Statistical Analysis Plan	
Study Code	D3252C00002
Edition Number	Version 3
Date	20 Jun 2024

A Multicentre, Randomised, Double-Blind, Parallel-Group, Placebo-Controlled Phase 3 Efficacy and Safety Study of Benralizumab in Patients with Eosinophilic Chronic Rhinosinusitis with Nasal Polyps (ORCHID)

TABLE OF CONTENTS PAGE

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
ACQ-6	Asthma control questionnaire 6
ADA	Anti-drug antibody
AE	Adverse event
AER	Asthma exacerbation rate
AERD	Aspirin exacerbated respiratory disease
ALT	Alanine aminotransferase
ANCOVA	Analysis of covariance
AST	Aspartate aminotransferase
ATC	Anatomical therapeutic chemical
CCI [redacted]	
BLQ	Below limit of quantification
CI	Confidence interval
CMH	Cochran–Mantel–Haenszel
COVID-19	Corona Virus Disease 2019
CRF	Case report form
CRSwNP	Chronic rhinosinusitis with nasal polyps
CSP	Clinical study protocol
CSR	Clinical study report
CT	Computed tomography
DAE	Discontinuation of investigational product due to adverse event
DB	Double-blind
DL	Direct likelihood
DNA	Deoxyribonucleic acid
DRMI	Dropout reason-based multiple imputation
DSS	Difficulty with sense of smell
ECG	Electrocardiogram
eCRF	Electronic case report form
CCI [redacted]	
EoDB	End of treatment double-blind period
EOS	Eosinophil
ePRO	Electronic patient reported outcome
FAS	Full analysis set
FU	Follow-up
GGT	Gamma-glutamyl transferase
HRQoL	Health-related quality of life
HRU	Healthcare resource utilisation
IM	Immunogenicity
INCS	Intranasal corticosteroids
IP	Investigational product
IPD	Investigational product discontinuation

ITT	Intent-to-treat
LLOQ	Lower limit of quantification
LMS	Lund-Mackay score
LSMEAN	Least squares mean
MAR	Missing at random
Max	Maximum
MedDRA	Medical dictionary for regulatory activities
MNAR	Missing not at random
Min	Minimum
nAb	Neutralizing antibodies
NBS	Nasal blockage score
NERD	Nonsteroidal anti-inflammatory drug exacerbated respiratory disease
NHP	Non-compliance handing plan
NP	Nasal polyps
NPS	Nasal polyp score
NPSD	Nasal polyps symptom diary
OLE	Open-label extension
PFAS	Primary full analysis set
PGI-C	Patient general impression of change
PGI-S	Patient general impression of severity
PK	Pharmacokinetic(s)
PRO	Patient reported outcome
PT	Preferred term
Q1	Lower quartile
Q3	Upper quartile
SABA	Short acting beta-agonist
SAE	Serious adverse event
SAP	Statistical analysis plan
SC	Subcutaneous
SCS	Systemic corticosteroids
SD	Standard deviation
SI	Standard international
SNOT-22	SinoNasal outcome test
SOC	System organ class
TSS	Total symptom score
ULN	Upper limit of normal
UPSIT	University of Pennsylvania smell identification test
V	Visit
WHO	World health organisation
WOCF	Worst observation carried forward
WP	Worst possible

AMENDMENT HISTORY

Category*: Change refers to	Date	Description of change	In line with the CSP?	Rationale
Other	11 JUN 2024	Language update in Section 4.1 and Section 4.1.4.1 to align with estimand framework and terminology outlined in ICH E9(R1)	NA	To align with ICH E9(R1) guideline
Other	25 APR 2024	Updated study population from PFAS to 'severe CRSwNP patients with asthma' in Table 11 and Section 8.10.	Yes (CSP version 5.0, 06MAY2024)	To align with the CSP amendment v5.0
Data presentation	25 APR 2024	Removed selected descriptive summaries in section 4 and section 8.7.	N/A	Not mandatory for BLA submission so removed to simplify analysis and reporting and to accelerate submission timeline.
Data presentation	25 APR 2024	Clarified duration of exposure in Section 4.2.4.1 to be calculated as duration of IP administration which ends on the date of last dose in the analysis period.	N/A	To add clarity on the reporting of the data
Other	22 APR 2024	Study periods definition updated in Section 3.1.1 and Section 3.7.1.	N/A	To use consistent language in different sections in SAP.
Other	22 APR 2024	Clarified the responder definition for NBS in Section 3.3.2, for SF-36 in Section 3.5.7, and for ACQ-6 in Section 3.7.1.	N/A	To add clarity on the reporting of the data
Other	22 APR 2024	Clarified how to assign AE to DB period vs OLE period based on AE onset date in Section 3.8.1	N/A	To add clarity on the reporting of the data

Data presentation	22 APR 2024	Removed related SAE and SAE by PT and intensity summaries. Added helminth infections summaries in Section 4.2.10.1.	N/A	To align with the latest safety reporting guidance and standards in Fasenra program.
Data presentation	19MAR2024	Added derivations for ADA response categories in Section 8.7	N/A	To add clarity on the reporting of the data
Data presentation	18MAR2024	Created a new section 3.2 listing demographics and baseline variables to analysed and clarified the analysis set for each demographics/baseline summary in section 4.2.2.	N/A	To add clarity on the reporting of the data
Other	13MAR2024	Removed Week 40 from subgroup analysis and responder analysis	N/A	Not mandatory for BLA submission so removed to simplify analysis and reporting and to accelerate submission timeline.
Data presentation	13MAR2024	Added description of DB and OLE data collection for primary, secondary, exploratory, safety variables, and PK and ADA assessments in respect to primary and final DBL throughout.	Yes (CSP version 5.0, 06MAY2024)	To add clarity on the reporting of the data and align with the CSP amendment v5.0
Data presentation	11MAR2024	Clarified visit window definitions in Section 3.1.2 to address differences in data collection across CSP versions 1.0 to 5.0.	Yes (CSP version 5.0, 06MAY2024)	To add clarity on the reporting of the data.
Other	11MAR2024	Added Section 2.1.6 to specify the exclusion of Japanese patients enrolled in study site PPD and PPD due to GCP breach.	N/A	To address significant deviations from Good Clinical Practice at investigator sites
Data presentation	23FEB2024	Added definitions of exposure adjusted incidence rate (EAIR) and crude rate per 100 patient-years for adverse events. Added risk difference and confidence interval estimation using Miettinen and Nurminen method	N/A	To align with the latest safety reporting guidance

Other	14JAN2024	Removed selected supportive analyses, shift plots in Section 4.2.6 to 4.2.7	N/A	Not mandatory for BLA submission so removed to simplify analysis and reporting and to accelerate submission timeline.
Data presentation	14JAN2024	Added reporting rules for PK concentration values <LLOQ in Section 4.2.8.1	N/A	To add clarity on the reporting of the data
Other	21SEP2021	Some small text updates throughout the document, correcting typos or adding clarity in the analysis	N/A	To add clarity on the reporting of the data
Data presentation	21SEP2021	Updated ADA analysis in Appendix 8.7	N/A	To add clarity on the analysis and reporting of the ADA data in OLE period and on-study
Statistical analysis method for the primary or secondary endpoints	21SEP2021	Added additional supplementary analysis for LMS in Section 4.2.6.2	N/A	To add clarity on the reporting of the data
Other	21SEP2021	Added clarity that analysis for China/Japan submission will not be reported in CSR in Section 4.1.1	N/A	To add clarity on the reporting of the data
Statistical analysis method for the primary or secondary endpoints	21SEP2021	Added additional responder analysis for NBS in Section 4.2.5.4 and additional supplemental analysis for LMS in Section 4.2.6.2	N/A	To be consist with the global NP program
Data presentation	21SEP2021	Removed the analysis of Hy's law from Section 4.2.10.2	Yes (CSP version 4.0, 16MAR2021)	Data no longer collected
Other	12MAY2021	Updated Section 1 by adding OLE treatment period and other updates for the changes in the study design and number of patients	Yes (CSP version 4.0, 16MAR2021)	To be aligned with the CSP amendment v4.0
Other	12MAY2021	Added OLE analysis set in Section 2.1	Yes (CSP version 4.0, 16MAR2021)	To be aligned with the CSP amendment v4.0

Data representation	12MAY2021	Updated the corresponding subsections of 3.1 to address the OLE period	Yes (CSP version 4.0, 16MAR2021)	To be aligned with the CSP amendment v4.0
Primary or Secondary Endpoints	12MAY2021	Added key secondary endpoints: DSS and SNOT-22)	Yes (CSP version 4.0, 16MAR2021)	To be aligned with the CSP amendment v4.0
Statistical analysis method for the primary or secondary endpoints	12MAY2021	Analysis of continuous endpoints changed from MMRM to ANCOVA with MI. Imputation approach for intercurrent events changed from censoring (for treatment discontinuation, surgery for CRSwNP, or SCS use for CRSwNP) to composite Worst Possible after rescue by surgery for CRSwNP. WOCF after SCS use for CRSwNP is retained Changed the covariates in the ANCOVA model as well to region CCI Added Week 40 to responder analyses and subgroup analyses	Yes (CSP version 4.0, 16MAR2021)	To be aligned with the CSP amendment v4.0 Week 40 is added to be consistent with the global NP program.
Data presentation	12MAY2021	Changed baseline characteristics/efficacy subgroups: removed prior OCS CCI CCI	Yes (CSP version 4.0, 16MAR2021)	To be aligned with the CSP amendment v4.0
Data presentation	12MAY2021	Added summary of SNOT- 22 individual items. Added CDF curve of SNOT-22 total score.	Yes (CSP version 4.0, 16MAR2021)	To be aligned with the CSP amendment v4.0
Data presentation	12MAY2021	Added negative binomial, regression of number of courses of SCS use for CRSwNP.	Yes (CSP version 4.0, 16MAR2021)	To be aligned with the CSP amendment v4.0
Primary or Secondary Endpoints,	12MAY2021	New section added “Impact on analyses due to COVID-19”	Yes (CSP version 4.0, 16MAR2021)	Response to COVID-19 pandemic due to inability of sites to continue dosing and/or primary endpoint assessment to 56 weeks.
Primary or secondary endpoints; Statistical analysis method for the	12MAY2021	Additional secondary endpoints added: Time to Surgery and/or SCS use for CRSwNP, Difficulty with sense of smell (DSS)	Yes (CSP version 4.0, 16MAR2021)	To be aligned with the CSP amendment v4.0

primary or secondary endpoints				
Multiple testing procedure	12MAY2021	Testing order of secondary endpoints: DSS, LMS, SNOT-22	Yes (CSP version 4.0, 16MAR2021)	To be aligned with the CSP amendment v4.0
Other (Sensitivity analyses)	12MAY2021	Sensitivity/supplementary analyses updated for new primary analysis method.	N/A	Most appropriate sensitivity analyses for primary estimand and analysis.
Other (Safety analyses)	12MAY2021	Updated “extended follow up” to OLE period for safety endpoints, when applicable	Yes (CSP version 4.0, 16MAR2021)	To be aligned with the CSP amendment v4.0

*Pre-specified categories are:
Primary or secondary endpoints; Statistical analysis method for the primary or secondary endpoints; Derivation of primary or secondary endpoints; Multiple Testing Procedure; Data presentations; Other
N/A = Not applicable

1. STUDY DETAILS

1.1 Introduction

This is the statistical analysis plan (SAP) for study D3252C00002 and outlines the analyses to be generated for the global clinical study report (CSR). The SAP describes the statistical analyses specified in the latest version of the clinical study protocol (CSP) in more detail. Any changes to what is specified in the CSP will be described. Additional analyses required for regional submissions may be prespecified in separate documents and will be submitted to the appropriate authorities.

1.2 Study objectives

Table 1 Objectives and endpoints during the double-blind (DB) treatment period

Primary objective:	**Estimand Description/Endpoints**
To evaluate the effect of benralizumab on nasal polyp burden and patient-reported nasal blockage (NB).	• Population[a]: severe CRSwNP patients with asthma • Co-primary endpoints: Change from baseline in endoscopic total nasal polyp score (NPS) Change from baseline mean nasal blockage score (NBS) • Intercurrent event strategy: Treatment discontinuation – treatment policy (included in analysis regardless of treatment discontinuation) Chronic rhinosinusitis with nasal polyps (CRSwNP) surgery – composite (Worst Possible Carried Forward (WPCF)) Systemic corticosteroids (SCS) for CRSwNP – composite (Worst Observation Carried Forward (WOCF)) • Summary Measure: differences in least squares mean change from baseline in NPS and NBS between benralizumab and placebo. Week 56 is the primary timepoint.
Secondary Objectives:	**Endpoint/Variable:**
To evaluate the effect of benralizumab on:	

Sense of smell	• Change from baseline in mean difficulty with sense of smell (DSS) score[b] • Change from baseline in University of Pennsylvania Smell Identification Test (UPSIT) score
Sinus opacification by Computed Tomography (CT) scan	• Change from baseline in Lund-Mackay Score[b] (LMS) • Change from baseline in sinus severity score by Quantitative CT analysis • Change from baseline in Zinreich score (modified LMS)
Disease specific health-related quality of life (HRQoL)	• Change from baseline in SinoNasal Outcome Test (SNOT-22) score[b]
Nasal polyp surgery and/or systemic corticosteroids (SCS) use for relief of nasal symptoms	• Time to first surgery and/or SCS use for CRSwNP • Proportion of patients with surgery and/or SCS use for CRSwNP
Nasal polyp surgery	• Time to first surgery for CRSwNP • Proportion of patients with surgery for CRSwNP
Systemic corticosteroids (SCS) use for relief of nasal symptoms	• Proportion of patients with SCS use for CRSwNP • Time to first SCS course for CRSwNP • Number of courses of SCS for CRSwNP • Total SCS dose used and total duration of SCS use for CRSwNP
Symptoms associated with nasal polyps	• Change from baseline in nasal symptom score(s) as captured in the NP Symptom Diary (NPSD)
Patient-reported general health status	• Change from baseline in Short Form 36-item Health survey, Version 2 (SF-36v2),

	Physical Component Score (PCS), Mental Component Score (MCS) and domains
To characterize the Pharmacokinetics (PK) and immunogenicity of benralizumab	• PK: Serum trough concentrations • Immunogenicity: Anti-drug antibodies
Safety objective:	**Endpoint/variable**
To assess the safety and tolerability of benralizumab as compared to placebo in patients with Chronic Rhinosinusitis with Nasal Polyps	Safety and tolerability will be evaluated in terms of adverse events (AE)s, Vital signs, Clinical laboratory, and electrocardiogram (ECG) • Assessments related to AEs cover - Occurrence/Frequency - Relationship to investigational product (IP) as assessed by investigator - Intensity - Seriousness - Death - AEs leading to discontinuation of IP • Vital signs parameters include systolic and diastolic blood pressure, and pulse. Assessments cover - Observed value - Absolute change from baseline values over time • Laboratory variables • ECG
Exploratory objectives	**Endpoint/variable**
To assess the effect of benralizumab on asthma control	• Asthma Control Questionnaire (ACQ-6) • Annual asthma exacerbation rate (AAER)
To assess the effect of benralizumab on patient recognition of improvement	• Patient Global Impression of Severity (PGI-S) • Patient Global Impression of Change (PGI-C)
To assess the effect of benralizumab on exploratory biomarkers of inflammation	Exploratory biomarker parameters:

and nasal polyps disease and investigate biomarkers for predicting response to benralizumab	• Serum and plasma for protein biomarkers • Whole blood for transcriptomic profiling • Nasal secretions for protein biomarkers
To assess the effect of benralizumab on CCI as a biomarker of inflammation and nasal polyposis disease CCI	• CCI
To assess the effect of genetic variation on patient's response to therapy, susceptibility to, and severity and progression of disease	• A blood sample for deoxyribonucleic acid (DNA) isolation will be collected from patients who have consented to participate in the exploratory genetic analysis component of the study
To evaluate the effect of benralizumab on unplanned health care resource utilization	• Hospitalizations, emergency room and urgent care visits
[a] Treatment condition for primary estimand: Treatment with benralizumab versus placebo, regardless of compliance, where rescue with surgery and/or SCS for CRSwNP represents failure. [b] Key secondary efficacy endpoints. A similar estimand as outlined for the co-primary endpoints will be used for analyses of repeated measures secondary endpoints. Co-primary and key secondary endpoints are multiplicity protected.	

Table 2 Objectives and endpoints during the open-label extension (OLE)

Primary objective:	**Endpoint/Variable:**
To assess the safety and tolerability of benralizumab in patients with Chronic Rhinosinusitis with Nasal Polyps	Safety and tolerability will be evaluated in terms of adverse events (AE)s, Vital signs, and Clinical laboratory. • Assessments related to AEs cover - Occurrence/Frequency - Relationship to investigational product (IP) as assessed by investigator - Intensity - Seriousness - Death - AEs leading to discontinuation of IP

	• Vital signs parameters include systolic and diastolic blood pressure, and pulse. Assessments cover - Observed value - Absolute change from baseline values over time • Laboratory variables
Secondary objectives:	**Endpoint/Variable**
To characterize the PK and immunogenicity of benralizumab	• PK: Serum trough concentrations • Immunogenicity: Anti-drug antibodies
Exploratory objectives:	**Endpoint/Variable**
Nasal polyp surgery	• Proportion of patients with surgery for CRSwNP
SCS use for relief of nasal symptoms	• Proportion of patients with SCS use for CRSwNP • Number of courses of SCS for CRSwNP • Total SCS dose used and total duration of SCS use for CRSwNP
Effect of benralizumab on asthma control	• Asthma exacerbation

1.3 Study design

This is a randomised, double-blind, placebo-controlled, parallel-group, multicentre, Phase 3 study to evaluate the efficacy and safety of repeat dosing of benralizumab 30 mg administered subcutaneously (SC) versus placebo in patients with severe CRSwNP. Following completion of the double-blind period, patients have the option to enter an OLE.

Approximately 250 patients will be randomised to receive benralizumab 30 mg SC or matching placebo.

Patients will be stratified by region CCI CCI

After enrolment, and prior to screening/run in period (at V1), eligible patients will have their current INCS therapy switched to Mometasone Furoate (400 ug daily) or equivalent (highest local approved for CRSwNP), which should be maintained throughout the study until the last DB visit (V11). Patients must have been on a stable daily dose of INCS for at least 4 weeks prior to V1. Patients will be provided with an electronic patient-reported outcome (ePRO)

device to record symptoms and health-related quality of life (HRQL) data throughout the study until V11 (Week 56).

Patients who continue to meet eligibility criteria will be randomised 1:1 at Visit (V) 3 (Week 0) to receive either placebo or benralizumab 30 mg SC every 4 weeks for the first 3 doses (Weeks 0, 4 and 8) and every 8 weeks (Q8W) thereafter (Weeks 16, 24, 32, 40 and 48). All patients who complete the 56-week DB treatment period on investigational product (IP) may be eligible to continue into an OLE, during which all patients will receive 8 doses of investigational product, every Q4W for the first 3 doses (Weeks 56, 60, 64) and Q8W thereafter (Weeks 72, 80, 88, 96 and 104). Patients randomised to the benralizumab arm during DB period will receive one dose of placebo (dummy) injection at Visit 12 (week 60) in order to not get a loading dose again and to maintain the blind.

During the DB period, patients will return to the study site at dosing visits and at V11 (Week 56) for evaluation of efficacy, safety, PK, ADA, sample collection of blood and nasal secretion for exploratory biomarkers. Patients who enter the OLE period will receive their first OLE dose at V11, after completion of all visit-related assessments. Throughout the OLE period, patients will mainly have assessments related to safety (see Table 2). The last study visit will occur at 8 weeks after the last dose of IP (Week 112/Follow-up (FU)). Patients who do not enter OLE will have their last study visit at Week 56, end of DB period (EoDB), for follow-up and without administration of IP. If at any point a patient meets IP discontinuation (IPD) criteria, an early IPD visits will be performed. Details about IPD criteria are given in clinical study protocol (CSP), Section 7.1.

Throughout the entire study period, participants will remain on their currently prescribed treatment for CRSwNP.

CT scanning will be performed at V2 (Week -2) and EoDB V11 (Week 56), or IPD if at least 16 weeks since the previous CT scan. In case of surgery during the double-blind part of the study, the CT scan should be done prior to surgery instead of at Week 56 if at least 16 weeks since the previous CT scan. CT scan at V2 will be used both for inclusion criterion (for Asian countries only) and as baseline for efficacy (see CSP, Section 8.1.3).

The primary database lock will occur after all randomised patients have been followed up for the 56-week DB treatment period. The study will remain blinded until the primary database lock. The final database lock will occur after the last patient completes the OLE. All available data captured during the DB treatment period and selected safety data from the OLE period will be analysed at the primary DBL, while remaining data from the OLE period will be reported separately at the final DBL.

1.4 Number of patients

The primary analysis will compare the effect of benralizumab vs placebo on the change from baseline in total NPS at Week 56 and on the change from baseline in bi-weekly mean NBS at Week 56 using a hybrid of worst-possible/worst-observation carried forward and multiple imputation (MI), followed by an analysis of covariance (ANCOVA) with treatment arm,

baseline scores of the corresponding endpoint, region, CCI as covariates (See Section 4.2.5.1 for details).

Approximately 250 patients will be randomised into SC benralizumab 30 mg or placebo in a 1:1 ratio.

The sample size was estimated based on a 2-sided t-test at a significance level of 0.05. Assuming the population standard deviation (SD) of change from baseline is 2 for total NPS and 1 for NBS, this sample size will provide an overall 80% power to detect a true mean (population) treatment difference of 0.85 units in total NPS and a difference of 0.4 in NBS with a two-sided 0.05 alpha level.

Based on the assumptions above, the minimum observed mean difference of the changes from baseline between benralizumab and placebo that would be statistically significant at the 0.05 level is –0.50 in total NPS and –0.25 in NBS.

2. ANALYSIS SETS

2.1 Definition of analysis sets

Patients must have provided their informed consent. If no signed informed consent is collected (important protocol deviation), then the patient will be excluded from all analysis sets defined below.

2.1.1 All patients analysis set

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

2.1.2 Full analysis set

All patients randomised and receiving any IP will be included in the FAS, irrespective of their protocol adherence and continued participation in the study. Patients will be analysed according to their randomised treatment. Patients who withdraw consent, and assent when applicable, to participate in the study will be included up to the date of their study termination.

The primary full analysis set (PFAS) consists of all CRSwNP patients with asthma in the FAS.

The summary of demographics will be provided for the PFAS and FAS. Unless otherwise stated, all efficacy analyses will be performed using an Intent-to-treat (ITT) approach based on the PFAS. Efficacy data collected from patients without comorbid asthma at baseline will be listed only.

2.1.3 Safety analysis set

The safety analysis set consists of all patients who have received at least one dose of IP. Patients will be classified according to the treatment they received. A patient who received at least 1 dose of benralizumab will be classified as a patient in the benralizumab treatment group.

Unless otherwise stated, safety and immunogenicity analyses will be conducted based on the safety analysis set, regardless of comorbid asthma status at baseline.

2.1.4 Pharmacokinetic analysis set

All patients who received benralizumab and from whom PK blood samples are assumed not to be affected by factors such as protocol deviations and who had at least one measurable serum PK observation post first dose will be included in the PK analysis dataset.

Pharmacokinetic analyses will be conducted based on the PK analysis set.

2.1.5 OLE analysis set

The OLE analysis set will include all patients who enter the OLE part of the study and who received at least 1 dose of the IP during the OLE treatment period.

Analyses of the OLE period will be conducted based on the OLE analysis set.

2.1.6 Handling of Japan GCP Issue

The Japanese Ministry of Health, Labour and Welfare (MHLW) informed AstraZeneca of an unannounced inspection of a Site Management Organization (SMO). The MHLW identified significant deviations from Good Clinical Practice at the investigator sites supported by the SMO across multiple clinical trials by numerous organisations, including AstraZeneca. To ensure the integrity of clinical trial data, any Japanese patients enrolled from sites managed by the SMO (i.e. study sites PPD and PPD will be excluded from all analyses.

- In the disposition table, number of Japanese patients impacted by this issue will be presented in a separate category as “Excluded due to Japan GCP breach” under “Subjects randomised”, and will not be counted in any other category (i.e. completed, withdrawals, etc.).
- All AEs reported by impacted Japanese subjects will be documented in a separate listing.

2.2 Violations and deviations

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

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

- Inclusion Criteria Deviations
- Exclusion Criteria Deviations
- Discontinuation Criteria for study product met but patient not withdrawn from study treatment
- Discontinuation Criteria for overall study withdrawal met but patient not withdrawn from study
- Investigational Product (IP) Deviation
- Excluded Medications taken
- Deviations to study procedure
- Other Important Deviations

All-important PDs will be identified and documented by the study team prior to unblinding of the trial. Only important protocol deviations will be tabulated in the CSR. Potential important protocol deviations, either programmable or observable, will be reviewed periodically during the trial and at the time of blinded delivery reviews. Additional details for each of the above categories are provided in the Protocol Deviation Plan.

Protocol deviations associated with the COVID-19 pandemic will also be summarised and listed separately (See Section 4.2.11).

3. PRIMARY AND SECONDARY VARIABLES

3.1 General principles

3.1.1 Study periods for data summary

The following study periods will be derived for reporting purposes:

- Double-blind period: this is defined as the period starting on the date of randomization (for efficacy) or date of first dose of study treatment (for safety), and ending on the date of first OLE dose (for those who entered the OLE right after DB completion), or the date of the Week 56 Visit (for those who did not enter the OLE after DB, or those who entered the DB follow-up period per CSP v1.0 and v2.0). For patients who discontinued the study before the Week 56 Visit, the DB period ends on the last participation date.
    - Specifically, patients who enter the OLE period will receive their first OLE dose at the Week 56 Visit, after completion of all visit-related assessments. Therefore, all assessments at the Week 56 Visit are counted under the DB

period, and the first OLE dose (i.e. last procedure at the visit) indicates the start of the OLE period.

- Extended follow-up period, off treatment (applicable to CSP v1.0 or v2.0):
    - for patients who did not enter the OLE: this is defined as the period starting on the date after the end of DB period and ending on the date of last participation in the study.
    - for patients who were followed up off-treatment after DB period and then entered the OLE: this is defined as the period starting on the date after the end of DB period and ending on the date of first OLE dose.
- Open-label period: this is defined as the period starting on the date of first OLE dose and ending on the date of last participation in the study.
- On-study period: this is defined as the period starting on the date of randomization (for efficacy) or date of first dose of study treatment (for safety) and ending on the date of last participation in the study.

3.1.2 Visit window definitions

For endpoints that present visit-based data, the variables will be summarised based on the scheduled days with adjusted analysis-defined visit windows. The adjusted analysis-defined visit windows will be based on the collection schedule listed in the protocol and variables will be windowed to the closest scheduled visit for those variables.

Visit windows have been constructed so that every observation collected can be allocated to a particular visit. However, all values will be included in data listings. Of note, Day 1 indicates the beginning of analysis periods following the Clinical Data Interchange Standards Consortium (CDISC) standards, although it is labelled as Day 0 in the CSP.

For each analysis parameter, the windowing will be based on the protocol-specified schedule of events as defined in Table 1, Table 2, and Table 3 of the CSP. If multiple readings are recorded within a single visit window, the following rules will apply:

- If there are 2 or more valid, non-missing observations within the same visit window, then the non-missing one which is closest to the scheduled visit day will be used in the analysis.
- If 2 valid observations are equidistant from the scheduled visit, then the non-missing observation with the earlier collection date will be used in the analysis for the post-baseline observations and the non-missing observation with the later collection date will be used in the analysis for the screening observations.
- If 2 or more valid observations are collected on the same day after Day 1, then the non-missing observation with the earlier collection time will be included in the

analysis. And the non-missing observation with the later collection time will be included in the analysis if these records were collected during the screening.

- If 2 non-missing values (for continuous variables) are recorded on the same day and have no assessment time associated with at least one of them, or the same assessment time associated with both of them, the average of the two values will be selected for analysis at that visit. For categorical variables in this situation, the worst case will be used.

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

For assignment of data to adjusted analysis-defined visit windows, study day will be calculated relative to the reference start date, which is the date of randomization for efficacy variables and the date of first dose of study treatment for safety variables.

- Screening period: *Study day = Date of assessment – reference start date*
- After randomisation: *Study day = (Date of assessment – reference start date) + 1*

The adjusted analysis-defined windows for NPS are defined as in Table 3 below:

Table 3 Adjusted analysis-defined visit windows for NPS

Analysis-defined visit windows	Scheduled study day	Maximum windows
Week –6	–42	–49≤ Study day ≤–15
Week –2	–14	–14≤ Study day ≤1
Week 8	57	2≤ Study day ≤84
Week 16	113	85≤ Study day ≤140
Week 24	169	141≤ Study day ≤224
Week 40	281	225≤ Study day ≤336
Week 56	393	337≤ Study day ≤ min (434, first OLE dose); first OLE dose is applicable for those who entered the OLE after DB completion

The NBS, DSS, Total Symptom Score (TSS), and other items in the nasal symptom scores diary (NPSD) will be summarised every two weeks (bi-weekly) and the adjusted windows are defined as in Table 4 below.

Table 4 Adjusted analysis-defined visit windows for NPSD (including NBS, DSS, and TSS)

Analysis-defined visit windows	Scheduled study day	Maximum windows
Week 1, Day 1	1	–13≤ Study day ≤1
Week 2	15	2≤ Study day ≤15
Week 4	29	16≤ Study day ≤29
Week 6	43	30≤ Study day ≤43

…	…	…
Week 56	393	380≤ Study day ≤ min (393, first OLE dose); first OLE dose is applicable for those who entered the OLE after DB completion

The general adjusted analysis-defined windows for the visits following baseline for all other assessments in Table 2 and Table 3 of the CSP is summarised in Table 5. The windowing will only be performed for assessments within the appropriate periods, e.g. double-blind versus open label.

Table 5 General analysis-defined visit windows for assessments

Analysis-defined visit windows	Scheduled study day	Maximum windows
Week 1 Day 1	1	Study day=1
Week X	X*7+1=a	2≤ Study day ≤((b-a)/2+a)-1
Week Y[a]	Y*7+1=b	((b-a)/2+a) ≤ Study day ≤((c-b)/2+b)-1
Week Z	Z*7+1=c	((c-b)/2+b)≤ Study day

[a] For subjects enrolled from CSP v1.0 and v2.0 with efficacy endpoints (SNOT-22, UPSIT, SF-36) collected after Week 56, Week 56 is considered as Week Y in this table.

For visit-based safety summaries during the OLE period: There were 7 subjects who entered the OLE period after being followed up without IP dosing per CSP v1.0 or 2.0. Safety data collected from those subjects during the OLE period are grouped into each timepoint based on the relative duration to the first OLE dose. Specifically, Week 60 is the 2nd scheduled visit (~4 weeks) after the first OLE dose, and Week 64 is the 3rd scheduled visit (~8 weeks) after the first OLE dose, etc.

For local laboratory assessments (including vital signs, ECG), the nominal visit number reported by study sites in web-based data capture (WBDC) will be used in the visit-based summary. Similarly, nominal visit number recorded for scheduled PK assessments in the WBDC will be used for trough concentrations. Unscheduled PK assessments (where nominal visit number is not available) need to be 7-9 weeks (i.e. ±7 days of Visit X) after the previous IP dosing at Visit X-1 in order to be considered as trough concentration at Visit X.

For overall analyses not based on any particular study visit (e.g. shift tables or event-based analysis), no analysis window will be applied; all data will be listed and/or analysed during analysis periods specified in Section 3.1.1, including any repeat or unscheduled assessments, unless otherwise specified.

3.1.3 The definition of baseline

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

For the NPSD (which includes the NBS and DSS), the baseline is the mean of daily responses collected from Day –13 to Day 1. The mean is calculated as the sum of all non-missing daily scores if at least 8 days (>50%) over these 14 sequential days have available data, divided by the number of non-missing daily scores. Otherwise, the baseline is set to missing.

Baseline value for the OLE will be set to the last recorded value prior to starting the first dose of benralizumab in the OLE.

Summaries of demographics, safety, immunogenicity, etc. will be provided for the DB period and may be repeated for the OLE period using the OLE baseline values among all patients in the OLE analysis set including patients who were previously on benralizumab.

3.1.4 Definition of prior/concomitant medications

- For the DB period, a medication for CRSwNP/asthma will be classified as a maintenance medication at baseline if it started prior to or on the date of randomisation and is continued after randomisation.

- For the DB period, a medication will be regarded as prior if it was stopped on or before the date of randomisation (medication stop date ≤ date of randomisation).

- For the DB period, a medication will be regarded as concomitant if the start date is after the date of randomisation, or if it started on or prior to the date of randomisation and was ongoing after the date of randomisation. Medications with start date on or after the end of the analysis period (as defined in Section 3.1.1) will not be considered as concomitant.

- Maintenance/prior/concomitant during the OLE period will be defined similarly as above where the first dose in OLE will be compared instead of randomisation date.

3.1.5 Derivation for response variables

Response variables will be derived for selected efficacy endpoints as mentioned in the sections below. For the response variables, patients will be classified as responders or non-responders based on the specified criteria under each endpoint. Patients with missing or non-evaluable observation at the timepoint of interest will be counted as non-responders in the

analysis. Patients who have had surgery or SCS use for CRSwNP by the timepoint of interest will be considered as non-responders unless otherwise specified.

3.2 Demographics and baseline characteristics

Following demographics and key subject baseline variables will be provided. Baseline definition is specified in Section 3.1.3.

- Age
- Gender
- Country
- Region
- Race
- Ethnicity
- Weight
- Height
- BMI
- Nicotine use
- Nicotine consumption

Important baseline variables for CRSwNP and respiratory disease characteristics will be provided including but not limited to:

- number and percentage of subjects with prior surgery for CRSwNP
- number of prior surgery for CRSwNP
- number and percentage of subjects with prior SCS use for CRSwNP
- number of SCS use for CRSwNP in the past 12 months
- JESREC (<11 vs ≥11)
- any aspirin exacerbated respiratory disease (AERD)/nonsteroidal anti-inflammatory drug exacerbated respiratory disease (NERD)
- CCI
- baseline blood eosinophil count
- atopic disease CCI
- anosmia (based on UPSIT ≤18)

Important baseline variables for asthma disease characteristics will be provided including but not limited to:

- number of prior asthma exacerbations in the past 12 months
- number of prior SCS use for asthma in the past 12 months
- number of emergency room visit due to asthma in the past 12 months
- number of hospitalization due to asthma in the past 12 months
- baseline ACQ-6 score
- number and percentage of subjects using any ICS component (including maintenance and as needed use) at baseline
- number and percentage of subjects on maintenance ICS at baseline
- total daily ICS dose for asthma maintenance medications at baseline

3.3 Primary efficacy variables

According to the SoA Table 2 and Table 3 in the CSP, co-primary efficacy assessments will be collected during the DB period only. Therefore, derivation of co-primary endpoints applies to the DB period and will be reported at the primary DBL.

3.3.1 Nasal polyp score

The total NPS is the sum (maximum 8) of the right and left nostril scores, as evaluated by nasal endoscopy and the left and right score will be based on central read with scale from 0 to 4 as listed in Table 6. The total NPS and the changes from baseline to each post-baseline value will be calculated.

Table 6 Endoscopic nasal polyp score

Polyp score	Polyp size
0	No polyps
1	Small polyps in the middle meatus not reaching below the inferior border of the middle turbinate
2	Polyps reaching below the lower border of the middle turbinate
3	Large polyps reaching the lower border of the inferior turbinate or large polyps of score 2 with additional large polyps medial to the middle turbinate
4	Large polyps causing complete or near-complete obstruction of the inferior nasal cavity i.e. touching the floor of the nose

In addition, the NPS response will be derived as mentioned in Section 3.1.5, respectively. A patient will be classified as a responder at the time point of interest if the change from baseline

in total NPS is ≤-1 (improvement by at least 1 score) at that time point. A 1-grade reduction in bilateral nasal polyp burden has been used in clinical in trials of mometasone furoate nasal spray (Stjarne et al 2006) and sinus implants (Kern et al 2018) for NP.

3.3.2 Nasal blockage score

The NBS is an item in the NPSD. Patients were asked to rate the severity of their worst nasal blockage over the past 24 hours using the following response options: 0 – none; 1 – mild; 2 – moderate; 3 – severe. The NBS and the changes from baseline will be summarised every two weeks (bi-weekly). Baseline NBS is defined in Section 3.1.3 and bi-weekly mean will be calculated if at least 8 days in each 14-day period have evaluable data; otherwise, the bi-weekly mean is set to missing.

A responder definition of change in bi-weekly NBS ≤ -1 is based on the global study D3252C00001 (Shih et al 2023). NBS responder analysis will be conducted as described in Section 3.1.5.

3.4 Key secondary efficacy variables

According to the SoA Table 2 and Table 3 in the CSP, key secondary efficacy assessments will be collected during the DB period only. Therefore, derivation of relevant endpoints applies to the DB period and will be reported at the primary DBL.

3.4.1 Difficulty with sense of smell (DSS) score

The DSS is a key secondary endpoint and will be captured by an item in the NPSD. Patients are asked to rate the severity of their worst difficulty with sense of smell over the past 24 hours using the following response options: 0–none; 1–mild; 2–moderate; 3–severe. The DSS and the changes from baseline will be summarised every two weeks (bi-weekly). Baseline will be the average of daily responses from Day –13 to Day 1 (as defined in Section 3.1.3). Bi-weekly mean of DSS will be calculated if at least 8 days in each 14-day period have evaluable data; otherwise the bi-weekly mean is set to missing.

3.4.2 Sinus CT: Lund-Mackay score

The CT data will come from blinded central reader(s), and it will be performed at V2 (Week -2) and at EoDB (Week 56), or IPD or prior to a surgery for CRSwNP if at least 16 weeks passed since the previous CT scan.

The LMS evaluates the patency using a 0-2 scale (0 – no abnormality; 1 – partial opacification; and 2 – total opacification) of each sinus (maxillary, anterior ethmoid, posterior ethmoid, sphenoid, and frontal sinus on each side). The ostiomeatal complex on each side is graded as 0 – not occluded or 2 – occluded. The total CT score is the sum of the scores from all the sinus and ranges from 0 to 24. The total CT score based on LMS will be calculated for the observed values and the changes from baseline at each time point.

3.4.3 Health related quality life: SNOT-22 score

The SNOT-22 is a condition specific HRQL assessment which captures patient-reported physical problems, functional limitations, and emotional consequences of sinonasal conditions. Patient-reported symptom severity and symptom impact over the past 2 weeks are captured via a 6-point scale (0 – no problem to 5 – problem as bad as it can be). The total score is the sum of item scores and has a range from 0 to 110 (higher scores indicate poorer outcomes). The total score and the changes from baseline will be calculated.

A response to SNOT-22 total score will be defined as changes from baseline ≤–8.90 at the time point of interest. Responder analysis will be conducted according to the approach mentioned in Section 3.1.5.

3.5 Other secondary efficacy variables

According to the SoA Table 2 and Table 3 in the CSP, secondary efficacy assessments will be collected during the DB period only, except that surgery and SCS use for CRSwNP will be further collected during the OLE period.

3.5.1 Time to the first surgery and/or SCS use for CRSwNP

The time to the first surgery for CRSwNP during the DB period will be evaluated for all patients in the PFAS and it is calculated as follows:

Time to the first surgery for CRSwNP= Start date of the first surgery for CRSwNP – date of randomisation + 1

Time to first SCS use for CRSwNP and time to first SCS use or surgery for CRSwNP will be evaluated using the similar approach as time to first surgery for CRSwNP. Namely,

Time to the first SCS use for CRSwNP = Start date of the first SCS use for CRSwNP – date of randomisation + 1

Time to first surgery and/or SCS use for CRSwNP = earlier date of (start date of first surgery for CRSwNP, start date of first SCS use for CRSwNP) – date of randomisation + 1

For patients who do not undergo any surgery or do not experience any SCS use for CRSwNP, the time to the event will be censored at the date of first OLE dose (for those who entered the OLE right after DB completion), or the date of Week 56 Visit (for those who did not enter the OLE after DB, or those who entered DB follow-up period per CSP v1.0 and v2.0), or at the time point after which a surgery or SCS use could not be assessed (for patients discontinued before Week 56 Visit).

The time to recorded decision time to have surgery for CRSwNP will be a supportive endpoint and derived similarly.

3.5.2 SCS use for CRSwNP and proportion of surgery for CRSwNP

The following variables will be summarized during the DB period for all patients in the PFAS (reported at the primary DBL), and during the OLE period for all patients in the OLE analysis set (reported at the final DBL).

- Proportion of patients who had surgery for CRSwNP
- Proportion of patients who use SCS for CRSwNP
- Proportion of patients who had surgery or use SCS for CRSwNP
- The number of courses of SCS for CRSwNP: noted an SCS course can be considered as a new course if the start date is preceded by at least 7 days after the end date of the last SCS course for CRSwNP (i.e. start date of the new course - end date of the last course > 7)
- Total dose of SCS for CRSwNP (converted to prednisolone equivalents)
- Total duration of SCS use for CRSwNP

3.5.3 Nasal polyposis symptom diary (NPSD)

Patients will complete the NPSD each morning. Patients are asked to consider their experience with NP over the past 24 hours when responding to each question and report the severity of each symptom at its worst using a 4-point rating scale (0–none; 1–mild; 2–moderate; 3–severe). Questions to capture patient-reported difficulty with sleep and daily activities due to nasal symptoms use the same rating scale. A TSS will be calculated by taking the sum of the first 8 items in the NPSD.

Bi-weekly mean of each item in the NPSD and TSS, as well as the change from baseline in the bi-weekly mean will be calculated. Bi-weekly mean of each item in the NPSD will be calculated if at least 8 days in each 14-day period have evaluable data; otherwise the bi-weekly mean is set to missing.

3.5.4 University of Pennsylvania smell identification test score

The UPSIT is quantitative test of olfactory function which uses microencapsulated odorants that are released by scratching standardised odour-impregnated test booklets. Four booklets each with 10 odorants each are used for the test. Patients are asked to identify the odour using multiple choice format which lists different possibilities. The test is forced choice, i.e., the patient is required to mark one of the four alternatives even if no smell is perceived. Scores are based on number of correctly identified odours (score range 0 to 40). The UPSIT smell test will be performed in all countries except Thailand and Vietnam. The UPSIT total score and the changes from baseline will be calculated.

The olfactory diagnosis will be classified based on the test scores by gender (male vs female) as listed in Table 7. The categories of Probable Malingering (UPSIT score 0 - 5) and Total Anosmia (6-18) have been combined into a single Anosmia category. Presenting a combined category of anosmia is a conservative approach and ensures anosmia events will not be underreported. This is also consistent with the approach taken in other Phase III clinical trials such as D3252C00001 (OSTRO) and (Bachert et al, 2019).

Table 7 UPSIT olfactory diagnosis

Olfactory diagnosis	Test scores	
	Male	**Female**
Anosmia	00 – 18	00 – 18
Severe microsmia	19 – 25	19 – 25
Moderate microsmia	26 – 29	26 – 30
Mild microsmia	30 – 33	31 – 34
Normosmia	34 – 40	35 – 40

3.5.5 Sinus CT: sinus severity score

Quantitative assessment of sinus CT image data will be used to derive an objective measure of sinus disease burden called sinus severity score (Pallanch et al 2013), defined as:

Sinus severity score = (sinus mucosal volume) / (sinus mucosal volume + sinus air volume) ×100.

The sinus severity score and the change from baseline at all available time points will be calculated for patients in the PFAS.

3.5.6 Sinus CT: Zinreich score

In addition to the Lund-Mackay score described in Section 3.4.2, the same CT images will also be scored using the Zinreich (modified Lund-Mackay) scoring system (Okushi et al 2013, Likness et al 2014).

All five sinuses (maxillary, anterior ethmoid, posterior ethmoid, sphenoid, and frontal) on each side will be scored based on the percentage of opacification from mucosal thickening according to Table 8. Ostiomeatal complex has three categories: 0 - completely patent; 1 - partially obstructed, and 2 – completely obstructed. The maximum total Zinreich score is 50 (54 when including the Ostiomeatal complex score).

Only data from blinded central reader(s) will be used for analysis. The observed values and the changes from baseline at each time point will be calculated.

Table 8 Zinreich score

Score	Percent opacification
0	0%
1	1%-25%

2	26%-50%
3	51%-75%
4	76%-99%
5	100%

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

The short form 36-item health survey, version 2 (standard recall) (SF-36v2) is a 36-item, self-report survey of functional health and well-being, with 4-week recall period (Quality Metric 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 year 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 HRQL.

Categorical analyses will be conducted to evaluate treatment impact on the basis of the established threshold values for change for each scale. The SF-36v2 threshold listed in Table 9 is suitable for interpreting change at the individual level and is referred to as the responder threshold or responder definition (Quality Metric 2011). SF-36 responders, defined as change ≥ threshold noted in Table 9, will be derived using the approach mentioned in Section 3.1.5, regardless of surgery or SCS use for CRSwNP.

Table 9 Threshold values for the SF-36v2 scale and summary measures

	SF-36v2 score									
Threshold	**PCS**	**MCS**	**PF**	**RP**	**BP**	**GH**	**VT**	**SF**	**RE**	**MH**
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.

3.6 Pharmacokinetic and immunogenicity variables

According to the SoA Table 2 and Table 3 in the CSP, PK and ADA assessments will be collected during the DB period and the OLE period.

3.6.1 Pharmacokinetic variables

Serum samples for pharmacokinetic assessments will be collected according to visit schedule in Table 2 and Table 3 of the CSP. The drug concentration levels will be summarised using descriptive statistics.

3.6.2 Immunogenicity variables

Anti-drug antibodies (ADA) variables, such as ADA responses, will be generated and analysed as per the details in Appendix 8.7.

3.7 Exploratory efficacy variables

According to the SoA Table 2 and Table 3 in the CSP, exploratory efficacy assessments will be collected during the DB period only, except that asthma exacerbation will be further collected during the OLE period. Unless otherwise specified, derivation of relevant endpoints applies to the DB period and will be reported at the primary DBL. Additional data on asthma exacerbation from the OLE period will be reported at the final DBL.

3.7.1 Asthma control questionnaire 6

The ACQ-6 will be collected for patients in PFAS. Patients are asked to record their experience with 5 symptoms (night-time waking, symptoms on waking, activity limitation, shortness of breath, and wheezing) and use of short-acting β2 agonist over the previous week using a 7-point scale (0= no impairment; 6=maximum impairment). The ACQ-6 score is calculated by taking the mean of the 6 equally weighted items and ranges from 0 (well controlled) to 6 (extremely poorly controlled).

Individual score change of at least -0.5 is considered clinically meaningful and is used to support the responder definition (Juniper et al 2005, Juniper et al 2006). An ACQ-6 responder at Week 56 will be defined as a patient who had improvement of at least 0.5 on ACQ-6, i.e., an ACQ-6 responder at Week 56 variable takes value 1 if change from baseline to Week 56 in ACQ-6 $\leq$-0.5 and 0 otherwise. ACQ-6 responder variables at Week 24 are similarly defined based on change from baseline to Week 24.

ACQ-6 responses at Week 24 and Week 56 will be derived based on the classifications below and the approach mentioned in Section 3.1.5, regardless of surgery or SCS use for CRSwNP.

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

Furthermore, patients will also be categorised according to their ACQ-6 using the following score thresholds (Juniper et al 2005, Juniper et al 2006) and the approach mentioned in Section 3.1.5, regardless of surgery or SCS use for CRSwNP.

- ACQ-6 (Timepoint) $\leq$0.75 → Well controlled
- 0.75< ACQ-6 (Timepoint) <1.5 → Partly controlled

- ACQ-6 (Timepoint) ≥1.5 → Not well controlled

3.7.2 Asthma exacerbations rate (AER)

The number of exacerbations will be provided for 2 periods (i.e., DB period and OLE period as defined in Section 3.1.1), respectively. A patient will be considered as having asthma exacerbation if he/she starts a course of SCS (oral, parenteral) use due to asthma that lasts at least 3 days or have an emergency room/urgent care visit with SCS use due to asthma that lasts at least 3 days or a hospitalisation due to asthma.

The annual exacerbation rate in each treatment group will be calculated for DB period using the time-based approach by using the following formula:

Annual exacerbation rate = 365.25×total number of exacerbations/total duration of follow-up within the treatment group (days).

3.7.3 Patient global impression of severity and patient global impression of change

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; 1 – very mild; 2 – mild; 3 – moderate; 4 – severe; 5 – very severe.

The PGI-C captures the patient's overall evaluation of response to treatment. The patients are asked to report the degree to which they have changed since entering the treatment period using a 7-point scale: 1– much better; 2 – moderately better; 3 – a little better; 4 – about the same; 5 – a little worse; 6 – moderately worse; and 7 – much worse.

Patients will be categorized according to the following PGI-C responses post-baseline:

- About the same, a little worse, moderately worse, much worse → Not better
- Much better, moderately better, a little better → At least a little better
- Much better, moderately better → At least moderately better
- Much better → Much better

3.7.4 Healthcare resource utilisation (HRU)

The following unplanned/unscheduled healthcare resource use (HRU) due to (a) nasal polyps, (b) asthma exacerbation, and (c) other reasons will be collected.

- General and intensive care hospitalisations and lengths of stay
- Emergency room visits
- Urgent care visits

For each category above, the crude rates of HRU during the DB period (as defined in Section 3.1.1) will be calculated, where the *crude rate = (total days of HRU use/total days during double-blind period)*.

3.8 Safety outcome variables

According to the SoA Table 2 and Table 3 in the CSP, safety assessments (including AE, haematology, clinical chemistry, vital signs, etc.) will be collected during the DB period and the OLE period, except that ECG is collected only during the DB period.

No safety data will be imputed. The handling of partial/missing dates for AEs and prior/concomitant medications is detailed in Appendix 8.6.1. Duration of AEs and prior/concomitant medications will not be calculated using imputed dates and will instead be set to missing.

3.8.1 Adverse events

Adverse events experienced by 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).

AEs will be separated according to their onset date into the following study periods:

- AEs occurring during the DB period: date of first dose of IP in DB period ≤ AE onset date ≤ EoDB, except any AE that starts on or after the date of first OLE dose should only be counted in the OLE and not in the DB period.
- AEs occurring during the OLE period: date of first dose of IP in OLE ≤ AE onset date ≤ study completion or withdrawal date
- AE occurring during the extended follow-up (applicable to CSP v1.0 or v2.0):
  - for patients who did not enter the OLE: EoDB < AE onset date ≤ study completion or withdrawal
  - for patients who were followed up off treatment after DB period and then entered the OLE: EoDB < AE onset date < date of first dose of IP in OLE.

Similarly, if an AE has a partial onset date, then unless the partial onset date or the stop date indicates otherwise, it will be summarized in the DB or OLE period based on the imputed onset date (Section 8.6.1.1) compared to the first dose in the respective analysis period.

3.8.2 Laboratory variables

Laboratory assessment values and normal ranges will be presented in the International System (SI) unit. Eosinophil data will be presented in both SI and conventional units (cells/μL) in summaries.

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

Absolute values will be compared to the relevant reference range and classified as low (below range), normal (within range or on limits) or high (above range). All absolute values falling outside the reference ranges will be flagged. The maximum or minimum value post-baseline will be calculated over the entire DB period and OLE period, respectively, for all patients.

3.8.3 Vital signs

Pre-dose vital signs (pulse, systolic blood pressure, diastolic blood pressure, respiration rate, and body temperature) will be obtained in accordance with schedule provided in the protocol. Changes in vital signs variables between baseline and each post-baseline assessment will be calculated. There will be no imputation for missing values.

Absolute values will be compared to the reference ranges listed in Table 10 and classified as low (below lower limit), normal (within lower limit and upper limit, inclusive) or high (above upper limit). All absolute values falling outside the reference ranges will be flagged.

Table 10 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

CCI
CCI

3.8.4 ECG

ECG measurements will be assessed during the DB period in accordance with the protocol, with the baseline being defined as the last available non-missing measurement prior to first dose of randomised treatment (V3).

The outcome of the overall evaluation is to be recorded as normal/abnormal in the eCRF by the Investigator/authorised delegate, with any abnormalities being recorded as not clinically significant or clinically significant. Any new finding(s) or aggravated existing finding(s), judged as clinically significant by the Investigator, unless unequivocally related to the disease-under-study, will be reported as an AE.

3.8.5 Physical examination

Complete and brief physical examinations will be performed at time points specified in SOA Tables in the CSP. What is included in the assessment will be dependent on whether the examination is complete or brief, as described in Section 8.2.2 of the CSP. Only information on whether the assessment was performed or not is to be recorded. Any new finding(s) or aggravated existing finding(s), judged as clinically significant by the Investigator, unless unequivocally related to the disease-under-study, will be reported as an AE.

4. ANALYSIS METHODS

4.1 General principles

The primary database lock will occur after all randomised patients have been followed up for the 56-week double-blind treatment period. The final database lock will occur after the last patient completes the OLE.

Analyses at the primary DBL will include all data captured during the double-blind treatment period and selected safety endpoints during the OLE period. Unblinding and data analysis will occur after the primary database lock. Efficacy endpoints will be analysed using the PFAS according to their randomised treatment. Efficacy data collected from patients without comorbid asthma will be listed only. Safety endpoints will be analysed using the Safety Analysis Set (or the OLE set if specified) regardless of comorbid asthma status according to their actual treatment received.

In the primary analysis of co-primary endpoints (and key secondary endpoints unless otherwise stated), an ANCOVA model will be used with treatment arm, baseline score (for the respective endpoint), region, CCI [redacted] as covariates. Handling strategy of intercurrent events are specified as follows：

- *Nasal polyp surgery*: a nasal polyp surgery is considered as treatment failure. The composite strategy will be applied where the worst possible (WP) score for the respective endpoint will be used for post-surgery assessments.
- *SCS use for CRSwNP*: SCS use for CRSwNP is considered as treatment failure but less severe than surgery. The composite strategy will be applied where the worst observation prior to the SCS use will be carried forward (WOCF) for post-SCS assessments.
- *Treatment discontinuation*: the treatment policy strategy will be applied.

The analysis includes all the data collected during the study including data collected after discontinuation of study treatment except data collected after surgery and/or SCS use for CRSwNP. For participants who discontinue the study without any rescue event (i.e. NP surgery and/or SCS use for CRSwNP), a multiple imputation (MI) approach will be used to impute missing values assuming missing at random (MAR). The estimates of the treatment effects will be based on contrasts from the ANCOVA model.

Details regarding the estimand for primary and key secondary objectives are provided in Table 11, with additional details including supplementary and sensitivity analyses provided in Appendix 8.2 to 8.5 and Appendix 8.10.

Summary data will be presented in tabular format by treatment group. Categorical data will be summarised by the number and percentage of patients in each category. Continuous variables for parametric data will be summarised by descriptive statistics including number of patients in analysis (n), mean, SD, median, and range (minimum (Min), maximum (Max)).

All hypothesis testing will be reported using 2-sided. P-values will be rounded to 4 decimal places. The 95% CI will be reported for efficacy endpoints as appropriate.

The absolute change from baseline is computed as (*visit value – baseline value*). Percent change from baseline is computed as (*visit value – baseline value*) / *baseline value* × 100%. If either a visit value or the baseline value is missing, the absolute change from baseline value and the percent change from baseline will also be set to missing.

The data analyses will be conducted using the SAS® System (SAS Institute Inc., Cary, NC), version 9.4 or higher. All SAS® programs used to generate analytical results will be developed and validated according to AstraZeneca SAS® programming standards and validation procedures.

Table 11 Estimand for primary and key secondary objectives and safety objectives

	Estimand[1]			
Statistical Category & Section	**Treatment Condition[1]**	**Endpoint (Population)**	**Intercurrent Event Strategy[1]**	**Population Level Summary[1] (Analysis)**
Primary Objective: To evaluate the effect of benralizumab on nasal polyp burden and patient-reported nasal blockage				
Co- Primary/MCP Section 4.2.5.1	Treatment with benralizumab versus placebo, regardless of compliance, where rescue indicates treatment failure.	• CFB in endoscopic total NPS (severe CRSwNP patients with asthma) • CFB in bi-weekly mean NBS (severe CRSwNP patients with asthma)	• Treatment discontinuation –treatment policy • Surgery for CRSwNP– composite (Worst Possible) • SCS use for CRSwNP – composite (WOCF)	Mean difference between interventions at Week 56[2] (LSMD from CFB ANCOVA following hybrid WP/WOCF and MI).
Key Secondary Objective: To evaluate the effect of benralizumab on sense of smell				
Secondary/MCP Section 4.2.6.1	Treatment with benralizumab versus placebo, regardless of compliance, where rescue indicates treatment failure.	• CFB in bi-weekly mean DSS (severe CRSwNP patients with asthma)	• Treatment discontinuation –treatment policy • Surgery for CRSwNP – composite (Worst Possible) • SCS use for CRSwNP – composite (WOCF)	Mean difference between interventions at Week 56[2] (LSMD from CFB ANCOVA following hybrid WP/WOCF and MI).
Key Secondary Objective: To evaluate the effect of benralizumab on sinus opacification				
Secondary/MCP Section 4.2.6.2	Treatment with benralizumab versus placebo, regardless of compliance, where rescue indicates treatment failure.	• CFB in mean LMS (severe CRSwNP patients with asthma)	• Treatment discontinuation –treatment policy • Surgery for CRSwNP – composite (Worst Possible) • SCS use for CRSwNP – treatment policy	Mean difference between interventions at EoDB/IPD (LSMD from CFB ANCOVA following WP imputation).
Key Secondary Objective: To evaluate the effect of benralizumab on disease specific health-related quality of life				
Secondary/MCP Section 4.2.6.3	Treatment with benralizumab versus placebo, regardless of compliance, where rescue indicates treatment failure.	CFB in SNOT-22 score (severe CRSwNP patients with asthma)	• Treatment discontinuation –treatment policy • Surgery for CRSwNP – composite (Worst Possible) • SCS use for CRSwNP – composite (WOCF)	Mean difference between interventions at Week 56[2] (LSMD from CFB ANCOVA following hybrid WP/WOCF and MI).
Safety Objective: To evaluate the safety and tolerability of benralizumab				

Safety Section 4.2.10.1	Treatment with benralizumab, regardless of compliance, versus placebo, regardless of compliance.	• Presence of AEs and SAE (all severe CRSwNP patients regardless of co-morbid asthma status)	Treatment discontinuation –treatment policy	Categorical descriptive

MCP = Multiple comparisons procedure; CRSwNP = Chronic rhinosinusitis with nasal polyps; SCS = Systemic Corticosteroids; NPS = Nasal Polyps Score; NBS = Nasal Blockage Score; DSS = Difficulty with Sense of Smell; SNOT-22 = SinoNasal Outcome Test, 22 item; LMS = Lund-Mackay score; WP = Worst Possible; WOCF = Worst Observation Carried Forward; MI = Multiple Imputation; LSMD = Least Squares Mean Difference; CFB = Change from baseline; ANCOVA = Analysis of Covariance; EoDB = End of treatment in double-blind period; IPD = Investigational product discontinuation; AE = Adverse event; DB = Double-blind.

[1] All estimand attributes explicitly identified for primary and key secondary endpoints only.

[2] Week 56 is the multiplicity protected timepoint for co-primary and key secondary endpoints.

4.1.1 China and Japan sub-population

The China and Japan sub-populations are defined as all patients enrolled in sites in China (including mainland and Taiwan) and Japan, respectively.

To support registration in China and Japan, study summaries will be repeated for the China and Japan sub-populations, which includes but may not be limited to patient disposition, demographic and baseline characteristics, extent of exposure, prior and concomitant medications, efficacy, and safety analyses. Results in the China and Japan sub-populations may be reported separately with the objective of demonstrating consistency with the overall population.

Efficacy analyses in the China and Japan sub-population will proceed as described in Section 4.2.5, Section 4.2.6, and Section 4.2.7. Additional consideration should be given in case of a limited number of event(s) in a regional sub-population. Specifically, time to event analysis using Cox regression will not be conducted in case of < 10 events. Responder analysis using logistic regression will include only treatment group as model covariate in case of < 10 responders. Similarly, unadjusted estimate from negative binomial regression will be provided in case of < 10% subjects with event(s). If event number in either treatment group is 0 in a stratum, results from unstratified Cochran–Mantel–Haenszel analysis will be reported.

Exploratory and supplementary/sensitivity analyses may be replicated for the China and Japan sub-population, if deemed informative. All analyses based on the China and Japan sub-population will be considered exploratory. No adjustment for multiplicity will be made; thus the testing strategy detailed in 4.1.3 will not be employed directly to the China and Japan sub-population.

Safety analyses will include summaries of AE overview, AEs, and SAEs by System Organ Class (SOC) and Preferred Term (PT). Descriptive summary of PK concentrations will be repeated as well for the China and Japan sub-population.

The analyses for regional registration in China and Japan will be presented separately from the global CSR.

4.1.2 Hypothesis tests for co-primary endpoints

The co-primary efficacy endpoints are the change from baseline in total NPS at Week 56 and the change from baseline in bi-weekly mean NBS at Week 56. The primary analysis is to compare the changes from baseline in total NPS and in NBS of benralizumab with placebo. The following hypothesis will be tested for the co-primary endpoints:

- H_0: The change from baseline in total NPS and/or the change from baseline in NBS are similar between benralizumab and placebo.

- H_1: Both of the change from baseline in total NPS and the change from baseline in NBS are different between benralizumab and placebo.

The co-primary endpoints will be evaluated using a hybrid method of the worst-possible/worst-observation carried forward and multiple imputation, followed by an analysis of covariance with treatment arm, baseline scores of corresponding endpoint, region, CCI [redacted] as covariates.

4.1.3 Method for multiplicity control

To account for multiplicity to test the co-primary endpoints (change from baseline in total NPS at Week 56 and change from baseline in bi-weekly mean NBS at Week 56) and the three key secondary endpoints (the change from baseline in bi-weekly mean DSS, LMS, and SNOT-22 total score at Week 56), the type I error will be controlled across co-primary and key secondary endpoints at 0.05 (two-sided) statistical significance level. Both of the co-primary and the key secondary endpoints will be tested at 0.05 (two-sided) level. A testing strategy below will be as follows.

- Step 1: Perform the 2 tests of co-primary endpoints at a significant level of 0.05. If both p-values are less than 0.05, then proceed to Step 2. Otherwise no null hypothesis is rejected.
- Step 2: Test the three key secondary endpoints using a step-down approach following the hierarchical order: bi-weekly mean DSS, LMS, and SNOT-22 total score for at the significant level of 0.05. If at any time a null hypothesis cannot be rejected in favour of the benralizumab group, further testing will stop, and no subsequent null hypotheses in the testing hierarchy will be rejected.

4.1.4 Description of estimands used in the analyses

4.1.4.1 Primary estimand

The primary estimand quantifies the difference in outcomes for patients randomised to the benralizumab and placebo arms at the planned timepoints of the study, regardless of the treatment that patients actually received, where rescue by surgery and/or SCS use for CRSwNP indicates failure.

An ANCOVA model will be used with treatment arm, baseline score (for the respective endpoint), region, CCI [redacted] as covariates. Handling strategy of intercurrent events are specified as follows：

- *Nasal polyp surgery*: a nasal polyp surgery is considered as treatment failure. The composite strategy will be applied where the worst possible (WP) score for the respective endpoint will be used for post-surgery assessments.
- *SCS use for CRSwNP*: SCS use for CRSwNP is considered as treatment failure but less severe than surgery. The composite strategy will be applied where the worst

observation prior to the SCS use will be carried forward (WOCF) for post-SCS assessments.

- *Treatment discontinuation*: the treatment policy strategy will be applied.

The analysis includes all the data collected during the study including data collected after discontinuation of study treatment except data collected after surgery and/or SCS use for CRSwNP. For participants who discontinue the study without any rescue event (i.e. NP surgery and/or SCS use for CRSwNP), a multiple imputation (MI) approach will be used to impute missing values assuming missing at random (MAR). The estimates of treatment effects will be based on contrasts from the ANCOVA model.

A supplementary analysis will use a different strategy (treatment policy strategy) for SCS use for CRSwNP. In this supplementary analysis, a composite strategy will be used only for patients who undergo surgery for CRSwNP in which data collected post-surgery will be set to missing, and the patient's worst possible value will be imputed from that point through Week 56. For patients who discontinue the study without surgery, missing data will be imputed using a multiple imputation using all patients who did not have surgery. Patients who receive SCS for CRSwNP without having surgery for CRSwNP will be included in non-surgery patients. Details of this supplementary analysis are included in Appendix 8.2.

4.1.4.2 Effectiveness estimand

The effectiveness estimand quantifies the difference in outcomes for patients randomised to the study treatment and control arms, while on treatment and without having been rescued by surgery and/or SCS use for CRSwNP, at the planned endpoint of the study. In the effectiveness estimand, while-on-treatment strategy will be applied for all intercurrent events, and all the data up to the IP discontinuation visit will be included except the data observed after surgery and SCS use for CRSwNP. No imputation will be conducted for missing data as analysis will use an ANCOVA model for repeated measures. A supplementary analysis will be evaluated under the effectiveness estimand. Details of the supplementary analyses are included in Appendix 8.4.

4.1.4.3 Treatment policy estimand

The treatment policy estimand quantifies the difference in outcomes for patients randomised to the study treatment and control arms at the planned endpoint of the study, regardless of the occurrence of intercurrent events. The treatment policy estimand includes all the data collected during DB period including data collected after discontinuation of study treatment, surgery for CRSwNP, or SCS use for CRSwNP. No imputation will be conducted for the missing data as analysis will use an ANCOVA model for repeated measures. A supplementary analysis will be evaluated under the treatment policy estimand. Details of this supplementary analysis are included in Appendix 8.5.

4.1.5 Sensitivity analyses for missing data

Sensitivity analyses of the ANCOVA with missing data imputed will be performed for co-primary endpoints (and key secondary endpoints when specified) using controlled sequential

multiple imputation (MI) methods based on pattern mixture models, as described in the CHMP Guideline on Missing Data in Confirmatory Clinical Trials (CHMP 2010) and by Little et al (Little et al 2010). Details are included in Appendix 8.3.

4.2 Analysis methods

4.2.1 Patient disposition

Patient disposition will be summarised using the All Patients Analysis Set. The total number of patients will be summarised for the following groups: those who enrolled and those who were not randomised (and reason). The number and percentage of patients within each treatment group will be presented by the following categories: randomised, received treatment with study drug, did not receive treatment with study drug (and reason), completed treatment with study drug in DB period, discontinued treatment with study drug in DB period (and reason), discontinued treatment with study drug in DB period but completed DB follow-up, completed DB period, and withdrew from DB period (and reason).

For patients who completed the DB period, the number and percentage of patients will be presented by the following categories in the OLE analysis set: enrolled in the OLE period, completed treatment in OLE period, discontinued treatment in OLE period (and reason), completed OLE period, and withdrew from OLE period (and reason).

Screen failure information will be listed for the All Patients Analysis Aet.

The number of patients randomised by country and centre will be summarised by treatment group for patients in the PFAS.

4.2.2 Demographics and baseline characteristics

The demographics and key subject baseline variables (Section 3.2) will be descriptively summarised based on FAS and PFAS. This summary will be repeated for the OLE period using the OLE analysis set.

CRSwNP and respiratory disease characteristics (Section 3.2) and asthma disease characteristics (Section 3.2) at baseline will be descriptively summarised based on PFAS.

4.2.3 Prior and concomitant medications

The number and percentage of patients who take maintenance medications will be summarised.

The number and percentage of patients who take prior medications and patients who take allowed concomitant medications will be summarised over the DB period based on FAS, and patients who take disallowed concomitant medications will be based on the PFAS; these may be repeated over the OLE period based on the OLE analysis set. Concomitant medications

will be classified according to the WHODRUG dictionary. The summary tables will present data by generic term within Anatomical Therapeutic Chemical (ATC) code.

4.2.4 Study treatment administration

4.2.4.1 Exposure

Duration of the study drug exposure over the DB period and the OLE period will be summarised by treatment group for the safety analysis set and OLE analysis set, respectively. Duration of IP administration will be calculated in days as:

Duration of IP administration in the DB period = Date of last dose or early discontinuation in the DB – Date of first IP dose in the DB + 1.

Duration of IP administration in the OLE period = Date of last dose or early discontinuation in the OLE – Date of first IP dose in the OLE + 1.

4.2.4.2 Study treatment compliance

Study treatment compliance during the DB period and during the OLE period will be summarised by treatment group for the patients in the PFAS and OLE analysis set, respectively. In addition, compliance during the on-study period may be summarised using the PFAS. Study treatment compliance during the DB period will be reported at the primary DBL, and compliance during the OLE (and on-study) period will be reported at the final DBL.

Compliance will be calculated as:

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

Total number of doses expected includes all visits with protocol scheduled IP administration on or before a patient's IP discontinuation or treatment completion date during DB period (where the last IP is dosed at Week 48), OLE period (where the first IP starts at Week 56), and on-study, respectively.

4.2.5 Primary efficacy variables: nasal polyps score and nasal blockage score

Co-primary efficacy endpoints will be collected during the double blinded period according to the SoA Table 2 in the CSP. No data from the OLE period will be collected for co-primary endpoints. Therefore following analyses specified in Section 4.2.5.1 to 4.2.5.4 only apply to the double blind period at the primary DBL.

The co-primary efficacy endpoints are the change from baseline in total NPS at Week 56 and the change from baseline in bi-weekly mean NBS at Week 56.

The efficacy endpoints will be summarised by treatment and time point (i.e. week number) during DB period for all patients in PFAS.

4.2.5.1 Primary analysis

The primary analysis for both co-primary endpoints will use the primary estimand as described in Section 4.1.4.1. An ANCOVA model will be conducted adjusting for treatment arm, baseline scores (baseline total NPS for NPS model and baseline bi-weekly mean NBS for NBS model), region, CCI as covariates. Estimates of the treatment effects at Week 56 were based on contrasts from the ANCOVA model. The analysis includes all patients with baseline and at least one evaluable post-baseline assessment and all the patients with a baseline assessment who are rescued by surgery and/or SCS use for CRSwNP. For participants who discontinue the study without any rescue event (i.e. NP surgery and/or SCS use for CRSwNP), a multiple imputation (MI) approach will be used to impute missing values assuming missing at random (MAR).

Specifically, if a patient has surgery or SCS use for CRSwNP before Week 56, the data will be censored after the time of the first surgery and/or the time of having the first course of SCS use for CRSwNP, and the patient's worst-possible/previously worst-observed value will be imputed in its place. For patients rescued by SCS use for CRSwNP whose post-baseline values are all missing or for whom every post-baseline value is after rescue by SCS use for CRSwNP, the baseline will be used to impute. If a patient has surgery for CRSwNP before Week 56, the data will be censored after the time of the first surgery for CRSwNP and the worst possible value will be imputed in its place. If there is sufficient evaluable NBS data prior to rescue in the bi-weekly period in which rescue occurs, the bi-weekly mean for that period will be based on the data collected prior to rescue. Otherwise, the WP/WOCF will be imputed for that period as well. See Table 13 for a full list of WP values. For patients who discontinue the study without surgery or SCS use for CRSwNP, missing data will be imputed using multiple imputation using all patients who does not have surgery or receive SCS use for CRSwNP.

The following 5 steps will be used to build the imputation datasets and perform analyses:

1. 100 datasets with a monotone missing pattern will be obtained, induced by Markov Chain Monte Carlo (MCMC) method on all patients with at least one post-baseline and pre-rescue (if applicable) evaluable assessment (utilizing only data prior to surgery and/or SCS use for CRSwNP, or Week 56, whichever comes first for those rescued). The model for MCMC imputation will include adjustment for covariates including treatment group, region, CCI and baseline value of the response value. See Table 14 for random seeds to be used for MCMC step.

2. For each of the imputed datasets obtained in step 1, the remaining missing data for patients who are not rescued by surgery or SCS for CRSwNP by the end of double blind period will be imputed using the regression method for the monotone pattern with adjustment for covariates including treatment groups, region, CCI and baseline value of the response variable. No data from rescued patients is included in this step. See Table 14 for random seeds to be used for monotone regression step.

3. For patients who are rescued by surgery or SCS for CRSwNP by the end of double blind period, any remaining missing data following step 1 after the time of surgery or SCS use for CRSwNP will be replaced by the WP/WOCF approach described above. Any missing data between the last observed value and the date of rescue (for those patients that are rescued by Week 56) will be replaced using last observation carried forward (LOCF). Similarly for patients with only a baseline prior to rescue (who are excluded from step 1), any missing data after rescue will be replaced by WP/WOCF and any missing data between baseline and rescue will be replaced with the baseline value (baseline LOCF). This approach ensures complete data for all patients at all timepoints without forcing a poor outcome (i.e. WP/WOCF) prior to rescue.

4. For each of the 100 imputations, the datasets from the rescued and non-rescued patients will be combined to create 100 complete datasets which will then be analysed using the main statistical model. These 100 datasets will be saved.

5. Apply Rubin's rule to combine analysis results (point estimates and standard errors) from 100 imputations. The estimated least squares (LS) means, difference in LS means, and the corresponding 95% confidence intervals (CI) will be provided along with the nominal p-values for Week 56 and all earlier time points in turn.

4.2.5.2 Sensitivity/supplementary analyses

A sensitivity analysis under the primary estimand will be conducted with missing data imputed based on different withdrawal reasons. This model assumes that some pre-specified subset of patients who withdraw from the study have correlations with future unobserved visits similar to patients in the placebo arm.

- ANCOVA under primary estimand with WP/WOCF imputation after surgery and/or SCS use for CRSwNP and dropout reason based multiple imputation for non-rescued study discontinuation to test the MAR assumption (Section 8.3).

Additional supplementary analyses will be conducted with a different handling strategy for surgery and/or SCS use for CRSwNP.

- ANCOVA with composite strategy for surgery for CRSwNP only, treatment policy strategy for other intercurrent events, and multiple imputation for missing data (including those after rescue SCS use) assuming MAR (Appendix 8.2).
- ANCOVA for repeated measures (without imputation) with while-on-treatment strategy for all intercurrent events, where outcomes after treatment discontinuation, surgery for CRSwNP, or SCS use for CRSwNP will be set to missing (Appendix 8.4 effectiveness estimand).
- ANCOVA for repeated measures (without imputation) with treatment policy strategy for all intercurrent events, where all data as observed through Week 56 will be included

regardless of treatment discontinuation, surgery for CRSwNP, or SCS use for CRSwNP (Appendix 8.5 treatment policy estimand).

4.2.5.3 Subgroup analyses

To explore the uniformity of the detected overall treatment effect, subgroup analyses will be performed for the co-primary endpoints, i.e. change from baseline in total NPS and in bi-weekly mean NBS, and for the key secondary endpoints, bi-weekly mean DSS, LMS and SNOT-22 total score, using an ANCOVA at Week 56 for the following subgroup factors:

- Gender (male vs female)
- Age (18 - 65 vs ≥65 years)
- Geographical Region
  - North America includes the US
  - Europe includes Belgium, Bulgaria, France, Hungary, Italy, Poland, Russia, Turkey
  - Asia includes China mainland, Taiwan, Japan, Thailand, Vietnam
  - Rest of the World includes Argentina, Australia, Chile
- CCI [redacted]
- Prior surgery for CRSwNP (yes vs no)
- Number of prior surgery for CRSwNP (0, 1, 2 or more)
- Prior SCS use (yes vs no)
- Baseline AERD (yes vs no)
- CCI [redacted]
- Atopic status by atopic medical history (yes vs no)
- CCI [redacted]
- Quartiles of baseline blood eosinophil (EOS) counts

The quartile of baseline blood EOS counts CCI [redacted] will be calculated using the respective measurements at baseline from all the patients in the PFAS, regardless of treatment groups.

For each of the subgroup factors in turn, a separate ANCOVA model will be fitted using the same model terms as used for the primary analysis (defined in Section 4.2.5.1) and for the key secondary endpoints (defined in Sections 4.2.6.1, 4.2.6.2, and 4.2.6.3), including WP/WOCF and MI components, with additional terms for the subgroup main effect and the treatment×subgroup interaction.

Similar outputs will be presented for each subgroup as for the primary analysis. The p-value for the interaction term will be presented in the summary tables.

In addition, subgroup analyses will also be conducted at Week 56 by cumulative subgroups of EOS counts at baseline (≥150 cells/μL, ≥200 cells/μL, …, ≥500 cells/μL) and the other direction (<150 cells/μL, <200 cells/μL, …, <500 cells/μL). For these subgroup analyses, data will be analysed using the ANCOVA model as mentioned in Section 4.2.5.1 and Sections 4.2.6.1, 4.2.6.2, and 4.2.6.3 in each subgroup category.

If there is strong evidence for a differential treatment effect for NPS and/or NBS across subgroups, additional analyses based on the NPS and/or NBS responder analysis may be conducted. Additional terms for the subgroup main effect and treatment×subgroup interaction will be included in the logistic regression model defined in Section 4.2.5.4.

Subgroup analyses based on PK will also be performed for the change from baseline at Week 56 for co-primary and key secondary endpoints. All patients with at least one post-baseline PK sample collected will be included in this analysis. Patients on Benra will have their median PK concentration calculated based on all observed trough concentrations at steady state during the 56-week analysis period. Values collected after treatment discontinuation or without IP dosing at the previous visit will not be used to derive the median. Patients on Benra will then be grouped to PK quartiles based on their median concentration values and compared against the overall placebo group. A similar ANCOVA model, as described in Section 4.2.5.1 and Section 4.2.6.1 to Section 4.2.6.3, will be used with treatment group (i.e. Benra vs placebo) replaced by a 5-category PK variable (4 PK quartiles for benralizumab and 1 for placebo). Estimates of the treatment effects will be reported based on contrasts between each Benra quartile and placebo.

4.2.5.4 Supportive analyses: responder analyses

Patients with missing or non-evaluable observation at the timepoint of interest will be counted as non-responders for in the analysis. Patients who have had surgery or SCS use for CRSwNP by the timepoint of interest will be considered as non-responders.

The proportion of NPS responders with improvement by at least 1 score (defined as change from baseline in total NPS ≤-1) at Week 24 and Week 56 will be analysed using a logistic regression model with treatment arm, baseline total NPS score, region, CCI [redacted] CCI as covariates. Results of the analyses will be presented using odds ratio, together with associated 95% CI and 2-sided p-value.

Cumulative distribution function of absolute changes from baseline in NPS at Week 24 and at Week 56 will be plotted in a figure.

Similar responder analyses will be conducted for NBS responders (defined as change from baseline in NBS ≤ -1) at Week 24 and Week 56. Cumulative distribution function of absolute changes from baseline in NBS at Week 24 and at Week 56 will also be plotted in a figure.

4.2.6 Key secondary efficacy variable

Key secondary efficacy endpoints will be collected during the double blinded period according to the SoA Table 2 in the CSP. No data from the OLE period will be collected for key secondary endpoints. Therefore following analyses specified in Section 4.2.6.1 to 4.2.6.3 only apply to the double blind period at the primary DBL.

4.2.6.1 Difficulty with sense of smell (DSS) score

The change from baseline in bi-weekly mean DSS score will be analysed using a similar ANCOVA as the primary endpoints described in Section 4.2.5.1. DSS scores and the change from baseline will be summarised using descriptive statistics.

The cumulative distribution function of absolute changes from baseline in bi-weekly mean DSS at Week 24 and Week 56 will also be plotted in figures.

Subgroup analyses as mentioned in Section 4.2.5.3 will also be conducted for the change from baseline in bi-weekly mean DSS, including the subgroup analysis by PK exposure quartile.

4.2.6.2 Sinus CT: LMS

The total LMS score and the change from baseline will be summarised by treatment group and by time point using descriptive statistics for all patients in the PFAS. The change from baseline in total LMS will be analysed using ANCOVA model as the co-primary endpoints described in Section 4.2.5.1 but a different intercurrent event strategy (treatment policy) to handle rescue SCS use for CRSwNP. The analyses will use the data collected during the DB period, regardless of whether patients remained on treatment or not. In order to limit the radiation exposure, there is only a single CT scan post-baseline that occurs at Week 56 (or IPD or prior to surgery for CRSwNP), and hence the composite (WOCF) strategy used after SCS use for CRSwNP in the primary estimand is considered less appropriate as for co-primary endpoints. Instead, the analysis will use data collected after SCS use for CRSwNP. A composite strategy will be used for surgery for CRSwNP. If a patient has surgery for CRSwNP before EoDB, the data will be censored after the time of the first surgery for CRSwNP and the worst possible value will be imputed in its place. See Table 13 for a full list of WP values. The Multiple Imputation components described in Section 4.2.5.1 will not be necessary for analysis of LMS scores.

A supplementary analysis of LMS scores will be conducted in which the composite (WOCF) strategy is used for SCS use for CRSwNP. Patients who receive SCS for CRSwNP prior to EoDB/IPD CT scan will have the worst observed score (or baseline score if no post-baseline score available) imputed in place of the observed or missing values. Patients who have surgery for CRSwNP prior to EoDB/IPD CT scan will have the worst possible value imputed whether or not they also receive SCS for CRSwNP. An additional supplementary analysis of LMS scores will be conducted in which patients who have surgery for CRSwNP prior to EoDB/IPD CT scan will have the last observed score prior to surgery for CRSwNP imputed over data collected post-surgery.

Subgroup analyses as mentioned in Section 4.2.5.3 will also be conducted for the change from baseline in LMS, including the subgroup analysis by PK exposure quartile.

4.2.6.3 Health related quality life: SNOT-22 score

For the total score and individual item, SNOT-22 scores and changes from baseline will be summarised by treatment group and by time point using descriptive statistics for all patients in the PFAS. The change from baseline in the SNOT-22 total score will be analysed using a similar ANCOVA analysis as the primary endpoints described in Section 4.2.5.1. Results will be presented in terms of LSMEANS, treatment differences in LSMEANS, 95% CI and p-values for all time points.

Responder analyses at Week 24 and Week 56, where a responder is defined as change from baseline ≤–8.9, will be conducted using the same logistic regression model mentioned in Section 4.2.5.4. The results of the analyses will be presented using odds ratio, together with associated 95% CI and 2-sided p-value.

Subgroup analyses as mentioned in Section 4.2.5.3 will also be conducted for the change from baseline in SNOT-22 total score, including the subgroup analysis by PK exposure quartile.

The cumulative distribution function of absolute changes from baseline in SNOT-22 total score at Week 24 and at Week 56 will also be plotted in figures.

4.2.7 Other secondary efficacy variables

All other secondary efficacy endpoints will be collected during the double blinded period according to the SoA Table 2 in the CSP, except surgery and SCS use for CRSwNP will be further collected during the OLE period (CSP Table 3). Therefore following analyses specified in Section 4.2.7.1 and 4.2.7.6 only apply to the double blind period at the primary DBL. Surgery and SCS use for CRSwNP during the OLE period will descriptively summarized at final DBL.

4.2.7.1 Time to first surgery and/or SCS use for CRSwNP

At the primary DBL, the time to first surgery for CRSwNP, time to first course of SCS use for CRSwNP, and the time to first surgery or first course of SCS use for CRSwNP during the DB period will be analysed using a Cox proportional hazards model with treatment arm, region, CCI as covariates. Double blind period defined in Section 3.1.1 (as opposed to the target Day 393 for Week 56) will be used such that events (if any) reported after Day 393 but before the Visit 11 completion (or the first OLE dose) can be included in the respective time-to-event analysis. Results will be summarised as hazard ratios, 95% CI and p-values. Time to event will be displayed graphically using a Kaplan-Meier plot.

As another supportive analysis, the time to decision to have surgery for CRSwNP will be analysed using the same Cox proportional hazards model.

At the final DBL, additional descriptive summary will be provided for events reported during the OLE period; Cox proportional hazards model will not be conducted.

4.2.7.2 SCS use for CRSwNP and proportion of surgery for CRSwNP

At the primary DBL, surgery and SCS use for CRSwNP will be summarised during DB period for all patients in the PFAS. Specifically, the surgery for CRSwNP will be summarised by the surgery reason, and surgery procedures. The total number of courses of SCS use for CRSwNP, the total SCS dose for CRSwNP used, and total duration of SCS use for CRSwNP will be summarised using descriptive statistics. Each of these two events will be summarised regardless of if a patient experiences the other event (surgery for CRSwNP regardless of patient initiates SCS use for CRSwNP and vice versa).

The proportions of patients who had surgery and/or SCS CRSwNP listed in Section 3.5.2 will be summarised and analysed using the Cochran–Mantel–Haenszel (CMH) test stratified by region, CCI [redacted]. The results of the analyses will be presented as odds ratios with associated 95% CI and 2-sided p-value.

The total number of courses of SCS use for CRSwNP will be analysed using a negative binomial model. All courses of SCS use during the double blind period defined in Section 3.1.1 will be included in the analysis. The response variable in the model will be the number of courses of SCS use for CRSwNP for each patient. The model will include covariates of treatment group, region, CCI [redacted] and prior use of SCS for CRSwNP (yes vs no) as covariates. The logarithm of the patient's corresponding follow-up time will be used as an offset variable in the model. Marginal standardization methods will be used on the model estimates for all negative binomial analyses, unless otherwise specified.

As supportive analyses, the proportion of patients who potentially need surgery for CRSwNP (NPS≥5 and bi-weekly mean NBS≥1.5) during the DB period will be analysed using a similar CMH test as above.

At the final DBL, additional descriptive summary for surgery and SCS use for CRSwNP in the OLE period will be provided based on the OLE analysis set.

4.2.7.3 Nasal polyp symptom diary

The bi-weekly mean of NPSD (individual components) as well as the TSS and the corresponding changes from baseline will be summarised by treatment group and by time point using descriptive statistics for all patients in the PFAS.

The changes from baseline will be analysed using a similar ANCOVA analysis as described for the co-primary endpoints in Section 4.2.5.1. The ANCOVA analysis for the TSS and components of NPSD other than NBS and DSS will be done for the primary estimand only.

4.2.7.4 University of Pennsylvania smell identification test score

The UPSIT scores will be summarised and analysed overall as well as separately for men and women.

The change from baseline in UPSIT total score will be analysed using a similar ANCOVA analysis as the primary endpoints described in Section 4.2.5.1 for the primary estimand only. Shift table of the olfactory diagnosis categories between the baseline and Week 56 will be generated separately.

4.2.7.5 Sinus CT: Sinus severity score and Zinreich score

The sinus severity score (across all sinuses) and Zinreich total score and their changes from baseline will be summarised by treatment group and by time point using descriptive statistics for all patients in the PFAS.

The change from baseline in sinus severity score and Zinreich total score will be analysed using using a similar ANCOVA analysis as described in Section 4.2.6.2.

4.2.7.6 Short form 36-item health survey, version 2

For the SF-36v2, the eight subscale scores, the two component scores and their changes from baseline will be summarised by treatment group and by visit using descriptive statistics for all patients in the PFAS.

The responder (as defined in Section 3.5.7) rate at Week 24 and Week 56 will be summarised and analysed using the logistic regression with treatment arm, baseline from the corresponding SF-36v2 item, region, CCI as covariates. Observed data will be used; subjects with missing or non-evaluable observations at the timepoint of interest are counted as non-responders.

4.2.8 Pharmacokinetic and immunogenicity variables

PK and immunogenicity data will be collected during the double blinded period and the OLE period according to the SoA Table 2 and Table 3, respectively, in the CSP. PK and immunogenicity data from the DB period will be descriptively summarised at the primary DBL, while additional data from the OLE period will be reported at the final DBL.

4.2.8.1 Analysis of pharmacokinetic variables

Benralizumab serum concentrations will be summarised by using descriptive statistics during the DB period using the PK analysis set, and during the OLE period using the OLE analysis set. In addition, geometric mean, coefficient of variation (CV), lower (Q1), and upper (Q3) quartiles will also be reported.

Serum concentrations, that are below the LLOQ (<LLOQ) or if there are missing values, will be handled as follows:

- All values <LLOQ will be set to LLOQ/2. Missing values will be set to "NA".
- At a time point where mean is <LLOQ, Not Applicable (N/A) will be written in the field for SD, CV%, and 95% CI. The maximum value will be reported from individual data. If median is <LLOQ then shall be reported as <LLOQ, otherwise reported from individual data. The minimum value will be reported as <LLOQ.
- If all values are <LLOQ at a time point, no descriptive statistics will be calculated for that time point. Not Applicable (N/A) will be written in the field for SD, CV%, and 95% CI, and <LLOQ will be written in fields for the mean, geometric mean, minimum, median and maximum.

The LLOQ of Benralizumab in serum will be 3.86 ng/mL.

4.2.8.2 Analysis of immunogenicity variables

Anti-drug antibody assessment will be conducted and analysed as per the details in Appendix 8.7.

4.2.9 Exploratory efficacy variables

All exploratory efficacy endpoints will be collected during the double blinded period according to the SoA Table 2 in the CSP, except asthma exacerbation will be further collected during the OLE period (CSP Table 3). Therefore following analyses specified in Section 4.2.9.1 to Section 4.2.9.5 only apply to the double blind period at the primary DBL. Asthma exacerbation during the OLE period will descriptively summarized at final DBL.

4.2.9.1 Asthma control questionnaire 6

The observed mean ACQ-6 and the changes from baseline will be summarised by treatment group and by visit using descriptive statistics for all patients in PFAS in DB period.

The changes from baseline in mean ACQ-6 will be analysed under the treatment policy estimand where all data observed through Week 56 is included regardless of treatment discontinuation, surgery or SCS use for CRSwNP. The analysis will be conducted using ANCOVA for repeated measures (assuming missing at random) with treatment arm, baseline ACQ-6 scores, visit, region, CCI [redacted] and treatment×visit as covariates. See Appendix 8.5 for more details.

The ACQ-6 responder at Week 24 and at Week 56 will be analysed using the logistic regression with treatment arm, baseline ACQ-6, region, CCI [redacted] as covariates.

Observed data will be used; subjects with missing or non-evaluable observations at the timepoint of interest are counted as non-responders.

The number and proportion of responders and the number and proportion of patients by asthma control status will be summarised by treatment group for all patients in PFAS.

4.2.9.2 Asthma exacerbation rate

The number of exacerbations associated with SCS use, hospitalisation, ER visit, and urgent care visit will be summarised by treatment group during DB period for all patients in PFAS, and during OLE period for patients included in OLE analysis, respectively.

The annual exacerbation rate (as defined in Section 3.7.2) during the DB period will be analysed using a negative binomial model. The response variable in the model will be the number of asthma exacerbations experienced by a patient during the DB period. The model will include covariates of treatment group, region, CCI [redacted] and number of prior exacerbations as covariates. The logarithm of the patient's corresponding follow-up time during the DB period will be used as an offset variable in the model to adjust for patients having different exposure times during which exacerbations could occur. Marginal standardisation methods will be used on the model estimates for all negative binomial analyses, unless otherwise specified.

4.2.9.3 PGI-S and PGI-C

PGI-C response categories are described in in Section 3.7.3. The number and percentage of patients in each category of PGI-S and PGI-C responses, as well as PGI-C improvement categories will be summarised by treatment group and by time point during the DB period using the PFAS. Non-rescued subjects whose post-baseline observations are all missing are excluded from this analysis. Data after surgery for CRSwNP are imputed by WP and data after SCS use for CRSwNP are imputed by WOCF.

4.2.9.4 Healthcare resource utilisation

The following unplanned/unscheduled healthcare resource use (HRU) due to (a) nasal polyps, (b) asthma exacerbation (c) other reasons will be collected during the DB period.

- General and intensive care hospitalisations and lengths of stay
- Emergency room visits
- Urgent care visits

The frequency, crude rate (as defined in Section 3.7.4), and annualised rate (defined as *crude rate * 365.25*) of the HRU listed above and the total number of day hospitalised during the DB period will be summarised by reasons and treatment group using descriptive statistics for all patients in the PFAS.

4.2.9.5 Biomarkers

Biomarker values including blood eosinophils CCI as well as changes from baseline will be summarised by standard summary statistics for all patients in the FAS. Blood eosinophils will be presented for the DB period at the primary DBL and for the OLE period at the final DBL. CCI is collected during the DB period only so will be reported at the primary DBL.

Subgroup analyses by quartiles of blood eosinophils will be conducted as mentioned in Section 4.2.5.3 in order to assess the impact of blood eosinophils on efficacy endpoints. In addition, the associations between the efficacy endpoints and baseline blood eosinophil counts CCI will be evaluated by Loess plots against primary endpoint variables NPS and NBS for all patients in the PFAS.

The analyses of other exploratory biomarkers may be conducted and reported outside the CSR.

4.2.10 Safety outcome variables

All safety variables will be summarised by treatment group using descriptive statistics as appropriate. Adverse events (AEs) will be summarised during DB period using safety analysis set, as well as during OLE period (as defined in Section 3.1.1) using the OLE analysis set. All available safety data during the DB period and selected AE endpoints during the OLE period will be reported at the primary DBL. Remaining safety data on AE, laboratory, vital sign from the OLE period will be summarised and reported at the final DBL.

4.2.10.1 Adverse events

Summary of adverse events over the DB period

AEs will be summarised over the DB period by treatment group for all patients in the safety analysis set. An overview of AEs will present for each treatment group the number and percentage of patients with any AE, serious AEs (SAEs), SAEs with outcome of death, and AEs leading to discontinuation of IP (DAEs).

Adverse events, AEs with outcome of death, SAEs, DAEs, and possibly related AE (assessed by investigators) will be summarised by SOC and PT assigned to the event by MedDRA. For each PT, the number and percentage of patients reporting at least one occurrence will be presented, i.e., for a patient multiple occurrences of an AE will only be counted once.

Additional summaries of all AEs and common AE (PT Frequency of ≥ 3% in any treatment arm) will be presented by PT. AEs will also be summarised by PT and maximum intensity. If a patient reports multiple occurrences of the same AE within the same analysis 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 (MedDRA High Level Term of injection site reactions), hypersensitivity (standardised MedDRA query (SMQ) [Narrow] of hypersensitivity), malignancy (SMQ of malignant tumours) and helminth

infections (MedDRA High Level Group Term of helminthic disorders) will be summarised by preferred term.

The crude rate and exposure adjusted incidence rate (EAIR) (PHUSE 2017) will be reported for any AE, common AE, SAE, AEs leading to IP discontinuation, adverse event of interest. Both crude rate and EAIR will be expressed in terms of events per 100 patient-years. Crude rate per 100 patient-years is defined as number of subjects with AEs divided by total patient years across all subjects in a given treatment group, multiplied by 100. EAIR per 100 patient-years is calculated based on each AE term, defined as number of subjects with an AE divided by the total number of days at risk for that AEs across all subjects in given group, multiplied by 365.25 multiplied by 100. Total number of days at risk for an AE is defined as the duration from the first dose of IP to the start date of the first AE for the subjects with that AE, or the duration from the first dose of IP to end of double-blind period for subject without that AE. In case AE onset date is completely missing, time at-risk will be imputed as 1 day assuming the AE starts on the same date of the first dose of study treatment.

EAIR comparisons between treatment groups (i.e. risk difference in exposure-adjusted incidence rates) and 95% CI based on Miettinen and Nurminen (M-N) method (Liu et al. 2006, Miettinen and Nurminen 1985) will be reported. The risk difference scale is chosen to directly reflect the difference in magnitude of patients that could be affected by a risk.

For common AEs and AEs of interest (injection site reactions, hypersensitivity, malignancy and helminth infections), comparisons between treatment arms will be presented graphically in a panel plot with the left panel showing a dot plot of the proportion of patients reporting each AE by treatment arm, and the right panel showing a corresponding forest plot of estimated risk differences with M-N associated 2-sided 95% CIs.

Key patient listings for AEs with outcome of death, SAEs, or DAEs will be presented.

Summary of adverse events over the OLE period

The analysis of AEs during the DB period as described above will be repeated for the OLE period in the OLE analysis set. AE overview and summaries by SOC and PT for AE, AEs with outcome of death, SAEs and DAEs over the OLE period will be reported at the primary DBL, while remaining AE summaries for OLE period will be reported at the final DBL.

4.2.10.2 Laboratory data

All protocol-specified continuous laboratory parameters will be summarised descriptively for absolute value at each time point by treatment group, together with the corresponding changes from baseline. This will be provided for the DB period at primary DBL as well as for the OLE period at final DBL. All parameters will be summarised in SI units unless otherwise specified.

Central laboratory reference ranges will be used for the identification of abnormalities, and a shift table will be produced for each laboratory parameter to display low, normal, and high

values. The shift tables will present baseline and maximum/minimum post-baseline value, as applicable for each parameter and will include patients with both baseline and post-baseline data. Shift tables will use all available data during DB period and during OLE period, respectively.

4.2.10.3 Vital signs

Vital signs will be summarised using descriptive statistics for absolute value at each time point by treatment group, together with the corresponding changes from baseline during DB period and during OLE period, respectively.

Baseline to maximum post-baseline and baseline to minimum post-baseline value shift tables will be generated, as applicable for each parameter and will include patients with both baseline and post-baseline data.

4.2.10.4 ECG

A shift table will be produced to display the overall ECG evaluation of normal, abnormal – not clinically significant, abnormal – clinically significant, and not done. For this purpose, borderline (also recorded on the eCRF) will be grouped with normal. Shift tables will present baseline and last observed post-baseline value over safety analysis set during the DB period.

4.2.10.5 Physical examination

Any new or aggravated clinically relevant abnormal medical finding at a physical examination as compared with the baseline assessment will be reported as an AE unless unequivocally related to the disease under study and will be summarised as described in Section 4.2.10.1.

4.2.11 Impact on analyses due to COVID-19 pandemic

The study is conducted during the COVID-19 worldwide pandemic. Patient dosing, scheduled visits, and nasal endoscopies all may become difficult or impossible to perform according to protocol for patients being followed up in the study.

Efforts are ongoing to collect outstanding data via alternative means where possible, when on-site visits cannot be performed. Additionally, protocol deviations, including doses or visits missed due to COVID-19 related protocol deviations will be described separately in the CSR. These deviations will be identifiable in the database with a ‘COVID’ prefix.

5. INTERIM ANALYSES

No interim analysis is planned for this study.

6. CHANGES OF ANALYSIS FROM PROTOCOL

Not applicable.

7. REFERENCES

Bachert et al 2016
Bachert C, Mannent L, Naclerio RM, Mullol J, Ferguson BJ, Gevaert P, et al. Effect of subcutaneous dupilumab on nasal polyp burden in patients with chronic sinusitis and nasal polyposis: A randomised clinical trial. JAMA. 2016;315 (5):469-79.

Bachert et al, 2017
Bachert C, Sousa AR, Lund VJ, Scadding GK, Gevaert P, Nasser S, Durham SR, Cornet ME, Kariyawasam HH, Gilbert J, Austin D, Maxwell AC, Marshall RP, Fokkens WJ, Reduced need for surgery in severe nasal polyposis with mepolizumab: randomised trial, Journal of Allergy and Clinical Immunology (2017). doi: 10.1016/j.jaci.2017.05.044.

Bachert et al, 2019
Bachert C, Han JK, Desrosiers M, Hellings PW, Amin N, Lee SE, et al. Efficacy and safety of dupilumab in patients with severe chronic rhinosinusitis with nasal polyps (LIBERTY NPSINUS-24 and LIBERTY NP SINUS-52): results from two multicentre, randomised, double-blind, placebo-controlled, parallel-group phase 3 trials, Lancet. 2019;394 (10209):1638-1650

CHMP 2010
Committee for Medicinal Products for Human Use (CHMP). Guideline on Missing Data in Confirmatory Clinical Trials 2, July 2010. EMA/CPMP/EWP/1776/99 Rev. 1.

EMA/CHMP/EWP 2010
Guideline on Missing Data in Confirmatory Clinical Trials 2 July 2010. EMA/CHMP/EWP/1776/99 Rev. 1.

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.

Juniper et al 2005
Juniper EF, Svensson K, Mörk AC, Ståhl E. Measurement properties and interpretation of three shortened versions of the asthma control questionnaire. Respir Med 2005;99(5):553-8.

Juniper et al 2006
Juniper EF, Bousquet J, Abetz L, Bateman ED. Identifying ‘well-controlled’ and ‘not well-controlled’ asthma using the Asthma Control Questionnaire. Respir Med 2006;100:616- 621.

Kern et al 2018
Kern RC, Stolovitzky JP, Silvers SL, et al. A phase 3 trial of mometasone furoate sinus implants for chronic sinusitis with recurrent nasal polyps. Int Forum Allergy Rhinol 2018;8:471-481.

Likness et al 2014
Likness MM, Pallanch JF, Sherris DA, Kita H, Mashtare TL Jr, Ponikau JU. Computed tomography scans as an objective measure of disease severity in chronic rhinosinusitis. Otolaryngol Head Neck Surg. February 2014;150 (2):305–311.

Little et al 2010
Little RJA, Farrar JT, Frangakis C, Hogan JW, Molenberghs G, Murphy SA, et al. 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.

Liu et al 2006
Liu GF, Wang J, Liu K, Snavely DB. Confidence intervals for an exposure adjusted incidence rate difference with applications to clinical trials. Statistics in medicine. 2006 Apr 30;25(8):1275-86.

Miettenen and Nurminen 1985
Miettenen O & Nurminen M. Comparative analysis of two rates. Statistics in Medicine 1985; 4(2):213 - 226.

Okushi et al 2013
Okushi T, Nakayama T, Morimoto S, Arai C, Omura K, Asaka D, Matsuwaki Y, Yoshikawa M, Moriyama H, Otori N. A modified Lund-Mackay system for radiological evaluation of chronic rhinosinusitis. Auris Nasus Larynx 2013;40:548–553.

Pallanch et al 2013
Pallanch J, Yu L, Delone D, Robb R, Holmes DR, et al. 3-D volumetric computed tomographic scoring as an objective outcome measure for chronic rhinosinusitis: Clinical correlations and comparison to Lund-Mackay scoring. Int Forum Allergy Rhinol. 2013;3 (12): 963-972.

PHUSE 2017
Analysis and Displays Associated with Safety Topics of Interest: Focus on Phase II to IV Clinical Trials and Integrated Summary Documents

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

Shih et al 2023
Shih VH, Slagle AF, Ivanescu C, Flore G, Meyers O, Kreindler J, Martin UJ, Werkström V. Psychometric validation and meaningful change thresholds of the new Nasal Polyposis

Symptom Diary. Ann Otol Rhinol Laryngol. Published online June 4, 2023. doi:10.1177/00034894231177769.

Stjarne et al 2006
Stjarne P, Mosges R, Jorissen M, Passali D, Bellussi L, Staudinger H, et al. A randomised controlled trial of mometasone furoate nasal spray for the treatment of nasal polyposis. Arch Otolaryngol Head Neck Surg 2006;132:179–185.

8. APPENDIX

8.1 Missing data descriptions

Some patients may drop out prematurely, potentially leading to missing data. The amount of missing data is minimized as the protocol encouraged patients to attend study visits until they complete the overall study duration after they discontinue from randomized therapy. This section summarizes how we will describe the pattern of, and reasons for, missing data. It will also describe how we plan to account for missing data to assess the robustness of the treatment effect under different underlying missing data assumptions.

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 will be presented using Kaplan Meier plot. Dependent on these outputs additional exploratory analyses covering the DB period may be produced to further understand the pattern of missing data.

8.2 Supplementary analysis: primary estimand with WP for surgery for CRSwNP only

As mentioned in Section 4.2.5.1, the primary analyses will be based on primary estimand. A supplementary analysis with a different strategy for the SCS use for CRSwNP will be conducted. This analysis will use a composite strategy for surgery only and will impute the WP only for patients with any surgery for CRSwNP at the time of their first surgery for CRSwNP. The occurrence of SCS use for CRSwNP will not be considered rescue and a treatment policy strategy will be used instead. Data prior to any surgery for CRSwNP will be included regardless of the use of SCS for CRSwNP. The same multiple imputation strategy as described in Section 4.2.5.1 will be carried out, but patients who receive SCS for CRSwNP but never have surgery for CRSwNP will be included in the MI step for non-rescued patients. See Table 14 for random seeds to be used for this supplementary analysis.

8.3 Sensitivity analysis: primary estimand assuming dropout reason-based multiple imputation approach

As mentioned in Section 4.2.5.1, the primary analyses will be based on primary estimand. If a patient has a surgery and/or a course of SCS for CRSwNP, the data will be censored at the time of the first surgery and/or SCS use for CRSwNP, and the worst-possible value (after

surgery for CRSwNP) or the patient's worst-observed value (after SCS for CRSwNP) will be imputed from that point until Week 56. A sensitivity analysis under the primary estimand with missing data imputed based on different withdrawal reasons will be conducted. As in Section 4.2.5.1, a composite strategy will be used for patients who have surgery and/or receive SCS for CRSwNP. For patients that do not undergo surgery or receive SCS for CRSwNP, missing data will be imputed using controlled sequential multiple imputation methods based on pattern mixture models (EMA/CHMP/EWP 2010). See Table 14 for random seeds to be used for this sensitivity analysis. This model will assume that some pre-specified subset of patients who withdraw from the study have correlations with future unobserved visits similar to patients in the placebo arm. The assumptions that will be used to impute the missing data are as follows:

(a) Missing at Random (MAR): Assumes that the trajectory for patients who dropped out in each arm is similar to those observed in their own treatment arm. (The primary analysis already implements this approach. It will not be repeated.)

(b) Dropout Reason-based Multiple Imputation (DRMI): Assumes that the trajectory for patients in the benralizumab arm who dropped out for a treatment related reason and/or severe non-compliance of protocol is similar to that of patients in the placebo arm, whereas the remaining patients who has dropped out are imputed assuming MAR.

A summary of reasons for patients withdrawing from the benralizumab treatment arm and the corresponding treatment arm used to calculate the imputation under MAR, and DRMI are given in Table 12 .

Table 12 Parameters for calculating the imputation under MAR, and DRMI

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
Patient lost to follow up	Benralizumab	Benralizumab
Patient decision	Benralizumab	Based on review prior to study unblinding
COVID-19	Benralizumab	Benralizumab
Other	Benralizumab	Based on review prior to study unblinding

Note: Patients in the placebo arm are imputed using non-missing values in placebo arm.
*Development of study-specific criteria for discontinuation are based on the following: Anaphylactic reaction to the investigational product requiring administration of epinephrine; Development of helminth parasitic infestations requiring hospitalisation; A respiratory-related event requiring mechanical ventilation.

Some reasons for withdrawal are clearer to determine as treatment related (AEs, death, development of study-specific discontinuation criteria) or non-treatment related (patients lost

to follow up, eligibility criteria not fulfilled). Other reasons are less clear such as patient decision and 'Other'; a review of each patient who withdraws from the study will therefore be carried out prior to unblinding the study. Based on this review the default assumptions for DRMI as described in b), and Table 12 may be changed. A list of these patients and the assumptions made under DRMI will be documented prior to unblinding of the study.

8.4 Supplementary analysis under the effectiveness estimand

A supplementary analysis will be evaluated under the effectiveness estimand. This analysis includes data collected up until the time of discontinuation from study treatment and excludes data if a patient has a surgery and/or SCS use for CRSwNP. Because these data will be analysed by ANCOVA for repeated measures which implicitly accounts for missing data under the MAR assumption, no imputation will be done after discontinuation from treatment or after rescue by surgery and/or SCS use for CRSwNP. This analysis evaluates the effect of initially assigned randomised treatment as long as the patient remains on treatment without being rescued by a significant alternate course of therapy such as surgery or SCS for CRSwNP.

The analysis will be conducted using ANCOVA for repeated measures with treatment arm, baseline scores (baseline total NPS for NPS model and baseline NBS for NBS model), visit, region, CCI [redacted] and treatment×visit as covariates. The variance-covariance matrix will assume to be unstructured. If the procedure does not converge, then the Toeplitz, first-order autoregressive, and compound symmetric structure of variance-covariance matrices will be tried in that order. The estimate of the treatment effect will be based on a contrast from the repeated measures ANCOVA model.

8.5 Supplementary analysis under the treatment policy estimand

The co-primary endpoints will be evaluated under the treatment policy estimand during DB period. The treatment policy estimand includes data collected after discontinuation of study treatment regardless of rescue by surgery or SCS use for CRSwNP. Because these data will be analysed by ANCOVA for repeated measures which implicitly accounts for missing data under the MAR assumption, no imputation will be conducted for the missing data. The treatment policy estimand evaluates the treatment effect at the planned endpoint of the study, regardless of the treatments that patients actually received.

The analysis will be conducted using ANCOVA for repeated measures with treatment arm, baseline scores (baseline total NPS for NPS model and baseline NBS for NBS model), visit, region, CCI [redacted] and treatment×visit as covariates. The variance-covariance matrix will assume to be unstructured. If the procedure does not converge, then the Toeplitz, first-order autoregressive, and compound symmetric structure of variance-covariance matrices will be tried in that order. The estimate of the treatment effect will be based on a contrast from the repeated measures ANCOVA model.

It is noted that if the primary analysis is statistically significant, it is not necessarily expected that all sensitivity and supplementary analyses will also give statistically significant results. If the results of the sensitivity/supplementary 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 that are made about missing data in the different analyses will be considered and described in the clinical study report.

8.6 Handling incomplete dates for adverse events and medications

8.6.1 Partial dates for adverse events and prior/concomitant medication

Dates missing the day, or both the day and month of the year will adhere to the following conventions to classify AEs and to classify prior/concomitant medications.

8.6.1.1 Partial dates for adverse events

Onset date of AEs

If only the day of the AE onset date is missing, the missing day will be set to:

First day of the month that the event occurred, if the onset YYYY-MM is after the YYYY-MM of first study treatment

The day of the first study treatment, if the onset YYYY-MM is the same as YYYY-MM of the first study treatment

The date of informed consent, if the onset YYYY-MM is before the YYYYMM of the first treatment

If both of the day and month of the onset date of an AE are missing, the onset date will be set to:

January 1 of the year of onset, if the onset year is after the year of the first study treatment.

The date of the first treatment, if the onset year is the same as the year of the first study treatment

The date of informed consent, if the onset year is before the year of the first treatment

Resolution date of AEs

If only the day of the AE resolution date is missing, the missing day will be set to:

The last day of the month of the occurrence. If the patient died in the same month, then set the imputed date as the death date

If both day and month of the resolution date of an AE are missing, the date will be set to:

December 31 of the year of occurrence. If the patient died in the same year, then set the imputed date as the death date

8.6.1.2 Partial dates for prior/concomitant medication

Start date of prior/concomitant medication

If only the day is missing, then the start date of a therapy will be set to the first day of the month that the event occurred.

If both the day and month are missing, then the start date of a therapy will be set to January 1 of the year of onset.

If the start date of a therapy is completely missing, then the date will be set as following.

- If 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 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.

End date of prior/concomitant medication

If only the day is missing, then the end date of a therapy will be set to the last day of the month of the occurrence.

If both the day and month are missing, then the end date of a therapy will be set to December 31 of the year of occurrence.

If the end date of a therapy is completely missing, then the date will be set as following

- If 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 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.

8.7 Analysis plan for ADA data

Serum samples for ADA assessments will be conducted utilizing a tiered approach (screen, confirm, titre) and ADA data will be collected at scheduled visits shown in SOA Table 2 and Table 3 in the CSP. ADA result from each sample will be reported as either positive or negative. If the sample is positive, the ADA titre will be reported as well. In addition, the presence of neutralizing antibodies (nAb) will be tested in ADA-positive samples using a ligand binding assay.

In general, patients with a missing baseline ADA assessment will be assumed to be ADA negative at baseline as a conservative approach to ensure that all patients are included in all analyses. If a positive ADA titre result is reported as <15, then the titre will be imputed as 15

for titre summaries. ADA results from samples collected post-dose instead of pre-dose on an IP administration day are considered unreliable and should be excluded from all derivations.

For all patients, the following variables will be derived based on the data collected during the DB period. Results will be listed separately.

- Patients who are ADA positive at any time during the study, including baseline and/or post-baseline (also generally referred to as ADA positive). The proportion of ADA-positive patients in a population is known as ADA prevalence.
- Patients who are ADA negative at all assessments, including baseline and post-baseline (also generally referred to as ADA negative).
- Patients who are ADA positive at baseline only.
- Patients who are ADA positive at baseline and at least one post-baseline assessment.
- Treatment-emergent ADA positive (referred to as ADA incidence). A positive post-baseline result and either of the following statements holds:
    - Baseline is ADA negative and at least one post-baseline assessment is ADA positive. This is called treatment-induced ADA positive.
    - Baseline is ADA positive, and the baseline titre is boosted by greater than the variability of the assay (i.e. ≥ 4-fold increase) at ≥1 post-baseline timepoint. This is called treatment-boosted ADA positive.
- Non-treatment-emergent ADA positive: Patients who have a baseline and at least one post-baseline ADA assessment, and who are ADA positive but not fulling the conditions above for treatment-emergent ADA positive.
- Patients who are persistently ADA positive, which is defined as ADA negative at baseline and having at least 2 post-baseline ADA positive measurements with ≥ 16 weeks between first and last positive, or an ADA positive result at the last available post baseline assessment.
- Patients who are transiently ADA positive, defined as ADA negative at baseline and at least one post-baseline ADA positive measurement and not fulfilling the conditions for persistently positive.
- Patients who are treatment-emergent ADA positive with maximum post-baseline titre > median of maximum post-baseline titres. The median of maximum post-baseline titres will be calculated based on the maximum post-baseline titre of each ADA positive patient within each treatment group.
- nAb prevalence; defined as nAb positive at any visit including baseline and/or post-baseline (also referred to as nAb positive).
- nAb incidence; defined as nAb negative at baseline (or ADA negative at baseline) and nAb positive at any post-baseline visit. Patients who are ADA-negative at baseline are included to ensure that all patients who are nAb positive for the first-time post-baseline

satisfy this definition, given that all patients who are ADA negative at baseline do not have a nAb result reported.

Similar ADA responses described for the DB period will be derived for the OLE; specific considerations for the OLE period are provided as follows. Baseline value for the OLE period is based on the last available value collected prior to the 1st dose in OLE, not from the beginning of the DB period.

- Newly persistently positive: Patients who are ADA negative at baseline and positive at ≥2 post-baseline assessments (with ≥16 weeks between first and last positive) or positive at last post-baseline assessment.
- Stable persistently positive: Patients who are ADA positive at baseline and ≥2 post-baseline assessments (with ≥16 weeks between first and last positive).
- Newly treatment-emergent ADA-positive (ADA incidence): Patients who are either newly treatment-induced positive or treatment-boosted ADA positive.
    - Newly treatment induced ADA positive: Patients who are not ADA positive at any time in DB period but are ADA positive post-baseline for the first time in the OLE period.
    - Treatment-boosted ADA-positive: At least 1 of patient's post-baseline titres in the OLE period is ≥ 4-fold increase of the baseline titre.
- Non-treatment-emergent ADA positive: Subjects who have a baseline and at least one post-baseline ADA assessment in the OLE period, and who are ADA positive but not fulling the conditions above for newly treatment-emergent ADA positive.
- Transiently positive: Patients having at least 1 post-baseline ADA-positive assessment and not fulfilling the conditions of either newly persistently positive or stable persistently positive.

Similarly, ADA responses described for the DB period will be derived for the on-study period; specific considerations for the on-study period are provided as follows. Baseline value for the on-study period is based on the last available value collected prior to the 1st dose in DB period.

- Patients who are persistently ADA positive
    - 6-month persistently ADA positive: Subjects who are ADA negative at baseline and having ≥ 2 post-baseline ADA positive assessments (with ≥ 6 months between first and last positive) or positive at the last available post baseline assessment.
    - 12-month persistently ADA positive: Subjects who are ADA negative at baseline and having ≥ 2 post-baseline ADA positive assessments (with ≥ 12 months between first and last positive) or positive at the last available post baseline assessment.

The responses above will be summarised as counts and percentages by treatment group over the DB period, OLE period, and the on-study period as appropriate. The maximum ADA titre will also be summarised for patients in each of the ADA positive response categories listed above. The maximum titre will be derived based on all available ADA titres reported for each patient, including any unscheduled assessments.

ADA response (positive or negative) and titre will be summarised at baseline and at all scheduled post-baseline visits by treatment group using derived visit windows (refer to Section 3.1.2 for detailed definition of visit windows) in each study period. In the event a patient has more than one result within a given visit window, the maximum ADA titre will be used in the by-visit summary. In addition, the ADA response will be presented cumulatively. The cumulative ADA response is positive for a specific visit if a positive ADA result is detected at any time point up to and including the specific visit. If all ADA result are negative up to the specific visit, then the cumulative ADA response is negative for that visit. A summary of the number and percentage of patients who are ADA positive at a post-baseline assessment for the first time by visit will also be presented. A line plot of serum concentrations over time will be provided by ADA categories.

The proportion of patients with positive nAb response will be summarised by visit. Key patient information will be listed for patients with positive ADA results, including ADA status, nAb status, titer, benralizumab serum concentration, and blood eosinophil level.

All analyses will be conducted by treatment group on the safety analysis set for the DB period and the on-study period, and on the OLE analysis set for the OLE period, unless otherwise specified. All ADA results will be listed. Results from the DB period will be reported at the primary DBL while remining results from OLE period and the on-study period will be reported at the final DBL.

ADA and blood eosinophil levels

Eosinophil levels will be summarised by visit and for the following ADA response categories of patients:

For the DB period: ADA negative, ADA positive, treatment-emergent ADA positive, non-treatment-emergent ADA positive, ADA persistently positive, ADA transiently positive, nAb-positive, and treatment-emergent ADA positive with maximum post-baseline titer > median of maximum post-baseline titre.

For the OLE period: ADA negative, ADA positive, newly treatment-emergent ADA positive, non-treatment-emergent ADA positive, newly ADA persistently positive, stable ADA persistently positive, nAb-positive, ADA transiently positive, and treatment-emergent ADA positive with maximum post-baseline titer > median of maximum post-baseline titre.

For the on-study period: ADA negative, ADA positive, treatment-emergent ADA positive, non-treatment-emergent ADA positive, ADA persistently positive, ADA 6-month persistently positive, ADA 12-month persistently positive, ADA transiently positive, treatment-emergent

ADA positive with maximum post-baseline titer > median of maximum post-baseline titre and nAb-positive.

A line plot of blood eosinophil levels by visit and ADA status will also be presented.

ADA and efficacy

The effects of ADA on the co-primary endpoints calculated during the DB period will be evaluated through summary statistics by treatment group and ADA status (i.e. ADA negative, ADA positive, treatment-emergent ADA positive, non-treatment-emergent ADA positive, ADA persistently positive, ADA transiently positive, nAb-positive, and treatment-emergent ADA positive with maximum post-baseline titer > median of maximum post-baseline titre). Due to the expected small number of ADA positive patients in the placebo group, no formal statistical comparisons of benralizumab versus placebo by ADA status (positive/negative) are planned.

ADA and safety

Adverse events during the DB period and the OLE period will be summarised by treatment group and ADA status (i.e. ADA negative, (newly) TE-ADA positive, and non-TE-ADA positive) over safety analysis set and OLE analysis set, respectively. The potential impact of ADA on hypersensitivity will also be assessed.

ADA and PK

Similarly as blood eosinophil levels, benralizumab serum concentrations will be summarised by treatment group, visit and ADA status for patients in the PK analysis set for the DB period, OLE period, as well as on-study period.

8.8 Worst possible scores

The worst possible scores as given in the table below for each endpoint will be used for analyses in which the WP imputation method is used after surgery for CRSwNP.

Table 13 Worst possible scores

Endpoint	Worst Possible Score
NPS	8
NBS/DSS/NPSD individual item	3
SNOT-22 total score	110
SNOT-22 individual item	5
LMS	24
Zinreich total score	54
Sinus severity score	100
TSS	24
UPSIT	0
ACQ-6	6
SF-36 components	
PCS	7.32

Endpoint	Worst Possible Score
MCS	5.79
PF	19.26
RP	21.23
BP	21.68
GH	18.95
VT	22.89
SF	17.23
RE	14.39
MH	11.63
PGI-S	5 (very severe)
PGI-C	7 (much worse)

8.9 Random seeds

The starting random seeds to be utilised in the MCMC and monotone regression MI are provided in the table below. For each successive visit imputed in the monotone regression MI (step 2 in Section 4.2.5.1), the random seed increments by 1. For example, imputation of Week 8 of NPS in monotone regression MI uses a seed of 288263, imputation of Week 16 of NPS uses 288264, and so on.

Table 14 Random seeds for multiple imputation analyses

Endpoint	Primary/COVID MCMC	Primary /COVID MI	WP after Surgery MCMC	WP after Surgery MI	DRMI MCMC	DRMI MI
NPS	97348	288263	504986	438656	454805	536844
NBS (NPSD)	72766	714699	482798	804161	88386	459794
DSS (NPSD)	875052	519837	820852	680071	945412	172159
Nasal congestion (NPSD)	585248	518131	NA	NA	NA	NA
Runny nose (NPSD)	987759	285595	NA	NA	NA	NA
Postnasal drip (NPSD)	945217	86342	NA	NA	NA	NA
Headache (NPSD)	146002	930824	NA	NA	NA	NA
Facial pain (NPSD)	332834	604107	NA	NA	NA	NA
Facial pressure (NPSD)	280036	66016	NA	NA	NA	NA
Diff with sleeping (NPSD)	216628	431468	NA	NA	NA	NA
Diff with daily activities (NPSD)	483627	252059	NA	NA	NA	NA
TSS (NPSD)	135121	501498	NA	NA	NA	NA
SNOT-22	851498	608526	599733	263466	486856	513807
CT (LMS, SSS, ZS)	NA	NA	NA	NA	NA	NA
UPSIT	545559	925322	NA	NA	NA	NA
ACQ-6	655030	598573	NA	NA	NA	NA

8.10 Efficacy estimands

Statistical Category	Estimand[1]			SAP Section
	Endpoint (Population)	Intercurrent Event Strategy[1]	Population Level Summary[1] (Analysis)	
Primary objective: To evaluate the effect of benralizumab on nasal polyp burden and patient-reported nasal blockage				
Co-Primary/MCP	• CFB in endoscopic total NPS (severe CRSwNP patients with asthma) • CFB in bi-weekly mean NBS (severe CRSwNP patients with asthma)	Included in analysis regardless of treatment discontinuation; WP after surgery for CRSwNP, WOCF after SCS use for CRSwNP[3] (primary estimand)	Mean difference between interventions at Week 56[2] (LSMD from CFB ANCOVA following hybrid WP/WOCF and MI)	Section 4.2.5.1
Supplementary and Sensitivity	• CFB in endoscopic total NPS (severe CRSwNP patients with asthma) • CFB in bi-weekly mean NBS (severe CRSwNP patients with asthma)	• Included in analysis regardless of treatment discontinuation or SCS use for CRSwNP; WP after surgery for CRSwNP[4] • Disc of treatment assumed based on DRMI[3] • Remained adherent to treatment without surgery or SCS use for CRSwNP[5] (effectiveness) • Included in analysis regardless of treatment discontinuation; surgery or SCS use for CRSwNP[6], (treatment policy)	• Mean difference between interventions at Week 56[2] (LSMD from CFB ANCOVA following hybrid WP and MI) • Mean difference between interventions at Week 56[2] (LSMD from CFB ANCOVA following hybrid WP/WOCF and DRMI) • Mean difference between interventions at Week 56[2] (LSMD from CFB repeated measures ANCOVA) • Mean difference between interventions at Week 56[2] (LSMD from CFB repeated measures ANCOVA)	Appendix 8.2 Appendix 8.3 Appendix 8.4 Appendix 8.5
Key Secondary Objective: To evaluate the effect of benralizumab on sense of smell				
Secondary/MCP	CFB in bi-weekly mean DSS (severe CRSwNP patients with asthma)	Included in analysis regardless of treatment discontinuation; WP after surgery for CRSwNP, WOCF after SCS use for CRSwNP[3] (primary estimand)	Mean difference between interventions at Week 56[2] (LSMD from CFB ANCOVA following hybrid WP/WOCF and MI)	Section 4.2.6.1

Statistical Category	Estimand[1]			SAP Section
	Endpoint (Population)	**Intercurrent Event Strategy[1]**	**Population Level Summary[1] (Analysis)**	
Supplementary and Sensitivity	• CFB in bi-weekly mean DSS (severe CRSwNP patients with asthma)	• Included in analysis regardless of treatment discontinuation or SCS use for CRSwNP; WP after surgery for CRSwNP[4] • Disc of treatment assumed based on DRMI[3] • Remained adherent to treatment without surgery or SCS use for CRSwNP[5] (effectiveness) • Included in analysis regardless of treatment discontinuation; surgery or SCS use for CRSwNP[6], (treatment policy)	• Mean difference between interventions at Week 56[2] (LSMD from CFB ANCOVA following hybrid WP and MI) • Mean difference between interventions at Week 56[2] (LSMD from CFB ANCOVA following hybrid WP/WOCF and DRMI) • Mean difference between interventions at Week 56[2] (LSMD from CFB repeated measures ANCOVA) • Mean difference between interventions at Week 56[2] (LSMD from CFB repeated measures ANCOVA)	Appendix 8.2 Appendix 8.3 Appendix 8.4 Appendix 8.5
Key Secondary Objective: To evaluate the effect of benralizumab on sinus opacification				
Secondary/MCP	CFB in mean LMS (severe CRSwNP patients with asthma)	Included in analysis regardless of treatment discontinuation or SCS use for CRSwNP; WP after surgery for CRSwNP[4], (primary estimand)	Mean difference between interventions at EoDB/IPD (LSMD from CFB ANCOVA following WP at EoDB/IPD)	Section 4.2.6.2
Supplementary	CFB in mean LMS (severe CRSwNP patients with asthma)	Included in analysis regardless of treatment discontinuation; WP/WOCF after surgery[3] and SCS use for CRSwNP	Mean difference between interventions at EoDB/IPD (LSMD from CFB ANCOVA following WP/WOCF at EoDB/IPD)	Section 4.2.6.2
Key Secondary Objective: To evaluate the effect of benralizumab on disease specific health-related quality of life				
Secondary/MCP	CFB in mean SNOT-22 (severe CRSwNP patients with asthma)	Included in analysis regardless of treatment discontinuation; WP after surgery for CRSwNP, WOCF after SCS use for CRSwNP[3] (primary estimand)	Mean difference between interventions at Week 56[2] (LSMD from CFB ANCOVA following hybrid WP/WOCF and MI)	Section 4.2.6.3

Statistical Category	Estimand[1]			SAP Section
	Endpoint (Population)	**Intercurrent Event Strategy[1]**	**Population Level Summary[1] (Analysis)**	
Supplementary and Sensitivity	• CFB in mean SNOT-22 (severe CRSwNP patients with asthma)	• Included in analysis regardless of treatment discontinuation or SCS use for CRSwNP; WP after surgery for CRSwNP[4] • Disc of treatment assumed based on DRMI[3] • Remained adherent to treatment without surgery or SCS use for CRSwNP[5] (effectiveness) • Included in analysis regardless of treatment discontinuation, surgery or SCS use for CRSwNP[6] (treatment policy)	• Mean difference between interventions at Week 56[2] (LSMD from CFB ANCOVA following hybrid WP and MI) • Mean difference between interventions at Week 56[2] (LSMD from CFB ANCOVA following hybrid WP/WOCF and DRMI) • Mean difference between interventions at Week 56[2] (LSMD from CFB repeated measures ANCOVA) • Mean difference between interventions at Week 56[2] (LSMD from CFB repeated measures ANCOVA)	Appendix 8.2 Appendix 8.3 Appendix 8.4 Appendix 8.5
Safety Objective: To evaluate the safety and tolerability of benralizumab				
Safety	• Presence of AEs and SAE (all severe CRSwNP patients regardless of co-morbid asthma status)	Included in analysis regardless of treatment discontinuation[6] (treatment policy)	Categorical descriptive	Section 4.2.10.1

PFAS = Primary Full Analysis Set; MCP = Multiple comparisons procedure; CRSwNP = Chronic rhinosinusitis with nasal polyps; SCS = Systemic Corticosteroids; NPS = Nasal Polyps Score; NBS = Nasal Blockage Score; DSS = Difficulty with Sense of Smell; WP = Worst Possible; WOCF = Worst Observation Carried Forward; MI = Multiple Imputation; LSMD = Least Squares Mean Difference; CFB = Change from baseline; ANCOVA = Analysis of Covariance; AE = Adverse event; OLE = Open-label Analysis Set; EoDB = End of treatment during double-blind period; IPD = Investigational product discontinuation; LMS = Lund-Mackay score; SNOT-22 = SinoNasal Outcome Test, 22 item.

[1] All estimand attributes explicitly identified for primary and key secondary endpoints only.

[2] Week 56 is the multiplicity protected timepoint for co-primary and key secondary endpoints.

[3] Treatment Condition: Treatment with benralizumab versus placebo, regardless of compliance, where rescue with surgery and/or SCS use for CRSwNP indicates treatment failure.

[4] Treatment Condition: Treatment with benralizumab versus placebo, regardless of compliance, where rescue with surgery for CRSwNP indicates treatment failure.
[5] Treatment Condition: Treatment with benralizumab versus placebo, while on treatment and without rescue by surgery and/or SCS use for CRSwNP.
[6] Treatment Condition: Treatment with benralizumab versus placebo, regardless of compliance or occurrence of rescue by surgery and/or SCS use for CRSwNP.