



---

**Statistical Analysis Plan**

Study Code D3251C00003

Edition Number 3.0

Date 23 April 2018

---

---

**A Randomised, Double-blind, Chronic Dosing (56 week) Placebo-controlled, Parallel Group, Multicenter, Phase III Study to Evaluate the Efficacy and Safety of 2 Doses of Benralizumab (MEDI-563) in Patients with Moderate to Very Severe Chronic Obstructive Pulmonary Disease (COPD) with a History of COPD Exacerbations (GALATHEA)**

---

---

**A Randomised, Double-blind, Chronic Dosing (56 week) Placebo-controlled, Parallel Group, Multicenter, Phase III Study to Evaluate the Efficacy and Safety of 2 Doses of Benralizumab (MEDI-563) in Patients with Moderate to Very Severe Chronic Obstructive Pulmonary Disease (COPD) with a History of COPD Exacerbations (GALATHEA)**

---

**Study Statistician**



---

**A Randomised, Double-blind, Chronic Dosing (56 week) Placebo-controlled, Parallel Group, Multicenter, Phase III Study to Evaluate the Efficacy and Safety of 2 Doses of Benralizumab (MEDI-563) in Patients with Moderate to Very Severe Chronic Obstructive Pulmonary Disease (COPD) with a History of COPD Exacerbations (GALATHEA**

---

**Global Product Statistician**



<b>TABLE OF CONTENTS</b>	<b>PAGE</b>
TITLE PAGE.....	1
SIGNATURE OF STUDY STATISTICIAN.....	2
SIGNATURE OF GLOBAL PRODUCT STATISTICIAN .....	3
TABLE OF CONTENTS .....	4
LIST OF ABBREVIATIONS .....	8
AMENDMENT HISTORY.....	11
1. STUDY DETAILS.....	15
1.1 Study objectives.....	15
1.1.1 Primary Objective .....	15
1.1.2 Secondary Objectives.....	15
1.1.3 Safety Objective.....	16
1.1.4 Exploratory Objectives.....	17
1.2 Study design .....	17
1.3 Number of subjects .....	18
2. ANALYSIS SETS .....	19
2.1 Definition of analysis sets .....	19
2.1.1 All patients analysis set.....	19
2.1.2 Full analysis set.....	20
2.1.3 Safety analysis set.....	20
2.1.4 Pharmacokinetic analysis set.....	20
2.1.5 Sub-study analysis set .....	20
2.2 Violations and deviations .....	20
2.2.1 Important protocol deviations.....	20
2.2.2 Visit window definitions .....	21
2.2.3 The definition of baseline.....	24
2.2.4 Prior/concomitant medications .....	25
3. PRIMARY AND SECONDARY VARIABLES.....	25
3.1 Primary outcome variable .....	25
3.2 Secondary efficacy outcome variables.....	27
3.2.1 Time to first COPD exacerbation (days).....	27
3.2.2 Proportion of patients with $\geq 1$ COPD exacerbation during 56 weeks of treatment.....	27
3.2.3 Annual rate of COPD exacerbations that are associated with an emergency room visit or a hospitalization .....	28
3.2.4 FEV <sub>1</sub> (L) pre-bronchodilator measured at the study center.....	28

3.2.5	Healthcare resource utilization due to COPD.....	28
3.3	Patient reported outcome variables.....	29
3.3.1	St. George’s Respiratory Questionnaire (SGRQ).....	29
3.3.2	COPD Assessment Test (CAT).....	30
3.3.3	Baseline/Transitional Dyspnea Index (BDI/TDI).....	30
3.3.4	Exacerbations of Chronic Pulmonary Disease Tool – Patient-reported Outcome (EXACT-PRO).....	31
3.3.5	Evaluating Respiratory Symptoms in Chronic Obstructive Pulmonary Disease (E-RS: COPD).....	31
3.3.6	Total rescue medication use.....	32
3.3.7	Nights with awakening due to respiratory symptoms.....	33
3.3.8	European Quality of Life-5 Dimensions-5 Levels (EQ-5D-5L).....	33
3.3.9	Clinician Global Impression of Change (CGIC) and Patient Global Impression of Change (PGIC).....	33
3.4	Subset analysis outcome variables.....	34
3.4.1	6 Minute Walk Test (6MWT).....	34
3.4.2	Lung volumes.....	34
3.4.3	Diffusion capacity of the lung for carbon monoxide (DL <sub>CO</sub> ).....	34
3.4.4	Sputum biomarkers / sputum cell count.....	35
3.5	Safety outcome variables.....	35
3.5.1	Adverse events.....	35
3.5.2	Laboratory variables.....	36
3.5.3	Twelve-lead electrocardiogram.....	37
3.5.4	Physical examination.....	37
3.5.5	Vital signs.....	37
3.6	Pharmacokinetic variables.....	38
3.7	Immunogenicity variables.....	38
4.	ANALYSIS METHODS.....	38
4.1	General principles.....	38
4.1.1	Testing strategy to account for multiplicity considerations.....	39
4.2	Analysis methods.....	41
4.2.1	Patient disposition.....	41
4.2.2	Demography data and patient characteristics.....	42
4.2.3	Prior and concomitant medications.....	42
4.2.4	Study treatment administration.....	42
4.2.5	Primary outcome variable.....	43
4.2.5.1	Primary analysis.....	43
4.2.5.2	Subgroup analysis for the primary outcome variable.....	45
4.2.5.3	Supportive analysis for the primary outcome variable.....	45
4.2.6	Secondary efficacy outcome variables.....	46
4.2.6.1	Proportion of subjects with ≥1 COPD exacerbation.....	46
4.2.6.2	Time to first COPD exacerbation.....	46

4.2.6.3	FEV <sub>1</sub> (L) pre-bronchodilator measured at the study centre .....	47
4.2.6.4	St. George’s Respiratory Questionnaire (SGRQ) .....	48
4.2.6.5	COPD assessment tool (CAT) .....	49
4.2.6.6	Baseline/Transitional Dyspnea Index (BDI/TDI) .....	49
4.2.6.7	Exacerbations of Chronic Pulmonary Disease Tool – Patient-reported Outcome (EXACT-PRO) .....	50
4.2.6.8	Evaluating Respiratory Symptoms in Chronic Obstructive Pulmonary Disease (E-RS: COPD) .....	50
4.2.6.9	Medication Use .....	50
4.2.6.10	Nights with awakening due to respiratory symptoms .....	51
4.2.6.11	Healthcare resource utilization .....	51
4.2.7	Additional modelling for primary and key secondary endpoints .....	51
4.2.8	Exploratory Objectives .....	52
4.2.8.1	European Quality of Life-5 Dimensions-5 Levels (EQ-5D-5L) .....	52
4.2.8.2	Clinician Global Impression of Change (CGIC) and Patient Global Impression of Change (PGIC) .....	52
4.2.8.3	Blood eosinophils .....	52
4.2.8.4	Serum biomarkers .....	52
4.2.8.5	Pharmacogenetics – DNA Sampling .....	53
4.2.8.6	Mortality rate .....	53
4.2.8.7	Exploratory objectives in a subset of subjects .....	53
4.2.8.8	Sputum biomarkers and differential cell counts .....	54
4.2.9	Safety outcome variables .....	55
4.2.9.1	Adverse events (AEs) .....	55
4.2.9.2	Laboratory data .....	56
4.2.9.3	ECGs .....	57
4.2.9.4	Physical examination .....	57
4.2.9.5	Vital signs .....	57
4.2.10	Pharmacokinetic analyses .....	57
4.2.11	Immunogenicity analyses .....	57
5.	INTERIM ANALYSES .....	58
6.	CHANGES OF ANALYSIS FROM PROTOCOL .....	58
7.	REFERENCES .....	59
8.	APPENDIX .....	62
8.1	Accounting for missing data .....	62
8.2	Partial dates for adverse events and prior/concomitant medication .....	71
8.3	Analysis plan for immunogenicity data .....	73
8.4	Important protocol deviations .....	76

## LIST OF TABLES

Table 1 General analysis-defined visit windows for assessments .....	22
Table 2 Biweekly windows for daily diary assessments.....	23
Table 3 Vital signs reference ranges .....	38

## LIST OF ABBREVIATIONS

<b>Abbreviation or special term</b>	<b>Explanation</b>
6MWT	6 minute walk test
ADA	Anti-drug antibodies
AE	Adverse event
ALP	Alkaline phosphatase
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase
AZDD	AstraZeneca drug dictionary
BD	Bronchodilator
BDI/TDI	Baseline/transitional dyspnea index
BMI	Body Mass Index
BODE	Body-mass index, airflow obstruction, dyspnea, and exercise capacity index
CAT	COPD assessment tool
CGIC	Clinical global impression of change
CI	Confidence interval
CMH	Cochran–Mantel–Haenszel
COPD	Chronic obstructive pulmonary disease
COTE	COPD specific comorbidity test
CSP	Clinical study protocol
CSR	Clinical study report
DAE	Discontinuation of investigational product due to adverse event
DBP	Diastolic blood pressure
DL <sub>CO</sub>	Diffusion capacity of carbon monoxide
DNA	Deoxyribonucleic acid
ECG	Electrocardiogram
ECL	Electrochemiluminescent
eCRF	Electronic case report form
eDiary	Electronic diary
EOT	End of treatment
EQ-5D-5L	European quality of life-5 dimensions-5 levels
ePRO	Electronic patient reported outcome



<b>Abbreviation or special term</b>	<b>Explanation</b>
ER	Emergency room
E-RS: COPD	Evaluating respiratory symptoms in chronic obstructive pulmonary disease
EXACT-PRO	Exacerbations of chronic pulmonary disease tool – Patient-reported outcome
FAS	Full analysis set
FEV <sub>1</sub>	Forced expiratory volume in 1 second
FRC	Functional residual capacity
FU	Follow-up (visit)
FWER	Family-wise error rate
GCP	Good clinical practice
GGT	Gamma-glutamyltransferase
GOLD	Global Initiative for Chronic Obstructive Lung Disease
IC	Inspiratory capacity
ICS	Inhaled corticosteroids
IDMC	Independent data monitoring committee
IL-5	Interleukin-5
IL-5R	Interleukin-5 receptor
IM	Intramuscular
IP	Investigational product
IPD	Investigational product discontinuation
ITT	Intent-to-treat
IV	Intravenous
LABA	Long-acting beta agonist
LAMA	Long-acting muscarinic antagonist
LSMEANS	Least squares means
LOCF	Last observation carried forward
MACE	Major adverse cardiac events
MCID	Minimum clinically important difference
MedDRA	Medical dictionary for regulatory activities
MMRM	Mixed-effect model for repeated measures
N/A	Not applicable
nAb	Neutralizing antibodies

---

<b>Abbreviation or special term</b>	<b>Explanation</b>
PGIC	Patient global impression of change
PK	Pharmacokinetics
PN	Percent normal
PRO	Patient reported outcome
PT	Preferred term
RV	Residual volume
SAE	Serious adverse event
SAP	Statistical analysis plan
SBP	Systolic blood pressure
SC	Subcutaneous
SD	Standard deviation
SGRQ	St. George's respiratory questionnaire
SI	System international (units)
SOC	System organ class
TBL	Total bilirubin
TLC	Total lung capacity
ULN	Upper limit of normal
UK	United Kingdom
US	United States
WBDC	Web based data capture

---

## AMENDMENT HISTORY

Date	Brief description of change
14 December 2015 (v2.0)	<ul style="list-style-type: none"> <li>• SAP text amendment due to CSP amendment 4.0: lowering baseline eosinophil cutoff from <math>\geq 300/\mu\text{L}</math> to <math>\geq 220/\mu\text{L}</math>, corresponding sample size changes, and additional covariate of baseline eosinophil cohort (<math>220\text{-}299/\mu\text{L}</math> or <math>\geq 300/\mu\text{L}</math>) for <math>\geq 220/\mu\text{L}</math> inferential analyses.</li> <li>• Clarification around the process for identifying and summarizing important protocol deviations.</li> <li>• Updates to visit window definitions, taking into account the way in which daily diary metrics are recorded (daytime assessments recorded in the evening and night-time assessments recorded the morning of following calendar day) and ensuring that there are no gaps in visit windows. Update to definition of study day; relative to randomization to align with the way in which study visits are scheduled.</li> <li>• Clarification provided on the definition of baseline for efficacy versus safety endpoints.</li> <li>• Clarification on the definition of prior and concomitant medications.</li> <li>• Clarification provided on summaries of baseline demographic data and patient characteristics by baseline eosinophil groups (<math>&lt;220/\mu\text{L}</math>, <math>220/\mu\text{L}</math>).</li> <li>• Updated the method for primary endpoint from patient-based approach to time-based approach.</li> <li>• Additional details added for the definition of maximum follow-up time for exacerbations and added sensitivity analysis for on-treatment analysis, multiple imputation method details, and additional statistical modelling (primary and key secondary endpoints).</li> <li>• Updated the method for subgroup analyses (primary and key secondary endpoints) to a single model per subgroup factor with inclusion of subgroup main effect and treatment*subgroup interaction, to align with methods described in the protocol.</li> <li>• Updates to expand the list of endpoints to be analysed in patients with baseline blood eosinophil count <math>&lt;220/\mu\text{L}</math>; and also by baseline blood eosinophil count categories (<math>&lt;150/\mu\text{L}</math>, <math>150\text{-}219/\mu\text{L}</math>, <math>220\text{-}299/\mu\text{L}</math>, <math>300\text{-}449/\mu\text{L}</math>, <math>\geq 450/\mu\text{L}</math>), and cumulative baseline blood eosinophil count categories (<math>\geq 150/\mu\text{L}</math>, <math>\geq 200/\mu\text{L}</math>, <math>\geq 250/\mu\text{L}</math>, <math>\geq 300/\mu\text{L}</math>, <math>\geq 350/\mu\text{L}</math>, <math>\geq 400/\mu\text{L}</math>, <math>\geq 450/\mu\text{L}</math>).</li> <li>• Clarification provided on SGRQ total score response status and analysis method.</li> <li>• Analysis of SGRQ and TDI responders updated to evaluate patients with missing data at Week 56 as non-responders.</li> <li>• Clarification on the definition of an EXACT-PRO event, and rescue medication (incorporating morning/evening assessments).</li> <li>• Removal of ANCOVA analysis for E-RS: COPD total score.</li> </ul>

Date	Brief description of change
	<ul style="list-style-type: none"> <li>• Updated analysis of healthcare utilization from annual rates to summary statistics.</li> <li>• Added descriptions and statistical summary analyses on exploratory endpoints for subset analysis.</li> <li>• Updated categorization of adverse events into study periods (on-study, on-treatment, and post-treatment) and clarification of adverse event summaries above/below 220/<math>\mu</math>L.</li> <li>• Updates to incorporate protocol amendments, including: addition of CGIC and PGIC endpoints, removal of ADA analysis set (as analyses of ADA will be conducted in the safety set), removal of analysis of potential risks (included as part of adverse events).</li> <li>• Clarifications on sputum biomarkers / cell count reported separately in a scientific report or publication.</li> <li>• Added immunogenicity analyses and reference to the separate Statistical Analysis Plan for Benralizumab Anti-Drug Antibody data.</li> </ul>

Date	Brief description of change
23 April 2018 (v3.0)	<ul style="list-style-type: none"> <li>• Removed text on 10% drop out rate in sample size calculation erroneously included in the CSP. Actual sample sizes in CSP and SAP are correct and do not account for a 10% drop out rate.</li> <li>• Added definition for sub-study analysis set</li> <li>• Clarified ePRO data will be analyzed up to completion or discontinuation date and exclude data after ICF withdrawal</li> <li>• Updated baseline definition for EQ-5D-5L</li> <li>• Clarified baseline eosinophil levels for stratification are taken from central laboratory. Only if lab data is missing will IVRS stratification be used.</li> <li>• Added imputation rules for partially missing systemic corticosteroid or antibiotic exacerbation start/end dates</li> <li>• Added summary of healthcare resource utilization annual rates</li> <li>• Adding missing data rules for SGRQ and CAT scores</li> <li>• Changed nights with awakenings due to COPD to due to respiratory symptoms, and removed analysis of nights with awakenings requiring rescue medication to align with actual ePRO setup</li> <li>• Updated AE definitions for on-study, on-treatment, and post-treatment periods</li> <li>• Added imputation rules for inferential analyses in case of missing background therapy</li> <li>• Added details of disposition summary</li> <li>• Added marginal standardization method for analyzing negative binomial models</li> <li>• Added weights will be used for analysis by baseline cumulative eosinophil categories</li> <li>• Added supportive/sensitivity analyses to the primary endpoint: exacerbations associated with hospitalization or death, exacerbations with data truncated at the point when subjects switched background therapies, exacerbations</li> <li>• Added subgroup analyses to the primary endpoint: by prior exacerbations, smoking status</li> <li>• Added subgroup analyses for pre-BD FEV<sub>1</sub>: baseline eosinophil categories, cumulative eosinophil categories, and prior exacerbations</li> <li>• Added rules for MMRM covariance structures when convergence issues arise</li> <li>• Added subgroup analyses for SGRQ: baseline eosinophil categories and cumulative eosinophil categories</li> <li>• Added sensitivity analysis for SGRQ responder analysis using LOCF</li> <li>• Added summaries for switches in background therapy during the study</li> <li>• Added sensitivity analyses for primary and key secondary endpoints excluding potential data anomalies</li> <li>• Added plot comparing historical versus baseline eosinophil levels</li> </ul>

<b>Date</b>	<b>Brief description of change</b>
	<ul style="list-style-type: none"><li>• Added sub-study sputum analyses</li><li>• Changed cutoff for most common AEs from <math>\geq 5\%</math> to <math>\geq 3\%</math></li><li>• Added summaries of injection site reaction AEs</li><li>• Added summaries of hypersensitivity AEs causally related with IP</li><li>• Added SAE narratives will be provided</li><li>• Added final list of important protocol deviations in appendix</li><li>• Added all ADA analyses in appendix</li><li>• Removed pDRMI from all MI analyses and MAR from on-treatment MI analyses</li></ul>

## 1. STUDY DETAILS

This statistical analysis plan (SAP) outlines the analyses to be generated for the global clinical study report (CSR). Additional analyses required for regional submissions will be pre-specified in a separate analysis plan and will be submitted to the appropriate authorities.

### 1.1 Study objectives

#### 1.1.1 Primary Objective

<b>Primary Objective:</b>	<b>Outcome Measure:</b>
To evaluate the effect of 2 doses of benralizumab on COPD exacerbations in subjects with moderate to very severe COPD	Annual COPD exacerbation rate, where a COPD exacerbation is defined by symptomatic worsening of COPD requiring: <ul style="list-style-type: none"> <li>• Use of systemic corticosteroids for at least 3 days; a single depot injectable dose of corticosteroids will be considered equivalent to a 3-day course of systemic corticosteroids; and/or</li> <li>• Use of antibiotics; and/or</li> <li>• An inpatient hospitalization or death due to COPD</li> </ul>

#### 1.1.2 Secondary Objectives

<b>Secondary Objective:</b>	<b>Outcome Measure:</b>
To evaluate the effect of 2 doses of benralizumab on health status/ health-related quality of life	<ul style="list-style-type: none"> <li>• St. George's Respiratory Questionnaire (SGRQ)*</li> <li>• COPD assessment tool (CAT)</li> </ul>
To evaluate the effect of 2 doses of benralizumab on pulmonary function	Pre-dose/pre-bronchodilator Forced Expiratory Volume in 1 second (FEV <sub>1</sub> )* at the study center
To evaluate the effect of 2 doses of benralizumab on respiratory symptoms	<ul style="list-style-type: none"> <li>• Baseline/Transitional Dyspnea Index (BDI/TDI)</li> <li>• Evaluating Respiratory Symptoms in Chronic Obstructive Pulmonary Disease (E-RS: COPD)</li> </ul>
To evaluate the effect of 2 doses of benralizumab on rescue medication use	Total rescue medication use (average puffs/day)

<b>Secondary Objective:</b>	<b>Outcome Measure:</b>
To evaluate the effect of 2 doses of benralizumab on nocturnal awakenings	Number of nights with awakening due to COPD
To evaluate the effect of 2 doses of benralizumab on the severity, frequency and duration of EXACT-PRO defined events.	Exacerbations of Chronic Pulmonary Disease Tool – Patient-reported Outcome (EXACT-PRO)
To evaluate the effect of 2 doses of benralizumab on other parameters associated with COPD exacerbations	Time to first COPD exacerbation and proportion of subjects with $\geq 1$ COPD exacerbation
To evaluate the effect of 2 doses of benralizumab on COPD exacerbations involving emergency room visits and hospitalizations	Annual rate of COPD exacerbations that are associated with an emergency room visit or a hospitalization
To evaluate the effect of 2 doses of benralizumab on healthcare resource utilization due to COPD	Annual rate of hospitalizations due to COPD; annual rate of hospitalizations and emergency department visits combined due to COPD; annual rate of unscheduled outpatient visits including unscheduled visits to study centers due to COPD; and annual rate of unscheduled healthcare encounters due to COPD
To evaluate the pharmacokinetics and immunogenicity of 2 doses of benralizumab	<ul style="list-style-type: none"> <li>• Pharmacokinetic (PK) parameters</li> <li>• Anti-drug antibodies (ADA)</li> </ul>

\* Key secondary efficacy endpoints (multiplicity protected endpoints in patients with baseline blood eosinophil counts  $\geq 220/\mu\text{L}$ ).

A summary of PK analysis results will be reported either in the Clinical Study Report (CSR) itself or as an addendum to the CSR.

### 1.1.3 Safety Objective

<b>Safety Objective:</b>	<b>Outcome Measure:</b>
To evaluate the safety and tolerability of 2 doses of benralizumab	<ul style="list-style-type: none"> <li>• Adverse events/ Serious adverse events (AE/SAE)</li> <li>• Laboratory variables</li> <li>• 12 lead Electrocardiogram (ECG)</li> <li>• Physical Examination</li> <li>• Vital Signs</li> </ul>



### 1.1.4 Exploratory Objectives

<b>Exploratory Objective:</b>	<b>Outcome Measure:</b>
To evaluate the effect of 2 doses of benralizumab on general health status	European Quality of Life-5 Dimensions-5 Levels (EQ-5D-5L)
To evaluate the impact of 2 doses of benralizumab on blood eosinophil levels	Blood eosinophils
To evaluate the effect of 2 doses of benralizumab on blood biomarkers	Serum biomarkers
To collect and store DNA for exploratory research into genes/genetic variation that may influence response (ie, distribution, safety, tolerability and efficacy) to benralizumab	Pharmacogenetic sample (see CSP Appendix D)
To evaluate the effect of 2 doses of benralizumab on patient and physician global assessments	Patient Global Impression of Change (PGIC) and Clinical Global Impression of Change (CGIC)
To evaluate the effect of 2 doses of benralizumab on all cause and respiratory related mortality	Mortality rate
To evaluate the effect of 2 doses of benralizumab on post BD FEV <sub>1</sub> in a subset of subjects	Post Bronchodilator (BD) FEV <sub>1</sub>
To evaluate the effect of 2 doses of benralizumab on lung volumes in a subset of subjects	Inspiratory Capacity (IC), Total Lung Capacity (TLC), Residual Volume (RV)
To evaluate the effect of 2 doses of benralizumab on DL <sub>CO</sub> in a subset of subjects	Diffusion of carbon monoxide capacity (DLCO)
To evaluate the effect of 2 doses of benralizumab on exercise endurance in a subset of subjects	6 minute walk test (6MWT)
To evaluate the effect of 2 doses of benralizumab on sputum biomarkers/cell counts and microbiology in a subset of subjects	<ul style="list-style-type: none"> <li>• Sputum biomarkers</li> <li>• Sputum cell counts and differential, microbiome</li> </ul>

The results of the exploratory objectives may be reported outside of the CSR.

## 1.2 Study design

This is a randomised, double-blind, placebo-controlled, parallel group, multicentre, phase III study to evaluate the efficacy and safety of benralizumab (MEDI-563), 30mg and 100mg doses administered by subcutaneous (SC) injection every 4 weeks for the first 3 doses and then every 8 weeks thereafter, with the last dose of IP administered at Week 48 and last treatment evaluation performed at Week 56, in subjects with moderate to very severe COPD receiving standard maintenance therapy (inhaled corticosteroid/long-acting  $\beta$ 2 agonist (ICS/LABA), long-acting  $\beta$ 2 agonist /long acting muscarinic antagonist (LABA/LAMA) or ICS/LABA/LAMA) with a history of COPD exacerbation(s) in the year prior to enrolment (Week -4).

The study will recruit approximately 1626 subjects randomised 1:1:1 to the treatment arms stratified by country and blood eosinophil count ( $\geq 300/\mu\text{L}$  or  $< 300/\mu\text{L}$ ). Subjects will be recruited based on baseline blood eosinophil counts  $< 220/\mu\text{L}$ , 220-299/ $\mu\text{L}$ , and  $\geq 300/\mu\text{L}$ . The approximately 2:1 ratio between  $\geq 220/\mu\text{L}$  and  $< 220/\mu\text{L}$  allows analysis of the primary population for subjects most likely to respond to benralizumab (i.e.,  $\geq 220/\mu\text{L}$ ), while still including subjects below this threshold in order to help understand efficacy and safety in this group. The subject recruitment will also be capped at the study level for the cohorts with baseline eosinophil count ( $< 220/\mu\text{L}$ , 220-299/ $\mu\text{L}$ , or  $\geq 300/\mu\text{L}$ ) in order to keep predefined cohort sample size and approximately 2:1 ratio of subjects above and below the boundary of 220/ $\mu\text{L}$ . Instructions for screen failing these subjects from the study after a stratum is closed are noted in the protocol (Section 4.1.3 of the protocol). The estimated number of recruited subjects per arm in each of these cohorts ( $< 220/\mu\text{L}$ , 220-299/ $\mu\text{L}$ ,  $\geq 300/\mu\text{L}$ ) will be 174, 90, and 278, respectively.

Approximately 10% -15% of subjects (54-81 subjects in each arm of the study) will be included in a study subset that includes advanced induced sputum analysis and advanced pulmonary function testing. The aim of the subset analysis is to further explore the role of eosinophils in COPD and the effect of benralizumab on sputum and pulmonary physiology. Additional analysis on exercise endurance will also be performed on this subset of subjects using the 6 Minute Walk Test. The subset analyses will be performed in a limited number of sites in the US, Canada and UK.

The investigational product (IP) will be administered at the study centre every 4 weeks for the first 3 doses, and then every 8 weeks thereafter. The last dose will be given at Week 48 and the last treatment evaluation will be performed at Week 56.

After the initial enrolment and confirmation of the entry criteria, subjects will proceed to the screening/run-in period for 3 weeks to allow adequate time for all of the eligibility criteria to be evaluated. Subjects who meet the eligibility criteria will be randomised to the 56-week treatment period, with the last dose of benralizumab/placebo administered at Week 48. The end of treatment (EOT) visit will occur at Week 56. Subjects will be maintained on their currently prescribed maintenance therapies, from enrolment throughout the run-in and treatment period.

Final follow-up visit will be conducted at Week 60.

See CSP Section 4, Tables 2 and 3 for a detailed list of visits and assessments.

### **1.3 Number of subjects**

The study will recruit subjects with blood eosinophil counts  $\geq 220/\mu\text{L}$  and  $< 220/\mu\text{L}$  at a ratio of approximately 2:1 and the study is powered for the primary efficacy analysis of the subjects with baseline blood eosinophils  $\geq 220/\mu\text{L}$ . For the primary endpoint annual COPD exacerbation rate, 348 subjects with baseline blood eosinophil counts  $\geq 220/\mu\text{L}$  per treatment arm (1044 total) will need to be randomised to achieve 90% power of detecting a 30% reduction in both benralizumab dose groups versus placebo after multiplicity adjustment. This

calculation has assumed 2-sided 4% alpha level test, an annual placebo rate of 0.93 events/subject (1.0 events/subject/56 weeks), and a negative binomial shape parameter of 0.4. The sample size calculation is based on simulations and has accounted for the power loss from a proposed futility analysis. According to the approximately 2:1 ratio, the study will also enrol 174 subjects/arm (522 total) with baseline blood eosinophil counts  $<220/\mu\text{L}$ . An additional 20 subjects/arm will also be recruited to the 220-299/ $\mu\text{L}$  cohort to better characterize patients within this cohort. This addition results in an overall 368 subjects/arm in the  $\geq 220/\mu\text{L}$  cohort (1104 total). Therefore, a total of 1626 subjects are expected to be randomised in the study.

A futility analysis based on the primary endpoint will be conducted in this study. The futility analysis will be carried out using pre-defined rules by an external Independent Data Monitoring Committee (IDMC) in order to ensure the integrity of the blinded nature of the study. An IDMC charter and analysis plan will be created to detail the data to be reviewed, along with the roles and responsibilities of the IDMC members.

The sample size necessary to achieve a stated power (90% in our case) in this study is calculated based on the estimate of overall exacerbation rate and shape parameter from the negative binomial model. In order to better estimate the overall exacerbation rate and shape parameter, a blinded sample size re-estimation will be conducted before the last subject with baseline eosinophil counts  $\geq 220/\mu\text{L}$  is randomised. The sample size re-estimation will use blinded estimates of the overall exacerbation rate as well as the shape parameter from data pooled across all benralizumab doses and placebo for subjects with eosinophil counts  $\geq 220/\mu\text{L}$  and the potential for  $>180$  days of follow-up time, including prematurely withdrawn subjects. Strictly no treatment information will be used in the review. The exacerbation rate and shape parameter will be estimated using the maximum likelihood approach as proposed by [Friede and Schmidli 2010](#). Since this review will be performed in a blinded fashion, no adjustment for the type I error is needed. Additional blinded review(s) will be conducted if deemed necessary by the sponsor. The blinding will be strictly maintained and not be affected by the review(s) in any measure. Details of the sample size re-estimation is included in a separate blinded data review plan.

## **2. ANALYSIS SETS**

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

Five patient populations are defined below: all patient analysis set, full analysis set (FAS; effectively the ITT population), safety analysis set, pharmacokinetics (PK) analysis set, and sub-study analysis set. Subjects must have provided their informed consent. If no signed informed consent is collected (major protocol deviation), then the subject will be excluded from all analysis sets defined below.

#### **2.1.1 All patients analysis set**

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

### **2.1.2 Full analysis set**

All subjects randomised and who received at least 1 dose of investigational product will be included in the full analysis set, irrespective of their protocol adherence and continued participation in the study. Subjects will be analysed according to their randomised treatment, irrespective of whether or not they have prematurely discontinued, according to the ITT principle. Subjects who withdraw consent, and assent when applicable to participate in the study will be included up to the date of their study termination.

All efficacy analyses will be performed using an ITT approach based on the FAS. For consistency, demographic and baseline characteristics will be presented using the FAS.

### **2.1.3 Safety analysis set**

All subjects who received at least 1 dose of investigational product will be included in the safety analysis set. Subjects will be classified according to the treatment they actually received. A subject who has on 1 or several occasions received active treatment will be classified as active. If a subject has incorrectly received more than 1 active dose, then the subject will be classified as the higher active dose. All safety summaries and anti-drug antibodies (ADA) analyses will be based on this analysis set.

### **2.1.4 Pharmacokinetic analysis set**

All subjects who received benralizumab and from whom PK blood samples are assumed not to be affected by factors such as protocol violations and who had at least 1 quantifiable serum PK observation post first dose will be included in the PK analysis set. All PK summaries will be based on this analysis set.

### **2.1.5 Sub-study analysis set**

This analysis set comprises all subjects who participated in a sub-study that includes advanced induced sputum analysis and advanced pulmonary function testing. This sub-study was performed in approximately 10% -15% of subjects (54-81 subjects in each arm of the study) at a limited number of sites in the US, Canada and UK.

## **2.2 Violations and deviations**

Subjects who do not meet eligibility criteria but are still randomised will be analysed according to the analysis sets described in Section 2.1. There is no intention to perform a per-protocol analysis in this study.

### **2.2.1 Important protocol deviations**

Important protocol deviations are a subset of protocol deviations that may significantly impact the completeness, accuracy, and/or reliability of the study data or that may significantly affect a subject's rights, safety, or well-being.

The final list of important protocol deviations will be documented prior to unblinding the study data, and will include but is not limited to:

- Eligibility criteria not met (patients incorrectly randomised) – deviations from inclusion criteria
- Eligibility criteria not met (patients incorrectly randomised) – deviations from exclusion criteria
- Deviations from IP discontinuation procedures
- Received prohibited/restricted concomitant medication
- Deviations from requirements on background double/triple COPD therapy
- Deviations from visit schedule and study procedures
- IP management and treatment unblinding
- Other (such deviations will be clearly described in the CSR)

Only important protocol deviations will be summarised and listed in the CSR. Important protocol deviations will be reviewed and documented by the medical advisors and statisticians prior to unblinding. Additional details are provided in [Appendix 8.4](#).

Subjects for whom significant protocol deviations were recorded that impact the interpretation of the study safety outcomes will have a footnote added to applicable output to describe the deviation and its potential impact. Such subjects will be identified as part of the protocol deviation review process prior to database lock.

### **2.2.2 Visit window definitions**

For the exacerbation-related analyses, no windows will be applied. For local laboratory and all vital signs, the visit recorded in web based data capture (WBDC) will be used. For endpoints that present visit-based data, the variables will be summarised based on the scheduled days with adjusted analysis-defined visit windows. The adjusted analysis-defined visit windows will be based on the collection schedule listed in the protocol and variables will be windowed to the closest scheduled visit for that variable.

Visit windows have been constructed so that every observation collected can be allocated to a particular visit. No visit windows will be defined for screening visits.

The general adjusted analysis-defined visit windows for assessments are summarised in [Table 1](#).

**Table 1 General analysis-defined visit windows for assessments**

<b>Analysis-Defined Windows Visit</b>	<b>Scheduled Study Day</b>	<b>Maximum Windows</b>
Week 0 Day 1	1	Study Day=1
Week X	$X*7+1=a$	$2 \leq \text{Study Days} \leq ((b-a)/2+a)-1$
Week Y	$Y*7+1=b$	$((b-a)/2+a) \leq \text{Study Days} \leq ((c-b)/2+b)-1$
Week Z (Follow-up)	$Z*7+1=c$	$((c-b)/2+b) \leq \text{Study Days}$

Study day will be defined as follows:

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

For each analysis parameter, the windowing will be based on the protocol-specified schedule of events as defined in Section 4 of the protocol.

If multiple readings are recorded within a single visit window, please refer to the rules below.

- If there are 2 or more observations within the same visit window, then the non-missing 1 closest to the scheduled visit will be used in the analysis.
- If 2 observations are equidistant from the scheduled visit, then the non-missing observation with the earlier collection date will be used in the analysis.
- If 2 observations are collected on the same day, then the non-missing 1 with the earlier collection time will be included in the analysis.

For patient reported outcomes collected twice daily (i.e., rescue medication use), if 2 observations are collected on the same day during the same period (i.e., AM or PM) then the non-missing observation with the earlier collection time during the period will be included in the analysis.

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

For endpoints which are not collected every 4 weeks, such as EQ-5D-5L which is assessed every week, the above rules will be applied to derive adjusted analysis defined visit windows based on the protocol-defined visit schedule for that endpoint.

For pre-bronchodilator FEV<sub>1</sub> (L), the non-missing value with acceptable quality (acceptable or borderline quality grade) which is closest to the scheduled visit will be included in the analysis. For post-bronchodilator FEV<sub>1</sub> assessed in the subset of subjects, the highest value with acceptable quality from the same date as the pre-bronchodilator FEV<sub>1</sub> result will be used.

For the patient-report questionnaires collected by electronic patient reported outcomes (ePRO), the following rules will be applied:



Data for all ePRO assessments will be analysed up to the study completion or study discontinuation date. In the event that data is captured in the ePRO device after the patient has withdrawn consent, all results collected on or after the evening of the date of consent withdrawal will be excluded from analysis.

For daily diary assessments, biweekly means will be calculated using daily diary entries between the scheduled study days contained in Table 2. For daily assessments of rescue medication use, the daytime measurement is captured in the evening and the night-time measurement is captured in the following morning. Therefore, the total rescue medication use span across 2 calendar days. Night-time awakenings are also captured the following morning. All the measurements from the daily metrics must fall within specified AM/PM windows. Any observation recorded after the morning of Study Day 393 will not be included in the analysis but will be listed.

**Table 2 Biweekly windows for daily diary assessments**

Adjusted defined windows visit	Scheduled study day	Maximum windows
Baseline	1	The last 14 days from evening of Study Day -14 to the morning of Study Day 1
Week 2	15	Evening of Study Day 1 to the morning of Study Day 15
Week 4	29	Evening of Study Day 15 to the morning of Study Day 29
Week 6	43	Evening of Study Day 29 to the morning of Study Day 43
Week 8	57	Evening of Study Day 43 to the morning of Study Day 57
Week 10	71	Evening of Study Day 57 to the morning of Study Day 71
Week 12	85	Evening of Study Day 71 to the morning of Study Day 85
Week 14	99	Evening of Study Day 85 to the morning of Study Day 99
Week 16	113	Evening of Study Day 99 to the morning of Study Day 113
Week 18	127	Evening of Study Day 113 to the morning of Study Day 127
Week 20	141	Evening of Study Day 127 to the morning of Study Day 141
Week 22	155	Evening of Study Day 141 to the morning of Study Day 155
Week 24	169	Evening of Study Day 155 to the morning of Study Day 169
Week 26	183	Evening of Study Day 169 to the morning of Study Day 183
Week 28	197	Evening of Study Day 183 to the morning of Study Day 197
Week 30	211	Evening of Study Day 197 to the morning of Study Day 211
Week 32	225	Evening of Study Day 211 to the morning of Study Day 225
Week 34	239	Evening of Study Day 225 to the morning of Study Day 239
Week 36	253	Evening of Study Day 239 to the morning of Study Day 253
Week 38	267	Evening of Study Day 253 to the morning of Study Day 267
Week 40	281	Evening of Study Day 267 to the morning of Study Day 281
Week 42	295	Evening of Study Day 281 to the morning of Study Day 295
Week 44	309	Evening of Study Day 295 to the morning of Study Day 309
Week 46	323	Evening of Study Day 309 to the morning of Study Day 323
Week 48	337	Evening of Study Day 323 to the morning of Study Day 337
Week 50	351	Evening of Study Day 337 to the morning of Study Day 351

Week 52	365	Evening of Study Day 351 to the morning of Study Day 365
Week 54	379	Evening of Study Day 365 to the morning of Study Day 379
Week 56	393	Evening of Study Day 379 to the morning of Study Day 393

---

For overall analyses not based on any particular study visit, all data will be listed and/or analysed, including any repeat or unscheduled visits, unless otherwise specified. For safety endpoints, all post-baseline results will be included in the overall analysis up to and including the follow-up visit. For efficacy endpoints, the post-baseline treatment period will be included up to and including the end of treatment (EOT) visit (Visit 19, Week 56).

### 2.2.3 The definition of baseline

In general, the last recorded value on or prior to the date of randomization will serve as the baseline measurement for efficacy endpoints while the last recorded value prior to first dose of study treatment will serve as the baseline measurement for safety endpoints. If there is no value prior to randomization (or the first dose of study treatment, depending on the endpoint), then the baseline value will not be imputed and will be set to missing. No data known to be collected post first dose will be used in determining the baseline value, unless otherwise specified.

For baseline respiratory and cardiovascular disease characteristics, the latest screening assessment will be used.

For pre-bronchodilator FEV<sub>1</sub> (L), the last non-missing value with acceptable quality (acceptable or borderline quality grade) on or prior to the date of randomization will be used as baseline. The first dose of study treatment is scheduled to be administered on the date of randomization (Visit 4), however if the first dose of study treatment is delayed until after the date of randomization, the last recorded value with acceptable quality prior to first dose of study treatment will be used as baseline measurement for spirometry parameters.

Baseline EQ-5D-5L will be the Visit 4 assessment if available; Visit 3 assessment if Visit 4 assessment is missing; or set to missing if both Visit 3 and Visit 4 assessments are missing.

For EXACT-PRO, the baseline total score will be the mean score for each subject over the 7 days prior to randomization. A minimum of 4 days of data (>50%) is required for calculating the baseline total score. To allow for improvement or deterioration in disease state over the course of the trial, the baseline total score will be reset every 4 weeks in the absence of an EXACT-PRO defined event.

For E-RS: COPD, data collected in the last 14 days prior to randomization will be used to calculate the individual E-RS: COPD total and subscale baseline means. If more than 7 daily measures/scores (>50%) are missing, then baseline is set to missing.

For rescue medication use and nights with awakenings due to respiratory symptoms, baseline is defined as the mean from the last 14 days prior to the date of randomisation and will be



derived from data collected on the evening of Study Day -14 to the morning of Study Day 1. The biweekly mean is calculated as the sum of all non-missing daily measures/scores over 14 sequential days divided by the number of non-missing daily measures/scores. If more than 7 daily measures/scores (>50%) within the baseline period are missing, then baseline is set to missing.

For summaries of laboratory results by visit and in the calculation of change from baseline, the last non-missing assessment prior to first dose of study treatment will be used as baseline. For ECG, the last non-missing measurement at Week -1 (Visit 3) will be the baseline value.

For all data summaries and analyses by baseline blood eosinophil count strata (<220/ $\mu$ L vs.  $\geq$ 220/ $\mu$ L), baseline blood eosinophil count category (<150/ $\mu$ L, 150-219/ $\mu$ L, 220-299/ $\mu$ L, 300-399/ $\mu$ L,  $\geq$ 400/ $\mu$ L), baseline blood eosinophil count cumulative categories (<150/ $\mu$ L,  $\geq$ 150/ $\mu$ L, <220/ $\mu$ L,  $\geq$ 220/ $\mu$ L, <300/ $\mu$ L,  $\geq$ 300/ $\mu$ L, <400/ $\mu$ L,  $\geq$ 400/ $\mu$ L) or continuous baseline blood eosinophil count (i.e., modeling activities), the blood eosinophil count result from the central laboratory will be used. If no data is available, then the stratum recorded in the IVRS system will be used to classify subjects into baseline blood eosinophil count strata (<220/ $\mu$ L vs.  $\geq$ 220/ $\mu$ L), while for patients in the  $\geq$ 220/ $\mu$ L stratum, the IVRS data will be used to further classify subjects (220-299/ $\mu$ L vs.  $\geq$ 300/ $\mu$ L) for inferential analyses where this is included as a covariate in the model. The other baseline blood eosinophil count categories will be set to missing.

#### **2.2.4 Prior/concomitant medications**

Background COPD medication (i.e., double/triple therapy) will be classified as a 'COPD medication at baseline' if it started on or prior to randomization and was ongoing after randomization.

A medication will be regarded as 'prior' if it was stopped on or before the date of randomization (medication stop date  $\leq$  date of randomization).

A medication will be regarded as 'concomitant' if the start date is after the date of randomization, or if it started on or prior to the date of randomization and was ongoing after the date of randomization.

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

#### **3.1 Primary outcome variable**

The annual COPD exacerbation rate will be used as the primary efficacy variable. A COPD exacerbation will be defined as a change in the subject's baseline 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, and leads to any of the following:

- Use of systemic corticosteroids for at least 3 days; a single depot injectable dose of corticosteroids will be considered equivalent to a 3-day course of systemic corticosteroids
- Use of antibiotics
- An inpatient hospitalization or death due to COPD.

Exacerbations will be considered moderate if they require treatment with systemic steroids and/or antibiotics, and do not result in hospitalization or death. Exacerbations will be considered severe if they result in hospitalization or death.

In order to calculate the number of exacerbations experienced by a patient during the 56-week treatment period, the following rule will be applied:

*The start of an exacerbation is defined as the start date of systemic corticosteroids or antibiotic treatment or hospital admission, whichever occurs earlier, and the end date is defined as the last day of systemic corticosteroids or antibiotic treatment or hospital discharge, whichever occurs later. The last date of depot corticosteroid injection will be calculated as the day of injection plus 2 days.*

In the event of partial or missing start or stop dates recorded in the eCRF for systemic corticosteroids, antibiotic treatment, or hospital admission, the partial or missing start or stop dates will be imputed based on the overall exacerbation start and end dates as specified above.

In the primary analysis, the number of exacerbations observed for a patient during the 56-week double-blind treatment period will be used as the response variable.

A COPD exacerbation that occurs  $\leq 7$  days of the last dose of systemic steroids (oral, intramuscular (IM), intravenous (IV)), last dose of antibiotics, or hospital discharge will be counted as the same exacerbation event. Inpatient hospitalization due to COPD occurring during an exacerbation should not be regarded as a new exacerbation. In order to be counted as a new exacerbation it must be preceded by at least 7 days in which none of the exacerbation criterion is fulfilled.

Maximum follow-up time for calculation of the annual exacerbation rate is approximately 56 weeks, defined as the time from randomization up to and including the date of Visit 19 (EOT visit at Week 56). For patients who discontinue study treatment and remain in the study after the IPD visit, exacerbations will be counted from the time of randomization up to and including the date of Visit 19 (Week 56). For a patient lost to follow-up before Visit 19 (Week 56), the follow-up time will be defined as the time from randomization to the time point after which an exacerbation could not be assessed. Exacerbations that start after Visit 19 (Week 56) will not be included in the efficacy assessments but will be listed. If a subject misses Visit 19 (Week 56), then any exacerbations that start after the scheduled Visit 19 (Week 56) date (i.e., Study Day 393) will be excluded from efficacy assessments. If an exacerbation is ongoing at Visit 19 (Week 56), the exacerbation will be counted in the calculation of annual exacerbation

rate, however the maximum follow-up time will be truncated at the date of Visit 19 (Week 56), as will the duration of the exacerbation.

For the production of summary statistics, the annual exacerbation rate in each treatment group will be calculated using the time-based approach below:

*Annual Exacerbation Rate = 365.25\*Total Number of Exacerbations / Total duration of follow-up within the treatment group (days)*

The on-treatment annual exacerbation rate will be calculated similarly, as a sensitivity analysis, using only exacerbations and follow-up occurring during the on-treatment period as defined in Section 3.5.1.

An additional sensitivity analysis will be performed to evaluate the impact of out-of-window EOT visits on the annual exacerbation rate. For this analysis, the exacerbation follow up period will be truncated at Day 406 (according to the analysis visit windowing described in Section 2.2.2 for the EOT visit at Week 56). Exacerbations recorded with start date after Day 406 will be excluded from this sensitivity analysis.

## **3.2 Secondary efficacy outcome variables**

### **3.2.1 Time to first COPD exacerbation (days)**

Time from randomization to the first COPD exacerbation will be used as a supportive variable to the primary outcome variable, and is calculated as follows:

*Start Date of first COPD exacerbation – Date of Randomization + 1.*

An exacerbation event will be defined in the same way as outlined in Section 3.1. The time to first COPD exacerbation for patients who do not experience a COPD exacerbation during the treatment period will be censored at the date of their last visit for the 56-week double-blind treatment period, and for lost-to-follow-up patients, at the time point after which an exacerbation could not be assessed or at 56 weeks, whichever occurs first.

Time to first COPD exacerbation associated with a hospitalization or ER visit will be derived using the same approach.

### **3.2.2 Proportion of patients with $\geq 1$ COPD exacerbation during 56 weeks of treatment**

The proportion of patients with  $\geq 1$  COPD exacerbation during the 56 weeks of treatment will also be a supportive variable to the primary outcome variable.

An exacerbation event will be defined in the same way as outlined in Section 3.1. In the statistical analysis, a binary variable taking on the value 1 if a patient has experienced 1 or more exacerbations during the 56-week double-blind treatment period and 0 otherwise, will be used as response variable.

### **3.2.3 Annual rate of COPD exacerbations that are associated with an emergency room visit or a hospitalization**

The annual rate of COPD exacerbations that are associated with an emergency room (ER) visit or a hospitalization will be a secondary efficacy variable, based on data recorded in the eCRF.

The number of COPD exacerbations associated with a hospitalization or ER visit experienced by a patient during the 56-week treatment period and maximum follow-up time used in the calculation of this annual exacerbation rate will be derived according to the rule for the primary outcome variable in Section 3.1.

### **3.2.4 FEV<sub>1</sub> (L) pre-bronchodilator measured at the study center**

FEV<sub>1</sub> (L) pre-bronchodilator is a key secondary efficacy endpoint of this study, and the change from baseline to Week 56 is included in the multiple testing strategy.

The change from baseline to each of the post-randomization visits (post Week 0, Visit 4) up to and including the end of the 56-week double-blind treatment visit (Week 56, Visit 19) will be used as the key secondary efficacy variable. Details regarding the definition of baseline for FEV<sub>1</sub> pre-bronchodilator is provided in Section 2.2.3.

### **3.2.5 Healthcare resource utilization due to COPD**

COPD related healthcare utilization information will be collected by the Investigator/authorized delegate at each visit as specified in the protocol and recorded in the appropriate eCRF module.

The number of days/times the following resources were utilized and corresponding annual rates will be presented for each subject:

- Ambulance transport
- Hospitalization, intensive care (days in intensive care)
- Hospitalization, general care (days in general care)
- Emergency room visit
- Visit to specialist
- Visit to primary health care physician
- Other health care visit
- Home visit, physician
- Home visit, other health care

- Telephone call to physician
- Telephone call to nurse
- Telephone contact with other physician/health care provider

If multiple healthcare encounters are associated with 1 COPD exacerbation all the encounters will be counted for this endpoint.

### 3.3 Patient reported outcome variables

COPD patient reported outcomes (PROs) are captured during site visits or using an electronic patient diary (i.e., ePRO). PROs at site visits include SGRQ, CAT, BDI/TDI, and PGIC. ePROs include EXACT-PRO, E-RS: COPD, daily diary metrics (rescue medication use and night-time awakenings), and EQ-5D-5L. Daily diary metrics will be recorded each day from the evening of Visit 2 to the morning of Visit 19.

The EQ-5D-5L will be completed weekly (7±2 days) starting from Week 0 (Visit 4) throughout Week 56 (Visit 19).

Baseline is defined as the last non-missing value before randomization for the COPD PROs and as the average of the last 14 days before randomization for the daily metrics, as defined in Section 2.2.3. Post-randomization periods for the daily metrics will be defined for the calculation of biweekly means using the analysis-defined visit windows described in Section 2.2.2 and listed in Table 2.

The post-randomization biweekly means for daily metrics are calculated as the sum of all non-missing daily measures/scores over the 14 day window divided by the number of non-missing daily measures/scores. If more than 7 daily measures/scores (>50%) within that window are missing, then the mean daily measure/score for that period is set to missing. The change from baseline to each post-randomization period will be used as secondary efficacy variables. No imputations will be made for missing values in the endpoint derivation although a multiple imputation sensitivity analysis is planned for SGRQ total score (see Section 4.2.6.4).

#### 3.3.1 St. George's Respiratory Questionnaire (SGRQ)

The change from baseline in SGRQ is a key secondary efficacy endpoint of this study, and the change from baseline to Week 56 is included in the multiple testing strategy.

The SGRQ is a 50-item PRO instrument developed to measure the health status of subjects with airway obstruction diseases (Jones et al. 1991). The questionnaire is divided into 2 parts: part 1 consists of 8 items pertaining to the severity of respiratory symptoms in the preceding 4 weeks; and part 2 consists of 42 items related to the daily activity and psychosocial impacts of the individual's respiratory condition. The SGRQ yields a total score and 3 domain scores (symptoms, activity, and impacts). The total score indicates the impact of disease on overall health status. This total score is expressed as a percentage of overall impairment, in which 100 represents the worst possible health status and 0 indicates the best possible health status.

Likewise, the domain scores range from 0 to 100, with higher scores indicative of greater impairment. Specific details on the scoring algorithms are provided by the developer in a user manual (Jones et al. 2009).

The symptoms domain score will be set to missing if there are >2 missing items; activity domain score will be set to missing if there are >4 missing items; impacts domain score will be set to missing if there are >6 missing items; and total score will be set to missing if one of the 3 domain scores is missing.

Potential health status treatment benefits of benralizumab will be evaluated by comparing the change from baseline at Week 56 in SGRQ total score. A 4-point threshold will be used to define the response. If SGRQ total score at Week 56 has a  $\geq 4$ -point decrease from baseline, it is defined as “improvement”; if it has a  $\geq 4$ -point increase, it is defined as “worsening”. If the SGRQ total score change at Week 56 is not more than 4-points, it will be defined as “no change”. Missing SGRQ total score change at Week 56 will be considered as “not evaluable”. For the responder analysis of SGRQ, a responder will be defined as an individual who had “improvement” at Week 56 (ie,  $\geq 4$ -point decrease in SGRQ total score at Week 56). Patients who had SGRQ total score change defined as “no change” or “worsening” will be considered as non-responders. If SGRQ total score change at Week 56 is not evaluable due to missing data, then the patient will also be treated as non-responder.

### **3.3.2 COPD Assessment Test (CAT)**

The CAT is an 8-item PRO developed to measure the impact of COPD on health status (Jones et al. 2009). The instrument uses semantic differential 6-point response scales, which are defined by contrasting objectives to capture the impact of COPD. Content includes items related to cough, phlegm, chest tightness, breathlessness going up hills/stairs, activity limitation at home, confidence leaving home, sleep, and energy. The CAT total score is the sum of all item responses. Scores range from 0 to 40 with higher scores indicative of greater COPD impact on health status. A score cannot be calculated if >2 responses are missing; when one or two items are missing, their scores can be set to the average of the non-missing item scores.

### **3.3.3 Baseline/Transitional Dyspnea Index (BDI/TDI)**

The BDI/TDI is an instrument developed to provide a multidimensional measure of dyspnea in relation to activities of daily living (Mahler et al. 1984). The Baseline Dyspnea Index (BDI) provides a measure of dyspnea at a single state, the baseline, and the Transitional Dyspnea Index (TDI) evaluates changes in dyspnea from the baseline state. The instrument consists of three components: functional impairment, magnitude of task, and magnitude of effort. For the BDI, each of these three components are rated in five grades from 0 (severe) to 4 (unimpaired), and are summed to form a baseline focal (total) score from 0 to 12. The BDI will be captured at Week -4 and Week 0 only. For the TDI, changes in dyspnea are rated for each component by seven grades from -3 (major deterioration) to +3 (major improvement), and are added to form a focal (total) TDI score from -9 to +9. Positive scores indicate an improvement, and a change from the BDI or a difference between treatments of 1 point has



been estimated to constitute the minimum clinically important difference (MCID) (Mahler et al. 2005).

A TDI responder will be defined as a patient who had a change from the BDI of at least 1 positive point at Week 56. For a patient who had a change of less than 1 positive point or non-evaluable change due to missing measurement at Week 56, the patient will be considered as a non-responder.

### **3.3.4 Exacerbations of Chronic Pulmonary Disease Tool – Patient-reported Outcome (EXACT-PRO)**

The EXACT-PRO is a 14-item PRO instrument developed to assess the frequency, severity and duration of COPD exacerbations (Jones et al. 2011, Leidy et al. 2011). The instrument was developed for daily, at home, administration using a handheld electronic device. Respondents are instructed to complete the electronic diary (eDiary) each evening just prior to bedtime and to answer the questions while considering their experiences “today”. The daily EXACT-PRO total score has a range of 0-100 with higher scores indicative of greater severity. Total score changes are used to identify the onset and recovery from an EXACT-PRO defined exacerbation event. In identifying event onset and recovery, the EXACT-PRO can provide information on event frequency and duration as well as event severity.

The number, average duration, and severity of EXACT-PRO defined events will be evaluated. EXACT-PRO daily total scores will be calculated according to the developer approved scoring algorithm. The total score will be used to identify event onset and recovery as well as the magnitude (severity) of the event. The baseline total score will be the mean score within subject over the 7 days prior to randomization. A minimum of 4 days of data is required for calculating the baseline total score. To allow for improvement or deterioration in disease state over the course of the trial, the baseline total score will be reset every 4 weeks in the absence of an EXACT-PRO defined event. Event frequency is calculated by comparing the baseline with daily total scores. An increase in EXACT-PRO total score  $\geq 9$  for 3 days or  $\geq 12$  for 2 days indicate an event has occurred. Calculating event duration requires identification of the following 5 parameters: 1) onset; 2) 3-day rolling average; 3) maximum observed value; 4) threshold for improvement; and 5) recovery. The severity of an event is indicated by the worst (highest) EXACT-PRO total score during an event. Complete details concerning variable calculation are provided in the scoring manual (Evidera, Inc. 2014).

For the production of summary statistics, the annual EXACT-PRO defined exacerbation rate in each treatment group will be calculated using the same time-based approach as specified in Section 3.1.

### **3.3.5 Evaluating Respiratory Symptoms in Chronic Obstructive Pulmonary Disease (E-RS: COPD)**

The E-RS: COPD is an 11-item PRO developed to evaluate the severity of respiratory symptoms of COPD (Sexton et al. 2010, Sexton et al. 2011). The E-RS: COPD is a subset of items from the EXACT-PRO. The E-RS: COPD was designed to be captured as part of the daily EXACT-PRO assessment. Summation of E-RS: COPD item responses produces a total

score ranging from 0 to 40, with higher scores indicating greater severity. In addition to the total score, symptom domain scores can be calculated for breathlessness (5 items; score range: 0–17), cough and sputum (3 items; score range: 0–11) and chest symptoms (3 items; score range: 0–12) by summing the responses of items within a respective domain. As with the total score, higher domain scores indicate greater severity.

Individual daily E-RS: COPD total and subscale scores will be calculated and summarised as a biweekly (14 day) mean.

The biweekly mean is calculated as the sum of all non-missing daily measures/scores over 14 sequential days divided by the number of non-missing daily measures/scores. If more than 7 daily measures/scores (>50%) within that period are missing, then the mean daily measure/score for that period is set to missing.

### **3.3.6 Total rescue medication use**

The number of rescue medication inhalations and nebulizer treatments taken will be recorded by the patient in the eDiary twice daily. Daytime use is recorded in the evening, and night-time use is recorded in the morning of the next calendar day. Total rescue medication use is the sum of daytime and night-time use. If either daytime or night-time use is missing, then total rescue medication use for that day is set to missing.

Rescue medication usage will be summarised as the number of puffs with 1 instance of nebulizer use converted to 2 puffs. Inhaler usage will be reported as the number of puffs in a given period whereas nebulizer use will be reported as the number of times. The number of rescue medication inhalations and nebulizer treatments captured in the eDiary each day will be calculated per patient.

Total rescue medication use (inhaler and/or nebulizer), defined as the number of puffs per day will be calculated as follows:

*Number of daytime inhaler puffs (recorded in the evening) + 2 x [number of daytime nebulizer times (recorded in the evening)] + number of night-time inhaler puffs (recorded the next morning) + 2 x [number of night-time nebulizer times (recorded the next morning)]*

Total reliever inhaler puffs per day will be calculated as:

*Number of daytime inhaler puffs (recorded in the evening) + number of night-time inhaler puffs (recorded the next morning)*

Total nebulizer use (number of times) per day will be calculated as:

*Number of daytime nebulizer times (recorded in the evening) + number of night-time nebulizer times (recorded in the next morning)*

Biweekly mean rescue medication use (average puffs/day in total use) and change from baseline in the biweekly mean rescue medication use will be calculated.



### **3.3.7 Nights with awakening due to respiratory symptoms**

Nocturnal awakenings due to respiratory symptoms will be recorded by the patient in the eDiary each morning by answering a question about whether he/she woke up during the night due to respiratory symptoms by a “yes” or “no” response.

The biweekly proportion of nights with nocturnal awakenings due to respiratory symptoms with non-missing night-time awakening data, and the corresponding change from baseline for each post-randomization period will be calculated.

### **3.3.8 European Quality of Life-5 Dimensions-5 Levels (EQ-5D-5L)**

The EQ-5D-5L questionnaire assesses 5 dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has 5 response options (no problems, slight problems, moderate problems, severe problems, and extreme problems) that reflect increasing levels of difficulty.

The subject will be asked to indicate his/her current health state by selecting the most appropriate level in each of the 5 dimensions. The questionnaire also includes a visual analogue scale, where the subject will be asked to rate current health status on a scale of 0 to 100, with 0 being the worst imaginable health state.

The questionnaire will be completed by the patients using the ePRO device every week throughout the 56-week treatment period.

### **3.3.9 Clinician Global Impression of Change (CGIC) and Patient Global Impression of Change (PGIC)**

Clinician Global Impression of Change (CGIC) and Patient Global Impression of Change (PGIC) instruments are used for an overall evaluation of response to treatment, conducted separately by the Investigator and by the patient using a 7-point scale: 1=Very Much Improved; 2=Much Improved; 3=Minimally Improved; 4=No Change; 5=Minimally Worse; 6=Much Worse; and 7=Very Much Worse.

The Investigator (clinician) and the patient will be asked to rate the degree of change in the overall COPD status compared to the randomization visit. The CGIC and PGIC assessments were added through a protocol amendment, therefore not all subjects in the full analysis set will have these assessments. Calculation of percentages will be based on the number of subjects in the full analysis set with a completed assessment. There will be no imputation for missing values.

The questionnaire will be completed by the patients and clinicians at Weeks 16, 24, 32, and EOT.

Patients will also be categorized according to the following responses post-baseline, separately for CGIC and PGIC:

- Very much improved, much improved, minimally improved → ‘Improved’
- Very much improved, much improved → ‘Much improved’
- Very much improved → ‘Very much improved’

Agreement between CGIC and PGIC will be assessed at each visit, where agreement is achieved when both the patient and clinician provide the same response (e.g., if both the patient and clinician indicate a response of 1 (very much improved) at a particular visit, agreement is achieved for that visit). Agreement will also be assessed for categorized responses at each visit.

### **3.4 Subset analysis outcome variables**

Additional assessments will be performed for a subset of subjects at designated sites, with approximately 10-15% (54-81 subjects) in each treatment group.

#### **3.4.1 6 Minute Walk Test (6MWT)**

The 6 Minute Walk Test (6MWT) measures the distance that a patient can quickly walk on a flat, hard surface in a period of 6 minutes. It evaluates the global and integrated responses of all the systems involved during exercise, including the pulmonary and cardiovascular systems. The self-paced 6MWT assesses the submaximal level of functional capacity. The Borg scales for breathlessness and fatigue assesses a subject’s level of shortness of breath and fatigue, respectively, on a scale of 0 (no symptoms at all) to 10 (very, very severe (maximal) symptoms). The breathlessness and fatigue scores will be measured before and after a subject performs the 6MWT. For subjects who discontinue IP, the last study assessment of 6MWT will be performed at the IPD visit.

#### **3.4.2 Lung volumes**

Lung volume subdivisions which include total lung capacity (TLC), residual volume (RV), vital capacity (VC), functional residual capacity (FRC), and inspiratory capacity (IC) will be performed at the study center by the investigator or qualified designee according to the study schedule. Lung volumes will be determined by body plethysmography. At least 3 FRC values that agree within 5% (i.e., the difference between the highest and lowest value divided by the mean is  $<0.05$ ) will be obtained and the mean value reported. For subjects who discontinue IP, the last study assessment of lung volumes will be performed at the IPD visit.

#### **3.4.3 Diffusion capacity of the lung for carbon monoxide (DL<sub>CO</sub>)**

Diffusion capacity for carbon monoxide (DL<sub>CO</sub>) will be performed at the study center according to the study schedule. The single breath technique will be used to determine the DL<sub>CO</sub>. Acceptable test criteria include an inspiratory volume of more than 85% of VC; a stable breath hold of 10±2 seconds with no leaks, Valsalva or Mueller maneuvers; and expiration in less than 4 seconds with appropriate clearance of dead space. The average of the 2 best acceptable maneuvers will be used. There will be a minimum of 4 minutes between the

performance of each test. For subjects who discontinue IP, last study assessment of DL<sub>CO</sub> will be performed at the IPD visit.

#### **3.4.4 Sputum biomarkers / sputum cell count**

Sputum will be collected according to the study schedules. Sputum will be analysed to determine the differential cell count as well as to assess the levels of benralizumab related pharmacodynamic biomarkers, biomarkers of eosinophil and basophil activation, eosinophil and basophil growth factors, eosinophil recruitment, and inflammation associated with COPD including cytokines, chemokines, acute phase response proteins, and other inflammatory mediators. For subjects who discontinue IP but agree to continue with regular on-site visits, sputum biomarkers and sputum cell count assessments will be performed according to study schedule until the end of the study (Week 56).

### **3.5 Safety outcome variables**

The following safety data will be collected: reported AEs (including SAEs), clinical chemistry, haematology, urinalysis, 12-lead electrocardiogram (ECG), physical examination and vital signs.

All safety measurements will use all available data for analyses, including data from unscheduled visits and repeated measurements.

Change from baseline to each post-treatment time point where scheduled assessments were made will be calculated for relevant measurements. AEs will be summarised by means of using descriptive statistics and qualitative summaries.

No safety data will be imputed. The handling of partial/missing dates for AEs and prior/concomitant medications is detailed in Appendix 8.2.

#### **3.5.1 Adverse events**

Adverse events experienced by the patients will be collected throughout the entire study and will be coded using the latest version of the Medical Dictionary for Regulatory Activities (MedDRA) per the Data Management Plan.

Adverse event data will be categorized according to their onset date into the following study periods:

- AEs in the on-treatment period are defined as those with onset date between day of first dose of study treatment and the day of the scheduled end of treatment (EOT) visit (defined by the CSP as Week 56 +/- 7 days) for patients who complete study treatment. For instances where a patient does not attend EOT visit within the protocol defined window or prematurely discontinues IP, AEs occurring on or before the last dose of study medication +56 days will be assigned to the on-treatment period, while AEs with onset date after this date will be assigned to the post-treatment period.

- AEs in the on-study period are defined as those with onset date on or after the day of first dose of study treatment.
- AEs in the post-treatment period are defined as those with onset date after the on-treatment period defined above.

If an AE has a missing onset date then unless the stop date of the AE indicates otherwise, this will be considered an ongoing AE. Similarly, if an AE has a partial onset date, then unless the partial onset date or the stop date indicates otherwise, this will be considered an ongoing AE.

### 3.5.2 Laboratory variables

Blood and urine samples for determination of clinical chemistry, haematology and urinalysis parameters will be taken at the times detailed in the CSP, and will be assessed in a central laboratory. The parameters outlined in Section 5.2.6, Tables 4 and 5 of the CSP will be collected.

In summaries, figures, and listings, lab results and normal ranges will be presented in System International (SI) units. Eosinophils data will be presented in both SI and conventional units (cells/ $\mu$ L) in summaries.

For the purposes of clinical chemistry, haematology, and urinalysis shift tables, baseline will be defined as the last available non-missing assessment prior to first dose of randomised treatment, and maximum or minimum value post-baseline will be calculated over the entire post-baseline period, including the post-treatment period.

Changes in haematology and clinical chemistry variables between baseline and each post-baseline assessment will be calculated. The change from baseline is defined as the post-baseline visit value minus the baseline visit value. There will be no imputation for missing values. For values recorded with a leading greater than or less than ('>', '<') symbol, the reported numeric value will be used for analysis and the value with the symbol will be included in the listings, unless otherwise specified. For example, a value of <0.01 will be analysed as 0.01 and listed as <0.01.

Absolute values will be compared to the relevant reference range and classified as low (below range), normal (within range or on limits) or high (above range). The central reference ranges will be used for laboratory variables. All values (absolute and change) falling outside the reference ranges will be flagged.

Urinalysis data categorized as negative (0), positive (+), or strongly positive (++, +++, or >+++) or equivalent categories as reported by the central lab.

For the liver function tests: aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase (ALP), gamma-glutamyltransferase (GGT) and total bilirubin (TBL), the multiple of the central laboratory upper limit of the normal (ULN) range will be calculated for each data point.

$$\text{Multiple} = \text{Value} / \text{ULN}$$

That is, if the ALT value was 72 IU/L (ULN=36) then the multiple would be 2.

Patients who meet any of the following criteria at any point during the study will be flagged:

- $\text{AST} \geq 3 \times \text{ULN}$
- $\text{ALT} \geq 3 \times \text{ULN}$
- $\text{TBL} \geq 2 \times \text{ULN}$

### **3.5.3 Twelve-lead electrocardiogram**

Twelve-lead electrocardiogram (ECG) measurements will be recorded in accordance with the protocol, with the baseline visit being defined as the last available non-missing measurement prior to first dose of randomised treatment.

The outcome of the overall evaluation is to be recorded as normal/abnormal in the eCRF by the Investigator/authorized delegate, with any abnormalities being recorded as not clinically significant or clinically significant.

The ECG will also be transferred to a central reader.

### **3.5.4 Physical examination**

Complete and brief physical examinations will be performed at time points specified in Tables 2 and 3 of the CSP. Assessments will be dependent on whether the examination is complete or brief, as described in Section 5.2 of the CSP. For the brief physical examination, only information on whether the assessment was performed or not is to be recorded.

Each component of the baseline complete physical examination will be recorded as normal or abnormal. Each component of the follow-up complete physical examinations will be recorded as normal, same as baseline, or new/aggravated.

Any new finding(s) or aggravated existing finding(s), judged as clinically significant by the Investigator, will be reported as an AE.

### **3.5.5 Vital signs**

Pre-dose vital signs (pulse, systolic blood pressure, diastolic blood pressure, respiration rate, and body temperature) will be obtained in accordance with the visit schedule provided in the CSP.

Changes in vital signs variables between baseline and each subsequent scheduled assessment will be calculated. Baseline is defined as the last value prior to the first dose of randomised treatment. The change from baseline is defined as the post-baseline visit value minus the baseline visit value. There will be no imputation for missing values.

Absolute values will be compared to the reference ranges in [Table 3](#) and classified as low (below range), normal (within range or on limits) or high (above range). All values (absolute and change) falling outside the reference ranges will be flagged.

**Table 3 Vital signs reference ranges**

Parameter	Standard Units	Lower Limit	Upper Limit
Diastolic Blood Pressure (DBP)	mmHg	60	120
Systolic Blood Pressure (SBP)	mmHg	100	160
Pulse Rate	Beats/min	40	120
Respiratory Rate	Breaths/min	8	28
Body Temperature	Celsius	36.5	38
Weight	kg	40	200

Body mass index (BMI) will be calculated from the height and weight as follows:

$$\text{BMI (kg/m}^2\text{)} = \text{weight (kg)} / (\text{height (m)})^2$$

### 3.6 Pharmacokinetic variables

Blood samples (processed to serum) for pharmacokinetic assessments will be collected from all subjects at baseline prior to first benralizumab administration at Week 0 Day 1, at Weeks 4, 8, 16, 24, 32, 40, 48, and 56 before benralizumab administrations during the treatment period, and at the Week 60 follow-up visit. Serum concentrations of benralizumab will be determined using a validated electrochemiluminescent (ECL) immunoassay. Results below the lower limit of quantification (BLQ) will be set to LLOQ/2 for analysis and will be listed as <LLOQ.

### 3.7 Immunogenicity variables

Anti-drug antibodies (ADA) variables, such as ADA responses, will be generated and analysed as per the details in [Appendix 8.3](#).

## 4. ANALYSIS METHODS

### 4.1 General principles

The analysis of the primary and secondary efficacy endpoints will include all data captured during the 56-week double-blind treatment period, defined as the period after administration of randomised investigational product at Visit 4 (Week 0) and the conclusion of Visit 19 (Week 56), inclusive. This includes data regardless of whether study treatment was prematurely discontinued, or delayed, and/or irrespective of protocol adherence, unless the patient withdraws consent to study participation. The statistical analyses will compare each dosing regimen of benralizumab to placebo. The analysis of secondary safety endpoints will

include all data captured during the on-study period, defined as the period after first administration of randomised investigational product at Visit 4 (Week 0) and the conclusion of the scheduled post-treatment follow-up visit (Week 60), inclusive.

The data analyses will be conducted using the SAS® System (SAS Institute Inc., Cary, NC). All SAS® programs used to generate analytical results will be developed and validated according to AstraZeneca SAS® programming standards and validation procedures. Pharmacokinetic analyses will be performed using NONMEM or other appropriate software.

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

Following ongoing review of protocol deviations during the trial conduct, a small number of randomised subjects were identified as not on protocol-specified double (ICS/LABA, LABA/LAMA) or triple (ICS/LABA/LAMA) therapy at baseline. As such, these subjects have missing values for the background therapy covariate used in the statistical analyses. In order to ensure all patients in the full analysis set contribute to the primary and secondary endpoint inferential analyses, the following imputation rules will be applied. Each of these patients will be documented as having an important protocol deviation and summarised as described in Section 2.2.1.

- Subjects on ICS alone will be categorized as ICS/LABA
- Subjects on LABA alone will be categorized as LABA/LAMA
- Subjects on LAMA alone will be categorized as LABA/LAMA
- Subjects on ICS/LAMA alone will be categorized as ICS/LABA

All hypothesis testing will be reported using 2-sided tests. All p-values will be nominal and will be displayed in SAS pvalue6.4 format.

The absolute change from baseline is computed as (*visit value – baseline value*). Percent change from baseline is computed as  $((\text{visit value} - \text{baseline value}) / \text{baseline value}) \times 100\%$ . If either a visit value or the baseline visit value is missing, the absolute change from baseline value and the percent change from baseline will also be set to missing. If baseline value is zero, the percent change will be set to missing.

#### **4.1.1 Testing strategy to account for multiplicity considerations**

To account for multiplicity to test the primary (annual COPD exacerbation rate) and 2 key secondary endpoints (change from baseline in FEV<sub>1</sub> (L) pre-bronchodilator at Week 56 and change from baseline in SGRQ total score at Week 56), for each of the 2 dosing regimens (for patients with baseline blood eosinophils  $\geq 220/\mu\text{L}$ ), a testing strategy will be followed to



control the overall type I error rate. The testing strategy will be according to the following gatekeeping procedure:

Step 1: Perform the 2 tests of annual COPD exacerbation rate (1 test for each dose regimen vs. placebo) at the family wise error rate (FWER) of 0.04 using a Hochberg Procedure (

[Hochberg 1988](#)). If both p-values are less than 0.04, then proceed to Step 2; else if the smaller p-value is less than 0.02 then proceed to Step 2a; otherwise no null hypothesis is rejected.

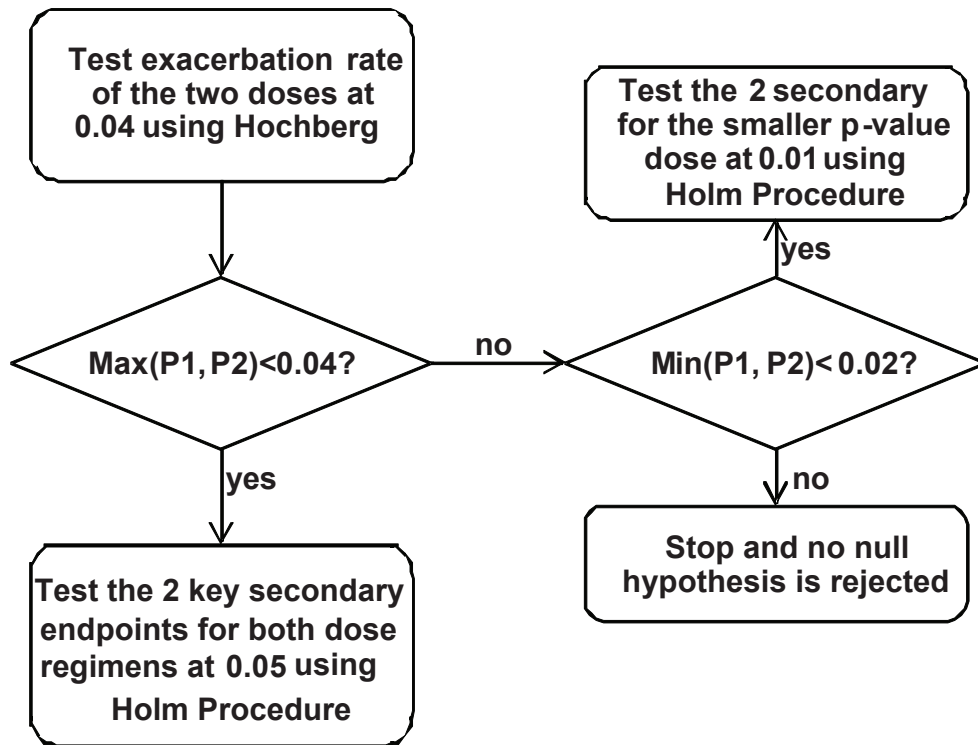
Step 2: Test the 2 key secondary endpoints for both dose regimens as 1 family at the FWER of 0.05 using a Holm Procedure ([Holm 1979](#)).

Step 2a: Test the 2 secondary endpoints for the smaller-p-value dose at the FWER of 0.01 using a Holm Procedure.

Since the correlation of the 2 test statistics for the annual COPD exacerbation rate in Step 1 is positive, due to the common placebo group, the FWER of the Hochberg Procedure in Step 1 is strongly controlled at 0.04. The overall FWER of the gatekeeping procedure is strongly controlled at 0.05.



The gatekeeping procedure is also shown in the diagram below:



## 4.2 Analysis methods

### 4.2.1 Patient disposition

Patient disposition will be summarised using the all patients analysis set and separately for patients in the full analysis set with baseline eosinophil count  $\geq 220/\mu\text{L}$  and  $< 220/\mu\text{L}$ . The total number of patients will be summarised for the following groups: those who enrolled and those who were not randomised (and reason). The number and percentage of patients within each treatment group will be presented by the following categories: randomised, did not receive treatment with study drug, received treatment with study drug, completed treatment with study drug, completed treatment but discontinued study, discontinued treatment with study drug (and reason), discontinued treatment with study drug but completed study, completed study, and withdrawn from study (and reason).

Screen failure information will be listed for the all patients analysis set.

The number of patient randomised by country and centre will also be summarised by treatment group in the full analysis set.

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

Demography data such as age, gender, race, and ethnicity will be summarised by treatment group for all patients in the full analysis set, and for patients in the full analysis set with baseline blood eosinophil counts  $\geq 220/\mu\text{L}$  and baseline blood eosinophil counts  $< 220/\mu\text{L}$ . Age will be derived from the date of informed consent-date of birth, rounded down to the nearest integer. For patients in countries where date of birth is not recorded, the age as recorded in the electronic case report form (eCRF) will be used.

Medical, surgical, and cardiovascular histories (including comorbidities of arrhythmia, ischemic or structural cardiac disease, and systemic inflammatory and peripheral vascular disorder) will be summarised by MedDRA PT within MedDRA SOC in the full analysis set

The following will also be summarised for patients in the full analysis set and repeated for patients with baseline blood eosinophil counts  $< 220/\mu\text{L}$  and blood eosinophil counts  $\geq 220/\mu\text{L}$ :

- Patient characteristics (weight, height, BMI, baseline eosinophil count as used for stratification for randomization, and historic eosinophil count)
- Baseline lung function data (pre- and post-BD FEV<sub>1</sub> (L), FEV<sub>1</sub> (% PN), FVC (L), FVC (% PN), FEV<sub>1</sub>/FVC, and reversibility)
- Respiratory disease characteristics
- Nicotine use and consumption

#### **4.2.3 Prior and concomitant medications**

The number and percentage of patients taking maintenance COPD medications (double or triple therapy) at baseline will be summarised based on actual medication therapy group, not the imputed therapy group used for the statistical analysis (Section 4.1).

The number and percentage of patients who take prior medications, those who take allowed concomitant medications, and those who take disallowed concomitant medications during the study will be presented by treatment group. Concomitant medications will be classified according to the AstraZeneca Drug Dictionary (AZDD). The summary tables will present data by generic term within ATC code.

#### **4.2.4 Study treatment administration**

Duration of investigational product administration will be calculated in days as:

*Last dose date of IP-first dose date of IP+1*

and will be summarised by treatment group for the safety analysis set.

Study treatment compliance will be summarised by treatment group for the full analysis set and calculated as:

*Study treatment compliance = (total doses administered/total doses expected) x 100.*

Patients who received no study treatment will have zero compliance. Total number of doses expected includes all visits with protocol scheduled IP administration on or before a subject's IP discontinuation or treatment complete date.

#### **4.2.5 Primary outcome variable**

##### **4.2.5.1 Primary analysis**

The primary efficacy variable is the annual COPD exacerbation rate and the primary analysis is to compare the annual COPD exacerbation rate of each benralizumab dose regimen with placebo in patients with baseline blood eosinophil counts  $\geq 220/\mu\text{L}$ . Patients will be analysed using the full analysis set according to randomised treatment.

For each of the 2 benralizumab dose regimens, the null hypothesis is that the exacerbation rate on benralizumab is equal to the exacerbation rate on placebo. The alternative hypothesis is that the exacerbation rate on benralizumab is not equal to the exacerbation rate on placebo, ie,

$H_0$ : Rate ratio (benralizumab vs Placebo) = 1

$H_a$ : Rate ratio (benralizumab vs Placebo)  $\neq 1$

Annual exacerbation rate in each of the 2 benralizumab dose regimen groups will be compared to annual exacerbation rate in the placebo group using a negative binomial model for the primary analysis. The response variable in the model will be the number of COPD exacerbations experienced by a patient over the 56-week double-blind treatment period. The model will include covariates of treatment group, eosinophil cohort (220-299/ $\mu\text{L}$  or  $\geq 300/\mu\text{L}$ ), region, background therapy (ICS/LABA, LABA/LAMA, or ICS/LABA/LAMA), and the number of exacerbations in the previous year. The logarithm of the patient's corresponding follow-up time will be used as an offset variable in the model to adjust for patients having different exposure times during which the events occur.

The COPD exacerbation rate and the corresponding 95% confidence interval (CI) within each treatment group will be presented. The estimated treatment difference from placebo (ie, the absolute difference between benralizumab and placebo, and the rate ratio of benralizumab versus placebo), corresponding 95% CI, and 2-sided p-values for the rate ratio will be presented. Marginal standardization methods will be used on the model estimates for all negative binomial analyses, unless otherwise specified.

The above negative binomial analysis will be repeated for the following:

- Baseline blood eosinophil counts  $< 220/\mu\text{L}$ . Eosinophil cohort (220-299/ $\mu\text{L}$  or  $\geq 300/\mu\text{L}$ ) will not be a covariate in this model.
- Cumulative baseline blood eosinophil count categories ( $< 150/\mu\text{L}$ ,  $\geq 150/\mu\text{L}$ ,  $< 220/\mu\text{L}$ ,  $\geq 220/\mu\text{L}$ ,  $< 300/\mu\text{L}$ ,  $\geq 300/\mu\text{L}$ ,  $< 400/\mu\text{L}$ ,  $\geq 400/\mu\text{L}$ ). Eosinophil cohort (220-299/ $\mu\text{L}$  or  $\geq 300/\mu\text{L}$ ) will not be a covariate in this model. Non-marginal estimates

will be weighted by the number of patients with baseline blood eosinophil count  $\geq 220/\mu\text{L}$  and  $< 220/\mu\text{L}$  to adjust for the approximately 2:1 randomization ratio in cumulative baseline blood eosinophil count categories. Patients with missing baseline blood eosinophil values from central laboratory are excluded.

- Cumulative historical blood eosinophil count categories ( $< 150/\mu\text{L}$ ,  $\geq 150/\mu\text{L}$ ,  $< 220/\mu\text{L}$ ,  $\geq 220/\mu\text{L}$ ,  $< 300/\mu\text{L}$ ,  $\geq 300/\mu\text{L}$ ,  $< 400/\mu\text{L}$ ,  $\geq 400/\mu\text{L}$ ). Eosinophil cohort (220-299/ $\mu\text{L}$  or  $\geq 300/\mu\text{L}$ ) will not be a covariate in this model.
- On-treatment exacerbations in patients with baseline blood eosinophil counts  $\geq 220/\mu\text{L}$  (i.e., excluding assessments after the EOT or IPD visit).
- Exacerbations that result in hospitalization or death due to COPD in patients with baseline eosinophil count  $\geq 220/\mu\text{L}$ . In the statistical model, the continuous covariate number of exacerbations in the previous year will be replaced by a binary covariate: exacerbation associated with a hospitalization in the previous year (Yes, No).
- Exacerbations in patients with baseline blood eosinophil counts  $\geq 220/\mu\text{L}$ , excluding exacerbations outside of the EOT visit window.
- Exacerbations in patients with baseline blood eosinophil counts  $\geq 220/\mu\text{L}$ , using the data truncated at the point when subjects switched their background therapies.
- Exacerbations by number and/or severity of exacerbations during the previous year in patients with baseline blood eosinophil counts  $\geq 220/\mu\text{L}$ 
  - $\leq 2$  exacerbations
  - $\geq 3$  exacerbations
  - Only severe exacerbations
  - $\geq 1$  severe exacerbation
  - $\geq 2$  severe exacerbations
  - Only moderate exacerbations
  - $\geq 2$  moderate exacerbations only
  - $\geq 3$  moderate exacerbations only

Further subgroup analyses for the COPD exacerbation rate are outlined in Section 4.2.5.2. Additional analyses to assess the robustness of the primary analysis results to missing data are outlined in Appendix 8.1. Marginal standardization methods will not be applied to missing data analyses.

COPD exacerbation summary statistics will be presented based on the full analysis set by treatment group for patients with baseline blood eosinophil counts  $\geq 220/\mu\text{L}$  and  $< 220/\mu\text{L}$ . The individual exacerbation criteria (use of systemic corticosteroids, use of antibiotics, and inpatient hospitalization or death due to COPD) will also be summarised descriptively.

#### **4.2.5.2 Subgroup analysis for the primary outcome variable**

To explore the uniformity of the detected overall treatment effect on the primary efficacy variable, subgroup analyses and statistical modelling including testing for interaction between treatment and covariates will be performed in patients with baseline blood eosinophil counts  $\geq 220/\mu\text{L}$  for the following factors: background therapy (ICS/LABA, LABA/LAMA, or ICS/LABA/LAMA) including imputation, baseline ICS use (Yes, No), baseline theophylline use (Yes, No), gender, age group categories ( $< 65$ ,  $\geq 65$  years), geographic region (Asia, Eastern Europe, Europe, North America, and rest of world), smoking status (current, former), BMI ( $\leq 35$ ,  $> 35$   $\text{kg}/\text{m}^2$  and  $\leq 21$ ,  $> 21$   $\text{kg}/\text{m}^2$ ), the number of exacerbations during the previous year ( $\leq 2$ ,  $\geq 3$  exacerbations), race (White, Asian, or Other), baseline IgE concentration ( $< 150\text{KU}/\text{L}$ ,  $\geq 150\text{KU}/\text{L}$ ), atopic status (Yes, No), COTE index ( $< 4$ ,  $\geq 4$ , [Divo et al. 2012](#)), and baseline GOLD classification of severity of airflow limitation in COPD (moderate, severe, very severe) based on [GOLD 2014](#).

An additional subgroup analysis using the methodology outlined above will be conducted using baseline blood eosinophil count categories ( $< 150/\mu\text{L}$ ,  $150\text{-}219/\mu\text{L}$ ,  $220\text{-}299/\mu\text{L}$ ,  $300\text{-}399/\mu\text{L}$ ,  $\geq 400/\mu\text{L}$ ). Eosinophil cohort ( $220\text{-}299/\mu\text{L}$  or  $\geq 300/\mu\text{L}$ ) will not be a covariate in these models. Patients with missing baseline blood eosinophil values from central laboratory are excluded.

For each of the subgroup factors, a separate negative binomial regression model will be fitted using the same model terms as used for the primary analysis (described in Section 4.2.5.1, with additional terms for the subgroup main effect and the treatment by subgroup interaction). Region is not a covariate if it is the subgroup; background therapy is not a covariate if background therapy or baseline ICS use is the subgroup; and number of exacerbations is not a covariate if it is the subgroup.

Similar outputs will be presented for each subgroup as for the primary analysis. The p-value for the interaction term by each treatment group will be presented in the summary tables and forest plots.

It is important to note that the study has not been designed or powered to assess efficacy within any of these pre-defined subgroups, and as such these analyses are considered as exploratory.

#### **4.2.5.3 Supportive analysis for the primary outcome variable**

As a supportive analysis, the annual exacerbation rate will be calculated based on exacerbations that are associated with a hospitalization or ER visit and will be analysed using the same method as in Section 4.2.5.1, for patients with baseline eosinophil count  $\geq 220/\mu\text{L}$ . In the statistical model, the continuous covariate number of exacerbations in the previous year

will be replaced by a binary covariate: exacerbation associated with a hospitalization in the previous year (Yes, No).

#### **4.2.6 Secondary efficacy outcome variables**

All secondary efficacy endpoints will be summarised separately by baseline blood eosinophil counts  $\geq 220/\mu\text{L}$  and  $< 220/\mu\text{L}$ , with formal statistical analyses conducted in the baseline blood eosinophil counts  $\geq 220/\mu\text{L}$  group.

Additional formal statistical analyses in the  $< 220/\mu\text{L}$  group will be conducted for the multiplicity protected key secondary endpoints (change from baseline in FEV<sub>1</sub> (L) pre-bronchodilator at Week 56 and change from baseline in SGRQ total score at Week 56), where presentation by baseline eosinophil count categories and cumulative baseline eosinophil count categories will also be conducted.

In addition, the key secondary endpoints will be analysed pooling together data from patients with baseline blood eosinophil counts  $< 220/\mu\text{L}$  from this study and study D3251C00004. The pooled analysis will be elaborated in a separate SAP.

##### **4.2.6.1 Proportion of subjects with $\geq 1$ COPD exacerbation**

The proportion of subjects with  $\geq 1$  COPD exacerbation during the 56 weeks treatment period will be addressed as a supportive variable to the primary objective for patients with baseline blood eosinophil counts  $\geq 220/\mu\text{L}$ . The proportion in each of the 2 benralizumab dose regimen groups will be compared with the proportion in the placebo group using a Cochran–Mantel–Haenszel test controlling for eosinophil cohort ( $\geq 220$ - $299/\mu\text{L}$  or  $300/\mu\text{L}$ ), region, and background therapy (ICS/LABA, LABA/LAMA, or ICS/LABA/LAMA) as covariates.

The results of the analyses will be presented using odds ratio, together with associated 95% CI and 2-sided p-values for each active dose regimen versus placebo. The number and percentage of patients with  $\geq 1$  COPD exacerbation will also be summarised by randomised treatment.

##### **4.2.6.2 Time to first COPD exacerbation**

Time to first COPD exacerbation will be analysed as another supportive efficacy variable to the primary objective to explore the extent to which treatment with benralizumab delays the time to first exacerbation compared with placebo. A Cox proportional hazard model will be fitted to the data with the covariates of treatment, eosinophil cohort ( $220$ - $299/\mu\text{L}$  or  $\geq 300/\mu\text{L}$ ), region, background therapy (ICS/LABA, LABA/LAMA, or ICS/LABA/LAMA), and number of exacerbations in the previous year. This analysis will also be repeated for the time to first COPD exacerbation associated with a hospitalization or ER visit where exacerbations associated with a hospitalization in the previous year (Yes, No) will be included as a covariate instead of number of exacerbations in the previous year.

Results of these analyses will be summarised as hazard ratios, 95% CI and p-values for patients with baseline blood eosinophil counts  $\geq 220/\mu\text{L}$ .



Time to first COPD exacerbation will be displayed graphically using a Kaplan-Meier plot, separately for patients with baseline blood eosinophil counts  $\geq 220/\mu\text{L}$  and  $< 220/\mu\text{L}$ .

The percentage of patients who are exacerbation free at Week 26 and Week 52 will be estimated for each treatment group. These percentages will be estimated using the Kaplan-Meier technique.

#### 4.2.6.3 FEV<sub>1</sub> (L) pre-bronchodilator measured at the study centre

Change from baseline in FEV<sub>1</sub> (L) pre-bronchodilator at Week 56 in the baseline blood eosinophil counts  $\geq 220/\mu\text{L}$  group is a multiplicity protected key secondary endpoint (Section 4.1.1).

Summary statistics of the change from baseline in FEV<sub>1</sub> (L) pre-bronchodilator at the study center will be produced by treatment group and visit.

Change from baseline in FEV<sub>1</sub> (L) pre-bronchodilator through Week 56 will be compared between each of the benralizumab treatment groups and the placebo group using a mixed-effect model for repeated measures (MMRM) analysis on patients with a baseline FEV<sub>1</sub> (L) pre-bronchodilator and at least 1 post-baseline FEV<sub>1</sub> (L) pre-bronchodilator assessment.

The dependent variable will be the change from baseline in FEV<sub>1</sub> (L) pre-bronchodilator at post-baseline protocol-specified visits (up to the EOT visit). Treatment group will be fitted as the explanatory variable, and baseline FEV<sub>1</sub> (L) pre-bronchodilator, region, eosinophil cohort (220-299/ $\mu\text{L}$  or  $\geq 300/\mu\text{L}$ ), background therapy (ICS/LABA, LABA/LAMA, or ICS/LABA/LAMA), visit, and treatment\*visit interaction as fixed effects. The covariance matrices will be assumed in the following order, unstructured, Toeplitz, first-order autoregressive, compound symmetric, variance components. If the procedure doesn't converge, the next covariance matrix will be used in the order. The model is:

*Change from baseline in FEV<sub>1</sub> (L) pre-bronchodilator = Treatment group + baseline FEV<sub>1</sub> (L) pre-bronchodilator + region + eosinophil cohort (220-299/ $\mu\text{L}$  or  $\geq 300/\mu\text{L}$ ) + background therapy (ICS/LABA, LABA/LAMA, or ICS/LABA/LAMA) + visit + treatment\*visit*

Results will be presented in terms of least square means (LSMEANS), treatment differences in LSMEANS, 95% CI and p-values. The results at Week 56 will be of primary interest.

The above MMRM analysis will be repeated for the following:

- Baseline blood eosinophil counts  $< 220/\mu\text{L}$ . Eosinophil cohort (220-299/ $\mu\text{L}$  or  $\geq 300/\mu\text{L}$ ) will not be a covariate in this model.
- Cumulative baseline blood eosinophil count categories ( $< 150/\mu\text{L}$ ,  $\geq 150/\mu\text{L}$ ,  $< 220/\mu\text{L}$ ,  $\geq 220/\mu\text{L}$ ,  $< 300/\mu\text{L}$ ,  $\geq 300/\mu\text{L}$ ,  $< 400/\mu\text{L}$ ,  $\geq 400/\mu\text{L}$ ). Eosinophil cohort (220-299/ $\mu\text{L}$  or  $\geq 300/\mu\text{L}$ ) will not be a covariate in this model. Estimates will be weighted by the number of patients with baseline blood eosinophil count  $\geq 220/\mu\text{L}$

and  $<220/\mu\text{L}$  to adjust for the approximately 2:1 randomization ratio in cumulative baseline blood eosinophil count categories.

- On-treatment pre-bronchodilator FEV<sub>1</sub> in patients with baseline blood eosinophil counts  $\geq 220/\mu\text{L}$  (ie, excluding assessments after the EOT or IPD visit).
- By number of exacerbations during the previous year ( $\leq 2$ ,  $\geq 3$ ) in patients with baseline blood eosinophil counts  $\geq 220/\mu\text{L}$ .

Additional analyses to assess the robustness of the repeated measures analysis to missing data are outlined in Appendix 8.1.

A subgroup analysis, using the same MMRM model defined above with additional terms for the subgroup main effect and the treatment by subgroup interaction, will be conducted for the following factor based on the full analysis set:

- Baseline blood eosinophil count categories ( $<150/\mu\text{L}$ ,  $150\text{-}219/\mu\text{L}$ ,  $220\text{-}299/\mu\text{L}$ ,  $300\text{-}399/\mu\text{L}$ ,  $\geq 400/\mu\text{L}$ ). Eosinophil cohort ( $220\text{-}299/\mu\text{L}$  or  $\geq 300/\mu\text{L}$ ) will not be a covariate in this model.

#### 4.2.6.4 St. George's Respiratory Questionnaire (SGRQ)

Change from baseline in SGRQ total score at Week 56 in the baseline blood eosinophil counts  $\geq 220/\mu\text{L}$  group is a multiplicity protected key secondary endpoint (Section 4.1.1).

Summary statistics for change from baseline in SGRQ total score and the domain scores will be produced by treatment group and visit.

Change from baseline in SGRQ total score and the 3 domain scores (symptoms, activity, and impacts) over 56 weeks will be analysed separately using a similar model as the above model for change from baseline in FEV<sub>1</sub> (L) pre-bronchodilator, described in Section 4.2.6.3. Results will be presented in terms of LSMEANS, treatment differences in LSMEANS, 95% CI and p-values for all visits. The results at Week 56 will be of primary interest.

The above MMRM analysis will be repeated for the following:

- Baseline blood eosinophil counts  $<220/\mu\text{L}$ . Eosinophil cohort ( $220\text{-}299/\mu\text{L}$  or  $\geq 300/\mu\text{L}$ ) will not be a covariate in this model.
- Cumulative baseline blood eosinophil count categories ( $<150/\mu\text{L}$ ,  $\geq 150/\mu\text{L}$ ,  $<220/\mu\text{L}$ ,  $\geq 220/\mu\text{L}$ ,  $<300/\mu\text{L}$ ,  $\geq 300/\mu\text{L}$ ,  $<400/\mu\text{L}$ ,  $\geq 400/\mu\text{L}$ ). Eosinophil cohort ( $220\text{-}299/\mu\text{L}$  or  $\geq 300/\mu\text{L}$ ) will not be a covariate in this model. Estimates will be weighted by the number of patients with baseline blood eosinophil count  $\geq 220/\mu\text{L}$  and  $<220/\mu\text{L}$  to adjust for the approximately 2:1 randomization ratio in cumulative baseline blood eosinophil count categories.



- On-treatment SGRQ total score in patients with baseline blood eosinophil counts  $\geq 220/\mu\text{L}$  (i.e., excluding assessments after the EOT or IPD visit).

Additional analyses to assess the robustness of the repeated measures analysis for SGRQ total score to missing data are outlined in Appendix 8.1.

A subgroup analysis, using the same MMRM model defined above with additional terms for the subgroup main effect and the treatment by subgroup interaction, will be conducted for the following factor based on the full analysis set:

- Baseline blood eosinophil count categories ( $<150/\mu\text{L}$ ,  $150\text{--}219/\mu\text{L}$ ,  $220\text{--}299/\mu\text{L}$ ,  $300\text{--}399/\mu\text{L}$ ,  $\geq 400/\mu\text{L}$ ). Eosinophil cohort ( $220\text{--}299/\mu\text{L}$  or  $\geq 300/\mu\text{L}$ ) will not be a covariate in this model.

The cumulative distribution function of absolute changes from baseline in SGRQ total score at Week 56 will be also plotted in a figure.

The proportion of subjects in terms of SGRQ total score response status (improvement, worsening, no change, and not evaluable) at Week 56 will be summarised descriptively by treatment group. The proportion of subjects with  $\geq 4$ -point decrease (improvement) in SGRQ total score at Week 56 in each of the 2 benralizumab dose groups will be compared with the proportion in the placebo group using a Cochran–Mantel–Haenszel (CMH) test controlling for eosinophil cohort ( $220\text{--}299$  or  $\geq 300/\mu\text{L}$ ), region, and background therapy (ICS/LABA, LABA/LAMA, or ICS/LABA/LAMA). The results of the analyses will be presented as odds ratios with associated 95% CI and 2-sided p-value for each active dose regimen versus placebo. A similar CMH analysis will be performed using last observation carried forward (LOCF) for patients with non-evaluable response status at Week 56.

#### **4.2.6.5 COPD assessment tool (CAT)**

The CAT total scores and changes from baseline will be summarised by treatment and visit.

Change from baseline in CAT score through Week 56 will be analysed using a similar model as for the change from baseline in FEV<sub>1</sub> (L) pre-bronchodilator described in Section 4.2.6.3. Results will be presented in terms of LSMEANS, treatment differences in LSMEANS, 95% CI and p-values for all visits.

#### **4.2.6.6 Baseline/Transitional Dyspnea Index (BDI/TDI)**

The BDI focal (total) score and component scores will be summarised by treatment group, and the TDI focal (total) score and component scores will be summarised by treatment group and visit.

Change from baseline in TDI focal (total) score through Week 56 will be analysed using a similar MMRM model as for the analysis of change from baseline in FEV<sub>1</sub> (L) pre-bronchodilator described in Section 4.2.6.3, with BDI focal (total) score as the baseline value

for TDI focal (total) score. Results will be presented in terms of LSMEANS, treatment differences in LSMEANS, 95% CI and p-values for all visits.

TDI responder (yes/no) at Week 56 will be analysed using a logistic regression model with covariates of treatment group, BDI score, eosinophil cohort (220-299/ $\mu\text{L}$  or  $\geq 300/\mu\text{L}$ ), region, and background therapy (ICS/LABA, LABA/LAMA, or ICS/LABA/LAMA).

#### **4.2.6.7 Exacerbations of Chronic Pulmonary Disease Tool – Patient-reported Outcome (EXACT-PRO)**

Summary statistics of the EXACT-PRO event frequency, duration, and severity will be produced by treatment group.

Annual rate of EXACT-PRO defined exacerbation will be analysed using the method described for the primary analysis in Section 4.2.5.1 for patients in the full analysis set with baseline eosinophil count  $\geq 220/\mu\text{L}$ .

#### **4.2.6.8 Evaluating Respiratory Symptoms in Chronic Obstructive Pulmonary Disease (E-RS: COPD)**

The E-RS: COPD total scores and domain scores as well as changes from baseline will be summarised by treatment and visit.

Change from baseline in E-RS: COPD total score and domain scores will be analysed separately using a similar MMRM model as for the analysis of change from baseline in FEV<sub>1</sub> (L) pre-bronchodilator described in Section 4.2.6.3. Results will be presented in terms of LSMEANS, treatment differences in LSMEANS, 95% CI and p-values for all visits.

#### **4.2.6.9 Medication Use**

##### **Total rescue medication use**

Total rescue medication use and change from baseline will be summarised by treatment and visit.

Change from baseline in total rescue medication use (average puffs/day) will be analysed using the same MMRM method for FEV<sub>1</sub> (L) pre-bronchodilator as described in Section 4.2.6.3, for patients with baseline blood eosinophil count  $\geq 220/\mu\text{L}$ .

##### **Background medication use**

Descriptive statistics of patients who switched their background medication will be summarised for the on-treatment and post-treatment periods as defined in Section 3.5.1. A shift table will be produced to summarise the types of therapies at baseline and at the end of treatment. In addition, the number of patients who switched their background medication during the on-treatment period will be plotted over time by treatment group for each of the following switches: ICS/LABA  $\rightarrow$  LABA/LAMA, LABA/LAMA  $\rightarrow$  ICS/LABA, ICS/LABA  $\rightarrow$  ICS/LABA/LAMA, LABA/LAMA  $\rightarrow$  ICS/LABA/LAMA, ICS/LABA/LAMA  $\rightarrow$  ICS/LABA, and ICS/LABA/LAMA  $\rightarrow$  LABA/LAMA.

## **Glucocorticoid use**

The total number of days patients were on systemic glucocorticoid due to COPD exacerbation during the treatment period will be summarised by treatment group.

### **4.2.6.10 Nights with awakening due to respiratory symptoms**

The number of nights with awakening and change from baseline will be summarised by treatment and visit.

Change from baseline in the proportion of nights with awakening due to respiratory symptoms will be analysed using the same MMRM method for FEV<sub>1</sub> (L) pre-bronchodilator as described in Section 4.2.6.3, for patients with baseline blood eosinophil count  $\geq 220/\mu\text{L}$ .

### **4.2.6.11 Healthcare resource utilization**

The number and rate of patients with COPD specific resource utilization during the treatment period will be summarised by treatment group.

## **4.2.7 Additional modelling for primary and key secondary endpoints**

Additional modelling will be performed to assess the relationship in each treatment group between the primary and key secondary endpoints with baseline blood eosinophil count (as a continuous variable) and number of exacerbations in the previous year. The following plots will be produced:

- Annual COPD exacerbation rate based on baseline blood eosinophil counts
- Annual COPD exacerbation rate based on number of exacerbations in the previous year
- Change from baseline in FEV<sub>1</sub> (L) pre-bronchodilator at end of treatment based on baseline blood eosinophil counts
- Change from baseline in SGRQ total score at end of treatment based on baseline blood eosinophil counts

Additional sensitivity analyses will be conducted on the primary and key secondary endpoints to investigate the sensitivity of the efficacy conclusions upon removal of patient data from sites with data anomalies. The primary assessment of benefit-risk will be based on the full analysis set in patients with baseline eosinophils  $\geq 220/\mu\text{L}$ . Data anomalies may include but are not restricted to: sites with multiple benralizumab PK concentration level anomalies (e.g. patients with PK levels persistently below LLOQ or excessive PK concentrations); sites with high correlations in entry time of ePRO data; sites with available ePRO data post-death; or sites closed due to quality issues.

## **4.2.8 Exploratory Objectives**

### **4.2.8.1 European Quality of Life-5 Dimensions-5 Levels (EQ-5D-5L)**

The EQ-5D-5L responses from each dimension and the visual analog scale, and changes from baseline, will be summarised by treatment group and visit.

Shift tables will also be produced for each dimension.

### **4.2.8.2 Clinician Global Impression of Change (CGIC) and Patient Global Impression of Change (PGIC)**

The CGIC and PGIC responses will be summarised by treatment group and visit. The number and percentage of patients will be presented for CGIC, PGIC, and for agreement in CGIC and PGIC.

The number and percentage of patients defined as responders based on categorized responses for CGIC and PGIC (improved, much improved, very much improved) will also be presented by treatment group and visit.

PGIC responses will be excluded from all analyses if only caregiver or other is the respondent.

### **4.2.8.3 Blood eosinophils**

An important marker of the biological activity of benralizumab is the reduction in eosinophils in the peripheral blood.

Absolute blood eosinophil count assessed by a central laboratory as a part of safety haematology testing at Week -4 (Visit 1) will be used as the baseline blood eosinophil measure to inform primary and secondary efficacy of benralizumab. Absolute eosinophil counts along with their absolute changes from baseline, and percentage changes from baseline will be summarised by treatment group and visit using conventional units (cells/ $\mu$ L). Baseline is defined as the last non-missing value prior to the first dose of study treatment. Percent change from baseline in eosinophil counts will be compared between each benralizumab group and the placebo group using a MMRM analysis, as defined for analysis of FEV<sub>1</sub> pre-bronchodilator in Section 4.2.6.3. Results will be presented separately by baseline blood eosinophil count (<220/ $\mu$ L and  $\geq$ 220/ $\mu$ L).

A shift table will be produced for the placebo treatment group to display change from baseline eosinophil count ( $\geq$ 220/ $\mu$ L and <220/ $\mu$ L). A scatterplot comparing historical eosinophil counts and baseline eosinophil counts will also be produced for patients with available historical eosinophil data.

### **4.2.8.4 Serum biomarkers**

Serum biomarker samples will be collected at Week 0, 4, 24 and 56. This will be done to evaluate the pharmacology of benralizumab as well as biomarkers of eosinophil recruitment, activation, and survival, and eosinophil- and inflammation-related biomarkers of COPD. This will be done in order to retrospectively analyse predictive biomarkers to further understand

patient populations responsive to benralizumab, and pharmacodynamic biomarkers to further evaluate the pharmacology of benralizumab. These biomarkers will include previously identified pharmacodynamic markers of benralizumab, additional markers of eosinophil and basophil recruitment, activation and survival, as well as biomarkers of COPD and inflammation.

As collection of biological samples is an integral part of this study. No separate optional biomarker consent form was used, and so it is expected that all patients will have samples available.

The results of any investigation will be reported separately in a scientific report or publication.

#### **4.2.8.5 Pharmacogenetics – DNA Sampling**

The pharmacogenetic sample will be collected at Week 0 or at any other visit after randomization. Only 1 sample is to be collected per patient for genetic research during the study. As there is a separate optional informed consent for this aspect, it is anticipated that not all patients will have data available. These results will be presented outside of the CSR.

#### **4.2.8.6 Mortality rate**

The number and percentage of patients who die from any cause will be presented by treatment group. Similarly, the number and percentage of respiratory related deaths will be presented by treatment group.

#### **4.2.8.7 Exploratory objectives in a subset of subjects**

Approximately 10% -15% of subjects (54-81 subjects in each arm of the study) will be included in a subset study that will be used to assess the exploratory objectives described in Section 1.1.4.

The summary statistics and the change from baseline for these exploratory objectives will be produced by treatment group and visit:

- FEV<sub>1</sub> (L) post-bronchodilator at study center
- Difference between FEV<sub>1</sub> (L) post-bronchodilator change from baseline and FEV<sub>1</sub> (L) pre-bronchodilator change from baseline
- Lung volume: total lung capacity (TLC)
- Lung volume: vital capacity (VC)
- Lung volume: residual volume (RV)
- Lung volume: functional residual capacity (FRC)
- Lung volume: inspiratory capacity (IC)

- Diffusion capacity of the lung for carbon monoxide (DL<sub>CO</sub>)
- 6 Minute Walk Test: distance
- 6 Minute Walk Test: Borg pre- and post-test scores for breathlessness and fatigue

The BODE index (Celli et al. 2004) at Week 0 (Visit 4) will also be summarised by treatment group.

Due to the relatively small number of patients in the sub-study, no formal statistical analyses will be performed.

#### 4.2.8.8 Sputum biomarkers and differential cell counts

Key baseline characteristics, including subject demographics, baseline disease characteristics, cardiovascular history, and baseline efficacy measures will be summarised overall for the sub-study analysis set and by baseline sputum eosinophil subgroups ( $\geq 3\%$  vs.  $< 3\%$ ).

Sputum differential cell count change from baseline and baseline scaled ratio over time will be summarised overall for the sub-study analysis set and by baseline sputum eosinophil subgroups ( $\geq 3\%$  vs.  $< 3\%$ ) using descriptive statistics. For the baseline scaled ratio, summary statistics will include the geometric mean and coefficient of variation (CV) in addition to the N, median, minimum, maximum and the upper and lower quartiles. The baseline scaled ratio is defined as the post-baseline value/baseline value.

Sputum differential cell count absolute values over time (weeks 24 and 56) will be compared between the combined benralizumab treatment group (30mg and 100mg groups combined) and placebo group using a mixed model for repeated measures analysis with treatment group, baseline value, visit and the treatment\*visit interaction effects included in the model for patients in the sub-study analysis set overall and by baseline sputum eosinophil subgroups ( $\geq 3\%$  vs.  $< 3\%$ ). This analysis will be repeated for the baseline scaled ratios on the log-scale, with estimates presented as geometric least squares means. Any results equal to zero will be set to 0.001 prior to log-transformation. In addition, the sputum differential cell count change from baseline at week 56 will be compared between the combined benralizumab treatment group and placebo using a Wilcoxon rank sum test.

The Spearman's rank correlation coefficient will be calculated to evaluate the correlation between baseline blood eosinophil counts (cells/ $\mu$ L) and baseline sputum eosinophils (%).

Formal comparisons of efficacy endpoints by baseline sputum eosinophil subgroups are not planned due to the small sample sizes, however the annual COPD exacerbation rates, pre-bronchodilator FEV<sub>1</sub> change from baseline and SGRQ total score change from baseline will be presented graphically by baseline sputum eosinophil subgroups.

Additional analyses of sputum biomarkers may be performed and will be presented outside of the CSR.



#### **4.2.9 Safety outcome variables**

All safety variables will be summarised using the safety analysis set and data presented according to actual treatment received.

##### **4.2.9.1 Adverse events (AEs)**

Adverse events (AEs) will be summarised separately for the on-treatment, on-study, and post-treatment periods, as defined in Section 3.5.1. All AEs will be listed for each subject. All summaries will be presented by treatment group.

An overall summary table will be produced showing the number and percentage of patients with at least 1 AE in any of the following categories: AEs, serious adverse events (SAEs), AEs with outcome of death, and AEs leading to discontinuation of investigational product (DAEs). This summary will also be repeated by baseline blood eosinophil count ( $<220/\mu\text{L}$ ,  $\geq 220/\mu\text{L}$ ) for the on-study, on-treatment, and post-treatment periods.

AEs, AEs with outcome of death, SAEs and DAEs will be summarised by System Organ Class (SOC) and Preferred Term (PT) assigned to the event by MedDRA. For each PT, the number and percentage of patients reporting at least 1 occurrence will be presented (ie, multiple occurrences of an AE for a patient will only be counted once). A summary of the most common (frequency of  $\geq 3\%$ ) AEs will be presented by PT. AEs and SAEs causing discontinuation of the study treatment and SAEs causing discontinuation from the study will also be summarised. Presentations for AEs and SAEs will also be repeated for baseline blood eosinophil count ( $<220/\mu\text{L}$ ,  $\geq 220/\mu\text{L}$ ) for the on-study, on-treatment, and post-treatment periods.

The rate of AEs per person-years at risk, calculated as (number of patients reporting AE)/(total period with patients at risk for AE), will also be reported for the on-study and on-treatment periods. The total period at risk for each patient will be the duration of the on-treatment and on-study periods as defined in Section 3.5.1. Rates will be expressed in terms of events per 100 patient-years.

AEs and SAEs will be summarised by preferred term and investigator's causality assessment (related vs. not related) and maximum intensity. If a patient reports multiple occurrences of the same AE within the same study period, the maximum intensity will be taken as the highest recorded maximum intensity (the order being mild, moderate, and severe) and will be summarised for the on-study, on-treatment, and post-treatment periods. Summaries of AEs by investigator's causality assessment will also be repeated for baseline blood eosinophil count ( $<220/\mu\text{L}$ ,  $\geq 220/\mu\text{L}$ ) for the on-study and on-treatment periods.

Adverse events of injection site reactions (high level term of administration and injection site) and hypersensitivity (standardized MedDRA query of hypersensitivity) will be summarised by preferred term. The summary of injection site reactions will be summarised by injection site location and number of IP administrations. The summary of AEs of hypersensitivity (overall and causally related to IP) will be repeated for baseline blood eosinophil count ( $<220/\mu\text{L}$ ,  $\geq 220/\mu\text{L}$ ) for the on-study, on-treatment, and post-treatment periods.

Adjudicated events (major adverse cardiac events (MACE) and malignancies) will be summarised by treatment group for the on-treatment and on-study periods and listed.

#### 4.2.9.2 Laboratory data

All protocol-specified continuous laboratory parameters will be summarised descriptively by absolute value at each visit by treatment group, together with the corresponding changes from baseline. All parameters will be summarised in SI units, with the exception of blood eosinophil counts which will be summarised in both SI and conventional units. Results which are reported from the central laboratory in conventional units will be converted to SI units for reporting.

Central laboratory reference ranges will be used for the identification of abnormalities, and a shift table will be produced for each laboratory parameter to display low, normal, and high values. The shift tables will present baseline and maximum/minimum on-treatment value, as applicable for each parameter and will include patients with both baseline and post-baseline data.

Shift plots showing each individual patient's laboratory value at baseline and at maximum/minimum post-baseline will be produced for each continuous laboratory variable. If any laboratory variables show any unusual features (high or low values or a general shift in the data points) at other time points, then shift plots of these data may be produced. Data for patients who have treatment-emergent changes outside central laboratory reference ranges will be presented.

Maximum post-baseline TBL elevations by maximum post-baseline ALT and AST will be presented, expressed as multiples of ULN. TBL will be presented in multiples of the following ULN:  $\leq 1.5$ ,  $>1.5-2$ ,  $>2$ . AST and ALT will be presented in multiples of the following ULN:  $\leq 1$ ,  $>1-3$ ,  $>3-5$ ,  $>5-10$ ,  $>10$ .

Maximum post-baseline TBL will be presented ( $<2$  and  $\geq 2$  x ULN) and plotted against maximum post-baseline ALT ( $<3$ ,  $\geq 3-5$ ,  $\geq 5-10$ , and  $\geq 10$  x ULN), expressed as multiples of ULN. This will be repeated to show maximum post-baseline TBL against maximum post-baseline AST.

Data for patients with ALT or AST  $\geq 3$  x ULN, and TBL  $\geq 2$  x ULN will be presented, which will include all visits for this subset of patients. A line plot of liver biochemistry test results (including ALP, ALT, AST, TBL, and GGT) over time will also be presented for this subset of patients.

For all patients who meet the biochemical criteria for confirmed Hy's law, a SAE narrative will be produced.

For urinalysis data, a shift table will be generated to present changes from baseline to maximum post-baseline value for selected parameters and will include patients with both baseline and post-baseline data.



Descriptive statistics and change from baseline at each visit will be presented for IgE in kU/L by treatment group.

Any data outside the central laboratory reference ranges will be explicitly noted on the listings that are produced.

#### **4.2.9.3 ECGs**

A shift table will be produced for each ECG parameter to display normal, abnormal – not clinically significant, abnormal – clinically significant, and not done. The shift tables will present baseline and last observation post-baseline value.

In addition, the number and percentage of patients with QTcF changes from baseline will be presented by QTcF value >450 ms, >480 ms and >500 ms and QTcF increase of >30 ms and >60 ms. A shift plot showing maximum post-dose QTcF change from baseline (with reference lines at 30 ms and 60 ms) will be also be produced.

#### **4.2.9.4 Physical examination**

Shift tables (normal, abnormal, not done) of baseline versus last observation post-baseline (normal, abnormal (same as baseline), abnormal (new or aggravated), not done) will be generated, presenting the assessment for each component of the complete physical examination separately.

#### **4.2.9.5 Vital signs**

Descriptive statistics and change from baseline for vital signs data will be presented for each treatment group by visit. Baseline to maximum post-baseline and baseline to minimum post-baseline value shift tables will be generated, as applicable for each parameter and will include patients with both baseline and post-baseline data.

#### **4.2.10 Pharmacokinetic analyses**

The PK analyses will be performed at or under the guidance of AstraZeneca Research and Development.

Benralizumab serum concentrations will be summarised using descriptive statistics at each visit by treatment group, and will be listed in the CSR. The population modeling will be presented in a separate pharmacometrics report.

#### **4.2.11 Immunogenicity analyses**

Anti-drug antibody (ADA) assessments will be conducted and analysed as per the details in Appendix 8.3.

## 5. INTERIM ANALYSES

A futility analysis for study efficacy will be performed by an Independent Data Monitoring Committee (IDMC). The futility analysis will be based on the pooled data of Study D3251C00003 and Study D3251C00004 when approximately 15% patients in the  $\geq 220/\mu\text{L}$  eosinophil stratum have completed the studies. The percentage of patients is based on the total number of patients expected to be recruited in the  $\geq 220/\mu\text{L}$  eosinophil stratum of the studies.

In the futility analysis, a decision will be based on the comparison of the futility boundary and the estimated exacerbation rate reductions of benralizumab doses vs. placebo in patients with eosinophil counts  $\geq 220/\mu\text{L}$  from the negative binomial model as documented for the final analysis in Section 4.2.5.1. Futility will be declared if the higher exacerbation rate reduction of benralizumab 30mg vs. placebo and benralizumab 100mg vs. placebo is less than 8%. A futility boundary of 8% was chosen based on the operating characteristics of ~55% chance of stopping the study for futility under the null hypothesis of equal exacerbation rates across all treatments and ~2% chance of stopping the study for futility assuming an exacerbation rate reduction of 30% for both benralizumab doses vs. placebo.

Full details about the futility analysis decision rules and procedures will be specified in an IDMC charter and the charter will also specify the roles and responsibilities of the IDMC members. A firewall will be established to ensure the maintenance of the study blind for the sponsor, the investigational site staff and patients. The IDMC charter will also specify the timing of analysis and the expected extent of data available at the time of interim analysis. Conducting a futility analysis results in power loss thus the sample size calculation in Section 1.3 has accounted for this loss and the study has 90% of power under the powering assumptions and the futility boundary of 8%.

## 6. CHANGES OF ANALYSIS FROM PROTOCOL

The protocol indicated that country would be among the baseline covariates adjusted for in formal statistical models. Based on small sample sizes within certain countries and consistency with the two Phase 3 asthma exacerbation studies where convergence issues were encountered for some of the statistical models, the decision was made to replace the country covariate effect with region in all analyses where this effect is included.

Additional sensitivity analyses on the primary and key secondary endpoints were added to investigate the sensitivity of the efficacy conclusions upon removal of patient data from sites with data anomalies.

For MMRM analysis of repeated measures, covariance matrices will be assumed in the following order, unstructured, Toeplitz, first-order autoregressive, compound symmetric, variance components. If the procedure doesn't converge, the next covariance matrix will be used in the order.

Exacerbations of Chronic Pulmonary Disease Tool – Respiratory Symptoms (E-RS) was renamed to Evaluating Respiratory Symptoms in Chronic Obstructive Pulmonary Disease (E-RS: COPD) after it was qualified as an exploratory endpoint (Evidera, 2016). The score range for E-RS: COPD for chest symptoms is correctly portrayed in SAP as 0-12 per the E-RS Manual (Evidera, 2016).

To align with actual ePRO setup, the proportion of nights with awakenings due to respiratory symptoms and the change from baseline will be analysed instead of number of nights with awakenings due to COPD and requiring rescue medication.

Additional sub-study analysis set was added for sub-study patients as well as corresponding sputum analyses.

Adverse events will also be summarised using the on-study period (in addition to pre-planned on-treatment and post-treatment periods).

## 7. REFERENCES

### **Burman et al 2009**

Burman C.F. et al. A recycling framework for the construction of Bonferroni-based multiple tests. *Statist. Med.* 2009; 28:739–761.

### **Celli et al. 2004**

Celli B.R. et al. The Body-Mass Index, Airflow Obstruction, Dyspnea, and Exercise Capacity Index in Chronic Obstructive Pulmonary Disease. *The New England Journal of Medicine.* 350;10.

### **Divo et al. 2012**

Divo M, Cote C et al. Comorbidities and Risk of Mortality in Patients with Chronic Obstructive Pulmonary Disease. *American Journal of Respiratory and Critical Care Medicine* Vol 186: 2012

### **Evidera, Inc. 2014**

Evidera, Inc. The Exacerbations of Chronic Pulmonary Disease Tool (EXACT) Patient-Reported Outcome (PRO) USER MANUAL (Version 7.0).

### **Evidera, Inc. 2016**

Evidera, Inc. Evaluating Respiratory Symptoms (E-RS™) in COPD (E-RS™: COPD) USER MANUAL (Version 5.0). Bethesda, MD: Evidera; March 2016.

### **Friede and Schmidli 2010**

Friede T, Schmidli H. Blinded Sample Size Reestimation with Negative Binomial Counts in Superiority and Non-inferiority Trials. *Methods of Information in Medicine.* 2010; 49(6): 618-624.

**GOLD 2014**

Global Strategy for the Diagnosis, Management and Prevention of COPD, Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2014. Available from:  
<http://www.goldcopd.org>.

**Hochberg 1988**

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

**Holm 1979**

Holm, S. A simple sequentially rejective multiple test procedure. *Scand.J. Statist.* 1979. 6, 65-70.

**Jones et al. 1991**

Jones PW, Quirk FH, Baveystock CM. The St George's Respiratory Questionnaire. *Respir Med*. 1991; 85 Suppl B:25-31.

**Jones et al. 2009**

Jones PW, Harding G, Berry P, Wiklund I, Chen W-H, Leidy NK. Development and first validation of the COPD Assessment Test. *Eur Respir J* 2009, 34: 648–654.

**Jones et al. 2011**

Jones PW, Chen WH, Wilcox TK, Sethi S, Leidy NK. Characterizing and quantifying the symptomatic features of COPD exacerbations. *Chest* 2011; 139(6):1388–1394.

**Leidy et al. 2011**

Leidy NK, Wilcox TK, Jones PW, et al. Standardizing measurement of chronic obstructive pulmonary disease exacerbations. Reliability and validity of a patient-reported diary. *Am J Respir Crit Care Med* 2011; 183(3):323–329.

**Mahler et al. 1984**

Mahler DA, Weinberg DH, Wells CK, Feinstein AR.; The measurement of dyspnea: contents, inter-observer agreement, and physiologic correlates of two new clinical indexes, *Chest* 1984; 85:751–758.

**Mahler et al. 2005**

Mahler DA, Witek Jr TJ.; The MCID of the transition dyspnea index is a total score of one unit., *COPD*, 2005; 2: 99-103.

**Sexton et al. 2010**

Sexton C, Notte S, Murray L, Leidy NK. PRO evidence dossier to support the use of the E-RS to evaluate respiratory symptoms in patients with COPD. United BioSource Corporation; Bethesda, MD: May 2010.

Statistical Analysis Plan  
Study Code D3251C00003  
Edition Number 3.0  
Date 23April 2018

**Sexton et al. 2011**

Sexton C, Leidy NK, Notte SM, et al. Quantifying the severity of respiratory symptoms of COPD: reliability and validity of a patient diary. Poster presented at the American Thoracic Society International Meeting; May 2011: Denver, CO.

## 8. APPENDIX

### 8.1 Accounting for missing data

#### Accounting for missing data for recurrent events (exacerbation rate endpoint)

In this study some patients dropping out of the study potentially leads to unobserved events. The amount of missing data is minimized in this study as patients are allowed in the protocol to switch to an alternative treatment or treatments after they discontinue from randomised treatment and are encouraged to complete visits until they withdraw from the study.

This section summarises how we will describe the pattern of and reasons for missing data from the study. It will also describe how we plan to account for missing data, including both the primary and sensitivity analyses to assess the robustness of the treatment effect under different underlying assumptions to account for missing data.

#### Missing data descriptions

Tabular summaries for the percentage of patients by the reason for discontinuation of randomised treatment as well as for withdrawal from the study will be presented by treatment to describe why patients discontinue from randomised treatment or withdraw from the study. The time to discontinuation of randomised treatment and withdrawal from the study will be presented using Kaplan Meier plots (overall and split by treatment related/not treatment related reason for discontinuation, as defined in Tables 1 and 2). Dependent on these outputs additional exploratory analyses may be produced as deemed necessary to further understand the pattern of missing data.

#### Primary analysis under the Treatment Policy Estimand using the Missing at Random (MAR) assumption

The primary analysis is under the treatment policy estimand which allows for differences in outcomes over the entire study treatment period to reflect the effect of initially assigned randomised treatment as well as if subsequent treatments are taken. This primary analysis includes all data until patients withdraw from the study regardless of if they discontinue from randomised treatment. The primary analysis uses the negative binomial regression model with (logarithm of) the observation period as an offset term and assumes that missing data is missing at random (MAR) and is a direct likelihood approach (DL).

#### Sensitivity analyses under the Treatment Policy Estimand using both MAR and MNAR assumptions

To examine the sensitivity of the results of the primary analysis to departures from the underlying assumptions, additional analyses will be performed using controlled multiple imputation method introduced in [1] and further developed at AstraZeneca [2,3] which allows for different underlying assumptions to be used. As with the primary analysis the sensitivity

analyses includes all data until patients withdraw from the study regardless of if they discontinue from randomised treatment.

For this method an underlying negative binomial stochastic process for the rate of exacerbations is assumed and post study withdrawal counts will be imputed conditional upon the observed number of events prior to the withdrawal. This allows various assumptions about the missing data to be analysed by modifying the post-withdrawal model assumption.

The method involves first fitting the primary analysis, ie, negative binomial regression model to the observed data and then imputing post-withdrawal counts by sampling from the conditional negative binomial probability relating post-withdrawal counts and observed prior-withdrawal counts based on various assumptions.

$$\Pr(Y_{i,j,2} = y_2 | Y_{i,j,1} = y_1) = \frac{\Gamma(r + y_1 + y_2)}{\Gamma(y_2 + 1)\Gamma(r + y_1)} p_j^{y_2} (1 - p_j)^{r + y_1} \quad (1)$$

Here  $y_1$  is number of counts before withdrawal from the study,  $y_2$  is number of counts after withdrawal from the study,  $r$  is the dispersion parameter and which is assumed to be the same for different treatment arms,  $j$  denotes the treatment arm and  $i$  denotes the subject identifier. Furthermore

$$p_j = \frac{p_{j,z} - p_{j,1} p_{j,z}}{1 - p_{j,1} p_{j,z}} \quad (2)$$

where  $p_{j,1}$  is the negative binomial distribution (NBD) rate parameter before withdrawal from the study, and  $p_{j,z}$  is the rate parameter after withdrawal from the study as determined based on various assumptions.

The imputed number of exacerbations that would have been seen is then combined with the observed exacerbations and data is analysed using the primary analysis methodology (DL). This analysis is repeated multiple times and the results combined using Rubin's formulae [7, 8].

The following default assumptions that will be used to impute the missing data who withdraw early from the study are as follows:

- a) MAR: Missing counts in each arm are imputed assuming the expected event rate within that arm.
- b) Dropout Reason-based Multiple Imputation (DRMI): Missing counts will be imputed differently depending on the reason for dropout; counts for patients in the benralizumab arms who dropped out for a treatment related reason are imputed based on the expected event rate in the placebo arm, whereas the remaining patients who have dropped out are imputed assuming MAR. Treatment related reasons include (1) AEs, (2) Death, (3)

development of study specified reasons to stop active treatments, and (4) severe non-compliance of protocol.

Some reasons for withdrawal are clearer to determine as treatment related (*Adverse Events, Death, Development of study-specific discontinuation criteria*) or non-treatment related (*Subject lost to follow up, eligibility criteria not fulfilled*). Other reasons are less clear such as severe non-compliance of protocol, subject decision and ‘Other’; a review of each patient who withdraws from the study will therefore be carried out prior to unblinding the study. The review will include assessment of the reason for discontinuation of randomised treatment for those patients who discontinued randomised treatment and then withdrew from the study and also free text for when the reason for withdrawal or discontinuation of randomised treatment is subject decision or other. Based on this review the default assumptions for DRMI as described in b) and [Table 1](#) may be changed. A list of these patients and the assumptions made under DRMI will be documented prior to unblinding of the study.

A summary of reasons for patients withdrawing from the benralizumab treatment arm and the corresponding treatment arm used to calculate the imputation exacerbation rate under MAR and DRMI is given in [Table 1](#).

**Table 1**

<b>Reason for withdrawal</b>	<b>MAR</b>	<b>DRMI</b>
Adverse Event	Benralizumab	Placebo
Development of study-specific discontinuation criteria*	Benralizumab	Placebo
Death	Benralizumab	Placebo
Severe non-compliance to protocol	Benralizumab	Placebo
Eligibility criteria not fulfilled	Benralizumab	Benralizumab
Subject lost to follow up	Benralizumab	Benralizumab
Subject decision	Benralizumab	Based on review prior to study unblinding
Other	Benralizumab	Based on review prior to study unblinding

Note all patients on exacerbation rate in the placebo arm are imputed using the placebo arm rate

\*Development of study-specific discontinuation criteria are based on the following: anaphylactic reaction to the IP requiring administration of epinephrine, development of helminth parasitic infestations requiring hospitalization, 2 consecutive doses of IP missed or more than 2 scheduled doses of IP are missed during course of the study, a COPD-related event requiring mechanical ventilation.

Together with the primary analysis the sensitivity analyses are considered to cover the range from realistic to plausible worst case assumptions about missing data. The MAR multiple imputation approach is expected to correspond closely to the primary analysis, and is included to allow for comparisons with MNAR assumptions (specifically methods b and c) using the same multiple imputation methodology.



The dropout reason-based multiple imputation (DRMI) approach was selected as the most conservative approach based on the fact that placebo patients are receiving standard of care and are not expected to change to a substantially more effective treatment after withdrawing from study or study treatment. For patients receiving benralizumab who withdraw from the study due to treatment related reasons it is assumed that at worst they would be on the standard of care treatment, ie, the placebo arm. For patients receiving benralizumab who withdraw from the study due to non-treatment related reasons it seems reasonable to assume they would be similar to those patients who complete treatment.

#### On-Treatment Analyses (Effectiveness estimand)

In addition primary and sensitivity analyses described previously, an alternative estimand will be estimated using only the on initial randomised treatment data:

- Effectiveness estimand with assumed loss of effect post discontinuation of benralizumab: This will be estimated using the DRMI controlled imputation approaches including only data from patients whilst on treatment.

Therefore the primary analyses and sensitivity analyses will be repeated including only data from patients whilst being on initial randomised treatment, ie, excluding data once patients discontinue from randomised treatment.

A summary of reasons for patients withdrawing from the benralizumab treatment arm and the corresponding treatment arm used to calculate the imputation exacerbation rate under MAR and DRMI are given in [Table 2](#). As for patients who withdraw from the study, a review of each patient who discontinued randomised treatment will be carried out prior to unblinding the study where the default assumptions for DRMI as described in [Table 2](#) may be changed. Again a list of these patients and the assumptions made under DRMI will be documented prior to unblinding of the study.

**Table 2**

<b>Reason for discontinuation of randomised treatment</b>	<b>MAR</b>	<b>DRMI</b>
Adverse Event	Benralizumab	Placebo
Development of study-specific discontinuation criteria*	Benralizumab	Placebo
Severe non-compliance to protocol	Benralizumab	Placebo
Subject lost to follow up	Benralizumab	Benralizumab
Subject decision	Benralizumab	Based on review prior to study unblinding
Other	Benralizumab	Based on review prior to study unblinding

Note all patients on exacerbation rate in the placebo arm are imputed using the placebo arm rate  
 \*Development of study-specific discontinuation criteria are based on the following: anaphylactic reaction to the IP requiring administration of epinephrine, development of helminth parasitic infestations requiring hospitalization, 2 consecutive doses of IP missed or more than 2 scheduled doses of IP are missed during course of the study, a COPD-related event requiring mechanical ventilation.

Using on treatment data is easier to interpret as it is not impacted by any subsequent pattern of alternative treatments once patients discontinue from randomised treatment. Sensitivity analyses using the effectiveness estimand under the DRMI allow for alternative assumptions to be made based on reasons for discontinuation.

*Overall summary of analyses to account for missing data*

A summary of the different analyses to be carried out under different estimands and assumptions are described in [Table 3](#).

Table 3

	Treatment Policy Estimand			On-Treatment Analyses (Effectiveness estimand)
	On-treatment + post-discontinuation of randomised treatment			On-treatment
Population	Treatment policy (MAR)		Treatment policy (MNAR)	Treatment policy (MNAR)
Estimand	DL	MAR	DRMI	DRMI
Exacerbation rate for imputation in Benra arm <sup>b</sup>	No explicit imputation <sup>a</sup>	Benra for all reasons for withdrawal assumed to be on Benra	Placebo rate for AEs, Death, development of study specified reasons to stop active treatments and Severe non-compliance to protocol assumed to be on, otherwise Benra or based on review prior to study unblinding	Placebo rate for AEs, Death, development of study specified reasons to stop active treatments and Severe non-compliance to protocol assumed to be on, otherwise Benra or based on review prior to study unblinding
Default definition for $P_{i,1}$ and $P_{i,2}$ based on formula (2) <sup>c</sup> .		$P_{i,2} - P_{i,1}$ For all treatment arms j=B and P	$P_{B,2} - P_{B,1}$ $P_{P,2} - P_{P,1}$ for reasons above otherwise $P_{B,2} = P_{B,1}$	$P_{B,2} - P_{B,1}$ $P_{P,2} - P_{P,1}$ for reasons above otherwise $P_{B,2} = P_{B,1}$

<sup>a</sup> Implicitly assumes unobserved rate the same as observed

<sup>b</sup> All patients on exacerbation rate in the placebo arm are imputed using the placebo arm rate (ie,  $P_{P,2} = P_{P,1}$ )

<sup>c</sup> Note can be over written by review prior to study unblinding

B Benralizumab 30mg or 100mg; P Placebo

Forest plots will be used to show the primary analysis results along with the missing data sensitivity and alternative estimand analysis results.

It is noted that if the primary analysis is statistically significant, it is not necessarily expected that all sensitivity analyses will also give statistically significant results. If the results of the sensitivity analyses provide reasonably similar estimates of the treatment effect to the primary analysis, this will be interpreted as providing assurance that neither the lost information nor the mechanisms which cause the data to be missing have an important effect on primary analysis conclusions. Based on these outputs and the drug's mechanism of action, the plausibility of the assumptions we make about missing data in the different analyses will be considered and described in the clinical study report.

## **Accounting for missing data for continuous endpoints (FEV1 and SGRQ total score)**

### Missing data descriptions

In addition to the tables and figures suggested above, plots of change from baseline vs time, by dropout pattern (eg, completers vs non-completers, split by reason for dropout and/or split by last available visit) will also be produced.

### Primary analysis under the Treatment Policy Estimand using the MAR assumption

As for the primary variable, the primary analysis of the FEV1 and SGRQ total score key secondary endpoints includes all data captured during the trial and is therefore considered to be under the treatment policy estimand. The Mixed Model Repeated Measures model (MMRM) used is a DL approach which is valid under the MAR assumption.

### Sensitivity analysis under the Treatment Policy Estimand using MNAR assumptions

Sensitivity analyses of the repeated measures analyses will be performed for the FEV1 and SGRQ total score using controlled sequential multiple imputation methods based on pattern mixture models, as described in [5].

The method is analogous to the multiple imputation of exacerbation events and the imputation process consists of a sequence of MI steps, where each step is intended to impute missing values at 1 time-point only. This model will assume that some pre-specified subset of subjects who withdraw from the study have correlations with future (unobserved) visits similar to subjects in the placebo arm. As for the exacerbation events, this allows us to assess various deviations from the MAR assumption.

The assumptions that will be used to impute the missing data who withdraw early are as follows:

- (a) MAR: Assumes that the trajectory for patients who dropped out in each arm is similar to those observed in their own treatment arm
- (b) DRMI: Assumes that the trajectory for patients in the benralizumab arms who dropped out for a treatment related reasons (according to the same classification as for the DRMI analysis of the primary endpoint) is similar to that of the placebo subjects, whereas the remaining patients who has dropped out are imputed assuming MAR.

Approach b) can be considered more conservative than the approach for the primary analysis because the assumptions mean that as soon as subjects withdraw for a treatment related reason, they begin to worsen immediately.

The MNAR imputation is achieved by only using appropriate data at each stage of the imputation. Imputation will be done in 2 steps, the non-monotone (intermediate) missing FEV1 values will be imputed first (Markov chain Monte Carlo (MCMC) method is used to

partially impute the data using SAS PROC MI) and then the missing value at each visit will be imputed using a sequential regression method (using MONOTONE REG option of SAS PROC MI).

For example, to impute missing values at time t for subjects in the benralizumab arms, that dropped out due to an AE, include only placebo observations up to and including time t, plus observations from subjects in the benralizumab arms, that dropped out due to an AE, up to and including time t-1. This is done for each visit, 1 at a time using observed data, and missings just imputed. Placebo missing observations and benralizumab observations that are not missing due to AEs are imputed assuming missing at random (MAR) and follow the pattern of observed placebo observations in each treatment arm respectively. 100 imputations will be carried out, and a seed of 784088 will be used for the monotone imputation step and a seed of 409345 will be used for the sequential regression imputation step. The analysis of each of the imputed dataset will be as described for the primary analysis in Section 4.2.5.1 and these will be combined using SAS procedure PROC MIANALYSE.

To avoid possible convergence issues when fitting MMRM models to 100 imputed datasets, the estimated unstructured covariance parameters from the first imputation where the model converges will be used as the starting values to fit the models for all imputed datasets.

#### On-Treatment Analyses (Efficacy and Effectiveness estimands)

Analogously to the approach for the primary endpoint, efficacy and effectiveness estimands will be estimated using on-treatment data and the methods described above.

Results for continuous endpoints will be presented as per the recurrent event sensitivity analyses.

#### **References**

1. Keene ON, Roger JH, Hartley BF, Kenward MG. Missing data sensitivity analysis for recurrent event data using controlled imputation. *Pharmaceutical Statistics* 2014, 13 258–264.
2. Wan R, Hirsch I, Gottlow M, Hollis S, Darilay A, Weissfeld L, France L. Controlled imputation approach for analyzing missing data in recurrent events due to early discontinuations. *DIA/FDA Statistics Forum* 2015.
3. Gottlow M, Hollis S, Wan R, Hirsch I, Darilay A, Weissfeld L, France L. A Simulation study of a controlled imputation approach for analyzing missing data in recurrent events due to early discontinuations. *PSI Annual Conference* 2015.
4. The Panel on Handling Missing Data in Clinical Trials; National Research Council. *The Prevention and Treatment of Missing Data in Clinical Trials*. National Academies Press, 2010.

5. Guideline on Missing Data in Confirmatory Clinical Trials 2 July 2010  
EMA/CPMP/EWP/1776/99 Rev. 1
6. AZ guidance (clinical OPI): Guidance on Minimizing the Loss of Patient Data in AstraZeneca Clinical Trials, ed 2.0. (LDMS\_001\_00102309)
7. Fleming, TR. Addressing Missing Data in Clinical Trials. *Ann. Intern. Med.* 2011;154:113-117.
8. Ratitch B, O’Kelly M, Tosiello R. Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. *Pharmaceutical Statistics* 2013; 12: 337-347.
9. Rubin DB. *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons, Inc. 1987.
10. Barnard J, Rubin DB. Small-sample degrees of freedom with multiple imputation. *Biometrika* 1999; 86:948-955.
11. Mallinckrodt CH, Lin Q, Lipkovich I, Molenberghs G. A structured approach to choosing estimands and estimators in longitudinal clinical trials. *Pharmaceutical Statistics* 2012, 11:456–461.

## **8.2 Partial dates for adverse events and prior/concomitant medication**

Dates missing the day or both the day and month of the year will adhere to the following conventions in order to classify treatment-emergent AEs and to classify prior/concomitant medications:

### **Adverse Events**

- The missing day of onset of an AE will be set to:
  - First day of the month that the event occurred, if the onset YYYY-MM is after the YYYY-MM of first study treatment
  - The day of the first study treatment, if the onset YYYY-MM is the same as YYYY-MM of the first study treatment
  - The date of informed consent, if the onset YYYY-MM is before the YYYY-MM of the first treatment.
- The missing day of resolution of an AE will be set to:
  - The last day of the month of the occurrence. If the patient died in the same month, then set the imputed date as the death date.
- If the onset date of an AE is missing both the day and month, the onset date will be set to:
  - January 1 of the year of onset, if the onset year is after the year of the first study treatment
  - The date of the first treatment, if the onset year is the same as the year of the first study treatment
  - The date of informed consent, if the onset year is before the year of the first treatment
- If the resolution date of an AE or end date of a IP is missing both the day and month, the date will be set to:
  - December 31 of the year of occurrence. If the patient died in the same year, then set the imputed date as the death date.

### **Prior/concomitant medication**

- The missing day of start date of a therapy will be set to the first day of the month that the event occurred.

- The missing day of end date of a therapy will be set to the last day of the month of the occurrence.
- If the start date of a therapy is missing both the day and month, the onset date will be set to January 1 of the year of onset.
- If the end date of a therapy is missing both the day and month, the date will be set to December 31 of the year of occurrence.
- If the start date of a therapy is null and the end date is not a complete date then the start date will be set to the date of the first study visit.
- If the start date of a therapy is null and the end date is a complete date
  - and the end date is after the date of the first study visit then the start date will be set to the date of the first study visit.
  - otherwise the start date will be set to the end date of the therapy.
- If the end date of a therapy is null and the start date is not a complete date then the end date will be set to the study end date.
- If the end date of a therapy is null and the start date is a complete date
  - and the start date is prior to the study end date then the end date will be set to the study end date.
  - otherwise, the end date will be set to the start date of the therapy.



### 8.3 Analysis plan for immunogenicity data

The purpose of this appendix is to provide a general reference for the analysis of anti-drug antibody (ADA) data in this study. The complete set of presentations described in this appendix will be conducted.

All analyses will be conducted on the safety analysis set by treatment group unless otherwise specified. These analyses will also be presented for subjects with baseline blood eosinophils  $\geq 220/\mu\text{L}$  in the safety analysis set. All ADA results will be listed.

Serum samples for ADA assessments will be conducted utilising a tiered approach (screen, confirm, titre) and ADA data will be collected at scheduled visits shown in the CSP. ADA result from each sample will be reported as either positive or negative. If the sample is positive, the ADA titre will be reported as well. In the event that an ADA sample is collected post-IP dose instead of pre-dose (protocol deviation), the post-dose ADA sample will be included in listings however will be excluded from all derivations and summaries. In addition, the presence of neutralizing antibody (nAb) will be tested in all ADA-positive samples using a ligand-binding assay. The nAb results will be reported as positive or negative.

For each subject, the following ADA response variables will be evaluated:

- ADA prevalence; defined as ADA positive at any visit including baseline and/or post-baseline (also generally referred to as ADA positive)
- ADA incidence; defined as ADA negative at baseline and ADA positive at any post-baseline visit, or ADA positive at baseline and a boosted ADA post-baseline titre result ( $> 4$ -fold increase)
- ADA positive at baseline and at least one post-baseline positive assessment
- ADA positive post-baseline only
- ADA positive at baseline only
- ADA persistently positive; defined as ADA negative at baseline and at least 2 post-baseline ADA positive measurements with at least 16 weeks (112 days) between the first and last positive measurement or an ADA positive result at the last available assessment
- ADA transiently positive; defined as ADA negative at baseline and at least one post-baseline ADA positive measurement and not fulfilling the conditions for persistently positive
- nAb prevalence; defined as nAb positive at any visit including baseline and/or post-baseline (also generally referred to as nAb positive)
- nAb incidence; defined as nAb negative at baseline (or ADA negative at baseline) and nAb positive at any post-baseline visit. Patients who are ADA negative at baseline are included to ensure that all patients who are nAb positive for the first time post-baseline satisfy this definition, given that patients who are ADA negative at baseline would not have an nAb result reported
- ADA positive with maximum titre  $>$  median of maximum titres
- ADA positive with maximum titre  $\leq$  median of maximum titres

The responses above will be summarized across the on-study period, up to and including the post-treatment follow up visit. The proportion of subjects with the responses above will be summarized by treatment group. No comparisons or formal hypothesis testing will be conducted.

In ADA positive subjects, the ADA titre at each visit and the maximum titre across the different visits will be summarised by descriptive statistics [n, minimum, quartiles (Q1, median and Q3) and maximum]. The maximum titre will be derived based on all ADA titre results reported up to the time point of interest for each subject. If a titre result is reported as  $\leq 50$  at a specific visit, then the titre for this visit will be imputed as 50 for analysis.

In addition, descriptive statistics of the maximum ADA titre for each subject will be summarised in subjects who are ADA persistently positive with titre  $\leq$  median of maximum titres and in the subjects who are ADA persistently positive with titre  $>$  median of maximum titres. The median of maximum titres will be calculated based on the maximum titre for each ADA positive subject within each treatment group.

A summary of ADA positive subjects who had  $>4$  fold increase in titre from time of first post-baseline ADA positive result, as well as subjects with  $>75\%$  decrease in titre from the maximum titre (excluding the post-treatment follow-up assessment) will be produced for each treatment group. For a titre change from positive to negative to represent a  $>75\%$  decrease, the maximum titre must be  $\geq 200$ . Patients must have two post-baseline positive ADA results.

### **Demographic and patient characteristics by ADA subgroups**

Demographic and patient characteristics at baseline will be presented by ADA subgroups; ADA positive, ADA negative, ADA persistently positive, and nAb positive.

### **ADA results by visit**

The proportion of patients with a positive ADA result and the ADA titres will be summarised at baseline and at all scheduled post-baseline visits by treatment group. To classify the ADA responses at each visit, the ADA result reported within the derived analysis visit window will be included in the analyses. The analysis visit windows are defined in Section 2.2.2. In addition, the ADA response will be presented cumulatively. The cumulative ADA response is positive for a specific time point if a positive ADA result is detected at any time point up to the time point of interest. If a positive ADA result is not detected at all time points up to time point of interest, then the cumulative ADA response is negative. The first post-baseline positive ADA result will also be summarised by visit and treatment group.

### **nAb response by visit**

The proportion of patients with positive nAb response will be summarised by visit. The summary will be repeated for ADA persistently positive patients.

### **Efficacy by ADA subgroups**

The effect of ADA on the primary endpoint (annual COPD exacerbation rate) and two key secondary endpoints (change from baseline in pre-BD FEV<sub>1</sub> and change from baseline in SGRQ total score) will be evaluated in subgroups of ADA positive, ADA negative, ADA persistently positive, ADA positive with maximum titre > median of maximum titres, nAb positive, and ADA persistently positive and nAb positive. Due to the expected small number of ADA positive subjects in the placebo group, no formal statistical analysis on efficacy (benralizumab vs placebo) by ADA status (positive/negative) is planned.

For annual COPD exacerbation rates, a descriptive summary by treatment group will be presented instead of the rate estimated from the negative binomial model. The maximum follow-up time for calculation of the annual COPD exacerbation rate will be as defined in Section 3.1. The 95% confidence interval for the treatment group rates will be estimated based on the Poisson distribution. For FEV<sub>1</sub> and SGRQ total scores, the last non-missing observation on or prior to the EOT (Week 56) or IPD visit will be used to calculate arithmetic means by treatment group, rather than the estimated least squares means estimated from the mixed model for repeated measures (MMRM).

### **Safety by ADA subgroups**

Summaries of adverse events will be presented separately for the on-treatment and on-study periods by the subgroups of ADA positive, ADA negative, ADA positive with titre > median of maximum titre and ADA persistently positive. Summaries will include adverse events (AEs), serious adverse events (SAEs), causally related AEs and SAEs as assessed by the investigator, and adverse events of hypersensitivity.

Relative risk comparisons between ADA positive versus ADA negative patients for safety outcomes are not planned due to the low expected frequency of the potential adverse events (AE) related to the potential risk (at the preferred term level).

### **Pharmacokinetics by ADA subgroups**

Benralizumab serum concentrations will be summarised by visit and by ADA status (positive, negative) using the following descriptive statistics: n, geometric mean, geometric mean 95% CI, geometric mean coefficient of variation (CV), arithmetic mean, standard deviation, median, minimum and maximum.

### **Blood eosinophils by ADA subgroups**

Blood eosinophil levels will be summarised by visit and by subgroups of ADA positive, ADA negative, ADA positive with titre > median of maximum titres, ADA persistently positive, nAb positive, and both ADA persistently positive and nAb positive using the following descriptive statistics: n, mean, standard deviation, median, minimum and maximum.

## 8.4 Important protocol deviations

As specified in Section 2.2.1, important protocol deviations (PDs) are a subset of protocol deviations that may significantly impact the completeness, accuracy, and/or reliability of the study data or that may significantly affect a subject's rights, safety, or well-being. The following important PDs will be summarised and listed in the CSR (\*indicates an important PD which may impact the primary efficacy endpoint):

### A. Eligibility criteria not met (patients incorrectly randomised) – deviation(s) from inclusion criteria:

1. Provision of informed consent prior to any study specific procedures.
2. History of moderate to very severe COPD with a post-bronchodilator FEV1/FVC < 0.70 and a post-bronchodilator FEV1 > 20% and ≤ 65% of predicted normal value at screening (central spirometry will be used for this criteria assessment).\*
3. History of 2 or more moderate COPD exacerbations that required treatment with systemic corticosteroids and/or antibiotics, or 1 or more severe COPD exacerbation(s) that required hospitalization (defined as an inpatient admission ≥ 24 hours in the hospital, in an observation area, the emergency department or other equivalent healthcare facility depending on the country and healthcare system) within 2 to 52 weeks prior to enrolment. Prior use of antibiotics alone does not qualify as a moderate exacerbation unless the antibiotic was specifically prescribed for the treatment of worsening COPD symptoms.\*
4. mMRC score ≥ 1 at Visit 1.\*
5. Subjects should have evidence of having been treated with double (ICS/LABA or LABA/LAMA) or triple (ICS/LABA/LAMA) therapy for COPD throughout the year prior to enrolment (Visit 1). It is acceptable for subjects to have stepped up or stepped down during that period of time (from double to triple therapy and vice versa), but they have to be consistently treated with locally approved COPD medications and on approved doses for at least 2 weeks prior to enrolment (Visit 1). Subjects currently receiving background therapy that is not approved for COPD are not eligible for the study.\*
6. Ability to read, write and use electronic devices.\*
7. Blood eosinophils due to subject's stratification and cap for blood eosinophil levels. When any eosinophil cohort (< 220/μL, 220-299/μL, or ≥ 300/μL) is full, subjects in the completed cohort will not be randomised and will be withdrawn from the study (see CSP Section 3.7.2).\*
8. Provision of a signed and dated written informed consent for the pharmacogenetic sample and analysis. If a subject declines to participate in the pharmacogenetic research, there will be no consequence or loss of benefit to the subject. The subject will not be excluded from the other aspects of the study described in the CSP, as long as they consent to participate in the study.

### B. Eligibility criteria not met (patients incorrectly randomised) – deviation(s) from exclusion criteria:

1. Clinically important pulmonary disease other than COPD (e.g. active lung infection, clinically significant bronchiectasis, pulmonary fibrosis, cystic fibrosis,

- hypoventilation syndrome associated with obesity, lung cancer, alpha 1 anti-trypsin deficiency and primary ciliary dyskinesia) or another diagnosed pulmonary or systemic disease that is associated with elevated peripheral eosinophil counts (e.g. allergic bronchopulmonary aspergillosis/mycosis, Churg-Strauss syndrome, hypereosinophilic syndrome) and/ or radiological findings suggestive of a respiratory disease other than COPD that is contributing to the subject's respiratory symptoms.\*
2. Any disorder, including, but not limited to, cardiovascular, gastrointestinal, hepatic, renal, neurological, musculoskeletal, infectious, endocrine, metabolic, haematological, psychiatric, or major physical impairment that is not stable in the opinion of the Investigator and/or could:
    - Affect the safety of the subject throughout the study
    - Influence the findings of the study or their interpretation
    - Impede the subject's ability to complete the entire duration of studySubjects who have epilepsy must be on a stable dose of medication for 30 days prior to Visit 4.\*
  3. Unstable ischemic heart disease, arrhythmia, cardiomyopathy, heart failure, renal failure, uncontrolled hypertension as defined by the Investigator, or any other relevant cardiovascular disorder as judged by the Investigator or any ECG abnormality obtained during the screening/run-in period that in Investigator's judgement may put the patient at risk or negatively affect the outcome of the study.\*
  4. Pregnant, breastfeeding, or lactating women.
  5. Alanine aminotransferase (ALT) or aspartate aminotransferase (AST) level  $\geq 1.5$  times the upper limit of normal (ULN) confirmed by repeated testing during screening period.
  6. Risk factors for pneumonia (including, but not limited to: immunosuppression, neurological disorder with increased risk of aspiration).\*
  7. Known history of allergy or reaction to any component of the investigational product formulation.
  8. History of anaphylaxis to any other biologic therapy.
  9. Long term oxygen therapy (LTOT) with signs and/or symptoms of cor pulmonale and/or right ventricular failure. Subjects receiving long term treatment with oxygen  $> 4.0$  liters/minute (L/min). While breathing supplemental oxygen, subjects should demonstrate an oxyhemoglobin saturation  $\geq 89\%$ . In order to be admitted to the trial subjects on LTOT have to be ambulatory and be able to attend clinic visits.\*
  10. Use of any non-invasive positive pressure ventilation device (NIPPV). Note: Subjects using continuous positive airway pressure (CPAP) or bilevel positive airway pressure (BiPAP) for Sleep Apnea Syndrome are allowed in the study.\*
  11. Any clinically significant abnormal findings in physical examination, vital signs, haematology, clinical chemistry, or urinalysis during screening/run-in period, which, in the opinion of the Investigator, may put the subject at risk because of his/her participation in the study, or may influence the results of the study, or the subject's ability to complete entire duration of the study.
  12. Fever  $> 37.8^{\circ}\text{C}$  ( $100^{\circ}\text{F}$ ) measured using the tympanic temperature (or equivalent oral/rectal/axillary temperature) at Visit 4.

13. Use of immunosuppressive medication, including rectal corticosteroids and systemic steroids within 2 weeks prior to enrolment (Visit 1) and/or during the enrolment and screening/run-in period.\*
14. Receipt of any investigational non-biologic product within 30 days or 5 half-lives prior to Visit 1.
15. Receipt of any marketed (e.g. omalizumab) or any other monoclonal or polyclonal antibody therapy (e.g. gamma globulin) taken for any reason within 6 months or 5 half-lives prior to Visit 1, whichever is longer.\*
16. Receipt of live attenuated vaccines 30 days prior to Visit 4.
17. Subjects are excluded if they have any of the following:
  - A history of known immunodeficiency disorder including a positive test for human immunodeficiency virus, HIV-1 or HIV-2
  - Positive hepatitis B surface antigen, or positive hepatitis C virus antibody serology, or a positive medical history for hepatitis B or C. Patients with a history of hepatitis B vaccination without history of hepatitis B are allowed to enroll.
18. A helminth parasitic infection diagnosed within 24 weeks prior to Visit 1 that has not been treated with, or has failed to respond to standard of care therapy.
19. History of alcohol or drug abuse within the past year, which may compromise the study data interpretation as judged by Investigator or Study Physician.\*
20. Malignancy, current or within the past 5 years, except for adequately treated noninvasive basal cell and squamous cell carcinoma of the skin and cervical carcinoma-in-situ treated with apparent success more than 1 year prior to Visit 1.
21. Subjects who in the opinion of the investigator or qualified designee have evidence of active tuberculosis (TB). Subjects with a recent (within 2 years) first-time or newly positive purified protein derivative (PPD) test or Quantiferon test need to complete an appropriate course of treatment before being considered for enrolment. Evaluation will be according to the local standard of care and may consist of history and physical examinations, chest x-ray, and/or TB test as determined by local guidelines.\*
22. Subjects with lung volume reduction surgery within the 6 months prior to Visit 1. Subjects with history of partial or total lung resection (single lobe or segmentectomy is acceptable).\*
23. Asthma as a primary or main diagnosis according to the Global Initiative for Asthma (GINA) guidelines (GINA 2011) or other accepted guidelines. Subjects with a past medical history of asthma (e.g. childhood or adolescence) may be included.\*
24. Previous treatment with benralizumab (MEDI-563).\*
25. Previous allogeneic bone marrow transplant.\*
26. Non-leukocyte depleted whole blood transfusion within 120 days of the genetic sample collection.

**C. Deviations from IP discontinuation procedures:**

1. Subject developed criteria for discontinuation of IP (as defined in the CSP Section 3.6) but continued IP treatment without consultation with AZ Study Physician.



**D. Received prohibited/ restricted concomitant medication:**

1. Systemic corticosteroids (tablets, suspension or injections) administered outside of COPD exacerbation event.\*
2. Antibiotics taken for prevention of COPD exacerbation.\*
3. Use of any of the following medications:•
  - Roflumilast (Daxas, Daliresp) after Visit 1 and during the study.
  - Any monoclonal antibody (e.g. omalizumab, denosumab) or polyclonal antibody (e.g. gamma globulin) within 6 months prior to Visit 1 and during the study.
  - Allergen immunotherapy within 90 days prior to Visit 1 and during the study.
  - Live attenuated vaccines within 30 days prior to Visit 4, during the treatment period and for 16 weeks after the last dose of IP.
  - Potent Cytochrome P (CYP) 3A4 inhibitors (if used for >4 weeks)

**E. Deviations from requirements on background double/triple COPD therapy:**

1. Patient not on double/triple COPD medications for >4 weeks during the treatment period (e.g. interruptions and/or step-down to mono therapy). Note: any deviations from Inclusion criteria #6 in regards to background therapy must be coded as deviation from eligibility criteria.\*

**F. Deviations from visit schedule and study procedures:**

1. Multiple phone visits / phone follow-up visits (more than 2) missed during the treatment phase/ follow-up.\*
2. COPD exacerbation not assessed and reported as per protocol.\*
3. Central spirometry violations at Visit2/Visit4:
  - assessment not performed
  - unacceptable post-PD spirometry at visit 2 not consulted with Study Physicians and not repeatedOther important/ repeated violations in spirometry assessment (to be consulted with Study physicians)\*
4. Central spirometry assessment not performed or other important/ repeated violations in spirometry assessment not followed at multiple ( $\geq 3$ ) visits after randomization.\*
5. Any safety assessment (ECG, physical examination, vital signs, safety lab tests) not performed at screening, baseline and/or at multiple ( $\geq 3$ ) visits after randomization.
6. Pregnancy test not done prior to randomization and/or at  $\geq 2$  visits after randomization.
7. Central blood eosinophil testing not done/ results not received prior to randomization.\*
8. Any other study procedures not performed or incorrectly performed at multiple ( $\geq 3$ ) visits for the same subject.\*

**G. IP management and treatment blinding:**

1. IP is not administered according to the protocol\*
  - Incorrect route of administration
  - Incorrect dose
  - IP administered prior to completion of all visit procedures

2. IP administered in a presence of condition(s) contraindicating dosing.
3. Patient not observed for a minimum of 2 hours after IP administration.
4. Incorrect randomised treatment (wrong kit ID) administered.\*
5. IP administered with interval between doses <3 weeks or >10 weeks.\*
6. Use of expired or damaged IP kit/ syringe.\*
7. Use of IP with temperature excursion without any approval from Supply Chain manager (not given either prior to or after IP administration to subject).\*

**H. Other:** Important PDs that do not fall into any above categories. Details of these important PDs will be described in the CSR.\*