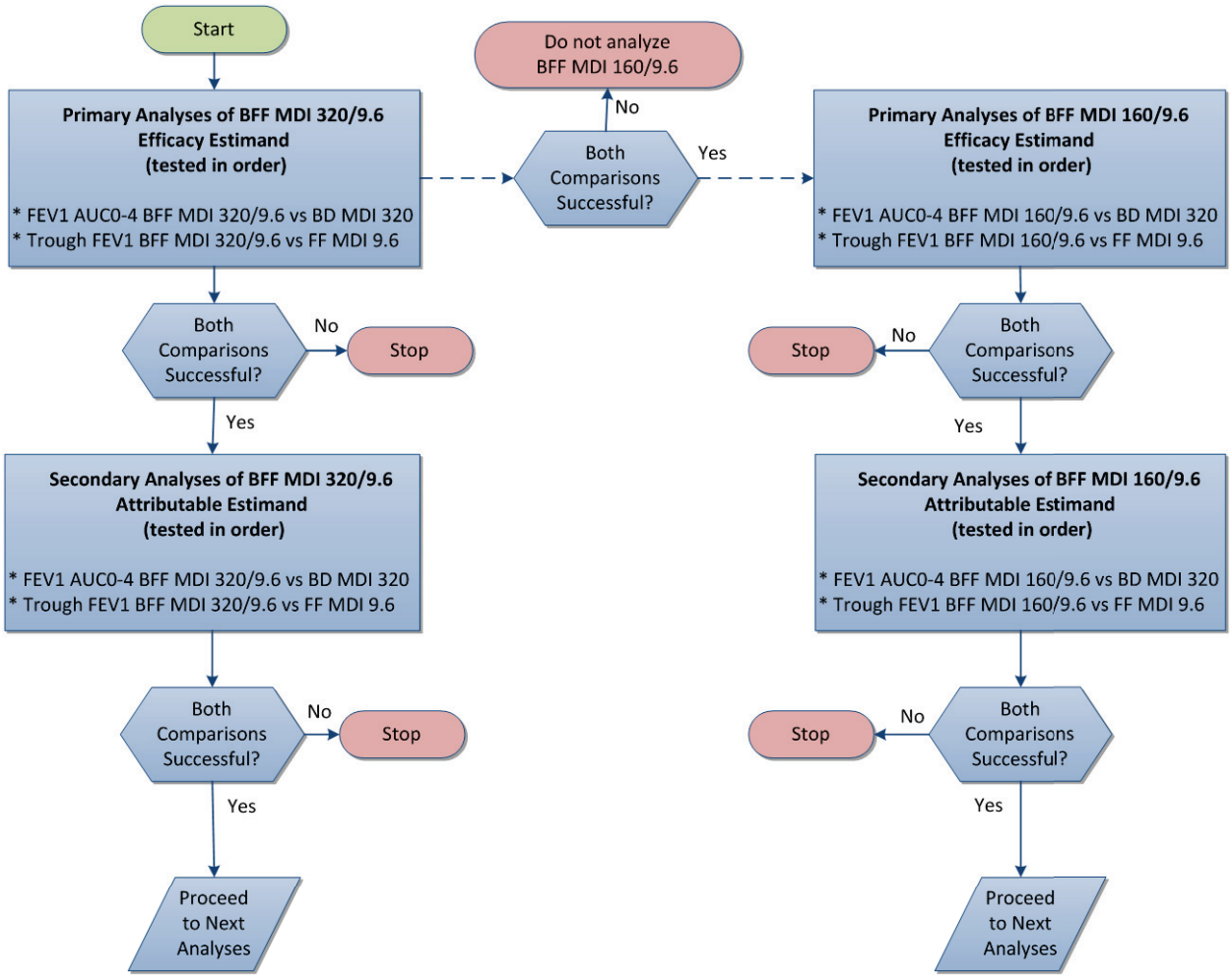
The Type I error rate will be controlled within the primary and secondary efficacy analyses. The primary analyses associated with BFF MDI 160/9.6 will proceed only if all of the primary analyses associated with BFF MDI 320/9.6 are successful. The secondary analysis of the primary endpoints under the attributable estimand and remaining secondary analyses for each BFF dose will proceed only if the primary analyses associated with that dose of BFF MDI are successful.

For ease of review, the set of planned analyses has been divided into two groups:

Group 1: The primary analyses under the efficacy estimand and the secondary analysis of the primary endpoints under the attributable estimand.

Group 2: The remaining secondary analyses under the efficacy estimand.

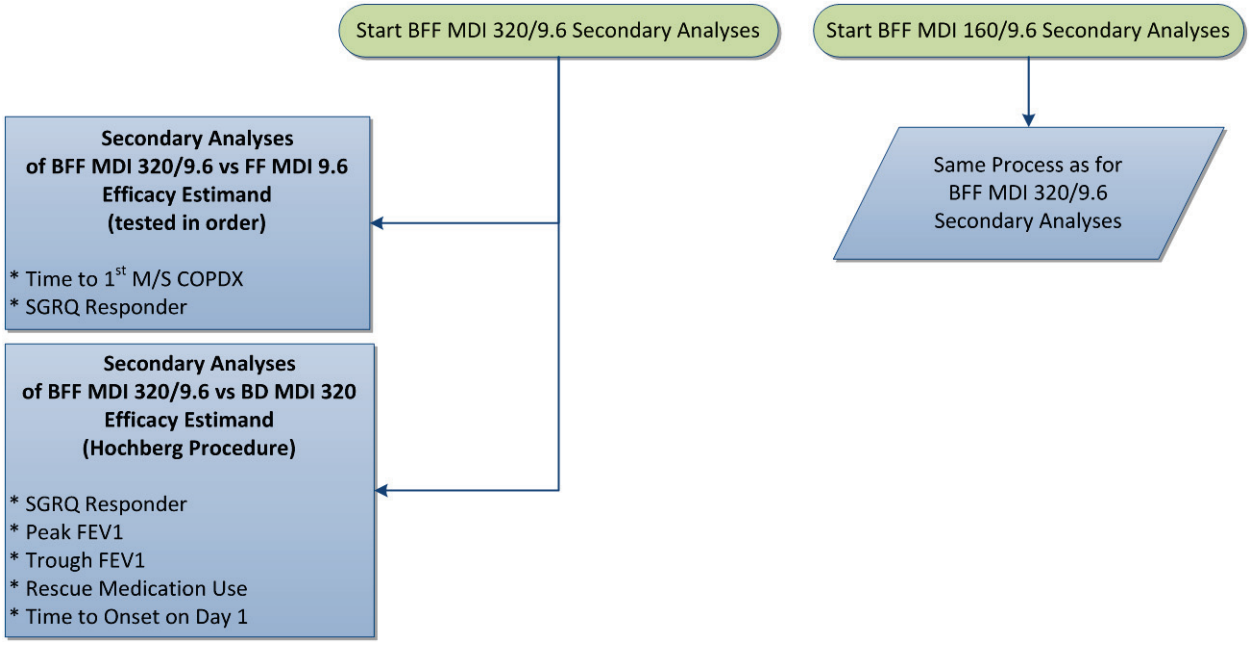
Figure 5 Group 1: Type I Error Control for the Analyses of the Primary Endpoints (US Approach)



In Group 1, a sequential multiplicity approach will be used in the analysis of the primary endpoints. In this approach, the analyses of the primary endpoints are listed in a pre-specified order in which they will be tested. Each hypothesis will be tested at the 2-sided 0.05 level. If a p-value is less than 0.05, then that hypothesis is rejected and the next hypothesis is tested. If a primary hypothesis is not rejected, then testing will stop. However, it is noted that p-values will still be calculated for all subsequent analyses for descriptive purposes.

Group 2 analyses are presented in Figure 6.

Figure 6 Group 2: The Analysis of the Secondary Endpoints (US Approach)

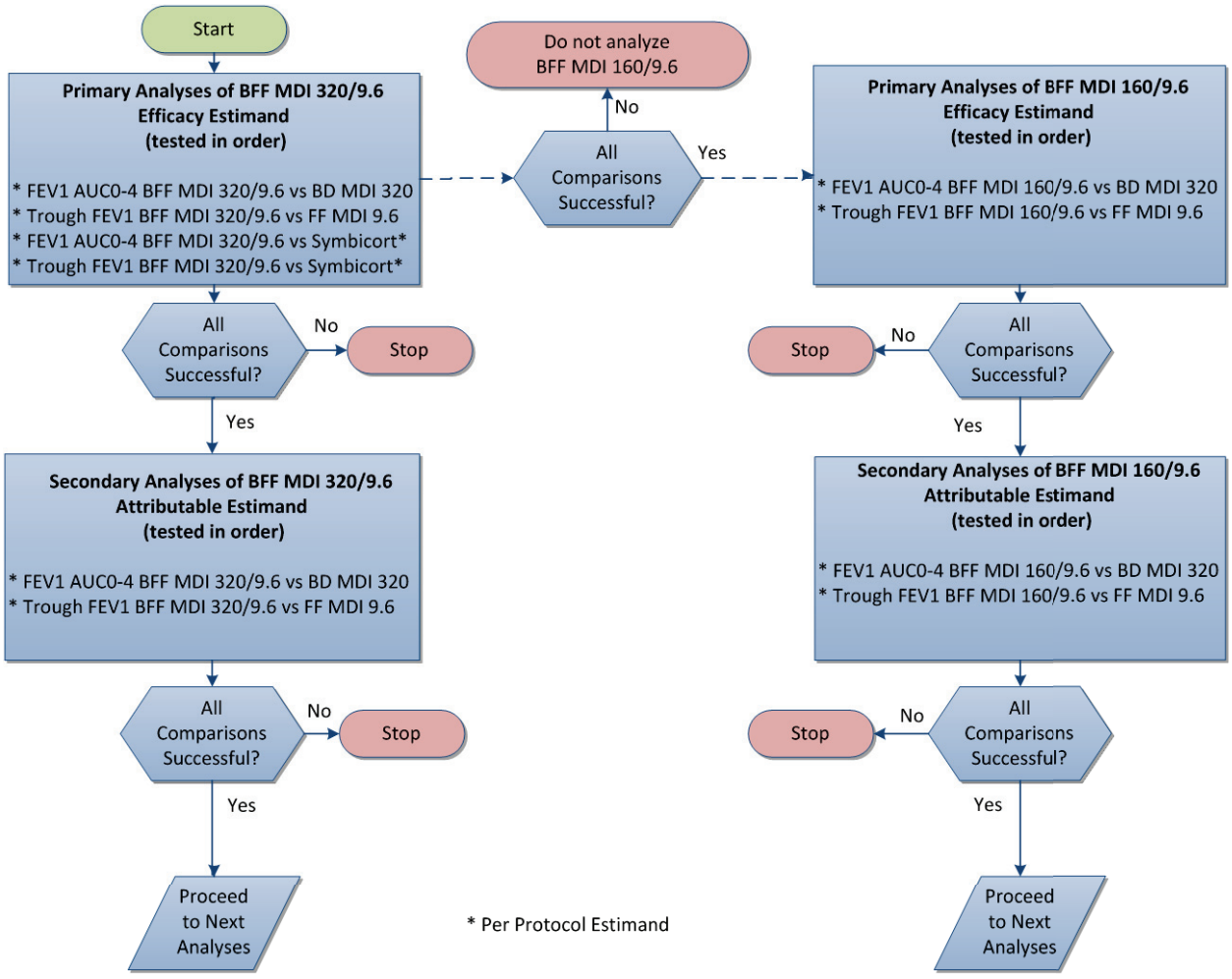


In Group 2, the secondary comparisons to the mono-component control arms will be treated as separate families of hypotheses. The Type I error in each family of secondary analyses will be controlled to alpha with either a sequential testing method or the Hochberg procedure.

6.4.7.2 EU Approach

The EU Approach will follow a similar general strategy as the US Approach. The control of Type I error in the EU Approach differs from the US Approach in the need to compare BFF MDI 320/9.6 µg to Symbicort® for non-inferiority and the number of secondary endpoints being evaluated. Otherwise, the Type I error control strategy is similar. The graphical representations of the primary, secondary, and subgroup analyses are presented below.

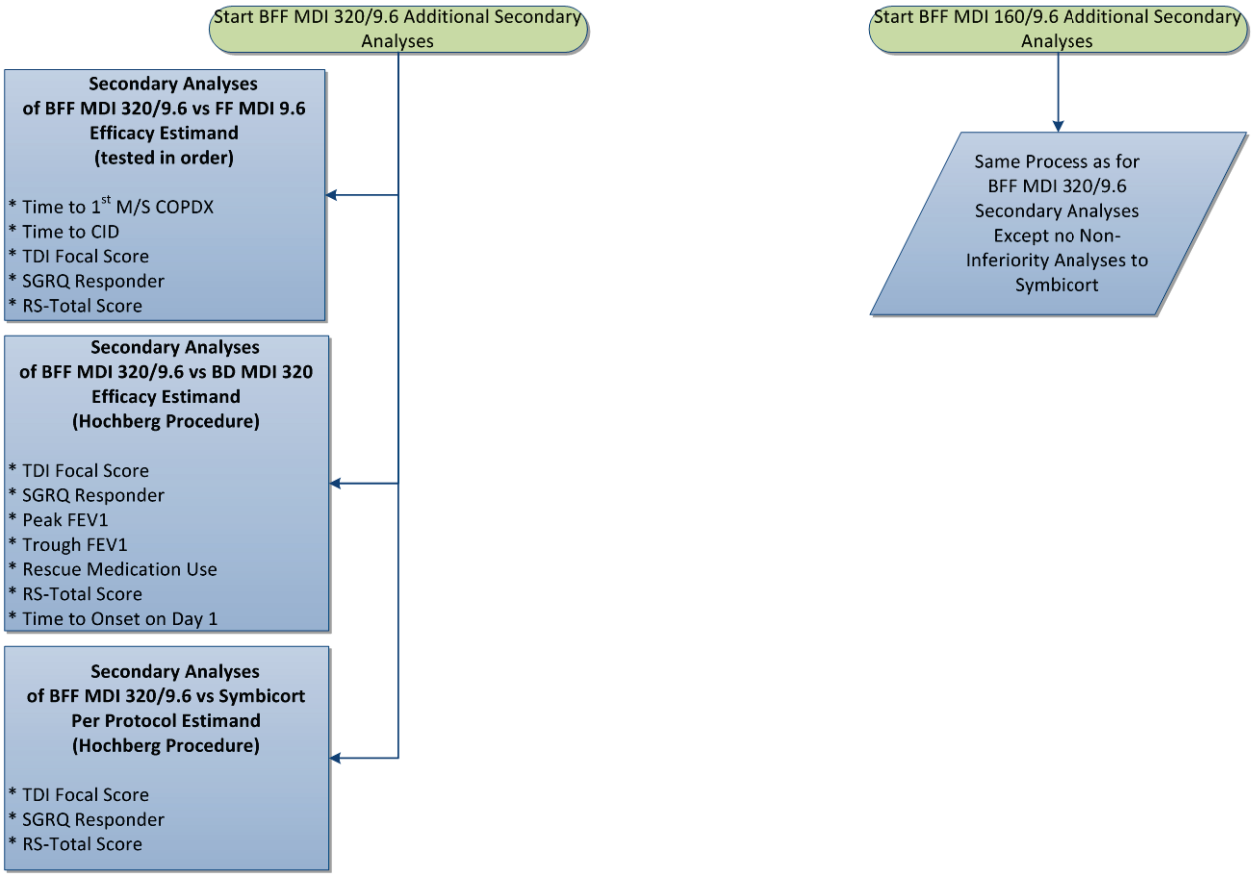
Figure 7 Group 1: Type I Error Control for the Analyses of the Primary Endpoints (EU Approach)



If the Group 1 analyses are successful within a BFF MDI dose, the analysis of the remaining secondary endpoints for that BFF MDI dose (Group 2) will proceed under the efficacy estimand.

Group 2 analyses are presented in Figure 8.

Figure 8 **Group 2: The Analysis of the Secondary Endpoints (EU Approach)**



6.4.8 Correlation Analysis

Pearson correlation coefficients will be generated between the primary and secondary continuous endpoints (from both US and EU approaches). 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, $AUC_{0\text{-}4}$ 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 2.18.1*), and also analyzed individually (*Tables 2.18.2 to 2.18.6*).

6.5 Safety Analysis

All safety analyses are based on the Safety Population. Hypothesis testing will not be performed for any safety analyses.

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 ("On-Treatment") are events with an onset date on or after the first date of dose and up to and including the last day of randomized treatment (for study drug completers) or the last day of randomized treatment + 1 day (for premature treatment discontinuation). 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 after the last day of randomized treatment (for study drug completers) or on or after the last day of randomized treatment + 2 days (for premature treatment discontinuation). The exception is that deaths are still considered to be during the Treatment Period if any adverse event that led to that death occurred during the Treatment Period.

Any AEs, clinically significant laboratory values, vital signs, and ECG values during the randomized-treatment period will be tabulated and listed. Beginning on the day after the date of discontinuation from or completion of study medication has passed, 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-randomized-treatment maintenance period will be tabulated and listed.

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 AEs reported in the study along with the associated Investigator's verbatim will be provided in *Listing 7.2*.

An AE is considered treatment-emergent if an event occurs after the first dose of randomized 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 last day of randomized treatment (for study drug completers) or the last day of randomized treatment + 1 day (for premature treatment discontinuation). AEs with onset date after the last day of randomized treatment (for study drug completers) or the last day of randomized treatment + 1 day (for premature treatment discontinuation) will not be considered treatment-emergent, but will be listed in adverse event data listings, and will be tabulated separately. For the purposes of this SAP, the terms "treatment-emergent AE" and "On-Treatment AE" are synonymous. A more detailed definition may be found in Appendix 1 (Data Handling Rules, Category 16). AEs that occur between the time the subject signs the informed consent form 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.

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 Medical Dictionary for Regulatory Activities (MedDRA) preferred term and the MedDRA system organ class. No hypothesis tests will be performed.

An overview table will be prepared for the Safety Population with the incidences of subjects with at least one treatment-emergent adverse event (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 treatment discontinuation, and a report of death (*Tables 3.1.1, 3.1.13, and 3.1.14*).

Events with Irregular Onset Dates

All treatment-emergent adverse events 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 system organ class, 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 *Table 3.8.2.1* and *Table 3.8.2.2*.

The listing of adverse events will provide the severity, maximum severity, relationship to study drug, action taken and outcome for each adverse event. Any SAEs reported will be listed for all subjects screened (*Tables 3.8.2.1).* Adverse events leading to permanent discontinuation of study treatment will be listed for the Safety Population *(Table 3.6.1)*. A listing of any reported deaths

during the study (prior to randomization, during the randomized-treatment period, or during the post-randomized-treatment maintenance period) 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, for each treatment, for each primary system organ class, and for each preferred term within a system organ class:

- The incidence of all treatment-emergent adverse events *(Tables 3.2.1.1, 3.2.1.2, Table 3.2.1.3)*
- The incidence of subjects with adverse events by SOC during the post-randomized-treatment maintenance period (*Tables 3.2.1.18, 3.2.1.19)*
- The incidence of treatment-emergent adverse events occurring in SMQs (Standard MedDRA Queries)/groupings of interest *(Tables 3.2.3.1, 3.2.3.2)*
- The incidence of non-serious treatment-emergent adverse events occurring in ≥5% of subjects in a treatment *(Tables 3.2.4.1, 3.2.4.2)*
- The incidence of all treatment-related treatment-emergent adverse events *(Tables 3.4.1, 3.4.2)*
- The incidence of discontinuation from study treatment due to a treatment-emergent adverse events *(Tables 3.5.1, 3.5.2)*
- The incidence of treatment-emergent serious adverse events *(Tables 3.7.1.1.1, 3.7.1.1.2, 3.7.1.1.3)*
- The incidence of subjects with serious adverse events by SOC during the post-randomized treatment maintenance period (*(Tables 3.7.2.3, 3.7.2.4)*
- The incidence of all treatment-related treatment-emergent serious adverse events *(Tables 3.9.1, 3.9.2)*
- The incidence of all treatment-emergent adverse events by highest severity to treatment *(Tables 3.11.1.1* through *3.11.4.2* for the four treatments)
- The incidence of treatment-emergent adverse events 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).
- 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 system organ class, and preferred term:

- Frequency and rate of AEs *(Tables 3.3.1, 3.3.2)*
- Frequency and rate of SAEs *(Tables 3.8.3.1, 3.8.3.2)*
- 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).

6.5.1.1 Adverse Events of Special Interest

Adverse events of special interest (AESIs) have been defined based on known effects of LABAs and ICS. These include but are not limited to cardiovascular, tremor effects, hyperglycemia, and hypokalemia for LABAs; and local (eg, candidiasis and voice effects) and systemic (eg, bone and skin effects, diabetes control, ocular and taste effects, adrenal suppression) steroid class effects and lung infection for ICS.

Standard MedDRA queries (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 for use in the assessment of AESIs associated with ICS and LABAs are listed in Table 6. SMQs will be utilized when possible and a selection of preferred terms in other situations (*Appendix 5*).

Table 6 Adverse Events of Special Interest

Medical Concept	Selection of MedDRA Terms
Adrenal suppression	Adrenal cortical hypofunctions HLT
Agitation or anxiety	Collection of PTs
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
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) Hypertrichosisa (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
Weight gain	Collection of PTs

Abbreviations: CNS=central nervous system; PT=preferred term.

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).

Adverse Events in MedDRA SMQs/Groupings of Interest by Term will be tabulated *(Tables 3.12.1, 3.12.2)*.

6.5.1.2 Major Adverse Cardiovascular Event (MACE) 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 Clinical Endpoint Committee

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 3.14.1.1, 3.14.1.2*). 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 Clinical Endpoint Committee

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, (4) mortality of cancer and (5) 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 one day *(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 randomized-treatment period *(Tables 3.15.1.1, 3.15.1.2)* and during the post-randomized-treatment

maintenance period (*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 system organ class, and preferred term (*Tables 3.15.3.1, 3.15.3.2*). Adjudicated death events will be listed in *Listing 7.4*.

6.5.1.5 **Paradoxical Bronchospasms**

During Visits 3, 4, 6 and 9, the paradoxical bronchospasm event is defined as a reduction in FEV1 of >20% from the morning trough, that occurs within 30 minutes post-dosing with associated symptoms of wheezing, shortness of breath, or cough. Spontaneous reporting of paradoxical bronchospasm will occur at Visits 5, 7, and 8.

All paradoxical bronchospasm events will be captured on the AE CRF page. For those events that occurred during Visits 3, 4, 6 and 9, a programmatic check will be done to verify whether they satisfy the condition on the change in FEV_1, and will be queried as necessary. Paradoxical bronchospasms will be summarized by treatment during the randomized-treatment period *(Table 3.2.3).*

6.5.2 Clinical Laboratory Measurements

Lab parameters collected include the following:

Table 7 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
Bilirubin, total	Chloride
Gamma-glutamyl transferase	Cholesterol
	Bicarbonate
	Creatinine
	Glucose
	Magnesium
	Potassium
	Phosphate
	Protein, total
	Sodium
	Triglycerides
Urinalysis	
Macroscopic examination including specific gravity, pH, protein, glucose, ketones, blood, and urobilinogen.	
Other Tests:	
Pregnancy test (women of childbearing potential only): serum hCG at Visit 1 (Screening) and Final Visit (Visit 9) or Treatment Discontinuation Visit	
Creatinine clearance will be estimated by the CKD-EPI formula [Levy, 2009]	
Abbreviations: CKD-EPI=Chronic Kidney Disease Epidemiology Collaboration; hCG=human chorionic gonadotropin	

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 laboratory abnormality that leads to a dose-limiting toxicity (e.g., an abnormality that results in study drug dose reduction, suspension or discontinuation)
- A laboratory abnormality that results in any therapeutic intervention (i.e., concomitant medication or therapy)
- 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, Covance *(Appendix 4)*.

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

If there are multiple laboratory values for the same parameter at pre-dose of 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 and unscheduled-visit 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 3.15.1* through *3.15.4,* 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 3.15.5 to 3.15.8* 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 Covance (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 (PCS) 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 8 Potentially Clinically Significant 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 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 3.15.9*). 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 3.15.10* and *3.15.11).* For all laboratory parameters other than glucose and potassium noted in Table 7, all laboratory data for the parameter identified as potentially clinically significant for a subject will be listed *(Table 3.15.12* - 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 average of all available pre-dose measurements taken prior to the start of dosing at the Randomization Visit (Visit 3). If there are no Visit 3 pre-dose values, the baseline will be defined as the average of pre-bronchodilator values at Visit 2. 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 9 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

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

Table 10 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, 1.4.2.1, 1.4.3.1, 1.4.4.1* and *1.4.5.1* for the mITT, ITT, PP, and Safety Populations, and all subjects not randomized respectively).

Summary statistics (n, mean, median, standard deviation and range) of the absolute value and change from baseline for systolic blood pressure, diastolic blood pressure, heart rate, and temperature will be tabulated by treatment, visit, and time point. 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 4 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 and unscheduled-visit data are candidates for clinically significant values and 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 11. No hypothesis tests will be performed.

Table 11 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 hrs	≥ 75 min. to < 4 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 3 or within the same post-dose study time window at a visit, the last value will be chosen for analysis.

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 9 and Table 10 (*Table 3.17.2.1*- 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 (Fridericia Corrected QT) interval will be calculated where baseline is defined as the average of the pre-dose measurements taken prior to the start of treatment at the randomization visit (Visit 3). If there are no Visit 3 pre-dose values, the baseline will be defined as the value obtained at Visit 1. The QTcF 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. These assessments will be tabulated for each treatment and each scheduled nominal time point at each visit and at end of treatment (*Table 3.18.1, Table 3.18.2*). End of Treatment is defined as the last non-missing on-treatment assessment available. Data from unscheduled visits will not be used for the by-visit summaries but both scheduled and unscheduled-visit data are candidates for clinically significant values and 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 3.18.1f)*. ECG data from subjects with pacemakers will not be included in analyses, but will be listed.

If there are multiple ECG values for the same parameter at pre-dose of a visit date (other than for Visit 3), the last value will be chosen for analysis.

Other than for the change from baseline analyses mentioned above, all available data post-baseline including data from unscheduled visits will be used for ECG parameter analyses.

Table 12 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

Potentially clinically significant ECG parameter values will be identified based on criteria listed in Table 12. 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), (*Tables 3.18.7, 3.18.8* for PR interval increase), and (*Tables 3.18.9, 3.18.10* for QRS prolongation). No hypothesis tests will be performed.

6.5.5 Healthcare Resource Utilization

Data on healthcare resource utilization (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 mean number of days in CCUs

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 per year, 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 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

Figures 7 and 8 (Type I error control charts for EU approach) were fixed to reflect that the non-inferiority testing is performed using the Per Protocol Estimand.

8. STATISTICAL SOFTWARE

Data processing, statistical screening, descriptive reporting and analysis of the efficacy and safety data will be performed using SAS (Version 9.3 or higher). Graphs may also be produced using R (R Development Core Team, 2003).

9. REFERENCES

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).

EuroQol Group. EQ-5D. Accessed at http://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.