AstraZeneca

Statistical Analysis Plan

Study Code	D3250C00045
Edition Number	3.0
Date	14 Oct 2019

A Multicenter, Randomized, Double-blind, Parallel Group, Placebo controlled, Phase 3b Study to Evaluate the Safety and Efficacy of Benralizumab 30 mg sc in Patients with Severe Asthma Uncontrolled on Standard of Care Treatment (ANDHI)

TABLE OF CONTENTS PAGE

LIST OF TABLES

LIST OF FIGURES

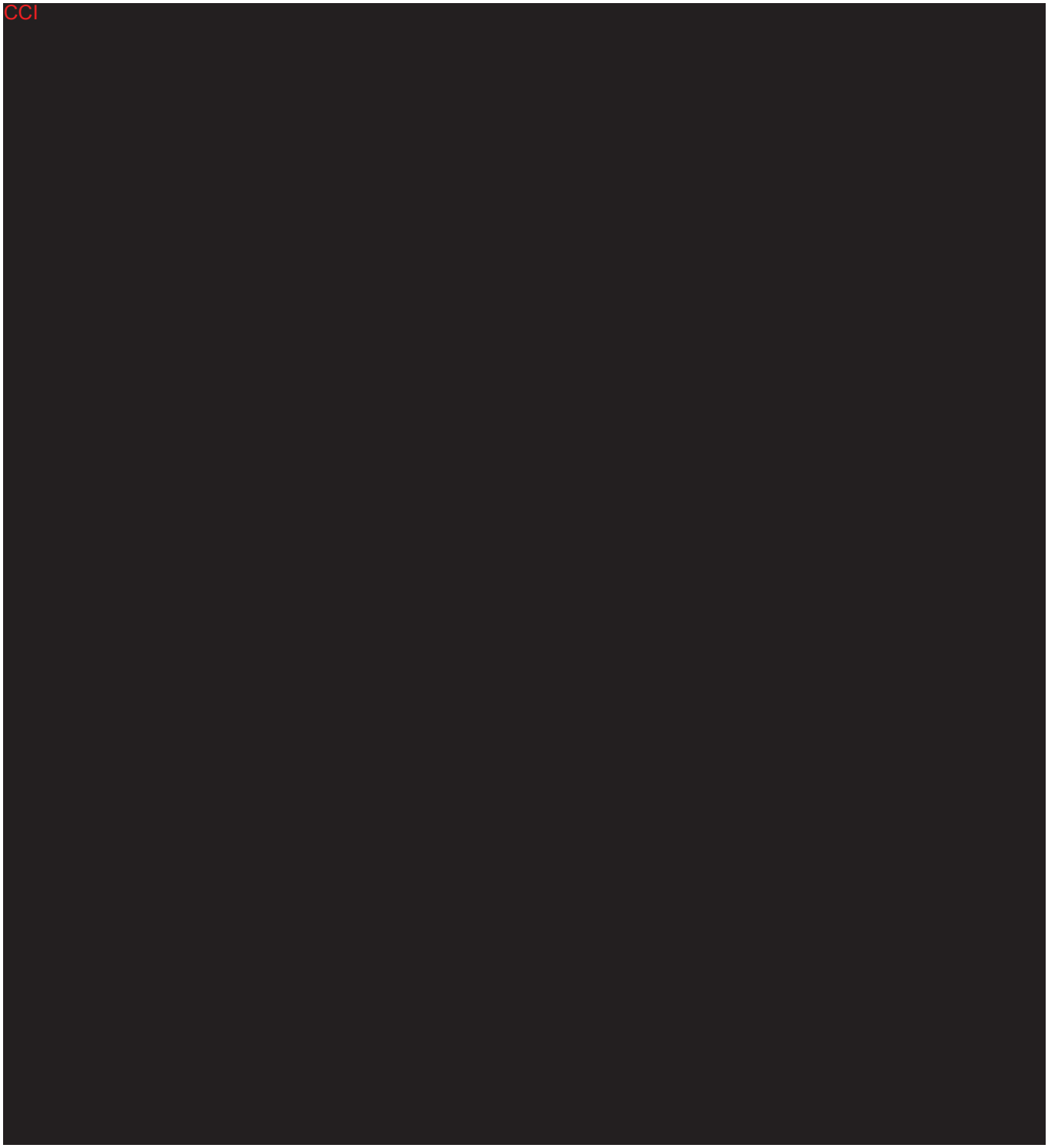
LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
ACQ-6	Asthma Control Questionnaire 6
ALT	Alanine Aminotransferase
AST	Aspartate Aminotransferase
ATC	Anatomical Therapeutic Chemical
ANDHI IP	ANDHI in Practice
AE	Adverse Event
BD	bronchodilator
BMI	Body Mass Index
BP	Bodily Pain
CGI-C	Clinician Global Impression of Change
CI	Confidence Interval
CRF	Case Report Form
CSP	Clinical Study Protocol
DAE	AE leading to discontinuation of IP
DL	Direct Likelihood
DRMI	Dropout Reason-based Multiple Imputation
ECG	Electrocardiogram
EOT	End of Treatment
ePRO	electronic PRO
FEV_1	Forced Expiratory Volume in first second
GGT	Gamma-GT
GH	General Health Perceptions
ICS	Inhaled Corticosteroids
IP	Investigational Product
IPD	IP Discontinuation
ITT	Intent To Treat
IVRS	Interactive Voice Response system
LSMEANS	Least Square Means
MAR	Missing at Random

Abbreviation or special term	Explanation
MCID	Minimal Clinically Important Difference
MCMC	Markov Chain Monte Carlo
MCS	Mental Health Component Summary
MedDRA	Medical Dictionary for Regulatory Activities
MH	Mental Health
MMRM	Mixed-effect Model for Repeated Measures
MNAR	Missing Not at Random
NBS	Norm-Based Scoring
NRS	Numeric Rating Scale
OCS	Oral Corticosteroids
PCS	Physical Component Summary
PEF	Peak Expiratory Flow
PF	Physical Functioning
PGI-C	Patient Global Impression of Change
PGI-S	Patient Global Impression of Severity
PRO	Patient Reported Outcome
PSIA	Predominant Symptom and Impairment Assessment
PT	Preferred Term
RE	Role Limitations due to Emotional Problems
REML	Restricted maximum likelihood
RP	Role Limitations due to Physical Health
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
sc	subcutaneously
SF	Social Functioning
SF-36v2	Short Form 36-item Health survey, version 2
SGRQ	Saint George Respiratory Questionnaire
SNOT-22	Sino-Nasal Outcome Test 22 Item
SI	Standard International
SOC	System Organ Class
TBL	Total Bilirubin
ULN	Upper Limit of the Normal

Abbreviation or special term	Explanation
VT	Vitality
WHO	World Health Organization

AMENDMENT HISTORY



CCI

[redacted]

CCI

[redacted]

CCI

[redacted]

1. STUDY DETAILS

This is the statistical analysis plan (SAP) for the reporting of the double-blind on-study period of study D3250C00045 which is from Visit 1 (enrolment) to Visit 12 (follow-up) inclusive. The SAP describes the statistical analyses specified in the clinical study protocol (CSP) for this period in more detail; any changes with regards to what is already specified in the CSP will be described in Section 6.

The reporting of the ANDHI in Practice (ANDHI IP) substudy which is from Visit 13 to Visit 27 (end of study visit) inclusive will be described in a separate analysis plan. This separate analysis plan will describe in detail how ANDHI IP will be analyzed standalone. It will also include how additional integrated presentations involving the double-blind period of the study will be analyzed with ANDHI IP as needed.

In this current SAP the ANDHI IP will only be referenced to show where the cut-off of data is between the end of the double-blind on-study period and the start of the ANDHI IP and in disposition to show how many patients have entered ANDHI IP.

1.1 Study objectives

1.1.1 Primary objective

Primary Objective:	**Outcome Measure:**
To determine the effect of benralizumab on the rate of asthma exacerbations	The annualized rate of asthma exacerbations between benralizumab and placebo (treatment period 24 weeks)

1.1.2 Secondary objectives

Key Secondary Objectives:	**Outcome Measure:**
To determine the effect of benralizumab on patient-reported disease-specific quality of life	The change from baseline (Visit 4) in Saint George Respiratory Questionnaire (SGRQ) to the EOT (Day 168/Week 24)

Secondary objectives:	Outcome measure:
To determine the effect of benralizumab on lung function	The change from baseline (Visit 4) in forced expiratory volume in first second (FEV_1) over the treatment period (up to and including Day 168/Week 24)
To determine the effect of benralizumab on patient-reported asthma control	The change from baseline (Visit 4) in Asthma Control Questionnaire 6 (ACQ6) to the EOT (Day 168/Week 24)
To determine the effect of benralizumab on time to first asthma exacerbation	Time to first asthma exacerbation (treatment period 24 weeks)
To determine the effect of benralizumab on lung function at home	The change from run-in baseline morning peak expiratory flow (PEF) to the EOT (Day 168/Week 24)
To determine the effect of benralizumab general quality of life and health status	The change from baseline (Visit 4) Short Form 36-item Health survey, version 2 (SF-36v2) to the EOT (Day 168/Week 24)
To evaluate patient impression of overall asthma severity (patient global impression of severity [PGI-S])	The change from baseline (Visit 4) in PGI-S to the EOT (Day 168/Week 24)
To evaluate patient impression of change in the overall asthma status from baseline as reported by the patient (patient global impression of change [PGI-C]) and clinician (clinician global impression of change [CGI-C])	The degree of change reported by the patient (PGI-C) and Investigator (CGI-C) expressed as a proportion of each of the 7 possible responses to the EOT (Day 168/Week 24)
To determine the effect of benralizumab on the patient's predominant symptoms (Predominant Symptom and Impairment Assessment; PSIA)	The degree of change reported by the patient in their predominant symptom to the EOT (Day 168/Week 24)[1]
To determine the effect of benralizumab on disease specific health-related quality of life in patients with doctor diagnosed chronic sinusitis with nasal polyposis	The change from baseline (Visit 3) in the sino-nasal outcome test (SNOT-22) score to the EOT (Day 168/Week 24).

[1] PSIA analyses updated from CSP description, see section 6

1.1.3 Safety objectives

Safety Objectives:	Outcome Measure:
To assess the safety and tolerability of benralizumab	Adverse events (AEs), laboratory variables, physical examination

1.1.4 Exploratory objectives

Exploratory objectives:	Outcome measure:
To determine the effect of eosinophil depletion with benralizumab on: • Biomarker components of known asthma inflammatory pathways or airway remodeling (including periostin, DPP4, YKL-40 and MMPs) • Biomarker surrogates of eosinophilic inflammation/activation (including, eosinophil granule proteins)	The change from baseline (Visit 4) in circulating biomarkers to each pre-specified scheduled assessment during the treatment period.
To determine the effect of benralizumab on the patient's level of asthma control based standard asthma guidance recommendations	The proportion of patients with well-controlled asthma, based on composite diary measures, over time.

1.2 Study design

This is a Phase IIIb, randomized, double-blind, placebo-controlled, parallel-group study designed to evaluate the efficacy and the safety of repeat dosing of benralizumab 30 mg subcutaneously (sc) versus placebo on top of standard of care asthma therapy in patients with severe uncontrolled asthma. The target patient population will consist of male and female patients with at least 2 asthma exacerbations while on maintenance inhaled corticosteroids (ICS) plus another asthma controller, that required treatment with systemic corticosteroids (intramuscular, intravenous, or oral) in the 12 months prior to study enrollment.

Approximately 630 patients with peripheral blood eosinophil counts ≥150 cells/μL will be randomized at approximately 275 study centers globally, in a 2:1 ratio to receive benralizumab 30 mg sc or matched placebo. Patients will be stratified at randomization by previous exacerbations (2 exacerbations in 12 months prior to Visit 1; ≥3 exacerbations in 12 months prior to Visit 1), maintenance oral corticosteroids (OCS) use (use, non-use), and region (North America, rest of world). Approximately 40% of randomized patients are expected to have ≥3 exacerbations in 12 months prior to Visit 1.

Therefore, enrollment of patients with only 2 exacerbations in the 12 months prior to Visit 1 may be halted if this subgroup within a site or region reaches approximately 60% of randomized patients.

After initial enrollment and confirmation of entry criteria, patients will proceed to a screening/run-in period of up to 42 days to allow adequate time for all of the eligibility criteria to be evaluated. Patients who meet specific exclusion criteria (2, 12, 13, 15, 17 or 18 per CSP version 1) can be re-screened once at the discretion of the investigator. Patients who meet the

eligibility criteria will be randomized to a 24-week treatment period. Either benralizumab or placebo will be administered at the study center every 56 days (every 8 weeks) through Week 16, with a single loading dose at Week 4. End of treatment (EOT) visit is at Week 24. Changes to the patient's usual asthma controller medications are discouraged during the treatment period. A follow-up telephone visit for final safety assessments will be conducted at Week 26.

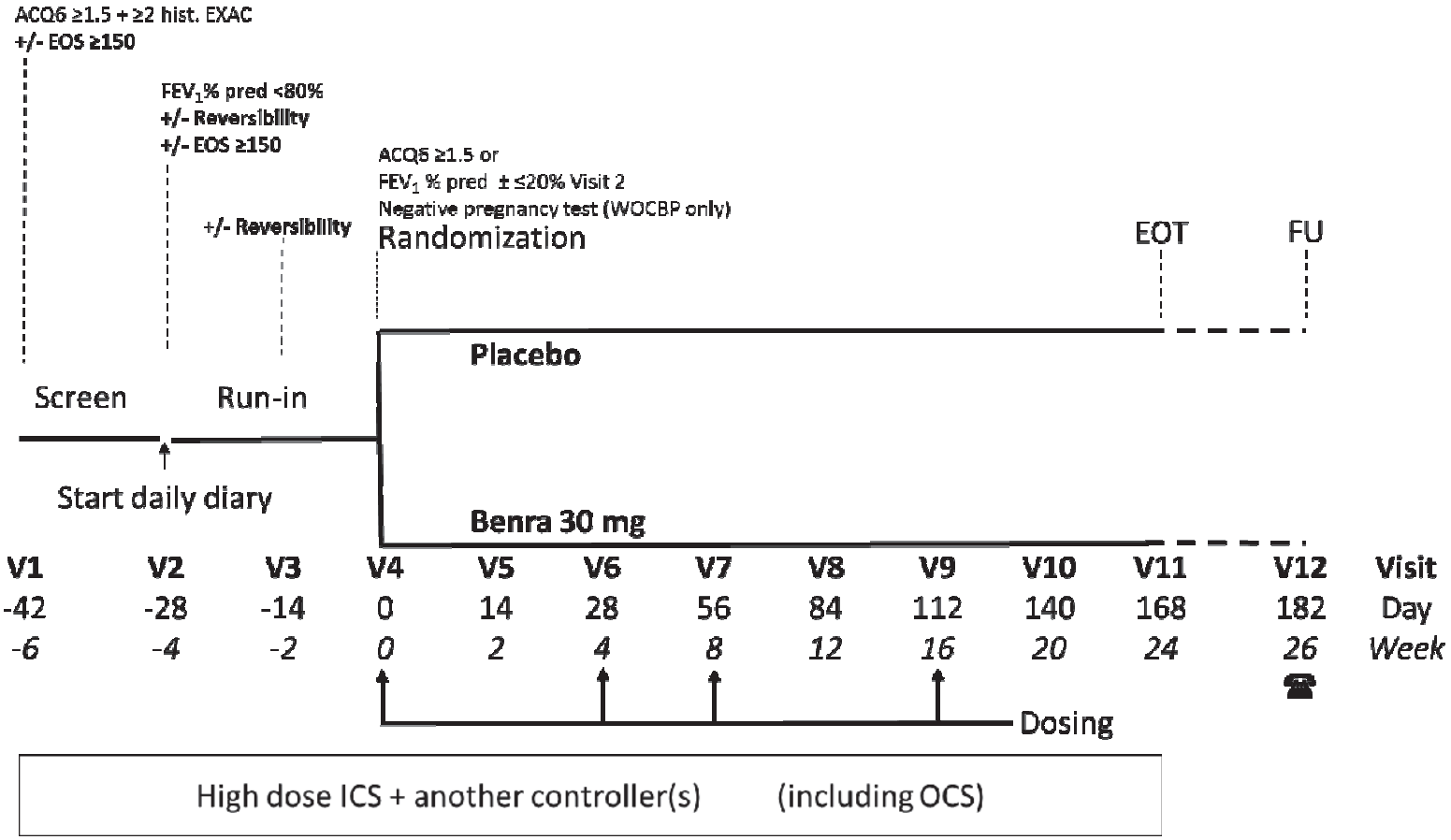
Patients with baseline chronic rhinosinusitis with nasal polyposis may elect to participate in a sub-study. The rhinosinusitis health status and quality of life of these patients will be assessed using the 22-item SNOT 22 completed at on-site visits.

All patients who prematurely discontinue investigational product (IP) should return to the study center and complete the procedures described for the premature IP Discontinuation (IPD) Visit within 4 weeks ±3 days. At that visit, patients should be encouraged to remain in the study and complete all subsequent scheduled study visits, procedures, and assessments through study completion. If the patient is not willing to participate further in the study after the IPD visit, the patient should be withdrawn from the study and also complete the follow-up visit. Withdrawn patients will not be replaced.

All eligible patients who completed the EOT Visit 11 of the double-blind period may be enrolled in the open label ANDHI IP substudy. This ANDHI IP substudy design will be detailed in the ANDHI IP SAP.

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

Figure 1 Study flow chart



Patients with a peripheral blood eosinophil count of ≥150 to <300 cells/µL have to fulfill additional clinical criteria described in inclusion criterion 8 (Section 3.1 of the study protocol), patients with a peripheral blood eosinophil count of ≥300 cells/µL do not need to meet the additional criteria.
ACQ6 asthma control questionnaire 6; Benra benralizumab; EOS eosinophil; EOT end of treatment; EXAC exacerbation; FEV_1 forced expiratory volume in first second; FU follow-up; ICS inhaled corticosteroid; V visit; WOCBP women of childbearing potential.; ☎ telephone visit.

1.3 Number of subjects

The study is powered for the primary objective (to determine the effect of benralizumab on asthma exacerbations) through the primary endpoint (annualized rate of asthma exacerbations over 24 weeks) as well as for the key secondary objective (to determine the effect of benralizumab on patient reported disease-specific quality of life) through the key secondary endpoint (the change from baseline to EOT in SGRQ up until and including the Week 24 visit.).

Previous benralizumab Phase III asthma exacerbation studies (Bleeker et al 2016, Fitzgerald et al 2016) indicate that an annual exacerbation rate for the placebo group of 1.25, a 40% reduction in exacerbation rate for the benralizumab group, and a common negative binomial shape dispersion parameter of 1.2 may be expected. Results from previous studies also indicate that a greater than 4 point difference between treatment groups in the change from baseline SGRQ score may be expected, given improvements in related patient reported outcomes (PROs) seen in the exacerbation studies as well as the results from mepolizumab pivotal trials (Bel et al 2014, Ortega et al 2014). A difference of 5 points and common standard deviation of 19 points was assumed.

Under these assumptions, a 630-patient study randomized in a 2:1 ratio (420 benralizumab: 210 placebo) has approximately 91% power with respect to the primary endpoint (assessed over a 24-week period) and 87% power with respect to the key secondary endpoint, assuming a 2-sided 5% significance level in both cases.

The sample size was estimated using nQuery version 8.2.0.0.

The total estimated number of patients to be randomized was updated from approximately 800 with 1:1 randomization ratio (benralizumab : placebo) to approximately 630 with a 2:1 randomization (benralizumab : placebo) to mitigate early challenges in recruiting sufficient numbers of appropriately severe eosinophilic asthma patients to the study. The change preserves the number of patients receiving active benralizumab treatment and reduces the number of patients exposed to placebo, while retaining statistical power to detect a treatment difference for both asthma exacerbation reduction and SGRQ improvement. The revised powering remains consistent with a level that would assure robust conclusions for scientific exchange. The revised sample size was not calculated using any current study data to maintain the blind.

2. ANALYSIS SETS

2.1 Definition of analysis sets

The following analysis sets will be used in this study:

2.1.1 All patients analysis set

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

2.1.2 Full analysis set

All patients randomized and receiving any IP will be included in the full analysis set, irrespective of their protocol adherence and continued participation in the study. Patients will be analyzed according to their randomized treatment, irrespective of whether or not they have prematurely discontinued, according to the intent to treat (ITT) principle. Patients 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 data, except SNOT-22, will be summarized and analyzed using the full analysis set on an ITT basis. For consistency, demographic and baseline characteristics will also be presented using the full analysis set.

2.1.3 Chronic rhinosinusitis with nasal polyposis sub-study analysis set

The chronic rhinosinusitis with nasal polyposis sub-study analysis set is defined as the subset of patients with:

- Doctor-diagnosed chronic rhinosinusitis and nasal polyposis included in their medical history; who
- Signed informed consent to participate in the sub-study; and
- Received at least one dose of study IP.

2.1.4 Safety analysis set

The safety analysis set will include all patients randomized who received any IP. Patients will be classified according to the treatment they actually received. A patient who has on one, or several occasions, received active treatment will be classified as active (e.g. benralizumab 30 mg sc). Any deviations from the randomized treatment will be listed and considered when interpreting the safety data. All safety analyses will use this analysis set.

2.2 Violations and deviations

Only important protocol deviations will be listed and tabulated in the clinical study report. Protocol deviations that may greatly impact the completeness, accuracy, and/or reliability of the study data or that may significantly affect a patient's rights, safety, or well-being will include but may not be limited to:

- Patients who do not meet the inclusion criteria
- Patients who do not meet the randomization criteria
- Patients who meet any of the exclusion criteria
- Patients who use one or more disallowed medication (for any reason, unless otherwise specified) during the randomized treatment period. A list of concomitant medications for all patients will be provided for team review before each data review meeting, for the identification of disallowed medications.
- Patients who received the incorrect IP or study dose at any time during the 24 week treatment period
- Patients who developed withdrawal criteria (Section 3.11 of the CSP) during the study but were not discontinued from IP.

Important protocol deviations will be identified and documented by the study physicians and statisticians prior to unblinding of the data.

3. PRIMARY AND SECONDARY VARIABLES

3.1 General Definitions

3.1.1 Definition of baseline

In general, the last measurement prior to the first dose of study treatment will serve as the baseline measurement. If time is collected, the assessment performed the same day but at a time prior to the first dose of study treatment will be included in the baseline definition. For PROs with reflection times of 1 week or more (e.g. the SGRQ, ACQ-6), the time component of the definition of baseline (i.e. time of assessment relative to the first dose) will not be considered; assessments on the same day as the first dose of study treatment will be used for baseline. If there is no value prior to the first dose of study treatment (or the same day for PROs with reflection times $\geq$ 1 week), then the baseline value will not be imputed and will be set to missing.

For information, the *planned* baseline visit for SGRQ, ACQ-6, SF-36v2, PGI-S and PSIA is Visit 4 and for SNOT-22 is Visit 3. For FEV1, the pre-bronchodilator (pre-BD) FEV_1 measurement will be used as baseline FEV_1.

Baseline for the Asthma Daily Diary variables will be the mean for data collected over the last 7 days of the run-in period prior to randomization (as defined in Table 3). If more than 3 daily entries (>50%) within that period are missing, then the baseline will be set to missing.

PGI-C and CGI-C are used for an overall evaluation of change compared to randomization visit which will be conducted only at post-baseline visits subsequent to Visit 4 (hence there is no baseline for these measures).

For laboratory data and vital signs, baseline will be defined as the latest non-missing assessment prior to first dose. If no time is recorded for an assessment, and the assessment takes place at Visit 4, this will be assumed to be a pre-dose assessment.

For physical examination, the baseline visit will be Visit 1.

3.1.2 Change from baseline

Change from baseline outcome variables are computed as

(post-baseline value – baseline value).

Percent change from baseline outcome variables are computed as:

*((post-baseline value – baseline value) / (baseline value))*100*

If either the post-baseline value or the baseline value is missing, then both change from baseline and percent change from baseline value will also be set to missing. If the baseline value is zero, the percent change will be set to missing.

3.1.3 Visit windows

For the exacerbation-related analyses no windows will be applied.

Summaries of laboratory data will have analysis-defined windows per Table 1 below for haematology parameters scheduled to be collected at Visits 5-11. For laboratory parameters not collected at every visit, the adjusted analysis-defined windows will be based on the collection schedule in Table 2 of the study protocol, and as described below. Vital signs will be based on the case report form (CRF) visit designation.

For efficacy endpoints that present visit-based data (patient and clinician reported outcomes, spirometry), variables will be summarized based on the scheduled days with adjusted analysis-defined visit windows. The adjusted analysis-defined 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 following baseline will be constructed in such a way that the upper limit of the interval falls half way between the two visits (the lower limit of the first post-baseline visit will be study day 2). If an even number of days exists between two consecutive visits then the upper limit will be taken as the midpoint value minus 1 day. Visit windows are constructed so that every observation collected can be allocated to a particular visit. No visit windows will be defined for screening visits.

The adjusted analysis-defined windows for assessments conducted at every visit of Visit 5 - Visit 11 are summarized in Table 1.

Table 1 Visit windows for assessments conducted at every visit of Visit 5 - Visit 11

Adjusted defined window visit*	Scheduled study day	Maximum windows
Week 2	15	2 ≤ Study Days ≤ 21
Week 4	29	22 ≤ Study Days ≤ 42
Week 8	57	43 ≤ Study Days ≤ 70
Week 12	85	71 ≤ Study Days ≤ 98
Week 16	113	99 ≤ Study Days ≤ 126
Week 20	141	127 ≤ Study Days ≤ 154
Week 24	169	155 ≤ Study Days ≤ Visit 11 (EOT) assessment date**

*All data are rewindowed apart from any baseline data. This is to account for any delayed IP during visit 4 (Week 0) to still be assessed as baseline.

**If no Visit 11 (EOT) assessment then the upper limit is equal to the scheduled study day of 169.

Table 2 Visit windows for assessments conducted weekly until Visit 6 (ACQ-6, PGI-S and PGI-C)

Adjusted defined window visit*	Scheduled study day	Maximum windows
Week 1	8	2 ≤ Study Days ≤ 11
Week 2	15	12 ≤ Study Days ≤ 18
Week 3	22	19 ≤ Study Days ≤ 25
Week 4	29	26 ≤ Study Days ≤ 42
Week 8	57	43 ≤ Study Days ≤ 70
Week 12	85	71 ≤ Study Days ≤ 98
Week 16	113	99 ≤ Study Days ≤ 126
Week 20	141	127 ≤ Study Days ≤ 154
Week 24	169	155 ≤ Study Days ≤ Visit 11 (EOT) assessment date**

*All data are rewindowed apart from any baseline data. This is to account for any delayed IP during visit 4 (Week 0) to still be assessed as baseline.

**If no Visit 11 (EOT) assessment then the upper limit is equal to the scheduled study day of 169.

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

(Date of assessment – date of randomization) +1

By this definition, the day of randomization will be study day 1 and the planned date of Visit 5 (Week 2) will be study day 15 (=14+1), for example.

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

- If there are 2 or more observations within the same visit window, then the non-missing observation closest to the scheduled study day will be used in the analysis.
- If 2 observations are equidistant from the scheduled study day, 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 observation with the earlier collection time 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 at every visit of Visit 5 – Visit 11, such as SGRQ which is assessed only at Visit 6 (Week 4), Visit 8 (Week 12) and Visit 11 (Week 24), the above rules will be applied to derive adjusted analysis-defined visit windows based on the protocol-defined visit schedule for that endpoint. Similarly for laboratory parameters not collected at every visit.

For pre-BD FEV_1, the non-missing value with missing or acceptable quality (acceptable or borderline quality grade) which is closest to the scheduled visit will be included in the analysis.

Home PEF assessments and daily diary data will be analyzed as weekly means throughout the treatment period, using daily diary entries between the scheduled study days contained in Table 3. Any observation recorded after the morning of Study Day 169 will not be included in the analysis but will be listed.

Table 3 Weekly windows for home PEF assessments and daily dairy data

Adjusted defined window visit	Scheduled study day	Maximum windows
Baseline	1	Evening of Study Day -7 to the morning of Study Day 1
Week 1	8	Evening of Study Day 1 to the morning of Study Day 8
Week 2	15	Evening of Study Day 8 to the morning of Study Day 15
Week 3	22	Evening of Study Day 15 to the morning of Study Day 22

Adjusted defined window visit	Scheduled study day	Maximum windows
Week 4	29	Evening of Study Day 22 to the morning of Study Day 29
Week 5	36	Evening of Study Day 29 to the morning of Study Day 36
Week 6	43	Evening of Study Day 36 to the morning of Study Day 43
Week 7	50	Evening of Study Day 43 to the morning of Study Day 50
Week 8	57	Evening of Study Day 50 to the morning of Study Day 57
Week 9	64	Evening of Study Day 57 to the morning of Study Day 64
Week 10	71	Evening of Study Day 64 to the morning of Study Day 71
Week 11	78	Evening of Study Day 71 to the morning of Study Day 78
Week 12	85	Evening of Study Day 78 to the morning of Study Day 85
Week 13	92	Evening of Study Day 85 to the morning of Study Day 92
Week 14	99	Evening of Study Day 92 to the morning of Study Day 99
Week 15	106	Evening of Study Day 99 to the morning of Study Day 106
Week 16	113	Evening of Study Day 106 to the morning of Study Day 113
Week 17	120	Evening of Study Day 113 to the morning of Study Day 120
Week 18	127	Evening of Study Day 120 to the morning of Study Day 127
Week 19	134	Evening of Study Day 127 to the morning of Study Day 134
Week 20	141	Evening of Study Day 134 to the morning of Study Day 141
Week 21	148	Evening of Study Day 141 to the morning of Study Day 148
Week 22	155	Evening of Study Day 148 to the morning of Study Day 155
Week 23	162	Evening of Study Day 155 to the morning of Study Day 162
Week 24	169	Evening of Study Day 162 to the morning of Study Day 169

The cut-off for data to be considered as in the double-blind period but not in the ANDHI IP (and therefore the end of the on-study period for the double-blind) is as follows:

If a patient transitions directly into ANDHI IP from EOT visit (Week 24) then the cut-off is the date of the EOT visit (protocol visit 11) inclusive.

If a patient transitions into ANDHI IP after follow-up visit 12 then the cut-off is the date of the day before protocol visit 13.

If a patient does not enter ANDHI IP then for efficacy the cut-off is EOT visit and for safety the cut-off is protocol visit 12 inclusive.

Any data that are to be continued to be captured in ANDHI IP such as AEs and exacerbations that were ongoing at the time of this cut-off shall remain as ongoing. If an AE or exacerbation

shows as resolved after (and including) the date of Visit 13 but started prior to the date of Visit 13 it will be shown as ongoing in the double-blind period.

3.2 Calculation or derivation of efficacy variables

3.2.1 Primary outcome variable

The annualized rate of asthma exacerbation over the 24-week treatment period will be used as the primary efficacy variable.

An asthma exacerbation will be defined as a worsening of asthma that leads to any of the following:

- Use of systemic corticosteroids (or a temporary increase in a stable OCS background dose) for at least 3 days; a single depo-injectable dose of corticosteroids will be considered equivalent to a 3-day course of systemic corticosteroids.

- An emergency room/urgent care visit (defined as evaluation and treatment for <24 hours in an emergency department or urgent care center) due to asthma that required systemic corticosteroids (as per above).

- An inpatient hospitalization (defined as admission to an inpatient facility and/or evaluation and treatment in a healthcare facility for ≥24 hours) due to asthma.

In order to calculate the number of exacerbations experienced by a patient during the 24-week treatment period, the following rules will be applied:

- The start of an exacerbation is defined as the start date of systemic corticosteroids or start date of a temporary increase in a stable oral corticosteroid background dose, or start date of hospital admission, whichever occurs the earliest.

- The end date is defined as the last day of systemic corticosteroids or the last day of a temporary increase in a stable oral corticosteroid background dose, or the date of discharge from a hospital, whichever occurs the latest.

Two or more exacerbations with the same start date and end date will be counted as one exacerbation for the purposes of calculating the number and duration of exacerbations for a patient. In the case that one or more exacerbations are recorded as starting or ending during another exacerbation, these will be counted as one exacerbation, using the earliest exacerbation start date and the latest exacerbation stop date to calculate duration.

Additional systemic corticosteroid treatments, emergency room/urgent care visits requiring use of systemic corticosteroids, or inpatient hospitalization due to asthma 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 criteria are fulfilled. If 2 or more exacerbations are recorded less than 7 days apart, these will be counted as one

exacerbation, using the earliest exacerbation start date and the latest exacerbation stop date to calculate duration.

Maximum treatment period for a patient is approximately 24 weeks; defined as the time from randomization to the date of EOT visit (Week 24). 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 EOT visit. For a patient lost to follow-up before EOT visit, the treatment period will be defined as the time from randomization to the time point after which an exacerbation could not be assessed. Exacerbations that start after EOT visit will not be included in the efficacy assessments but will be listed. If a patient misses EOT visit, then any exacerbations that start after the scheduled EOT visit date (i.e. Study Day 169) will be excluded from efficacy assessments. If an exacerbation is ongoing at EOT visit, the exacerbation will be counted in the calculation of annual exacerbation rate, however the maximum treatment period will be truncated at the date of EOT visit, 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.

3.2.2 Key secondary outcome variable

The change from baseline in SGRQ total score to the EOT visit (Week 24) is the key secondary efficacy variable and is included in the multiple testing strategy.

The SGRQ is a 50-item PRO instrument developed to measure the health status of patients 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; 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. Based on empirical data and interviews with patients, a mean change score of 4 units is associated with a minimal clinically important difference (MCID). Specific details on the scoring algorithms are provided by the developer in a user manual (Jones et al 1991).

When considering the derivation of domain scores (as per Jones et al), and the impact of missing items on each domain score, the following items should not be considered as missing items if logically skipped (in the context of the patient's prior responses):

Question 6: 'Length of Worst Attack of Chest Trouble;

- If a patient reports no 'severe or very bad unpleasant attacks of chest trouble' (Q5), the length of worst attack (Q6) will be logically skipped and should be imputed as zero with no change to the denominator for the domain / total score(s).

Question 8: 'Wheeze Worse in the Morning';

- If Frequency of Wheezing Attacks (Q4) is 'Not at All', 'Wheeze Worse in the Morning' (Q8) will be logically skipped and should be imputed as zero with no change to the denominator for the domain / total score(s).

Question 14 (parts 1-4): About your medication: 'Medication Does Not Help Very Much', 'Embarrassed Using Medication in Public', 'Having Side Effects From Medication', 'Medication Interferes With Life a Lot';

- If a patient is not taking any relevant medication, all medication questions (Q14) will be logically skipped. Therefore if all 4 responses are missing, all will be imputed as zero and the denominator for the domain / total score(s) will not be changed. If at least 1 of the 4 are answered, any of the remaining questions not answered will be treated as missing.

Potential health status treatment benefits of benralizumab will be evaluated by comparing the change from baseline at Weeks 4, 12 and EOT visit in SGRQ total score. A 4-point threshold will be used to define the response:

- SGRQ total score (Week 4/12/EOT visit – baseline) $\leq -4 \rightarrow$ Improvement
- $-4 <$ SGRQ total score (Week 4/12/EOT visit – baseline) $< 4 \rightarrow$ No change
- SGRQ total score (Week 4/12/EOT visit – baseline) $\geq 4 \rightarrow$ Deterioration

For the responder analysis of SGRQ, a responder at Weeks 4, 12 and EOT will be defined as a patient who had improvement at Weeks 4, 12 or EOT visit respectively (i.e., ≥4-point decrease in SGRQ total score at Weeks 4, 12 or EOT visit respectively). Patients who had SGRQ total score change defined as no change or deterioration will be considered as non-responders. If SGRQ total score at Weeks 4, 12 or EOT visit is missing and the patient did not complete the study, then the patient will be treated as a non-responder at that visit. For patients who complete the study but are missing the SGRQ total score at Week 12 or EOT visit, their last evaluable post-baseline score prior to that visit (regardless of analysis-defined windows) will be used to define the responder status at that visit. For patients who complete the study but are missing the SGRQ total score at Week 4, their score will be treated as missing at Week 4. Patients with a missing baseline score will have missing responder status.

The SGRQ domain (symptoms, activity, impacts) scores will be calculated and summarized to evaluate the relative contribution of each domain to the total score.

3.2.3 Other secondary outcome variables

3.2.3.1 Pre-bronchodilator forced expiratory volume in first second (pre-BD FEV_1)

The pre-BD FEV_1 will be measured by spirometry at the study center. Section 5.1.2 of the CSP contains further details of the spirometry recordings.

The change from baseline and percent change from baseline to each of the post-randomization visits over the treatment period up to and including the EOT visit (Week 24) will be used as secondary outcome variables.

3.2.3.2 Asthma control questionnaire 6 (ACQ-6)

The ACQ-6 is a shortened version of the ACQ that assesses asthma symptoms (nighttime waking, symptoms on waking, activity limitation, shortness of breath, wheezing, and short acting β2 agonist use) omitting the FEV_1 measurement from the original ACQ score.

Patients are asked to recall the status of their asthma during the previous week and respond to the symptom and bronchodilator use questions of the ACQ-6, on a 7-point scale. The ACQ-6 questions are weighted equally and scored from 0 (totally controlled) to 6 (severely uncontrolled). The mean ACQ-6 score is computed as the mean of the responses from all the items in the questionnaire. If response to any of the questions is missing, the ACQ-6 mean score will be missing.

The questionnaire will be completed by the patients using the electronic PRO (ePRO) device at the site at Visits 1 and 4 then at the patients' home thereafter, every 7 days (±2 days) for 28 days (until Visit 6; Week 4) after which it will be completed every 28 days from Visit 7 (Week 8) to EOT visit (Week 24).

The outcome variable for the ACQ-6 will be the change in mean score from baseline to each of the post-randomization assessments up to and including the EOT visit.

Asthma control responder status will be evaluated as a supportive analysis. Patients will be categorized according to the following limits (Juniper et al 2006), where end of treatment is defined as EOT visit:

- ACQ-6 (End of treatment – baseline) $\leq$-0.5 → Improvement
- -0.5 < ACQ-6 (End of treatment – baseline) <0.5 → No change
- ACQ-6 (End of treatment – baseline) $\geq$0.5 → Deterioration

An ACQ-6 responder will be defined as a patient who had improvement on ACQ-6, i.e., an ACQ-6 responder variable takes value 1 if change from baseline to end of treatment in ACQ-6 $\leq$-0.5 and 0 otherwise. Patients with missing ACQ-6 score at EOT visit and not completing the study will be considered non-responders. For patients who complete the study but are missing the ACQ-6 score at EOT visit, their last evaluable post-baseline score (regardless of analysis-

defined windows) will be used to define the responder status. Patients with a missing baseline score will have missing responder status.

Furthermore, patients will be categorized according to their ACQ-6 defined asthma control status at the end of treatment using the following score thresholds (Juniper et al 2006):

- ACQ-6 (End of treatment) ≤0.75 → Well controlled
- 0.75< ACQ-6 (End of treatment) < 1.5 → Partly controlled
- ACQ-6 (End of treatment) ≥ 1.5 → Not well controlled

Similarly, for patients who complete the study but are missing the ACQ-6 score at EOT visit, their last evaluable post-baseline score (regardless of analysis-defined windows) will be used to define the asthma control status.

3.2.3.3 Time to first asthma exacerbation

Time from randomization to the first asthma exacerbation is a secondary efficacy variable and is derived as follows:

Start date of first asthma exacerbation − Date of randomization + 1.

An exacerbation event will be defined in the same way as outlined in Section 3.2.1. The time to first asthma exacerbation for patients who do not experience an asthma exacerbation during the treatment period will be censored at EOT visit (Week 24) for patients who complete the study. Patients who withdraw from the study or are lost to follow-up before EOT visit will be censored at the last visit date after which an exacerbation could not be assessed.

3.2.3.4 Peak expiratory flow (PEF) assessment at home

Home PEF testing will be performed by the patient each morning after awakening and before taking their morning asthma medications, and each evening. Measurements should be taken at approximately the same time each day, and recorded in the Asthma Daily Diary.

Weekly means and change from baseline in weekly means will be calculated for morning PEF and evening PEF. The maximum of the 3 measurements performed at every morning and evening will used in the calculation of the weekly means. A weekly mean is calculated as the sum of all non-missing daily measures/scores over the 7 sequential days divided by the number of non-missing daily measures/scores. If more than 3 daily measures/scores (>50%) within a period are missing, then the weekly mean for that period will be set to 'missing'.

3.2.3.5 Short form 36-item health survey, version 2 (SF-36v2)

The SF-36v2 is a 36-item, self-report survey of functional health and well-being, with 1-week recall period (QualityMetric 2011). Responses to 35 of the 36 items are used to compute an 8-domain profile of functional health and well-being scores. The remaining item, referred to as the 'Health Transition' item, asks patients to rate how their current state of health compared to

their state of health 1 week ago, and is not used to calculate domain scores. The 8-domain profile consists of the following subscales: Physical Functioning (PF), Role Limitations due to Physical Health (RP), Bodily Pain (BP), General Health Perceptions (GH), Vitality (VT), Social Functioning (SF), Role Limitations due to Emotional Problems (RE), and Mental Health (MH). Psychometrically-based physical and mental health component summary scores (PCS and MCS, respectively) are computed from subscale scores to give a broader metric of physical and mental health-related quality of life. Computation will be carried out by Optum.

Norm-based scoring (NBS) is used to calculate the eight SF-36v2 subscales and the two component scores. NBS standardizes scale and component scores using the means and standard deviations from a U.S. general population normative sample derived from responses to the internet-based 2009 QualityMetric PRO Norming Study.

The norm-based scores in the U.S. general population have been set to have a mean of 50 and a standard deviation of 10. By using the NBS method, the data in the current study will be scored in relation to U.S. general population norms; therefore, all scores obtained that are below 50 can be interpreted as being below the U.S. general population norm while scores above 50 can be interpreted as above the U.S. general population norm. PCS and MCS scores are each calculated through weighted sums of all 8 scale z-scores. A more detailed description of this scoring process can be found in the SF-36v2 manual (QualityMetric 2011).

Algorithms that allow for the evaluation of SF-36v2 scale and component summary scores in the presence of item-level missing data have been previously developed using a combination of Item Response Theory, mean substitution, and regression methods. A more detailed description of this scoring process, the Missing Score Estimation, can be found in the SF-36v2 manual (QualityMetric 2011).

Two types of thresholds have been developed for interpretation of SF-36v2 scores (Table 4). The first type is suitable for comparing group mean scores and is generally referred to as the MCID. The second type is suitable for interpreting change at the individual level and is referred to as the responder threshold or responder definition (QualityMetric 2011).

Table 4 Threshold values for the SF-36v2 scale and summary measures

	SF-36v2 score									
Threshold	**PCS**	**MCS**	**PF**	**RP**	**BP**	**GH**	**VT**	**SF**	**RE**	**MH**
Group difference	2	3	3	3	3	2	2	3	4	3
Individual change	3.4	4.6	4.3	3.4	6.2	7.2	6.2	6.9	4.5	6.2

BP Bodily Pain; GH General Health Perceptions; MCS mental health component summary; MH Mental Health; PCS physical component summary; PF Physical Functioning; RE Emotional Problems; RP Role Limitations due to Physical Health; SF Social Functioning; VT Vitality.

The questionnaire will be completed at on-site visits electronically at Visit 4, Visit 8 (Week 12) and EOT visit (Week 24).

The outcome variable for the SF-36v2 will be the change in subscale/component summary score from baseline to each of the post-randomization assessments up to and including the EOT visit.

Patients will also be categorized according to the domains and threshold values in Table 4, where end of treatment is defined as EOT visit:

- SF-36v2 (End of treatment – baseline) ≥ Threshold value → Improvement
- –Threshold value < SF-36v2 (End of treatment – baseline) < Threshold value → No change
- SF-36v2 (End of treatment – baseline) ≤ –Threshold value → Deterioration.

An SF-36v2 responder will be defined for each scale and component summary score separately, as a patient who had improvement on SF-36v2 using the responder threshold in Table 4, i.e. an SF-36v2 responder variable based on PF domain takes value 1 if change from baseline to end of treatment in SF-36v2 PF score ≥4.3 and 0 otherwise. Patients with missing SF-36v2 subscale score at EOT visit and not completing the study will be considered non-responders when categorizing according to that domain. For patients who complete the study but missing a SF-36v2 subscale score at EOT visit, their last evaluable post-baseline score of that domain (regardless of analysis-defined windows) will be used to define the corresponding responder status. Patients with a missing baseline score will have missing responder status.

3.2.3.6 Patient global impression of severity (PGI-S)

The PGI-S is a single item designed to capture the patient's perception of overall symptom severity at the time of completion using a 6-point categorical response scale (0=no symptoms, 5=very severe symptoms).

This questionnaire will be completed on the ePRO device on-site at Visit 4, then at home thereafter, every 7 days for 28 days (until Visit 6; Week 4) and every 28 days from Visit 7 (Week 8) to EOT visit (Week 24).

The shift in PGI-S response from baseline to post-randomization assessments up to and including the EOT visit will be used as a secondary efficacy variable.

Patients will also be categorized according to the following criteria, where end of treatment is defined as EOT visit:

- PGI-S (End of treatment – baseline) ≤ -1→ Improvement
- PGI-S (End of treatment – baseline) = 0 i.e. no change in severity → No change
- PGI-S (End of treatment – baseline) ≥ 1 → Deterioration
- PGI-S (Baseline) = 4 or 5 or 6 shifting to PGI-S (End of treatment) = 1 or 2 or 3 → Important improvement.

The proportion of patients with improvement and important improvement will be analyzed respectively. Patients with missing PGI-S response at EOT visit and not completing the study will be considered as no improvement and no important improvement at end of treatment. For patients who complete the study but missing PGI-S response at EOT visit, their last evaluable post-baseline response (regardless of analysis-defined windows) will be used to define the categories above. Patients with a missing baseline score will have missing responder status.

3.2.3.7 Clinician and patient global impression of change (CGI-C and PGI-C)

CGI-C and PGI-C instruments are used for an overall evaluation of response to treatment. The investigators (clinicians) and patients will be asked to rate the degree of change in the overall asthma status compared to the start of treatment, i.e., randomization visit. A 7-point rating scale will be used: 1=Very Much Improved; 2=Much Improved; 3=Minimally Improved; 4=No Changes; 5=Minimally Worse; 6=Much Worse and 7=Very Much Worse.

The CGI-C should be completed before other study assessments and IP administration at all on-site visits from Visit 5 (Week 2) to EOT visit (Week 24). PGI-C is completed at home by the patient on the ePRO device, every 7 days for 28 days (until Visit 6; Week 4) and every 28 days from Visit 7 (Week 8) to EOT visit (Week 24).

Patients will also be categorized according to the following responses post-baseline, separately for CGI-C and PGI-C:

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

Agreement between CGI-C and PGI-C will be assessed at each visit with both instruments planned, 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 those visits. Agreement for EOT visit will be assessed before the imputation for missing response, as detailed below, is made.

The proportion of patients categorized as much improved and very much improved at EOT visit will be analyzed for CGI-C and PGI-C. For patients who complete the study but missing

response at EOT visit, their last evaluable post-baseline response (regardless of analysis-defined windows) will be used instead.

3.2.3.8 Predominant symptom and impairment assessment (PSIA)

The objective of this assessment is to capture the degree to which patient-stated bothersome symptoms and impairments change over time. This is achieved by generating an individualized profile of symptoms and impairments ranked in order of importance by the patient (to be performed on-site at Visit 3). The initial PSIA, at Visit 3, (Part 1) asks patients to review a list of 8 concepts (including cardinal asthma symptoms, activities, awakenings, triggers) and select those which are typically bothersome. Part 2 of the initial PSIA asks patients to rank the concepts they selected in Part 1 in order of importance: from most important (i.e., value of 1) to least important. The initial PSIA assessment produces a rank order list of bothersome concepts which will be evaluated in subsequent administrations.

Part 3 of the PSIA, administered throughout the study period, will ask the patient to record the severity of each selected symptom or impairment using an 11-point numeric rating scale (NRS) where '0' = 'did not experience' and '10' = 'worst I can imagine'.

The PSIA administered throughout the study period will be individualized per the patient ranking. Every 7 days for the first 16 weeks of the treatment period patients will be asked to record the severity of the symptoms/impairments previously selected at Visit 3 using the 11-point NRS. During this 16-week period patients were also meant to record the degree to which each selected symptom changed since starting treatment using a 7 point change scale (e.g. much worse, worse, a little worse, no change, a little better, better, much better). However this symptom change was only recorded at week 20 (see Section 6 for further details).

The NRS response for the patient's top ranked symptom or impairment will be collected. The change in this score from baseline to each of the post-randomization assessments up to and including the EOT visit, will be used as an outcome variable for the PSIA to evaluate the degree of change during the treatment period. This will be repeated for the 2nd and 3rd ranked symptoms or impairments.

In addition, the average NRS responses from the patient's top 3 ranked symptoms and impairments will be computed at each visit; if any NRS response is missing, the average score will be missing. The change in this average score from baseline to each of the post-randomization assessments up to and including the EOT visit, will also be used as an outcome variable for the PSIA to evaluate the degree of change during the treatment period.

Additionally, for each of the patient's top 3 ranked symptoms and impairments, the 7-point change scale will be used to define an improvement status for that symptom/impairment at week 20 from baseline (week 0): 1=Much Better ; 2=Better; 3=A Little Better; 4=About The Same; 5=A Little Worse; 6=Worse and 7=Much Worse. If the degree of symptom change is recorded as "better" or "much better", it is defined as an "improvement". Otherwise it is defined as "no improvement". A combined response status will also be defined at each

assessment: if all the patient's top 3 ranked symptoms and impairments show an "improvement", the combined response is defined as an "improvement"; otherwise, "no/partial improvement".

3.2.3.9 Sino-Nasal outcome test 22 item (SNOT-22)

Patients with baseline chronic rhinosinusitis with nasal polyposis will participate in a sub-study as outlined in Section 2.1.3. The rhinosinusitis health status and quality of life of the patients will be assessed using the 22-item SNOT-22.

The SNOT-22 is a further modification of the SNOT-20 (Piccirillo et al 2002), where the scoring has been simplified by removing the importance rating. In addition to the normal 20-item version of the SNOT, 2 additional items were measured, nasal blockage, and loss of sense of taste and smell. Each of the 22-questions is on a 6-point scale ranging from 0 (no problem) to 5 (problem as bad as it can be). The total score is calculated as the sum of the responses to all questions answered, with a range from 0 to 110 (higher scores indicate poorer outcomes). If response to any of the questions is missing, the total score will be missing. A MCID of 8.90 has been established (Hopkins et al 2009).

The questionnaire will be completed at on-site visits electronically at Visit 3, Visit 6 (Week 4), Visit 8 (Week 12) and EOT visit (Week 24).

The outcome variable for the SNOT-22 will be the change in total score from baseline to each of the post-randomization assessments up to and including the EOT visit.

SNOT-22 responder status will be evaluated as a supportive analysis. Patients will be categorized according to the following limits, where end of treatment is defined as EOT visit:

- SNOT-22 (End of treatment – baseline) $\leq$ -8.9 $\rightarrow$ Improvement
- -8.9 < SNOT-22 (End of treatment – baseline) < 8.9 $\rightarrow$ No change
- SNOT-22 (End of treatment – baseline) $\geq$ 8.9 $\rightarrow$ Deterioration

An SNOT-22 responder will be defined as a patient who had improvement on SNOT-22, i.e., an SNOT-22 responder variable takes value 1 if change from baseline to end of treatment in SNOT-22 total score $\leq$ -8.9 and 0 otherwise. Patients within the nasal polyp sub-study analysis set, with missing SNOT-22 total score at EOT visit and not completing the study will be considered non-responders. For patients who complete the study but missing SNOT-22 total score at EOT visit, their last evaluable post-baseline total score (regardless of analysis-defined windows) will be used to define the responder status. Patients with a missing baseline score will have missing responder status.

3.3 Calculation or derivation of safety variables

The following safety data will be collected in the double-blind period: reported AEs (including serious AEs [SAEs]), hematology, clinical chemistry, urinalysis, vital signs, local ECGs, and physical examination.

All safety measurements will use all available data during this period for analyses, including data from unscheduled visits and repeated measurements. No safety data will be imputed.

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

3.3.1 Adverse events (AEs)

Adverse events (including SAEs) experienced by the patients will be collected throughout the entire study and will be coded by the AstraZeneca designee using the latest version of the Medical Dictionary for Regulatory Activities (MedDRA).

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

- AEs in the pre-treatment period are defined as those with onset before the day of first dose of study treatment.

- AEs in the on-treatment period are defined as those with onset between day of first dose of study treatment and scheduled EOT visit or IPD visit for those patients who prematurely discontinue study treatment, inclusive. If both EOT and IPD visits are missing, then the upper limit of the on-treatment period is defined as the day of last dose of study treatment + 56 days. If the upper limit is after the end of on-study period then set the upper limit to end of on-study period

- AEs in the on-study period are defined as those with onset between day of first dose of study treatment and up to

    (a) the day of the scheduled follow-up visit, inclusive if the patient does not enter ANDHI IP.

    (b) the day before the scheduled visit 13 if the patient delays transition into ANDHI IP.

    (c) the day of the scheduled EOT visit inclusive, if the patient transitions directly into ANDHI IP from EOT (visit 11).

3.3.2 Laboratory variables

Blood samples for determination of clinical chemistry and hematology parameters will be taken at the times detailed in the CSP, and will be assessed in a central laboratory. The parameters outlined in Section 5.2.1, Tables 4 of the CSP will be collected.

In summaries, listings and figures, lab results and normal ranges will be presented in the Standard International (SI) unit. Eosinophils data will be presented in both SI and conventional units (cells/μL) in summaries.

Changes in hematology and clinical chemistry variables between baseline and each post-baseline assessment will be calculated. Baseline is defined as the last available non-missing value measured prior to the first dose of study 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. 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 analyzed 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 laboratory reference ranges will be used for laboratory variables. All absolute values falling outside the reference ranges will be flagged.

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

For the liver function tests: Aspartate Aminotransferase (AST), Alanine Aminotransferase (ALT), Alkaline phosphatase, Gamma-GT (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:

Multiple = Value / ULN,

i.e. 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:

- AST ≥ 3x ULN
- ALT ≥ 3x ULN
- TBL ≥ 2x ULN

3.3.3 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 Table 2 of 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 study 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 5 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 5 Vital signs reference ranges

Parameter	Standard Units	Lower Limit	Upper Limit
Diastolic Blood Pressure	mmHg	60	120
Systolic Blood Pressure	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:

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

3.3.4 Local ECGs

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

3.3.5 Physical examination

Complete and brief physical examinations will be performed at time points specified in Section 4, Tables 1 and 2 of the CSP. For the physical examinations only information on whether the assessment was performed or not is to be recorded.

Baseline data will be collected at Visit 1. Any new finding(s) or aggravated existing finding(s), judged as clinically significant by the investigator or designee, will be reported as an AE.

3.4 Other asthma control variables

Daily diary metrics (rescue medication use, awakenings due to asthma symptoms requiring rescue medication, and asthma symptom scores) will be used as additional asthma control variables in support of primary analysis, and will be derived from the data collected through the ePRO device. Daily diary metrics will be recorded in the Asthma Daily Diary each day from the evening of Visit 2 to the morning of Visit 11.

Baseline is defined as the average of the last 7 days before randomization for the daily metrics, as defined in Section 3.1.1. Post-randomization periods for the daily diary metrics will be defined for the calculation of weekly means using the analysis-defined study windows described in Section 3.1.3 and listed in Table 3.

The post-randomization weekly means for daily diary metrics are calculated as the sum of all non-missing daily measures/scores over the 7 day window divided by the number of non-missing daily measures/scores. If more than 3 daily measures/scores (>50%) within that window are missing, then the mean daily measure/score for that period will be set to missing.

3.4.1 Asthma symptom score

Asthma symptoms during night-time and daytime will be recorded by the patient each morning and evening in the Asthma Daily Diary. Symptoms will be recorded using a scale of 0-3, where 0 indicates no asthma symptoms. Asthma symptom daytime score (recorded in the evening), night-time score (recorded in the morning of the next calendar day), and total score will be calculated and presented separately. The daily asthma symptom total score will be calculated by taking the sum of the daytime score recorded in the evening and the night-time score recorded the following morning. If a patient is missing a value for either the daytime or night-time asthma symptom score on a given day, then the total score for that day will be set to missing.

Weekly mean scores and change from baseline in weekly mean scores will be calculated for total asthma symptom score, daytime asthma symptom score and night-time asthma symptom score.

The number of asthma symptom-free days will be calculated for each patient as the total number of days in the 24-week treatment period where the total asthma symptom score is 0. The proportion of asthma symptom-free days will be calculated using the total number of days with completed asthma symptom score diary during the 24-week treatment period as the denominator.

3.4.2 Rescue medication

The number of times (occasions) rescue medication inhalations and nebulizer treatments taken will be recorded by the patient in the Asthma Daily Diary twice daily. Daytime use is recorded in the evening and night-time use is recorded in the morning of the next calendar day. Rescue medication inhaler usage will be captured as the number of times (occasions) that the patient used inhaler irrespective of number of puffs taken. Nebulizer use is the same as one occasion of inhaler use.

The number of times (occasions) rescue medication inhalations and nebulizer treatments captured in the Asthma Daily Diary each day will be calculated per patient. If a patient is missing a value for either daytime or night-time rescue medication use on a given day, then the total rescue medication use for that day will be set to missing.

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

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

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

Number of daytime inhaler times (recorded in the evening diary) + number of nighttime inhaler times (recorded in the morning diary for next calendar day).

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

Number of daytime nebulizer times (recorded in the evening diary) + number of nighttime nebulizer times (recorded in the morning diary for next calendar day).

Weekly mean rescue medication use (average number of times/day) and change from baseline in weekly mean rescue medication use will be calculated. If more than 3 daily entries within that period are missing, then the average will be set to missing.

3.4.3 Night-time awakening due to asthma

Night-time awakenings due to asthma symptoms and requiring rescue medication use will be recorded by the patient in the Asthma Daily Diary each morning by answering the question whether he/she woke up during the night due to asthma symptoms by a “yes” or “no” response, as well as the follow-up question whether he/she used rescue medication upon awakening during the night.

The weekly proportion of nights with nocturnal awakenings for the patient due to asthma and requiring rescue medication use out of the nights with non-missing nighttime awakening data and the corresponding change from baseline for each post-randomization period will be calculated.

The weekly proportion will be calculated as the number of times the answers to each of the questions ("Did your asthma casue you to wake up during the night?" and "Did you use your rescue medication upon awakening during the night?") over the 7 day window are non-missing and equal to "Yes", divided by the number of times the answer to the question "Did your asthma cause you to wake up during the night?" is non-missing, and multiplied by 100.

If more than 3 days have missing answers (>50%) within that window (excluding any missing that are logically skipped), then the weekly proportion for that period will be set to missing. (i.e. if ' Did your asthma casue you to wake up during the night?' is 'No', then 'Did you use your rescue medication upon awakening during the night?' should not be considered missing.)

In addition, the proportion of nights with awakenings independent of rescue medication use during the randomized treatment period will also be calculated similarly.

3.4.4 Composition of daily diary measures

Daily diary metrics will be used to determine asthma control status based on the GINA control definition (GINA 2018). Asthma control will be assessed:

- On a weekly basis using diary data from the previous 7 days (regardless of analysis-defined windows), from week 1 until week 24 (EOT) (or study discontinuation if sooner); and
- At weeks 4, 8, 12, 16, 20 and 24 (EOT) based on diary data from the previous 4 weeks.

Baseline asthma control status will be determined at week 0 (visit 4) based on the previous 7 days' / 4 weeks' diary data as appropriate.

Asthma control status will be determined in accordance with Table 6 for 7 day assessment and Table 7 for 4 week assessment:

Table 6 Asthma control status based on 7 days previous data

GINA Control Criterion	Asthma Control Status
Daytime symptom score > 0 no more than twice in the 7 previous days	**Well controlled**: All criteria met and no protocol-defined exacerbation **Partly controlled**: 1-2 of these criteria not met and no protocol-defined exacerbation **Uncontrolled**: 3-4 of these criteria not met and/or ≥ 1 protocol-defined exacerbation
No nocturnal awakenings due to asthma and night-time symptom score = 0 every night in the 7 previous days	
Total rescue medication use > 0 no more than twice in the 7 previous days	
Daytime asthma symptom score < 2 every day of the 7 previous days	

Table 7 Asthma control status based on previous 4 weeks data

GINA Control Criterion	Asthma Control Status
Daytime symptom score > 0 no more than twice in each of the 4 previous weeks	**Well controlled**: All criteria met and no protocol-defined exacerbation **Partly controlled**: 1-2 of these criteria not met and no protocol-defined exacerbation **Uncontrolled**: 3-4 of these criteria not met and/or ≥ 1 protocol-defined exacerbation
No nocturnal awakenings due to asthma and night-time symptom score = 0 every night in each of the 4 previous weeks	
Total rescue medication use > 0 no more than twice in each of the 4 previous weeks	
Daytime asthma symptom score < 2 every day in each of the 4 previous weeks	

If missing data prevent the asthma control status from being determined, control status will be set to missing. Note that a limited amount of missing data may still allow the control status (particularly an **uncontrolled** status) to be determined. Patients who withdraw from the study prior to EOT will have control status at EOT set to missing.

The time to the first well-controlled week (based on the previous 7 days' diary data) will also be recorded and defined as follows:

End date of first asthma well-controlled week (scheduled study date of week per Table 2) − Date of randomization + 1.

The time to first asthma well-controlled week for patients who do not experience an asthma well-controlled week during the treatment period will be censored at EOT visit (Week 24) for patients who complete the study. Patients who withdraw from the study or are lost to follow-up before EOT visit will be censored at the last week an asthma well-controlled week is assessed.

An EOT control indicator variable will be created to assess asthma control status based on the data collected in the 4 weeks prior to EOT. This variable will be set to 1 if asthma control status, assessed over the 4 weeks prior to EOT, is well-controlled and 0 otherwise. Patients with missing EOT control status who do not complete the study will have their control indicator variable set to 0. For patients who complete the study, but are missing EOT control status their last available assessment of control (regardless of analysis-defined windows) will be used to determine their control indicator variable..

A second EOT control indicator variable will be created following the rules above, except it will be set to 1 if asthma control status, assessed over the 4 weeks prior to EOT, is well-controlled or partly controlled.

4. ANALYSIS METHODS

4.1 General principles

The analysis of the efficacy endpoints will include all data captured during the 24-week double-blind treatment period, defined as the period after randomization at Visit 4 and the conclusion of EOT visit, 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 benralizumab to placebo.

The analysis of safety endpoints will include all data captured during the double-blind on-study period, defined as the period after first administration of IP at Visit 4 and up to

(a) For patients who do not enter ANDHI IP: the conclusion of the scheduled post-treatment follow up visit, inclusive for adverse events and concomitant medications. For all other safety endpoints the conclusion of the scheduled end of treatment visit.

(b) For patients who transition directly into ANDHI IP: the conclusion of the scheduled end of treatment visit for all safety endpoints.

(c) For patients who have delayed transition into ANDHI IP: the day before the date of first run-in visit for ANDHI IP, inclusive for adverse events and concomitant medications. For all other safety endpoints the conclusion of the scheduled end of treatment visit.

Statistical models will include stratification factors based on CRF data for previous exacerbations, maintenance OCS use at baseline, and region, and not data entered in IVRS unless specified otherwise.

If the number of previous exacerbations is missing from the CRF, the following imputation will be applied for the purposes of the statistical modelling:

- Previous exacerbations entered in IVRS =2, CRF previous exacerbations = 2;
- Previous exacerbations entered in IVRS >=3, CRF previous exacerbations = 3.

P-values will be rounded to 4 decimal places.

4.1.1 Testing strategy to account for multiplicity considerations

To account for multiplicity to test one primary variable and one key secondary variable, the following hierarchical testing strategy to control for the overall type-1 error (0.05) will be adopted:

- Initially, test the annualized asthma exacerbation rate at the two-sided 5% significance level
- If annualized asthma exacerbation rate is significant (i.e., the p-value for the primary analysis is less than 0.05) then test SGRQ total score change from baseline to EOT visit at the two-sided 5% significance level

4.2 Analysis methods

4.2.1 Patient disposition

Patient disposition will be summarized using the all patients analysis set. The total number of patients will be summarized for the following groups: those who enrolled, those who entered run-in, and those who were not randomized (and reason). The number and percentage of patients within each treatment group will be presented by the following categories: randomized, not randomized (and reason), received treatment with study drug, did not receive treatment with study drug (and reason), completed treatment with study drug, discontinued treatment with study drug (and reason), discontinued treatment with study drug but completed study follow-up, completed double-blind period of the study, and withdrawn from double-blind period of the study (and reason). The number of patients who transitioned into the open label ANDHI IP substudy will also be presented.

The number of patients randomized by region, country and center will be summarized by treatment group in the full analysis set.

4.2.2 Demography data and patient characteristics

Demography data such as age (as continuous and categorical variable in 3 categories: <50 years / ≥50 to <65 years / ≥ 65 years), country, gender, race, and ethnicity will be summarized

by treatment group and for all patients in the full analysis set. 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 CRF will be used. The table will be repeated for patients with screening blood eosinophil count:

- ≥150 -<300 cells/μL and
- ≥300 cells/μL.

Various baseline characteristics will also be summarized by treatment group and for all patients in the full analysis set. These include patient characteristics (weight, height and BMI), blood eosinophil count at baseline (as continuous and categorical variable in 3 categories: <300 cells/μL / ≥300 to <450 cells/μL / ≥450 cells/μL) and at screening (first assessment with blood eosinophil count ≥300 cells/μL before Visit 4 if achieved, otherwise first assessment ≥150 cells/μL before Visit 4 and categorized as ≥150 to <300 cells/μL / ≥300 cells/μL; and for the former category the additional clinical criteria met for study inclusion (see Inclusion Criterion #8 in CSP), and subjects meeting at least 1 of the additional clinical criteria), total IgE and Phadiatop, smoking status, medical and surgical histories, lung function data at screening (i.e. FEV_1, FEV_1/FVC, reversibility), PRO and diary data at baseline, home lung function (morning and evening PEF) at baseline, and respiratory disease characteristics including asthma duration, age at onset of asthma, the number of exacerbations in the previous 12 months, and the number of exacerbations requiring hospitalizations in the previous 12 months. Lung function data at screening, PRO and diary data at baseline and respiratory disease characteristics at study entry will be repeated for patients with screening blood eosinophil count:

- ≥150 - <300 cells/μL and
- ≥300 cells/μL.

Medical and surgical histories will be summarized by MedDRA Preferred Term (PT) within MedDRA System Organ Class (SOC).

4.2.3 Prior and concomitant medications

A medication will be classified as a maintenance asthma medication at baseline if it started prior to or on the date of randomization and was ongoing after randomization. Maintenance asthma medications with a stop date on the date of randomization will be considered as prior medication. ICS doses will be converted to their Fluticasone Propionate equivalent in micrograms and OCS doses will be converted to their Prednisolone equivalent in milligrams.

A medication will be regarded as prior if it was stopped on or before the date of randomization (medication stop date ≤ date of randomization). A medication will be regarded as concomitant if the start date is on or after the date of randomization, or if it started prior to the date of randomization and was ongoing after the date of randomization. Medications with start date after the on-treatment period (as defined for AEs in Section 3.3.1) will not be considered as

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

The number and percentage of patients taking maintenance asthma medications, including ICS/LABA fixed dose combinations, at baseline will be summarized and for those patients taking ICS and OCS at baseline, the converted dose will be summarized. The proportion of patients taking maintenance OCS at baseline will also be categorized by screening blood eosinophil count (≥150 - <300 cells/μL / ≥300 cells/μL). The number of patients treated with ICS at baseline will be summarized by ATC code and preferred term, with total daily dose (non-converted) at baseline summarized for each preferred term.

The number and percentage of patients who take prior medications, will be presented by treatment group. Medications will be classified according to the WHO Drug Dictionary (WHO DD). The summary table will present data by generic term within ATC code.

Total number and percentage of days on systemic corticosteroid treatment associated with asthma exacerbations and cumulative OCS dose associated with asthma exacerbations during the treatment period will be summarized descriptively by treatment group for the full analysis set.

4.2.4 Study treatments

4.2.4.1 Exposure

Exposure to IP will be calculated in days as:

Last dose date of IP in double-blind treatment period-first dose date of IP+1,

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

4.2.4.2 Compliance

Study treatment compliance in the double-blind treatment period will be summarized descriptively for the full analysis set and will be calculated as:

(Total doses administered/total doses expected) x 100.

Patients who received no study treatment in the double-blind treatment period will have zero compliance.

4.2.5 Analysis of the primary variable

The primary efficacy variable is the annualized rate of asthma exacerbations and the primary analysis is to compare the exacerbation rate of benralizumab with placebo based on the full analysis set.

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, i.e.:

H_0: Rate ratio (benralizumab vs placebo) = 1
H_a: Rate ratio (benralizumab vs placebo) ≠ 1

The exacerbation rate on benralizumab will be compared to the exacerbation rate on placebo using a negative binomial model for the primary analysis. The response variable in the model will be the number of asthma exacerbations experienced by a patient, over the 24-week double-blind treatment period. The model will include covariates of treatment group, region (North America/rest of world), number of exacerbations in previous year (count, as a continuous variable) , and maintenance OCS use at baseline (yes/no). Number of exacerbations in previous year (count) will not be used as a covariate in the subgroup analysis by number of exacerbations in previous year (categorical). The logarithm of the patient's corresponding treatment period will be used as an offset variable in the model to adjust for patients having different exposure times during which the events occur.

The estimated treatment effect (i.e., the rate ratio of benralizumab versus placebo), corresponding 95% CI, and 2-sided p-value for the rate ratio will be presented. In addition, the annual exacerbation rate and the corresponding 95% CI within each treatment group and the absolute difference between treatment groups with the corresponding 95% CI will be presented. Marginal standardization methods will be used on the model estimates for all negative binomial analyses, unless otherwise specified.

The above analysis will be repeated for the following:

- Modification of primary endpoint: Exacerbations associated with hospitalization and/or emergency room visit
- Patients with screening blood eosinophil count ≥150 - <300/ μL
- Patients with screening blood eosinophil count ≥300/ μL
- Patients with screening blood eosinophil count ≥150 - <400/ μL
- Patients with screening blood eosinophil count ≥400/ μL
- Patients with screening blood eosinophil count ≥150 - <500/ μL
- Patients with screening blood eosinophil count ≥500/ μL

Asthma exacerbation summary statistics will be presented based on the full analysis set by treatment group.

Sensitivity analyses

Sensitivity analyses to assess the robustness of the primary analysis results to missing data as outlined in Appendix I may be conducted depending on the amount of missing data due to patients who discontinue IP and/or withdraw from the study.

In addition, the following analyses will also be performed:

- A Poisson regression model taking over dispersion into account will be included as a sensitivity analysis for the primary analysis. The correction for potential over dispersion will be made by Pearson chi-square. The response variable, covariates and offset variable will be the same as described for the primary analysis.

- The primary analysis will be repeated where the time at risk (which is included in the model as an offset variable) excludes any time during which a patient is having an exacerbation (where exacerbation is as detailed in Section 3.3.1 plus 7 days).

Additional sensitivity analyses may 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; and/or

- Patients with important protocol deviations that may have a direct effect on efficacy results (for example: non-qualifying blood eosinophil counts during screening).

Subgroup analyses

To explore the uniformity of the detected overall treatment effect on the primary efficacy variable, subgroup analyses and statistical modeling including testing for interaction between treatment and covariates will be performed on the full analysis set for the following factors:

- Gender (male, female)

- BMI ($\leq$30 kg/m^2, >30 kg/m2)

- Geographic region (North America, rest of world)

- Number of exacerbations in previous year (2, $\geq$3 exacerbations)

- History of nasal polyps (yes, no)

- Age at asthma onset (<18, $\geq$18 years)

- Screening blood eosinophil counts ($\geq$150 - <300/ μL and $\geq$300)

- FVC % predicted at baseline (<65%, $\geq$65%)

- Maintenance OCS use at baseline (yes, no)

For each of the subgroup factors in turn, a separate negative binomial regression model will be fitted using the same model terms as used for the primary analysis, with additional terms for the subgroup main effect and the treatment by subgroup interaction. The p value of the

treatment by subgroup interaction will be presented. Similar output will be presented for each subgroup as for the primary analysis.

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. If any model does not converge, sub-groups may be collapsed appropriately.

4.2.6 Analysis of the key secondary variable

Change from baseline in SGRQ total score at EOT visit is a multiplicity protected key secondary efficacy variable (Section 4.1.1).

Change from baseline in SGRQ total score at EOT visit will be compared between the benralizumab group and the placebo group using a restricted maximum likelihood (REML) based on a mixed-effect model for repeated measures (MMRM) analysis on patients with a baseline SGRQ total score and at least one post-randomization SGRQ total score in the full analysis set.

The dependent variable will be the change from baseline in SGRQ total score at post-baseline protocol-specified visits (up to the EOT visit). Treatment group will be fitted as the explanatory variable, region (North America/rest of world), number of exacerbations in previous year, maintenance OCS use at baseline (yes/no), visit, and treatment*visit interaction as fixed effects and baseline SGRQ total score as a covariate. The variance-covariance matrix will be assumed to be unstructured. If the procedure does not converge then a compound symmetric variance-covariance matrix will be used instead. If the procedure still does not converge other structures for the matrix will be explored and/or the covariates will be explored to see if, for example, changing from counts to categorical data leads to convergence,if a covariate should be removed, or if the OM option below should be removed.

The model is:

*Change from baseline in SGRQ total score = Treatment group + baseline SGRQ total score + region + maintenance OCS use at baseline + number of exacerbations in previous year + visit + treatment*visit*

Results will be presented in terms of least square means (LSMEANS), treatment differences in LSMEANS, 95% CI and p-values for all visits. The results at EOT visit (Week 24) will be of primary interest. The LSMEANS will be calculated using the OM option in the LSMEANS statement and the visit by treatment interaction to provide an LSMEAN estimate for each scheduled visit.

Sensitivity analyses to assess the robustness of the repeated measures analysis to missing data as outlined in Appendix I may be conducted depending on the amount of missing data due to patients who discontinue IP and/or withdraw from the study.

Subgroup analyses, 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 factors previously specified in Section 4.2.5 based on the full analysis set.

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

The proportion of patients in terms of SGRQ total score response status (improvement, deterioration, no change, and missing) at Weeks 4, 12 and EOT will be summarized descriptively by treatment group. The corresponding SGRQ total score responder statuses will be analyzed using a logistic regression model with covariates of treatment, region (North America/rest of world), number of exacerbations in previous year, maintenance OCS use at baseline (yes/no) and baseline SGRQ total score. The results of the analyses will be presented as odds ratios with associated 95% CIs and 2-sided p-values.

Change from baseline in SGRQ domain scores (symptoms, activity, and impacts) during the 24-week treatment period will be analyzed separately using a similar model as the above MMRM model for change from baseline in SGRQ total score.

4.2.7 Analysis of the other secondary variables

4.2.7.1 Pre-bronchodilator forced expiratory volume in first second (pre-BD FEV_1)

Change from baseline and percent change from baseline in pre-BD FEV1 over the 24-week treatment period will be summarized. Change from baseline in pre-BD FEV1 over the 24-week treatment period will also be analyzed using the same method, MMRM, as for the primary analysis of SGRQ total score described in Section 4.2.6, for patients in the full analysis set. Included in the model will be the baseline pre-BD FEV1 measurement plus age (or age group [in 3 categories as specified in Section 4.2.2] dependent on model fit/convergence) and gender. No sensitivity analysis will be performed for this endpoint.

Subgroup analyses, using the same MMRM model as the primary analysis of this endpoint with additional terms for the subgroup main effect and the treatment by subgroup interaction, will also be conducted for the factors previously specified in Section 4.2.5 based on the full analysis set.

4.2.7.2 Asthma control questionnaire 6 (ACQ-6)

Change from baseline in ACQ-6 score will be summarized and analyzed using the same method, MMRM, as for the primary analysis of SGRQ total score described in Section 4.2.6, for patients in the full analysis set. Included in the model will be the baseline ACQ-6 mean score. Sensitivity analyses will not be performed.

Subgroup analyses, using the same MMRM model as the primary analysis of this endpoint with additional terms for the subgroup main effect and the treatment by subgroup interaction, will also be conducted for the factors previously specified in Section 4.2.5.

Asthma control responder status based on ACQ-6 at EOT, as defined in Section 3.2.3.2, will be analyzed using the same method, logistic regression model, as for SGRQ total score responder status described in Section 4.2.6.

The number and percentage of patients achieving an improvement, no change, or deterioration, and the number and percentage of patients achieving mean ACQ-6 ≤ 0.75 (Well controlled), >0.75 - < 1.5 (Partly controlled) and ≥ 1.5 (Not well controlled) at EOT as defined in Section 3.2.3.2 will be summarized by treatment group.

4.2.7.3 Time to first asthma exacerbation

Time to first asthma exacerbation will be analyzed as 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 data with the covariates of treatment, region (North America/rest of world), number of exacerbations in previous year, maintenance OCS use at baseline (yes/no). The model will include Efron method for handling ties (TIES=EFRON), and the confidence intervals for hazard ratios will be calculated using profile-likelihood confidence limits (RL=PL).

Results of these analyses will be summarized as hazard ratios, 95% confidence intervals and p-values for patients in the full analysis set.

Time to first asthma exacerbation will be displayed graphically using a Kaplan-Meier plot.

4.2.7.4 Peak expiratory flow (PEF) assessment at home

Change from baseline in weekly mean morning and evening PEF will each be summarized and analyzed using the same method, MMRM, as for the primary analysis of SGRQ total score described in Section 4.2.6, for patients in the full analysis set. Included in the model will be the baseline morning and evening PEF. Sensitivity and subgroup analyses will not be performed.

4.2.7.5 Short form 36-item health survey, version 2 (SF-36v2)

Change from baseline in SF-36v2 subscale and component summary scores (PF, RP, BP, GH, VT, SF, RE, MH, PCS, MCS) will each be summarized and analyzed using the same method, MMRM, as for the primary analysis of SGRQ total score described in Section 4.2.6, for patients in the full analysis set. Included in the model will be the corresponding baseline score. Sensitivity and subgroup analyses will not be performed.

The SF-36v2 responder statuses based on subscale and component summary scores at EOT and the responder threshold values, as defined in Section 3.2.3.5, will each be analyzed using the same method, logistic regression model, as for SGRQ total score responder status described in Section 4.2.6.

4.2.7.6 Patient global impression of severity (PGI-S)

The patients' improvement statuses (improvement, important improvement) based on PGI-S at EOT, as defined in Section 3.2.3.6, will each be analyzed using the same method, logistic regression model, as for SGRQ total score responder status described in Section 4.2.6, for patients in the full analysis set. Included in the model will be the baseline PGI-S response.

The number and percentage of patients in each category of the PGI-S responses will be summarized by treatment group and visit, along with a shift table to demonstrate the change from baseline in PGI-S categories throughout the treatment period.

4.2.7.7 Clinician and patient global impression of change (CGI-C and PGI-C)

The CGI-C and PGI-C responses will be summarized by treatment group and visit for patients in the full analysis set. The number and percentage of patients will be presented for CGI-C, PGI-C, and for agreement in CGI-C and PGI-C responses as described in Section 3.2.3.7.

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

The CGI-C's and PGI-C's responder statuses (much improved, very much improved) at EOT will each be analyzed using the same method, logistic regression model, as for SGRQ total score responder status described in Section 4.2.6. No baseline response will be included in the model.

4.2.7.8 Predominant symptom and impairment assessment (PSIA)

For each of the PSIA symptom / impairment concepts, the number and percentage of patients ranking the concepts 1st through to 8th as well as "Not Scored", as part of the initial PSIA, will be summarized descriptively by treatment group for the full analysis set.

The change from baseline (Visit 4) in severity of each patient's top ranked symptom/impairment, based on NRS responses as defined in Section 3.2.3.8, will be summarized and analyzed using the same method, MMRM, as for the primary analysis of SGRQ total score described in Section 4.2.6, for patients with a baseline score and at least one post-baseline rating in the full analysis set. The corresponding baseline score will be included in the model. Sensitivity and subgroup analyses will not be performed. The absolute severity at baseline and at each assessment will also be summarized. This will be repeated for each of the 2nd and 3rd ranked symptoms/impairments.

The change from baseline in the average of each patient's top 3 ranked symptoms and impairments will be summarized and analyzed using a MMRM, as above. The average of the patient's top 3 ranked symptoms and impairments at baseline will be included in the model. The absolute average severity at baseline and at each assessment will also be summarized.

The number and percentage of patients in each category of the change scale responses will be summarized descriptively by treatment group at week 20 for each ranked symptom/ impairment (ranked 1st, 2nd, 3rd through to 8th).

The number and proportion of patients recording an "improvement" in their top ranked symptom/impairment will be summarized by treatment group at week 20, for the full analysis set. 95% confidence intervals about the proportions (calculated using the exact Clopper-Pearson formula) will be presented. The number of patients in each treatment group with missing data at Week 20 will also be reported. The above analysis will be repeated for the 2nd and 3rd top ranked symptoms/impairments and for the combined response (based on the 3 top ranked symptoms and impairments: see Section 3.2.3.8). Note that when calculating these proportions, the denominator will be the number of patients with an assessment of 1, 2 or 3 ranked (respectively) bothersome symptoms/impairments at Week 20. This is not consistent with other "responder"-type analyses in this study but because change data were not collected prior to Week 20 (see Section 6), LOCF imputation for any missing data is not possible.

Improvement status at week 20 for the top ranked symptom/impairment will be analyzed using the same method, logistic regression model, as for the SGRQ total score responder status described in Section 4.2.6. No baseline response will be included in the model. This analysis will be repeated for the combined response (based on the 3 top ranked symptoms and impairments) at Week 20.

4.2.7.9 Sino-Nasal outcome test 22 item (SNOT-22)

Change from baseline in SNOT-22 total score will be summarized and analyzed using the same method, MMRM, as for the primary analysis of SGRQ total score described in Section 4.2.6, for patients in the chronic rhinosinusitis with nasal polyposis sub-study analysis set. Included in the model will be the baseline SNOT-22 total score. Sensitivity and subgroup analyses will not be performed.

The SNOT-22 responder status based on SNOT-22 total score at EOT, as defined in Section 3.2.3.9, will be analyzed using the same method, logistic regression model, as for SGRQ total score responder status described in Section 4.2.6.

The analyses described above will be repeated for patients with SNOT-22 total score at baseline >30.

4.2.8 Analysis of safety variables

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

4.2.8.1 Adverse events (AEs)

Adverse events will be summarized separately for the on-study and on-treatment periods, as defined in Section 3.3.1. All AEs will be listed for each patient, regardless of treatment period. 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, SAEs, AEs with outcome of death, and AEs leading to discontinuation of IP (DAEs). The total number of AEs in the different AE categories in terms of AE counts will also be presented (i.e., accounting for multiple occurrences of the same event in a patient).

Adverse events, AEs with outcome of death, SAEs, and DAEs will be summarized by SOC and PT assigned to the event by MedDRA. For each PT, the number and percentage of patients reporting at least one occurrence will be presented, i.e. for a patient, multiple occurrences of an AE will only be counted once. SAEs causing discontinuation of the study treatment and SAEs causing discontinuation from the study will also be summarized.

The number of AEs and SAEs will also be summarized by SOC and PT by MedDRA for the on-study and on-treatment periods. For each PT, the number of occurrences will be presented.

The incidence rate of AEs per person-years at risk, calculated as (number of patients reporting AE)/ (total period with patients at risk of AE), will also be reported for the on-study and on-treatment periods. The total period at risk for each patient will be defined as the period from first dose of study treatment to the date of the EOT or IPD visit, or day of last dose of study treatment + 56 days when both EOT or IPD visits are missing, or end of study date if day of last dose of study treatment + 56 days is after end of study date (see Section 3.3.1) for the on-treatment period; and as the period from first dose of study treatment to the EOT visit, follow-up visit or scheduled visit 13 (see Section 3.3.1) for the on-study period. Rates will be expressed in terms of events per 100 patient-years.

A summary of the most common (frequency of >5%) AEs and will be presented by PT. AEs, SAEs and DAEs will be summarized 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).

Adverse events of injection site reactions (high level term of administration and injection site) and hypersensitivity (standardized MedDRA query of hypersensitivity) will be summarized by preferred term. Hypersensitivity events will be listed.

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

4.2.8.2 Laboratory data

All continuous laboratory parameters will be summarized descriptively by absolute value at each visit by treatment group, together with the corresponding changes from baseline. All parameters will be summarized in SI units, with the exception of blood eosinophil counts which will be summarized 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, high, and missing values. The shift tables will present baseline and maximum/minimum post-baseline 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. A change is treatment-emergent if it occurs during on-treatment period as defined in Section 3.3.1. This data presentation will include all visits for this subset of patients.

Maximum post-baseline bilirubin elevations by maximum post-baseline ALT and AST will be presented, expressed as multiples of ULN. Bilirubin will be presented in multiples of the following ULN ≤1.5, >1.5-2, >2, and AST and ALT will be presented in multiples of the following ULN ≤1, >1-3, >3-5, >5-10, >10.

Maximum post-baseline total bilirubin will be presented (<2 and ≥2 x ULN) against maximum post-baseline ALT (<3, ≥3 - <5, ≥5-<10, and ≥10 x ULN), expressed as multiples of ULN. This will be repeated to show maximum post-baseline total bilirubin against maximum post-baseline AST.

Data for patients with ALT or AST ≥3 x ULN, and bilirubin ≥2 x ULN will be presented, which will include all visits for this subset of patients.

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

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

4.2.8.3 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 displaying low, normal, high, and missing values, as applicable for each parameter and will include patients with both baseline and post-baseline data.

All recorded vital signs data will be listed.

4.2.8.4 Local ECGs

The investigator's overall evaluation of ECG (normal or abnormal) will be listed for all patients, detailing whether any abnormalities were clinically significant or not.

A summary table will be produced for baseline ECG evaluation to display normal, abnormal – not clinically significant, abnormal – clinically significant and not done.

4.2.8.5 Physical examination

Physical examination data will not be summarized.

4.2.9 Exploratory analyses

4.2.9.1 Analysis of biomarker data

Summaries and analyses for exploratory biomarkers will be documented in a separate analysis plan and will be reported outside the clinical study report in a separate report.

4.2.9.2 Analysis of asthma control assessments

Exploratory asthma control variables as recorded in daily diaries and defined in Section 3.4 will be reported descriptively.

Summary statistics will be produced by treatment group and visit for absolute value and change from baseline in total asthma symptom score, daytime score, night-time score, total rescue medication use (average number of times/day), weekly proportion of nights with nocturnal awakenings requiring and independent of rescue medication use.

The proportion of asthma symptom-free days up to EOT visit will be summarized descriptively by treatment group. The number of asthma symptom-free days and asthma symptom assessment days will also be presented.

The number and proportion of patients:

- Well-controlled;
- Partly controlled; and
- Uncontrolled

will be summarized by treatment group and week, based on the previous 7 days of data, from week 0 (baseline) until week 24 (EOT), for the full analysis set. The number of patients in each treatment group with missing data at each timepoint will also be reported. In addition, the proportion of patients well-controlled and well-/partly controlled (based on the previous 7 days of data), with 95% confidence intervals about the proportions (calculated using the exact Clopper-Pearson formula) will be presented, by treatment group and week for the full analysis set.

The proportion of patients well-controlled (based on the previous 7 days of data) over time (week 0 - week 24) will be analyzed by treatment group using a local regression model (Loess method) and Loess curves (with 95% confidence intervals) by treatment group will be produced. This analysis will be repeated for the proportion of patients well-/partly controlled over time.

In addition, the number and proportion of patients:

- Well-controlled;
- Partly controlled; and
- Uncontrolled

at week 0 (baseline), weeks 4, 8, 12, 16, 20 and week 24 (EOT), based on the 4 weeks of data prior to each timepoint, will be summarized by treatment group, for the full analysis set. Frequency plots by treatment group over time (at weeks 0, 4, 12 and 24 only) will be produced.

Time to first well-controlled week (from week 0) based on the previous 7 days' data will be analyzed. A Cox proportional hazard model will be fitted to the data with the covariates of treatment, region (North America/rest of world), number of exacerbations in the previous year and maintenance OCS use at baseline (yes/no). The model will include Efron method for handling ties (TIES=EFRON), and the confidence intervals for hazard ratios will be calculated using profile-likelihood confidence limits (RL=PL). Results will be summarized as hazard ratios, 95% confidence intervals and p-values for patients in the full analysis set. Time to first well-controlled week will be displayed graphically using a Kaplan-Meier plot.

EOT asthma control status will be compared across treatment groups using two separate logistic regression models. The dependent variable in each model will be the EOT control indicator variables (based on well-controlled or well-/partly controlled status (see section 3.4.4), with covariates of treatment, region (North America/rest of world), number of exacerbations in the previous year and maintenance OCS use at baseline (yes/no)baseline asthma control status (based on the last 4 weeks prior to randomization).

Results will be presented as odds ratios with associated 95% CIs and 2-sided p-values.

5. INTERIM ANALYSES

No interim analyses are planned.

6. CHANGES OF ANALYSIS FROM PROTOCOL

The clarifications and changes that have been made are listed below:

- Section 1.1.4: the outcome measure for asthma control is updated to be the proportion of patients with well-controlled asthma (instead of the proportion of time that the patient's asthma is well controlled), based on composite diary measures over time.

- Section 1.3: while not a change to the analysis compared to the protocol, it was noted that sample size calculations for the primary endpoint in the protocol used a 12-month follow-up period. This has been corrected to use a 24-week follow-up and the power amended (from 97% to 91%) in Section 1.3.

- Section 3.1.1: For PROs with reflection times of 1 week or more (e.g. the SGRQ, ACQ-6), the time component of the definition of baseline (i.e. time of assessment relative to the first dose) will be waived provided the assessment occurs on the same day as the first dose of study treatment.

- The protocol describes that change for each of the patient's top 3 ranked symptoms and impairments using the 7-point change scale will be collected weekly from week 1 to week 16 for PSIA to define an improvement status for that ranked symptom/impairment. This data collection of patient-perceived change from baseline (Visit 4) is important as it is designed to show an early response to benralizumab if said early response exists. However, these data were only collected at Week 20 due to an error in the set-up of the electronic patient reported outcomes system which was not caught until the majority of patients had already enrolled and progressed through the study. Therefore, no local regression modeling of proportions showing an improvement (Loess method) and Loess curves (with 95% confidence intervals) for each of the top 3 symptoms/impairments and top 3 combined; or time to 1st improvement for each of the top 3 symptoms/impairments will be performed. A note to file has been implemented.

- The main analysis of PSIA data will be the change from baseline in severity scores for each of the top 3 ranked symptoms / impairments and for the average of the top 3 ranked symptoms / impairments.

- In addition, to investigate and quantify any early response to benralizumab that may exist (in the absence of the PSIA change data), SGRQ total score responder analyses at Weeks 4 and 12 have been included.

- The protocol includes responder analyses for the SGRQ subscales, but these will not be carried out (as there is no MCID defined for the subscales).

- SGRQ responder analyses will be performed at Weeks 4 and 12 in addition to EOT.

7. REFERENCES

Bel et al 2014
Bel EH, Wenzel SE, Thompson PJ, Prazma CM, Keene ON, Yancey SW, Ortega HG, Pavord ID. Oral glucocorticoid-sparing effect of mepolizumab in eosinophilic asthma. New England journal of medicine. 2014 Sep 25;371(13):1189-97.

Bleeker et al 2016
Bleecker ER, Fitzgerald JM, Chanez P, Papi A, Weinstein SF, Barker P, et al. Efficacy and safety of benralizumab for patients with severe asthma uncontrolled with high-dosage inhaled corticosteroids and long-acting β 2-agonists (SIROCCO): a randomized, multicentre, placebo-controlled phase 3 trial. The Lancet 2016;388(10056):2115-2127.

Fitzgerald et al 2016
Fitzgerald JM, Bleecker ER, Nair P, Korn S, Ohta K., Lommatzsch M, et al. Benralizumab, an anti-interleukin-5 receptor α monoclonal antibody, as add-on treatment for patients with severe, uncontrolled, eosinophilic asthma (CALIMA): a randomized, double-blind, placebo-controlled phase 3 trial. The Lancet 2016;388(10056):2128-2141.

GINA 2018
Global Initiative for Asthma. Global Strategy for Asthma Management and Prevention, 2018. Available from: www.ginasthma.org

Hopkins et al 2009
Hopkins C, Gillett S, Slack R, Lund VJ, Browne JP. Psychometric validity of the 22-item Sinonasal Outcome Test. Clinical otolaryngology. 2009;34(5):447-45.

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

Juniper et al 2006
Juniper EF, Bousquet J, Abetz L, Bateman ED; GOAL Committee. Identifying 'well-controlled' and 'not well-controlled' asthma using the Asthma Control Questionnaire. Respir Med. 2006 Apr;100(4):616-21.

Ortega et al 2014
Ortega HG, Liu MC, Pavord ID, Brusselle GG, FitzGerald JM, Chetta A, Humbert M, Katz LE, Keene ON, Yancey SW, Chanez P. Mepolizumab treatment in patients with severe eosinophilic asthma. New England Journal of Medicine. 2014 Sep 25;371(13):1198-207.

Piccirillo et al 2002
Piccirillo JF, Merritt MG, Richards ML. Psychometric and clinimetric validity of the 20-item Sino-Nasal Outcome Test (SNOT-20). Otolaryngology--Head and Neck Surgery. 2002; 126(1);41-47.

QualityMetric 2011
QualityMetric, I. User's manual for the SF-36v2 Health Survey (3 ed.). 2011. Lincoln, RI.

8. APPENDIX

Appendix I 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 encouraged to remain in the study after premature discontinuation of IP and complete visits according to the protocol.

This section summarizes 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 randomized treatment as well as for withdrawal from the study will be presented by treatment to describe why patients discontinue from randomized treatment or withdraw from the study. The time to discontinuation of randomized 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 randomized 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 randomized 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 include all data until patients withdraw from the study regardless of if they discontinue from randomized 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 analyzed by modifying the post-withdrawal model assumption.

The method involves first fitting the primary analysis i.e. 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.

$$\Pr\left(Y_{ij,2} = y_2 \middle| Y_{ij,1} = y_1\right) = \frac{\Gamma(\gamma + y_1 + y_2)}{\Gamma(y_2+1)\Gamma(\gamma + y_1)} p_j^{y_2} (1-p_j)^{\gamma + y_1} \qquad (1)$$

Here y_1 is number of counts before withdrawal from the study, y_2 is number of counts after withdrawal from the study, γ 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

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

where $p_{j,1}$ is the negative binomial distribution (NBD) rate parameter before withdrawal from the study, and $p_{j,2}$ 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 analyzed 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 and (3) development of study specified reasons (4) severe non-compliance of protocol to stop active treatments.

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 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 randomized treatment for those patients who discontinued randomized treatment and then withdrew from the study and also free text for when the reason for withdrawal or discontinuation of randomized 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 and documented in the BDR comments tracker.

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

Reason for withdrawal	**MAR**	**DRMI**
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, an asthma-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 i.e. 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 (Efficacy and Effectiveness estimands)

In addition, primary and sensitivity analyses described previously, two alternative estimands will be estimated using only the on initial randomized treatment data:

Efficacy estimand - what would have been the outcome if all patients had stayed on study treatment: This will be estimated using the primary analysis method but including only data from patients whilst being on initial randomized treatment, and assuming MAR subsequently.

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 randomized treatment i.e. excluding data once patients discontinue from randomized 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 randomized 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

Reason for discontinuation of randomized treatment	**MAR**	**DRMI**
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 unblindng
Other	Benralizumab	Based on review prior to study unblindng

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, an asthma-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 randomized treatment. The efficacy estimand together with the reason for and timing of why a patient might not tolerate the treatment allows for the simplest interpretation as it describes the treatment effect for patients who adhere to treatment together with why and when they might not adhere to treatment. Sensitivity analyses using the effectiveness estimands 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 (Efficacy and Effectiveness estimands)**		
	DL	**MAR**	**DRMI**	**DL**	**MAR**	**DRMI**
Population	**On-treatment + post-discontinuation of randomized treatment**			**On-treatment**		
Estimand	**Treatment policy (MAR)**		**Treatment policy (MNAR)**	**Efficacy (MAR)**		**Effectiveness (MNAR)**
Exacerbation rate for imputation in Benra arm**	No explicit imputation*	Benra rate assumed for all reasons for withdrawal	Placebo rate assumed for AEs, Death, development of study specified reasons to stop active treatments and Severe non-compliance to protocol, otherwise Benra rate assumed. Reasons are subject to review prior to study unblinding	No explicit imputation*	Benra rate assumed for all reasons for discontinued treatment	Placebo rate assumed for AEs, development of study specified reasons to stop active treatments and Severe non-compliance to protocol, otherwise Benra rate assumed. Reasons are subject to review prior to study unblinding
Default definition for $p_{j,1}$ and $p_{j,2}$ based on formula (2) ****.		$p_{j,2} = p_{j,1}$ For all treatment arms j=B and P	$p_{B,2} = p_{B,1}$ $p_{B,2} = p_{P,1}$ $p_{P,2} = p_{P,1}$ for reasons above otherwise $p_{B,2} = p_{B,1}$		$p_{j,2} = p_{j,1}$ For all treatment arms j=B and P	$p_{B,2} = p_{B,1}$ $p_{B,2} = p_{B,1} p_{B,2} = p_{P,1}$ $p_{P,2} = p_{P,1}$ for reasons above otherwise $p_{B,2} = p_{B,1}$

* Implicitly assumes unobserved rate the same as observed
** All patients on exacerbation rate in the placebo arm are imputed using the placebo arm rate (i.e. $p_{P,=,2} = p_{P,1}$)
*** P denotes Placebo, B denotes Benralizumab
**** Note can be over written by review prior to study unblinding`

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 (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 (e.g. 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 SGRQ total score key secondary endpoint 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 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 one 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 arm 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 patients, 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 patients 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 two steps, the non-monotone (intermediate) missing SGRQ total score 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 arm, 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, one 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 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. The analysis of each of the imputed dataset will be as described for the primary analysis in section xxx and these will be combined using SAS procedure PROC MIANALYZE.

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. 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.

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

11. Barnard J, Rubin DB. Small-sample degrees of freedom with multiple imputation. Biometrika 1999; 86:948-955.

12. Little RJ, Yosef M, Cain KC, Nan B, Harlow SD. A hot-deck multiple imputation procedure for gaps in longitudinal data on recurrent events. Statist. Med., 2008; 27:103-120.

13. 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.

Appendix II Partial dates for prior/concomitant medication and asthma history

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

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, and the year of occurrence is in the same year as the year of the end of study date, then set the end date to be equal to end of study date otherwise the date will be set to December 31 of the year of occurrence.

If the start date of a therapy has been checked in the eCRF as having started > 3 months prior to enrolment, the start date will be set to the earlier of 92 days prior to the date of enrolment and the therapy end date (if completed).

If the start date of a therapy is null and has not been checked in the eCRF as having started >3 months prior to enrolment, 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 has not been checked in the eCRF as having started >3 months prior to enrolment, 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 date of the last study visit.

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 date of the last study visit then the end date will be set to the date of the last study visit.

- otherwise, the end date will be set to the start date of the therapy.

The following conventions will be applied for partial dates of asthma diagnosis:

If the day of date of asthma diagnosis is missing it will be set to the first day of the month that the diagnosis occurred.

If the date of asthma diagnosis is missing both the day and month, and the year of diagnosis is in the same year as the year of birth, then set the date to be equal to date of birth (if both the day and month of birth are missing then set to January 1 of the year of birth; if the day of birth is missing then set to the first day of the month of birth); otherwise the date will be set to January 1 of the year of diagnosis.

SIGNATURE PAGE

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature

Document Name: d3250c00045-sap-ed-3		
Document Title:	Statistical Analysis Plan Edition 3	
Document ID:	CCI [redacted]	
Version Label:	3.0 CURRENT LATEST APPROVED	
Server Date (dd-MMM-yyyy HH:mm 'UTC'Z)	**Signed by**	**Meaning of Signature**
15-Oct-2019 14:32 UTC	PPD	Content Approval
15-Oct-2019 14:31 UTC	PPD	Author Approval
15-Oct-2019 14:36 UTC	PPD	Content Approval

Notes: CCI [redacted]