Statistical Analysis Plan

Study Code D3252C00001

Edition Number Version 3

Date 19 August 2020

A Multicenter, Randomized, Double-Blind, Parallel Group, Placebo Controlled Phase 3 Efficacy and Safety Study Of Benralizumab in Patients with Severe Nasal Polyposis (OSTRO)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation or special term	Explanation
ACQ-6	Asthma Control Questionnaire 6
ADA	Anti-Drug Antibody(ies)
AE	Adverse Event
AERD	Aspirin exacerbated respiratory disease
ALT	Alanine aminotransferase
ANCOVA	Analysis of covariance
AST	Aspartate aminotransferase
ATC	Anatomical therapeutic chemical
BMI	Body mass index
CI	Confidence interval
CSP	Clinical study protocol
CSR	Clinical study report
CT	Computed tomography
DRMI	Dropout reason-based multiple imputation
DSS	Difficulty with Sense of Smell
EFU	Extended follow-up (EFU)
EOS	Eosinophil
EOT	End of treatment
ER	Emergency room
FAS	Full analysis set
GGT	Gamma-glutamyltransferase
HRQoL	Health-related quality of life
HRU	Healthcare resource utilization
IgE	Immunoglobulin E
INCS	Intranasal corticosteroids
IPD	Investigational product discontinuation
ITT	Intent-to-Treat
LOCF	Last observation carried forward
LSMEAN	Least squares mean
MAR	Missing at random
MCID	Minimally Clinically Important Difference
MedDRA	Medical dictionary for regulatory activities
MMRM	Mixed-effects model for repeated measures

Abbreviation or special term	Explanation
MI	Multiple imputation
nAb	Neutralizing antibodies
NBS	Nasal blockage score
NP	Nasal polyposis
NPS	Nasal polyposis score
NPSD	Nasal polyposis symptom scores diary
PGIC	Patient general impression of change
PGIS	Patient general impression of severity
PK	Pharmacokinetic(s)
PRO	Patient reported outcome
PT	Preferred term
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SC	Subcutaneous
SCS	Systemic corticosteroids
SCS_NP	SCS use for NP
SD	Standard deviation
SI	Standard international
SNOT-22	SinoNasal outcome test
SOC	System organ class
TEAE	Treatment emergent adverse event
TESAE	Treatment emergent serious adverse event
TSS	Total Symptom Score
ULN	Upper limit of normal
UPSIT	University of Pennsylvania smell identification test
WHO	World Health organization
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	18AUG2020	Remove Visit from responder analysis (Section 4.2.6.3) as unnecessary.	Yes (CSP version 4.0, 05AUG2020)	Correction or error in specification
Primary or Secondary Endpoints	18AUG2020	Added key secondary endpoint: CT Lund Mackay Score (LMS)	Yes (CSP version 4.0, 05AUG2020)	Updates made following FDA feedback on SAP version 2 [redacted]
Derivation of primary or secondary endpoints	18AUG2020	Change from WOCF to Worst Possible after rescue by NP surgery. WOCF after SCS_NP is retained	Yes (CSP version 4.0, 05AUG2020)	Updates made following FDA feedback on SAP version 2 [redacted]
Data Presentation	18AUG2020	Change baseline characteristics/efficacy subgroups: remove prior OCS replace allergic rhinitis with atopic by phadiatop	Yes (CSP version 4.0, 05AUG2020)	Prior SCS is more appropriate as it includes other systemic use in addition to OCS, and SCS use is also consistent with the secondary endpoint. Atopic status considered a more reliable assessment than allergic rhinitis of atopy
Data Presentation	18AUG2020	Add summary of SNOT-22 individual items. Add CDF curve of SNOT-22 total score.	Yes (CSP version 4.0, 05AUG2020)	Updates made following FDA feedback on SAP version 2 [redacted]

Category*: Change refers to	Date	Description of change	In line with the CSP?	Rationale
Data Presentation	18AUG2020	Add negative binomial regression of number of courses of SCS_NP.	Yes (CSP version 4.0, 05AUG2020)	A formal analysis of number of courses of SCS was included in the other approved labeling for NP indication and therefore has been added for consistency
Primary or Secondary Endpoints,	8APR2020	Timepoint for NPS, NBS, SNOT-22, and DSS changed from Week 56 to Week 40 as primary and key secondary endpoints in the multiple testing procedure. Week 56 for the aforementioned endpoints moved to key secondary. New section added "Impact on analyses due to COVID-19"	No (CSP version 3.0, 20SEP2019)	Response to COVID-19 pandemic due to inability of sites to continue dosing and/or primary endpoint assessment to 56 weeks. Week 40 (the penultimate planned endoscopy) is the latest nasal polyp assessment time not substantially impacted by COVID-19.
Primary or Secondary Endpoints	8APR2020	Additional secondary endpoints added: Time to Surgery and/or SCS_NP, Difficulty with sense of smell	No (CSP version 3.0, 20SEP2019)	Based on regulatory review comments [redacted] CSP text to be aligned with amended SAP for next CSP amendment.

Category*: Change refers to	Date	Description of change	In line with the CSP?	Rationale
Statistical analysis method for the primary or secondary endpoints	8APR2020	Analysis of continuous endpoints changed from MMRM to ANCOVA with MI. Imputation approach for intercurrent events changed from censoring (for treatment discontinuation, NP surgery or SCS_NP) to composite WOCF (for NP surgery or SCS_NP) and not censoring data following treatment discontinuation.	No (CSP version 3.0, 20SEP2019)	Based on regulatory review comments [redacted]. CSP text to be aligned with amended SAP for next CSP amendment.
Multiple testing procedure	8APR2020	Testing order of secondary endpoints: SNOT-22, Time to Surgery and/or SCS_NP, Time to Surgery, then Difficulty with sense of smell	No (CSP version 3.0, 20SEP2019. CSP has UPSIT instead of DSS)	
Other (Sensitivity analyses)	8APR2020	Sensitivity analyses updated for new primary analysis method.	NA	Most appropriate sensitivity analyses for primary estimand and analysis.

* 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

1 STUDY DETAILS

1.1 Study objectives

Primary objective	Estimand Description / Endpoints
• To evaluate the effect of benralizumab on nasal polyp burden and patient-reported nasal blockage (NB).	• Population[a]: Full analysis set • Co-primary endpoints: Change from baseline in endoscopic total nasal polyp score (NPS) Change from baseline in mean nasal blockage score (NBS) • Intercurrent event strategy: Treatment discontinuation – treatment policy NP surgery – composite (Worst Possible) SCS_NP – composite (WOCF) • Summary Measure: differences in least squares mean change from baseline in NPS and NBS between benralizumab and placebo. Week 40 is the primary timepoint.[c]

Secondary objectives	Endpoint/variable
• Disease specific health-related quality of life (HRQoL)	• Change from baseline in SinoNasal Outcome Test (SNOT-22) score[bc]
• Nasal polyp surgery and/or SCS use	• Time to first NP surgery and/or SCS use for NP [b] • Time to first NP surgery [b]
• Sense of Smell	• Change from baseline in mean difficulty with sense of smell (DSS) score [bc] • Change from baseline in University of Pennsylvania Smell Identification Test (UPSIT) score
• Sinus opacification by Computed tomography (CT) scan (subset of patients)	• Change from baseline in Lund Mackay score[d] • Sinus severity score by Quantitative CT analysis[d]
• Proportion of NP surgery	• Proportion of patients with surgery for NP

Secondary objectives	Endpoint/variable
• Systemic corticosteroids (SCS) use for relief of nasal symptoms	• Proportion of patients with SCS use for NP • Time to first SCS course for NP, number of courses of SCS for NP, total SCS dose used and total duration of SCS use for NP
• Symptoms associated with nasal polyps	• Change from baseline in nasal symptom score(s) as captured in the daily diary
• 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

[a] Treatment Condition for primary estimand: Treatment with benralizumab versus placebo, regardless of compliance, where rescue indicates treatment 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.

[c] Week 56 will also be a multiplicity protected timepoint for co-primary and key secondary repeated measures endpoints.

[d] For CT endpoints: Treatment Condition = Treatment with benralizumab versus placebo, regardless of compliance, where rescue with NP surgery indicates treatment failure; Intercurrent event strategy = treatment discontinuation (treatment policy), NP surgery (Composite - WP), SCS_NP (treatment policy). Change from baseline in Lund Mackay Score is a key secondary endpoint.

Safety objective:	Endpoint/variable
• To assess the safety of benralizumab	• Adverse events (AEs) and serious adverse events (SAEs) • Laboratory variables • Physical Examination
• To assess the pharmacokinetics and immunogenicity of benralizumab	• Serum PK • Benralizumab anti-drug antibody (ADA)

[redacted]	[redacted]
[redacted]	[redacted]

[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
[redacted]	[redacted]
• To evaluate the effect of benralizumab on unplanned health care resource utilization	• Hospitalizations, emergency room and urgent care visits

1.2 Study design

This is a randomised, double-blind, placebo-controlled, parallel-group, international, multicentre, Phase III study to evaluate the efficacy and safety of repeat dosing of benralizumab 30 mg administered subcutaneously (SC) versus placebo in patients with severe nasal polyposis.

Approximately 400 patients will be randomized globally to receive benralizumab 30mg SC or matching placebo. Patients will be stratified by region (US vs non-US) and by the baseline comorbid asthma status (yes vs no). After enrolment, eligible patients on intranasal corticosteroids (INCS) will enter 5 weeks screening/run in period. Patients who meet eligibility criteria will be randomized 1:1 to receive either benralizumab or placebo SC every 4 weeks for the first 3 doses - Weeks 0, 4 and 8 and every 8 weeks thereafter - Weeks 16, 24,

32, 40 and 48. A total of 8 doses will be administered to those patients completing the treatment period. An end of treatment (EOT) visit will be conducted at Week 56.

The first approximately 200 patients who complete the treatment will receive the last dose of Investigational product (IP) at Week 48, have the EOT visit at Week 56, then be included in a 6 month extended follow-up period with last scheduled visit at Week 80. The remaining patients will receive the last dose at Week 48, have the EOT visit at Week 56 and have the last scheduled follow-up visit at Week 60.

1.3 Number of subjects

Approximately 400 patients will be randomized into SC benralizumab 30 mg or placebo in a 1:1 ratio.

[redacted]

[redacted] ased 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 0.01 level is –0.52 in total NPS and –0.26 in NBS.

2 ANALYSIS SETS

2.1 Definition of analysis sets

All efficacy analyses except Lund Mackay score (LMS), sinus severity score, asthma exacerbation rate and ACQ-6 will be performed using an Intent-to-Treat (ITT) approach based on the full analysis set (FAS). For consistency, demographic and baseline characteristics will be presented using the FAS. The LMS and sinus severity score will be analyzed based on the CT subset. The asthma exacerbation rate and ACQ-6 will be analyzed based on the comorbid asthma subset. Safety objectives and immunogenicity will be analyzed based on the safety analysis set. Pharmacokinetic (PK) analyses will be based on the PK analysis set.

Patients must have provided their informed consent. If no signed informed consent is collected (major 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 randomized and receiving any IP will be included in the full analysis set (FAS), irrespective of their protocol adherence and continued participation in the study. Patients will be analyzed according to their randomized treatment.

2.1.3 Safety analysis set

All patients who received at least 1 dose of IP will be included in the safety analysis set. Patients will be classified according to the treatment they actually received. A patient who received at least 1 dose of benralizumab will be classified as a patient in the benralizumab treatment group. All safety and immunogenicity analyses will be based on this analysis set.

2.1.4 Pharmacokinetic analysis set

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

2.2 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:

- Patients who do not have bilateral sinonasal polyposis, or patients who do not have a history of treatment with SCS (oral, parenteral) or prior surgery for NP or patients who do not have:
  - A minimum bilateral NPS of 5 out of a maximum score of 8 (with a unilateral score of at least 2 for each nostril) at visit 1 and visit 2, as determined by the study imaging core lab for the eligibility read;
  - Ongoing symptoms for at least 12 weeks prior to visit 1;

  - Patient-reported moderate to severe nasal blockage (score 2 or 3) over the 2-weeks prior to visit 1 (2-week recall assessment of symptoms, scores 0-none to 3-severe).
  - At randomization, a bi-weekly mean NBS ≥1.5.
- Patients with conditions or concomitant disease that makes them non-evaluable for the co-primary efficacy endpoint or with clinically important comorbidities that could confound interpretation of clinical efficacy results including, but not limited to: active upper or lower respiratory tract infection, cystic fibrosis, primary ciliary dyskinesia, eosinophilic diseases other than asthma (eg, allergic bronchopulmonary aspergillosis/mycosis, eosinophilic granulomatosis with polyangitis [Churg-Strauss syndrome], hypereosinophilic syndromes), granulomatosis with polyangiitis (Wegener's granulomatosis), Young's syndrome, etc.
- Patients who received the incorrect study treatment at any time during the 48-week double-blind treatment period.
- Patients who met IP discontinuation criteria but were not discontinued from the IP.
- Patients discontinued due to severe non-compliance to protocol.

Only important protocol deviations will be tabulated in the CSR. The important protocol deviations will be reviewed and documented by the medical advisors and statisticians prior to the database lock.

Protocol deviations associated with the COVID-19 pandemic will also be summarised and listed separately (see section 4.2.13).

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:

- On-treatment period: this is defined as from Day 1 (randomization date) to the EOT/IPD visit day.
- Main-study period: this is defined as from Day 1 to the Week 60 visit day.
- Extended follow-up (EFU) period: this is defined as the period after the Week 60 visit day until end of the study (only expected for the approximately 200 patients continuing follow-up beyond Week 60 visit day).
- On-study period: this is defined as from Day 1 to the end of a patient's expected participation in the study. The end of the study is the end of the extended follow-up period for EFU patients and the end of the main-study period for all other patients.

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 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. If multiple readings are recorded within a single analysis-defined 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. The non-missing observation with the later collection time will be included in the analysis if these records were collected during the screening.

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 defined as

- Screening period: Study day=Date of assessment–date of randomization
- After randomization: Study day=(Date of assessment–date of randomization)+1

The adjusted analysis-defined windows for NPS are defined as Table 1 below:

Table 1 Defined Visit Windows for NPS

Adjusted Defined Visit Windows	Scheduled Study Day	Maximum Windows
Week –5	–35	–42≤study day ≤–14
Week –1	–7	–13≤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 ≤434

Adjusted Defined Visit Windows	Scheduled Study Day	Maximum Windows
Week 68	477	435≤study day ≤518
Week 80	561	519≤study day

The NBS, DSS, and other items in the nasal polyposis symptom diary (NPSD) will be summarized every two weeks (bi-weekly) and the adjusted windows are defined as in Table 2 below.

Table 2 Defined Visit Windows for NPSD (including NBS and DSS)

Adjusted 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 ≤393
…		…
Week 80	561	548≤study day ≤561

3.1.3 The definition of baseline

In general, the last valid value on or prior to the date of randomization 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 randomization (or the first dose of study treatment, depending on the endpoint), then the baseline value will not be imputed and will be set to missing. No 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 average of daily responses from Day –13 to Day 1. The mean is calculated as the sum of all non-missing daily scores over these 14 sequential days divided by the number of non-missing daily scores. If more than 8 daily scores (>50%) within the baseline period are missing, then baseline is set to missing.

3.1.4 Definition of prior/concomitant medications

- A medication will be classified as a maintenance medication at baseline if it started prior to or on the date of randomization and is continued after randomization.
- A medication will be regarded as prior if it was stopped on or before the date of randomization (medication stop date ≤ date of randomization).
- A medication will be regarded as concomitant if the start date is on or after the date of randomization, or if it started prior to the date of randomization and was ongoing after the date of randomization.
- Medications with start date on or after the end of the on-study period (as defined in Section 3.1) will not be considered as concomitant.

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 a missing or non-evaluable observation at the timepoint of interest will be counted as non-responders for the main analysis. An additional sensitivity analysis of SNOT-22 will be performed in which any patient with a missing or non-evaluable observation at the timepoint of interest, the last post-baseline result will be used for patient who completed treatment, and patients who discontinued treatment will be counted as non-responders. In all analyses, patients who have had NP surgery or SCS_NP by the timepoint of interest will be considered as non-responders.

3.2 Primary efficacy variables

The efficacy endpoints will be summarized during the main study period for all patients and during on-study period for the patients who are included in the extended follow up period.

3.2.1 Nasal polyps score

The total nasal polyps score (NPS) is the sum 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 3.

Table 3 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

Polyp score	Polyp size
4	Large polyps causing complete or near-complete obstruction of the inferior nasal cavity i.e. touching the floor of the nose

The total NPS and the changes from baseline will be calculated. In addition, the NPS response will be derived as mentioned in Section 3.1.5, and analysed as a supportive analysis. A patient will be classified as a responder at the timepoint of interest if the change from baseline in total NPS ≤–1 at that timepoint. A 1-grade reduction in bilateral nasal polyp burden has been regarded as clinically meaningful in trials of mometasone furoate nasal spray (Stjarne et al 2006) and sinus implants (Kern et al 2018) for NP.

3.2.2 Nasal blockage score

The NBS will be captured by an item in the NPSD. Patients are 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 summarized every two weeks (bi-weekly). Baseline will be the average of daily responses from Day –13 to Day 1. Bi-weekly mean of NBS 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.

3.3 Key secondary efficacy variables

3.3.1 Health related quality life: SinoNasal Outcome Test

The SNOT-22 is a key secondary endpoint and is a condition specific health-related quality of life 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. An MCID of 8.90 has been established for SNOT-22 total score (Hopkins et al 2009). A SNOT-22 response will be defined as changes from baseline ≤–8.90 at the timepoint of interest. As a supportive analysis, a responder analysis will be conducted according to the approach mentioned in Section 3.1.5.

3.3.2 Time to the first nasal polyposis surgery and/or SCS use for NP

The time to first NP surgery and/or SCS use for NP up to Week 56 is another key secondary endpoint and will be evaluated for all patients. This time will be calculated based on the earliest occurrence of NP surgery and/or SCS use for NP and will be calculated as follows:

Time to first NP surgery and/or SCS use for NP = Earlier of (Start date of first NP surgery, Start date of first SCS use for NP) – date of randomization + 1

For patients who do not experience any surgery or SCS use for NP, the time to event will be censored at the date of their Week 56 visit, or at the time point after which a surgery or SCS use could not be assessed (for patients discontinued before Week 56). For those approximately 200 patients who are included in the extended follow-up period, the time to first NP surgery and/or SCS use for NP up to the end of study will also be calculated as a separate endpoint.

3.3.3 Time to the first nasal polyposis surgery

The time to the first NP surgery up to Week 56 is another key secondary endpoint and will be evaluated for all patients. The time to the first NP surgery is calculated as follows:

Time to the first NP surgery = start date of the first NP surgery–date of randomization+1

For patients who do not undergo any surgery, the time to the first NP surgery up to Week 56 will be right censored at the date of their Week 56 visit, or at the time point after which a surgery could not be assessed (for patients discontinued before Week 56). For those approximately 200 patients who are included in the extended follow-up period, the time to the first NP surgery up to the end of study will also be calculated as a separate endpoint.

3.3.4 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 summarized every two weeks (bi-weekly). Baseline will be the average of daily responses from Day –13 to Day 1. 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.

3.3.5 Sinus computed tomography: Lund-Mackay score

The CT data will come from a blinded central reader. The total CT score based on Lund-Mackay score (LMS) is evaluated at baseline and at EOT/IPD. Both the observed values and the change from baseline at EOT/IPD will be calculated. The LMS evaluates the patency using a 0-2 scale (0–normal; 1–partial opacification; and 2–total opacification) of each sinus (maxillary, anterior ethmoid, posterior ethmoid, sphenoid, frontal sinus on each side). The osteomeatal complex 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.

3.4 Other secondary efficacy variables

3.4.1 Nasal polyps surgery and systemic corticosteroids use for nasal polyposis

The following variables up to Week 56 will be derived for all patients.

- Proportion of patients who had NP surgery
- Proportion of patients who use SCS_NP
- Proportion of patients who had NP surgery or use SCS_NP.
- Time to first SCS_NP
- The number of courses of SCS_NP: noted an SCS_NP course can be considered as a new course if the start date is at least 7 days after the end date of the last SCS_NP course.
- Total SCS_NP dose (converted to prednisolone equivalents)
- Total duration of SCS_NP

For those approximately 200 patients who are included in the extended follow up period, the variables up to the end of the study will also be derived as an exploratory endpoint.

In addition, as a supportive endpoint, the proportion of patients who meet the randomization criteria for potential need for NP surgery (NPS $\geq$5 and bi-weekly mean NBS $\geq$1.5) will be calculated at Week 24, Week 40, Week 56 for all patients and at Week 80 for those 200 patients who are included in the extended follow up period.

The time to recorded decision time to have NP surgery will also be a supportive endpoint.

3.4.2 Nasal polyposis symptom diary

The patient will complete an 11-item NPSD each morning. The patient is asked to consider their experience with NP over the past 24 hours when responding to each question. Patients are asked to report their experience with NP symptoms (nasal blockage, nasal congestion, runny nose, postnasal drip (mucus drainage down the throat), headache, facial pain, facial pressure, and difficulty with sense of smell) and symptom impacts (difficulty with sleeping due to nasal symptoms and difficulty with daily activities due to nasal symptoms). Patients report the severity of each symptom and symptom impact at its worst using a 4-point verbal rating scale (0-None to 3-Severe). A single item to capture INCS compliance (yes or no) will be administered after the symptom and symptom impact items. Bi-weekly mean of each item in the NPSD will be calculated if at least 8 days in each 14-day period has evaluable data; otherwise the bi-weekly mean is set to missing. A Total Symptom Score (TSS) will be calculated by taking the sum of the first 8 items in the NPSD.

3.4.3 University of Pennsylvania Smell Identification Test

The UPSIT is quantitative test of olfactory function which uses microencapsulated odorants that are released by scratching standardized odor-impregnated test booklets. Four booklets each with 10 odorants each are used for the test. Patients are asked to identify the odor using multiple choice format which lists different possibilities. The test is forced-choice; ie, 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 odors (score range 0 to 40). The olfactory diagnosis will be classified based on the test scores by gender (male vs female) as listed in Table 4. The categories of Probable Malingering (UPSIT score 0 - 5) and Total Anosmia (6-18) have been combined into one 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 (Bachert et al, 2019).

Table 4 UPSIT olfactory diagnosis

	Test scores	
Olfactory Diagnosis	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

3.4.4 Sinus computed tomography: 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. Both of the observed values and the changes from baseline at EOT/IPD will be calculated. The sinus severity score is defined as (sinus mucosal volume)/(sinus mucosal volume+sinus air volume)×100.

3.4.5 Short Form 36-item Health survey, version 2

The SF-36v2 (standard recall) is a 36-item, self-report survey of functional health and well-being, with 4-week recall period (QualityMetric 2011). Responses to 35 of the 36 items are used to compute an 8-domain profile of functional health and well-being scores. The remaining item, referred to as the ‘health transition’ item, asks patients to rate how their current state of health compared to their state of health 1 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 health-related quality of life.

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 5 is suitable for interpreting change at the individual level and is referred to as the responder threshold or responder definition (QualityMetric 2011). SF-36 responders, defined as change ≥ threshold, will be derived using the approach mentioned in Section 3.1.5.

Table 5 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 Role Limitations due to Emotional Problems; RP Role Limitations due to Physical Health; SF Social Functioning; VT Vitality.

[redacted]

3.5.4 Healthcare resource utilization

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

- General and/or intensive care hospitalizations and lengths of stay
- Emergency room visits
- Urgent care visits

For each category above, the crude rates of HRU during the on-treatment period (as defined in Section 3.1) will be calculated, where the crude rate = (total days of HRU use/total days on treatment).

3.6 Safety outcome variables

The following safety data will be collected: reported AEs (including SAEs), hematology, clinical chemistry, physical examination and vital signs. Change from baseline to each post-treatment time point where scheduled assessments were made will be calculated for relevant measurements. AEs will be summarized by means of descriptive statistics and qualitative summaries.

No safety data will be imputed. The handling of partial/missing dates for AEs and prior/concomitant medications is detailed in Appendix 8.6. Duration of AEs and prior/concomitant medications will not be calculated using imputed dates and will instead be set to missing.

3.6.1 Adverse events

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

All safety analyses will be performed on the safety analysis set. Treatment emergent adverse events (TEAEs), treatment emergent serious adverse events (TESAEs) will be summarized over the main-study period for all patients. In addition, exposure adjusted summaries covering the on-study period will be considered for all patients. Separate presentation of adverse events may be considered for the extended follow-up period for patients who are

included in the extended follow-up period. If an AE has a missing onset date then unless the stop date of the AE indicates otherwise, this will be considered a TEAE. Similarly, if an AE has a partial onset date, then unless the partial onset date or the stop date indicates otherwise, this will be considered a TEAEs.

3.6.2 Laboratory variables

Lab results and normal ranges will be presented in the International System (SI) unit. Eosinophils data will be presented in both SI and conventional units (cells/µL) in summaries. Changes in hematology and clinical chemistry variables between baseline and each post-baseline assessment will be calculated. There will be no imputation for missing values. For values recorded with a leading greater than or less than ('>', '<') symbol, the reported numeric value will be used for analysis and the value with the symbol will be included in the listings, unless otherwise specified. For example, a value of <0.01 will be analyzed as 0.01 and listed as <0.01. Absolute values will be compared to the relevant reference range and classified as low (below range), normal (within range or on limits) or high (above range). All absolute values falling outside the reference ranges will be flagged. The maximum or minimum value post-baseline will be calculated over the entire main-study period for all patients and over the on-study period for the patients who are included in the extended follow-up period.

For the liver function tests: Aspartate Aminotransferase (AST), Alanine Aminotransferase (ALT), Alkaline phosphatase, Gamma-GT (GGT) and total bilirubin (TBL), the multiple of the central laboratory upper limit of the normal (ULN) range will be calculated for each data point. Multiple=Value/ULN, ie, if the ALT value was 72 IU/L (ULN 36) then the multiple would be 2.

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

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

3.6.3 Physical examination

Complete and brief physical examinations will be performed at time points specified in Tables 1 and 2 (Schedule of Activities) of 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. For all physical examinations 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.

3.6.4 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 subsequent scheduled assessment will be calculated. There will be no imputation for missing values.

Absolute values will be compared to the reference ranges listed in Table 6 and classified as low (below lower limit), normal (within lower limit and upper limit, inclusive) or high (above upper limit). All values (absolute and change) falling outside the reference ranges will be flagged.

Table 6 Vital signs reference ranges

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

Body mass index (BMI) will be calculated from the height and weight as follows:
BMI (kg/m^2) = weight (kg)/(height (m))^2

3.7 Pharmacokinetic and immunogenicity variables

3.7.1 Pharmacokinetic variables

Serum samples for pharmacokinetic assessments will be collected. The drug concentration levels will be summarized using descriptive statistics.

3.7.2 Immunogenicity variables

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

4 ANALYSIS METHODS

4.1 General principles

The primary estimand applied to the primary analysis of the co-primary endpoints (along with the main analysis of key secondary endpoints) will include all data up to Week 56, regardless of whether a patient remains on blinded study treatment or not. A composite strategy will be used for patients who undergo NP surgery or receive SCS for NP (SCS_NP). Data collected post-SCS_NP will be set to missing, and the patient's worst observed post-baseline value on or before the time of SCS_NP will be imputed from that point through Week 56. Data collected post-surgery will be set to missing, and the worst possible value (as outlined in Table 9) will be imputed from that point through Week 56. For repeated measures analyses, the primary timepoint of interest for the purpose of multiplicity protection (see section 4.1.2) is week 40, although the same statistical model will be used to estimate treatment effect at other timepoints as well.

For the purposes of identifying events of NP surgery and SCS use for NP through Week 56, the time through Week 56 is considered to be the time from first dose date through the later of day 393 and the Week 56 visit date. This also applies to the identification of patients who have been rescued by NP surgery and/or SCS use for NP.

Details regarding primary and key secondary estimands are provided in Table 7, with additional details including sensitivity analyses provided in Appendix 8.2, 8.3, 8.4, 8.5, and 8.10.

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

Table 7 Primary and key secondary efficacy and main safety estimands

	Estimand[1]			
Statistical Category & Section	**Treatment Condition[1]**	**Endpoint (Population[2])**	**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 4.2.6.1	Treatment with benralizumab versus placebo, regardless of compliance, where rescue indicates treatment failure.	• CFB in endoscopic total NPS (FAS) • CFB in mean NBS (FAS)	• Treatment discontinuation – treatment policy • NP surgery – composite (Worst Possible) • SCS_NP – composite (WOCF) (primary estimand)	Mean difference between interventions (LSMD from CFB ANCOVA following hybrid WP/WOCF, MI). Week 40 is the primary timepoint.[3]
Secondary Objective: To evaluate the effect of benralizumab on disease specific health-related quality of life				
Secondary/MCP Section 4.2.7.1	Treatment with benralizumab versus placebo, regardless of compliance, where rescue indicates treatment failure.	CFB in SNOT-22 score (FAS)	• Treatment discontinuation – treatment policy • NP surgery – composite (Worst Possible) • SCS_NP – composite (WOCF) (primary estimand)	Mean difference between interventions (LSMD from CFB ANCOVA following hybrid WP/WOCF, MI). Week 40 is the primary timepoint.[3]
Secondary Objective: To evaluate the effect of benralizumab on NP Surgery and/or SCS_NP				
Secondary/MCP Sections 4.2.7.3, 4.2.7.2	Treatment with benralizumab, regardless of compliance, versus placebo, regardless of compliance.	• Time to first NP Surgery and/or SCS_NP (FAS) • Time to first NP Surgery (FAS)	• Treatment discontinuation – treatment policy • SCS_NP – treatment policy (for time to first NP Surgery only) (primary estimand)	Hazard ratio from Cox proportional hazards model. Events beyond Week 56 will not be included.
Secondary Objective: To evaluate the effect of benralizumab on sense of smell				
Secondary/MCP Section 4.2.7.4	Treatment with benralizumab versus placebo, regardless of compliance, where rescue indicates treatment failure.	CFB in mean DSS (FAS)	• Treatment discontinuation – treatment policy • NP surgery – Composite (Worst Possible) • SCS_NP – Composite (WOCF) (primary estimand)	Mean difference between interventions (LSMD from CFB ANCOVA following hybrid WP/WOCF, MI). Week 40 is the primary timepoint.[3]

	Estimand[1]			
Statistical Category & Section	**Treatment Condition**[1]	**Endpoint (Population**[2]**)**	**Intercurrent Event Strategy**[1]	**Population Level Summary**[1] **(Analysis)**
Secondary Objective: To evaluate the effect of benralizumab on sinus opacification				
Secondary/MCP Section 4.2.7.5	Treatment with benralizumab versus placebo, regardless of compliance, where rescue with NP surgery indicates treatment failure.	CFB in mean LMS (CT subset)	• Treatment discontinuation – treatment policy • NP surgery – Composite (Worst Possible) • SCS_NP – treatment policy (WP after NP surgery estimand)	Mean difference between interventions (LSMD from CFB ANCOVA following WP imputation. EOT/IPD is the primary timepoint.
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 TEAEs during main-study period (Safety) • Presence of serious TEAEs during main-study period (Safety)	• Treatment discontinuation – treatment policy	Categorical descriptive

MCP = Multiple comparisons procedure; NP = Nasal Polyposis; SCS_NP = Systemic Corticosteroids used for Nasal Polyps; 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; TEAE = Treatment emergent adverse event.

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

[2] FAS = Full Analysis Set

[3] Week 56 is also a multiplicity protected timepoint for co-primary and key secondary repeated measures endpoints.

All hypothesis testing will be reported using 2-sided tests. P-values will be rounded to 4 decimal places. The 95% confidence interval (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 visit 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). All SAS® programs used to generate analytical results will be developed and validated according to AstraZeneca SAS® programming standards and validation procedures. Pharmacokinetic analyses will be performed using NONMEM or other appropriate software.

4.1.1 Hypothesis tests for co-primary endpoints

The co-primary efficacy endpoints are the change from baseline in total NPS at Week 40 and the change from baseline in bi-weekly mean NBS at Week 40. The primary analysis is to compare the changes from baseline of benralizumab with placebo.

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.

4.1.2 Method for multiplicity control

To account for multiplicity to test the co-primary endpoints (change from baseline in total NPS at Week 40 and change from baseline in NBS at Week 40) and the 9 key secondary endpoints (the change from baseline in SNOT-22 at Week 40, time to the first NP surgery and/or SCS use for NP up to Week 56, time to the first NP surgery up to Week 56, the change from baseline in DSS at Week 40, the change from baseline in NPS at Week 56, the change from baseline in NBS at Week 56, the change from baseline in SNOT-22 at Week 56, the change from baseline in DSS at Week 56, and the change from baseline in LMS at EOT/IPD), 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 endpoints will be tested at 0.01 (two-sided) and the key secondary endpoints will be tested hierarchically 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 0.01. If both p-values are less than 0.01 with treatment effects in favor of benralizumab, then proceed to step 2. Otherwise no null hypothesis is rejected.

- Step 2: Test the 9 key secondary endpoints at the significant level of 0.05 using hierarchical fixed-sequence testing in the order below.
  - Changes from baseline in SNOT-22 at Week 40.
  - Time to the first NP surgery and/or SCS use for NP up to Week 56.
  - Time to the first NP surgery up to Week 56.
  - Change from baseline in DSS at Week 40.
  - Change from baseline in NPS at week 56.
  - Change from baseline in NBS at week 56.
  - Change from baseline in SNOT-22 at week 56.
  - Change from baseline in DSS at week 56.
  - Change from baseline in LMS at EOT/IPD.

4.1.3 Description of estimands used in the analyses

4.1.3.1 Primary estimand

The primary estimand will be used for the primary analysis and quantifies the difference in outcomes for patients randomized to the benralizumab and placebo arms at the planned timepoints of the study, regardless of the treatments that patients actually received, where rescue by NP surgery and/or SCS_NP indicates failure. It includes all the data collected during main-study period including data collected after discontinuation of study treatment except data collected after NP surgery and/or SCS_NP. A composite strategy will be used for patients who undergo NP surgery or receive SCS for NP. Data collected post-SCS_NP will be set to missing, and the patient's worst observed post-baseline value on or before the time of SCS_NP will be imputed from that point through Week 56. Data collected post-surgery will be set to missing, and the worst possible value will be imputed from that point through Week 56. For patients who discontinue the study without surgery or SCS_NP, missing data will be imputed using multiple imputation using all patients who did not have surgery or receive SCS_NP. Details of the primary analysis are included in Section 4.2.6.1. Week 40 is the primary timepoint of interest. Week 56 is also a multiplicity protected timepoint for the co-primary endpoints and key secondary endpoints of change from baseline in SNOT-22 and change from baseline in DSS. Each of these endpoints will be evaluated at both Week 40 and Week 56.

A sensitivity analysis for the primary estimand will use a different composite strategy for surgery and SCS_NP. In this sensitivity analysis, a composite strategy will be used only for patients who undergo NP surgery in which data collected post-surgery will be set to missing, and the 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_NP without having NP surgery will be included with the non-surgery patients. Details of this sensitivity analysis are included in Appendix 8.2.

4.1.3.2 Effectiveness estimand

The effectiveness estimand quantifies the difference in outcomes for patients randomized to the study treatment and control arms, while on treatment and without having been rescued by NP surgery and/or SCS_NP, at the planned endpoint of the study. In the effectiveness estimand, all the data up to IPD/EOT visit will be included except the data observed after the SCS_NP and NP surgery. No imputation will be conducted for the missing data as analysis will follow an MMRM approach. A sensitivity analysis will be evaluated under the effectiveness estimand. Details of the sensitivity analyses are included in Appendix 8.4.

4.1.3.3 Treatment policy estimand

The treatment policy estimand quantifies the difference in outcomes for patients randomized to the study treatment and control arms at the planned endpoint of the study, regardless of the treatments that patients actually received and regardless of the occurrence of NP surgery and/or SCS_NP. The treatment policy estimand includes all the data collected during main-study period including data collected after discontinuation of study treatment, NP surgery, or SCS_NP. No imputation will be conducted for the missing data as analysis will follow an MMRM approach. A sensitivity analysis will be evaluated under the treatment policy estimand. Details of this sensitivity analysis are included in Appendix 8.5.

4.1.4 Sensitivity analyses for missing data

Sensitivity analyses of the ANCOVA with missing data imputed will be performed for co-primary endpoints 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 summarized using the all patients analysis set. The number and percentage of patients within each treatment group will be presented by the following categories; randomized, received treatment with study drug, did not receive treatment with study drug (and reason), completed treatment with study drug, discontinued treatment with study drug (and reason), discontinued treatment with study drug but completed study follow-up, complete main-study period, completed study, and withdrawn from study (and reason).

The number of patients randomized by country and center will be summarized. A listing of disposition of patients and the reasons for screen failures will be provided.

4.2.2 Demography data and patient characteristics

Demographic data and key baseline characteristics will be summarized for each treatment group using descriptive statistics on FAS. The demographic data and key baseline characteristics will include but not limited to the following variables.

- Age
- Gender
- Country
- Region (US vs Non-US)
- Race
- Ethnicity
- Baseline Characteristics (prior SCS_ NP, comorbid asthma status, AERD, prior NP surgery, blood eosinophil count, BMI, atopic status by phadiatop test, IgE)
- Baseline Characteristics for comorbid asthma subset (ACQ-6, number of prior asthma exacerbations, total daily ICS dose)

4.2.3 Prior and concomitant medications

The number and percentage of patients who take maintenance medications will be summarized. The number and percentage of patients who take prior medications, those who take allowed concomitant medications and those who take disallowed concomitant medications during the study, will be presented. The concomitant medication will be summarized over the main-study period and over the extended follow-up period, respectively. 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 Exposure

Durations of the study drug exposure and of on-treatment will be summarized for the safety analysis set, where the exposures are calculated in days as:

Study drug exposure = last dose date of IP - first dose date of IP+1.

On-treatment duration=date of EOT or IPD–first dose date of IP+1.

4.2.5 Study treatment compliance

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

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

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

The individual compliance rate will be summarised by descriptive statistics by treatment group.

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

The efficacy endpoints will be summarized during main-study period for all patients and during on-study period for the patients who are included in the extended follow up period.

The co-primary efficacy endpoints are the change from baseline in total NPS at Week 40 and the change from baseline in bi-weekly mean NBS at Week 40.

4.2.6.1 Primary analyses

The primary estimand will be applied to the co-primary endpoints using a hybrid method of the worst-possible/worst-observation carried forward (WP/WOCF) and multiple imputation (MI) followed by ANCOVA with treatment arm, baseline scores (baseline total NPS for NPS model and baseline NBS for NBS model), region (US vs non-US), and baseline comorbid asthma status (yes vs no) as covariates. The estimates of the treatment effects at Week 40 and Week 56 will be based on contrasts from this ANCOVA at the respective timepoints. The analyses will use the data collected up to Week 56 visit, regardless of whether patients remained on treatment or not except data collected after NP surgery and/or SCS_NP.

A composite strategy will be used for NP surgery and SCS_NP. If a patient had SCS_NP before Week 56, the data will be censored after the time of having the first course of SCS_NP and the patient's worst observed value will be imputed in its place. For patients rescued by SCS_NP whose post-baseline values are all missing or for whom every post-baseline value is after rescue by SCS_NP, the baseline will be used to impute. If a patient had NP surgery before Week 56, the data will be censored after the time of the first NP surgery and the worst possible value will be imputed in its place. If there is sufficient evaluable NBS data prior to rescue in the biweekly 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 9 for a full list of WP values.

Analyses will include all patients with baseline and at least one evaluable post-baseline assessment and all patients with baseline who were rescued by NP surgery and/or SCS_NP by Week 56.

The following 5 steps will be used to build the complete 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 NP surgery, SCS_NP, or Week 56, whichever comes first for those rescued). The model for MCMC imputation will include adjustment for covariates including treatment group, region, asthma status, and baseline value of the response value. See Table 10 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 were not rescued by NP surgery or receiving SCS_NP by Week 56 will be imputed using the regression method for the monotone pattern with adjustment for covariates including treatment groups, region, asthma status, and baseline value of the response variable. No data from rescued patients is included in this step. See Table 10 for random seeds to be used for monotone regression step.

3 For patients who were rescued by NP surgery or received SCS_NP by week 56, any remaining missing data following step 1 after the time of surgery or SCS_NP 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 were rescued by Week 56) will be replaced using last observation carried forward (LOCF). Similarly for patients with only a baseline prior to rescue (who were 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. 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.

The following sensitivity analyses using the same model as the primary analyses mentioned above will be conducted

1 ANCOVA under primary estimand with composite strategy for NP surgery only (WP imputation after NP surgery) and MAR imputation for non-surgery study discontinuation. (Appendix 8.2)
2 ANCOVA under primary estimand with WP/WOCF imputation after NP surgery and/or SCS_NP and Dropout Reason based Multiple Imputation (DRMI) for non-rescue study discontinuation to test MAR assumption. (Appendix 8.3)
3 MMRM under effectiveness estimand without imputation where outcomes after treatment discontinuation, NP surgery, or SCS_NP will be set to missing. (Appendix 8.4)
4 MMRM under treatment policy estimand without imputation where all data as observed through Week 56 is included regardless of treatment discontinuation, NP surgery, or SCS_NP. (Appendix 8.5)

[redacted]

4.2.6.3 Supportive analyses: responder analyses

The NPS responder at Week 40, which is defined as change from baseline in total NPS ≤–1, based on change from baseline in total NPS at Week 40, will be analyzed using a logistic regression model with covariates of treatment arm, baseline total NPS scores, region, baseline comorbid asthma status as covariates. The same analysis will be conducted for NPS response at Week 24 and Week 56, defined similarly based on change from baseline in total NPS at Week 24 and Week 56, respectively. Analyses of NPS responders at Week 24 and Week 40 by quartile of observed and model-predicted steady state trough concentrations will also be conducted.

Cumulative distribution function of absolute changes from baseline in NBS at Week 24, at Week 40, and at Week 56 will be plotted in a figure.

4.2.7 Key secondary efficacy variables

4.2.7.1 Health related quality life: SinoNasal Outcome Test

The change from baseline in the SNOT-22 total score will be analyzed using a similar ANCOVA as the primary endpoints described in Section 4.2.6.1. For the total score and individual item, SNOT-22 scores and the change from baseline will be summarized using descriptive statistics. Responder analyses at Week 24, Week 40, and Week 56, where a responder is defined as change from baseline ≤-8.9, will be conducted using the logistic regression mentioned in Section 4.2.6.3. Subgroup analyses as mentioned in Section 4.2.6.2 will also be conducted for the change from baseline in SNOT-22 total score, including the subgroup analysis with PK exposure quartile.

Cumulative distribution function of absolute changes from baseline in SNOT-22 total score at Week 24, at Week 40, and at Week 56 will be plotted in a figure.

4.2.7.2 Time to first NP surgery

The time to first NP surgery will be analyzed using a Cox proportional hazard model with treatment arm, region (US vs non-US) and baseline comorbid asthma status (yes vs no) as covariates. Time to the first NP surgery will be displayed graphically using the Kaplan-Meier estimate by treatment.

4.2.7.3 Time to first NP surgery and/or SCS use for NP

The time to first NP surgery and/or SCS use for NP will be analyzed in the same manner as Time to first NP surgery in section 4.2.7.2.

4.2.7.4 Difficulty with Sense of Smell (DSS) score

The change from baseline in DSS score will be analyzed using a similar ANCOVA as the primary endpoints described in Section 4.2.6.1. DSS scores and the change from baseline will be summarized using descriptive statistics.

The cumulative distribution function of absolute changes from baseline in DSS at Week 24, at Week 40, and at Week 56 will also be plotted in figures.

Subgroup analyses as mentioned in Section 4.2.6.2 will also be conducted for the change from baseline in DSS, including the subgroup analysis with PK exposure quartile.

4.2.7.5 Sinus computed tomography: Lund-Mackay score (LMS)

The change from baseline in LMS will be analysed using a similar ANCOVA model as the primary endpoints described in Section 4.2.6.1 but a different intercurrent event strategy. The analyses will use the data collected up to Week 56 visit, regardless of whether patients remained on treatment or not except data collected after NP surgery. Because there is only a single CT scan post-baseline that occurs at least six months after baseline, the composite (WOCF) strategy used after SCS_NP in the primary estimand is not as appropriate as for the main analysis. Instead, the analysis will use data collected after SCS_NP. A composite strategy will be used for NP surgery. If a patient had NP surgery before EOT/IPD CT scan, the data will be censored after the time of the first NP surgery and the worst possible value will be imputed in its place. See Table 9 for a full list of WP values. The Multiple Imputation components described in Section 4.2.6.1 will not be necessary for analysis of LMS scores. LMS scores at baseline and EOT/IPD and the change from baseline will be summarized using descriptive statistics.

A sensitivity analysis of LMS scores will be conducted in which the composite (WOCF) strategy is used for SCS_NP and patients who receive SCS_NP prior to EOT/IPD CT scan have their baseline score imputed in place of the observed or missing values. Patients who have NP surgery prior to EOT/IPD CT scan will have the worst possible value imputed whether or not they also received SCS_NP.

4.2.8 Other secondary efficacy variables

4.2.8.1 Nasal polyps surgery and systemic corticosteroids use for NP

The NP surgery and the SCS use for NP (SCS_NP) will be summarized up to Week 56 for all patients and up to the end of study for those patients in the extended follow up period. The proportions of NP surgery and/or SCS_NP listed in Section 3.4.1 will be summarized and analyzed using the Cochran–Mantel–Haenszel test stratified by region (US vs non-US) and baseline comorbid asthma status (yes vs no). The proportion of patients without each event type through 56 weeks will also be estimated by treatment group using Kaplan-Meier.

In addition, the NP surgery will be summarized by the surgery reason, surgery type and surgery procedures. Times to the first SCS_NP use will be analyzed using the Cox proportional hazard model mentioned in Section 4.2.7.2.

The total number of courses of SCS_NP, the total SCS_NP dose used, and total duration of SCS_NP will be summarized using descriptive statistics. The total number of courses of SCS_NP will be analyzed using a negative binomial model. The response variable in the model will be the number of courses of SCS_NP for each patient. The model will include covariates of treatment group, region (US vs non-US) and prior use of SCS for NP (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 meet the randomization criteria for potential need for NP surgery (NPS ≥5 and bi-weekly mean NBS ≥1.5) will be summarized and analyzed using the Cochran–Mantel–Haenszel test stratified by region (US vs non-US) and baseline comorbid asthma status (yes vs no).

As another supportive analysis, the time to decision to have NP surgery will be analyzed using the Cox proportional hazard model mentioned in Section 4.2.7.2.

4.2.8.2 Nasal polyposis symptom diary

The bi-weekly mean of NPSD (individual components) as well as the TSS and the corresponding changes from baseline will be summarized using descriptive statistics. The changes from baseline will be analyzed using a similar ANCOVA analysis as described for

the co-primary endpoints in Section 4.2.6.1. The ANCOVA analysis for the TSS and components of NPSD other than NBS and DSS will be done for the primary estimand only.

4.2.8.3 University of Pennsylvania Smell Identification Test

The UPSIT scores will be summarized and analyzed overall as well as separately for men and women. The change from baseline in UPSIT score will be analyzed using a similar ANCOVA analysis as the primary endpoints described in Section 4.2.6.1 for the primary estimand only.

The number and percent of patients in each of five olfactory diagnosis categories will be summarized. The proportion of patients in each olfactory diagnosis categories will be analyzed using proportional odds model with treatment, baseline scores, region, baseline comorbid asthma status, as covariates. Shift tables of the olfactory diagnosis categories between the baseline and Week 24, Week 40, and Week 56 will be separately generated.

4.2.8.4 Sinus computed tomography: Sinus severity score

The sinus severity score and change from baseline will be summarized using descriptive statistics. The change from baseline in sinus severity score will be analyzed using a similar ANCOVA analysis as described for Lund-Mackay score in Section 4.2.7.5.

4.2.8.5 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 summarized using descriptive statistics by treatment group and by visit.

The responder (as defined in Section 3.4.5) rate at Week 24 and at Week 56 will be summarized and analyzed using the logistic regression as mentioned in Section 4.2.6.3 with baseline from the corresponding SF-36v2 item included as a covariate in the model in place of baseline NPS

4.2.9.4 Healthcare resource utilization

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

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

The frequency, crude rate (as defined in Section 3.5.4), and annualized rate (defined as crude rate * 365.25) of the HRU listed above and the total number of day hospitalized during the treatment period will be summarized by reasons using descriptive statistics.

4.2.10 Safety outcome variables

All safety variables will be summarized using the safety analysis set. Treatment emergent adverse events (TEAEs) will be summarised over the main-study period for all patients. In addition, exposure adjusted summaries covering the on-study period will be considered for all patients. Separate presentation of adverse events may be considered for the extended follow-up period for patients who are included in the extended follow-up period.

4.2.10.1 Adverse events

Summary of adverse events over the main-study period

The treatment emergent AEs and SAEs will be summarized over the main-study period for all patients. An overall summary table will be produced showing the number and percentage of patients with at least 1 AE in any of the following categories; AEs, SAEs, AEs with outcome of death, and AEs leading to discontinuation of investigational product (DAEs). The total number of AEs in the different AE categories in terms of AE counts will also be presented (i.e., accounting for multiple occurrences of the same event in a patient).

Adverse events, AEs with outcome of death, SAEs and DAEs will be summarised by system organ class (SOC) and preferred term (PT) assigned to the event by MedDRA. For each PT, the number and percentage of patients reporting at least one occurrence will be presented, ie, for a patient multiple occurrence of an AE will only be counted once. The SAEs causing discontinuation of the study treatment and the SAEs causing discontinuation from the study will also be summarised.

A summary of the most common (frequency of ≥5%) AEs will be presented by PT. Adverse events and SAEs will be summarised by PT and investigator's causality assessment (related vs not related) and maximum intensity. If a patient reports multiple occurrences of the same AE within the same study period, the maximum intensity will be taken as the highest recorded maximum intensity (the order being mild, moderate, and severe). Adverse events of injection site reactions (high level term of injection site reactions), hypersensitivity (standardized MedDRA query (SMQ) [Narrow] of hypersensitivity), and malignancy (SMQ of malignant tumours) will be summarized by preferred term.

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

The rate of AEs per person-years at risk, calculated as (number of patients reporting AE)/(total period with patients at risk of AE), will also be reported. Rates will be expressed in terms of events per 100 patient-years.

Other adverse events summary

In addition, an overview of AEs, SAEs, AEs by SOC and PT (exposure adjusted), SAEs by SOC and PT (exposure adjusted) will be summarized covering the on-study period for all patients. Separate presentation of AEs including overview of AEs, AEs by SOC/PT, SAEs by SOC/PT over the extended follow-up period may be considered for the patients who are included in the extended follow-up period.

4.2.10.2 Safety Subgroups

Subgroup summaries for safety to support the submission will be outlined in a separate Summary of Clinical Safety SAP.

4.2.10.3 Laboratory data

All continuous laboratory parameters and the changes from baseline will be summarized using descriptive statistics. Shift plots showing each individual patient's laboratory value at baseline and at maximum/minimum post-baseline may be produced for continuous laboratory variable as applicable. If any laboratory variables show any unusual features (high or low values or a general shift in the data points) at other time points then shift plots of these data may be produced.

Data for patients who have treatment-emergent changes outside laboratory reference ranges will be presented. This data presentation will include all visits for this subset of patients.

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

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

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

For all patients who meet the biochemical criteria for confirmed Hy's law, a Patient Safety Narrative will be produced, and the relevant laboratory parameters will be tabulated showing all visits for these patients.

Blood eosinophils will also be presented for the on-treatment period.

4.2.10.4 Vital signs

Vital signs will be summarized using descriptive statistics. 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.

4.2.11 Pharmacokinetic and immunogenicity variables

4.2.11.1 Analysis of pharmacokinetic variables

Benralizumab serum concentrations will be summarized using descriptive statistics. The population PK analysis and pharmacodynamics analyses will be presented separately from the main clinical study report (CSR).

4.2.11.2 Analysis of immunogenicity variables

Anti-drug antibody (ADA) assessments will be conducted and analysed as per the details in Appendix 8.7.

[redacted]

4.2.13 Impact on analyses due to COVID-19 pandemic

While the study had already been fully recruited at the time of the COVID-19 worldwide pandemic, for ongoing patients, patient dosing, scheduled visits, and nasal endoscopies all became difficult or impossible to perform according to protocol for patients remaining in the study at that time.

Efforts are ongoing to collect outstanding data via alternative means where possible, when on-site visits cannot be performed. As a result, the following changes have been made to the planned analyses:

- The co-primary endpoints and key secondary repeated measures endpoints are first being tested at Week 40, followed by the same endpoints at Week 56. This timepoint is the latest visit not substantially affected by COVID-19 related study deviations.

Time to event (NP surgery and/or SCS_NP, NP surgery) will still be tested including all data through Week 56. See section 4.1.2 for hypothesis testing hierarchy.

- An additional sensitivity analysis of primary and key secondary endpoints at Week 56 will be conducted whereby data is excluded (censored) from a patient's first recorded protocol deviation related to COVID-19.

- 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.

- Confirmed or suspected cases of COVID-19 will be listed and included as AEs as appropriate.

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 randomized 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 NP SINUS-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.

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.

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

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

Wang 2014
Wang B, Fang Y, Jin M, "Combining Type-III Analyses from Multiple Imputations". *Proceedings of the SAS Global Forum 2014 Conference.* Cary, NC: SAS Institute Inc. Available at http://support.sas.com/resources/papers/proceedings14/1543-2014.pdf

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 and withdrawal from the study will be presented using Kaplan Meier plots (overall and split by treatment related/not treatment related reason for discontinuation, as defined in Table 8). Dependent on these outputs additional exploratory analyses covering the main study period may be produced to further understand the pattern of missing data.

8.2 Sensitivity analyses: Primary estimand with WP for NP surgery only

As mentioned in Section 4.2.6.1, the primary analyses will be based the primary estimand. A sensitivity analysis under the primary estimand to the strategy for imputing WOCF for SCS_NP rescue 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 NP at the time of their first NP surgery. The occurrence of SCS_NP will not be considered rescue and a treatment policy strategy will be used instead. Data prior to any NP surgery will be included regardless of the use of SCS_NP. The same multiple imputation strategy as described in Section 4.2.6.1 will be carried out, but patients who receive SCS_NP but never have NP surgery will be included in the MI portion for non-rescue patients. See Table 10 for random seeds to be used for this sensitivity analysis.

8.3 Sensitivity analyses: Primary estimand assuming dropout reason-based multiple imputation approach

As mentioned in Section 4.2.6.1, the primary analyses will be based on the primary estimand. If a patient had surgery and/or had course of SCS_NP, the data will be censored at the time of NP surgery and/or SCS_NP and the patient's worst possible (after NP surgery) or worst-observed value (after SCS_NP) will be imputed from that point until Week 56. A sensitivity analyses under the primary estimand with missing data imputed will be conducted. As in

section 4.2.6.1, a composite strategy will be used for patients who have NP surgery and/or receive SCS_NP. For patients that do not undergo NP surgery or receive SCS_NP, missing data will be imputed using controlled sequential multiple imputation (MI) methods based on pattern mixture models (EMA/CHMP/EWP 2010). See Table 10 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 arms who dropped out for a treatment related reason or severe non-compliance of protocol is similar to that of the placebo patients, 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 8.

Table 8 Parameters for calculating the imputation under MAR and DRMI

Reason for withdrawal or missing data	MAR	DRMI
Adverse Event	Benralizumab	Placebo
Development of study-specific discontinuation criteria*	Benralizumab	Placebo
Death	Benralizumab	Placebo
Severe non-compliance to protocol	Benralizumab	Placebo
Eligibility criteria not fulfilled	Benralizumab	Benralizumab
Subject lost to follow up	Benralizumab	Benralizumab
Subject decision	Benralizumab	Based on review prior to study unblinding
COVID-19	Benralizumab	Benralizumab
Other	Benralizumab	Based on review prior to study unblinding

Note: Patients in the placebo arm are imputed using the mean of the non-missing values in placebo arm.

*Development of study-specific discontinuation criteria are based on the following: Anaphylactic reaction to the investigational product requiring administration of epinephrine; Development of helminth parasitic infestations requiring hospitalization; If 2 doses of the IP are missed during course of the study; 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 subject 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 8 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 Sensitivity analyses under the effectiveness estimand

A sensitivity 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 NP surgery and/or SCS_NP. Because these data will be analysed by MMRM which implicitly accounts for missing data under the MAR assumption, no imputation will be done after discontinuation from treatment or after rescue by NP surgery and/or SCS_NP. This analysis evaluates the effect of initially assigned randomized treatment as long as the subject remains on treatment without being rescued by a significant alternate course of therapy such as NP surgery or SCS for NP.

The analysis will be conducted using a mixed effect repeated measures (MMRM) analysis with treatment arm, baseline scores, visit, region (US vs non-US), baseline comorbid asthma status (yes vs no), and treatment×visit as covariates. The variance-covariance matrix will be assumed to be unstructured. If the procedure does not converge, then the Toeplitz, first-order autoregressive, and compound symmetric, variance components variance-covariance matrices will be tried in that order. The estimate of the treatment effect will be based on a contrast from this MMRM model.

8.5 Sensitivity analyses under the treatment policy estimand

A sensitivity analysis of the co-primary endpoints will be evaluated under the treatment policy estimand during main-study period. The treatment policy estimand includes data collected after discontinuation of study treatment regardless of rescue by NP surgery or SCS_NP. Because these data will be analysed by MMRM which implicitly accounts for missing data under the MAR assumption, no imputation will be conducted for the missing data. The treatment policy estimand evaluate the treatment effect at the planned endpoint of the study, regardless of the treatments that patients actually received.

The analysis will be conducted using a MMRM analysis with treatment arm, baseline scores, visit, region (US vs non-US), baseline comorbid asthma status (yes vs no), and treatment×visit as covariates. The variance-covariance matrix will be assumed to be unstructured. If the procedure does not converge, then the Toeplitz, first-order autoregressive, and compound symmetric, variance components variance-covariance matrices will be tried in that order. The estimate of the treatment effect will be based on a contrast from this MMRM model.

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

8.6 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 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 YYYY-MM 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 subject died in the same month, then set the imputed date as the death date.

If both of the 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 subject died in the same year, then set the imputed date as the death date.

8.6.2 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 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 all ADA-positive samples using a ligand binding assay. The nAb results will be reported as positive or negative.

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 subjects are included in all analyses. If a positive ADA titre result is reported as ≤50, then the titre will be imputed as 50 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 on-treatment period (from Day 1 to EOT/IPD visit). Results will be listed separately.

- Subjects 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 subjects in a population is known as ADA prevalence.
- Subjects who are ADA negative at all assessments, including baseline and post-baseline (also generally referred to as ADA negative).
- Subjects who are ADA positive at baseline only.
- Subjects 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.
- Subjects 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.

- Subjects 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.
- Subjects who are ADA positive with maximum titre > median of maximum titres. The median of maximum titres will be calculated based on the maximum titre of each ADA positive subject within each treatment group (including both baseline and post-baseline measurements).
- 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.

The responses above will be summarized as counts and percentages by treatment group. The maximum ADA titre over the on-treatment period will also be summarized 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 subject, including any unscheduled assessments.

The immunogenicity will be summarized during the on-treatment period for all patients.

ADA response (positive or negative) and titre will be summarized 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 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 the proportion of subjects who are ADA positive at each visit will be provided.

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

Key patient information will be listed for patients with positive ADA results, including ADA status, nAb status, titer, and benralizumab serum concentration, and blood eosinophil level.

All analyses will be conducted on the safety analysis set by treatment group unless otherwise specified. All ADA results will be listed.

ADA and eosinophil levels

Eosinophil levels will be summarised by visit for the following ADA response categories of patients: ADA positive, treatment-emergent ADA positive, ADA negative, ADA persistently positive, nAb-positive, both ADA persistently positive and nAb positive and ADA positive with titer > median of maximum. 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 through EOT will be evaluated through summary statistics by treatment group and ADA status (ADA positive, treatment-emergent ADA positive, ADA negative, ADA persistently positive, nAb-positive, both ADA persistently positive and nAb positive and ADA positive with titre > median of maximum titre). Due to the expected small number of ADA positive subjects 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 study (for main study period only will be summarized by treatment group and ADA status (ADA positive, treatment-emergent ADA positive, ADA negative, ADA persistently positive and ADA positive with titre > median of maximum titre). The main study period is as defined in Section 3.1.1. The potential impact of ADA on hypersensitivity will also be assessed (overall and by causality as assessed by the investigator).

ADA and PK

Benralizumab serum concentrations will be summarised by treatment group, visit and ADA status (ADA positive, treatment-emergent ADA positive, ADA negative, ADA persistently positive, nAb-positive, both ADA persistently positive and nAb positive and ADA positive with titre > median of maximum titre) for patients in the PK analysis set.

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 NP surgery.

Table 9 Worst Possible Score

Endpoint	Worst Possible Score
NPS	8
NBS / DSS / NPSD individual item	3
SNOT-22 Total Score	110
SNOT-22 Individual Item	5
TSS	24
UPSIT	0
ACQ-6	6
SF-36	0
SF-36 Component	0
PGI-S	7 (very severe)
PGI-C	-3 (much worse)
Lund-Mackay Score	24
Sinus Severity Score	100

8.9 Random Seeds

The starting random seeds to be utilized 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 4.2.6.1), the random seed increments by 1. For example, imputation of Week 8 of NPS in monotone regression MI uses a seed of 385210, imputation of Week 16 of NPS uses 385211, and so on.

Table 10 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	264375	385210	504986	146085	658420	569078
NBS (NPSD)	185274	637105	819674	274189	620794	59428
DSS (NPSD)	649439	829542	928623	366750	934017	402264
Nasal congestion (NPSD)	137406	958605	NA	NA	NA	NA
Runny nose (NPSD)	995714	87609	NA	NA	NA	NA

Endpoint	Primary/COVID MCMC	Primary/COVID MI	WP after Surgery MCMC	WP after Surgery MI	DRMI MCMC	DRMI MI
Postnasal drip (NPSD)	277739	978153	NA	NA	NA	NA
Headache (NPSD)	920579	953780	NA	NA	NA	NA
Facial Pain (NPSD)	612791	673269	NA	NA	NA	NA
Facial pressure (NPSD)	127294	593028	NA	NA	NA	NA
Diff with sleeping (NPSD)	91744	906958	NA	NA	NA	NA
Diff with daily activities (NPSD)	736567	968982	NA	NA	NA	NA
TSS (NPSD)	262534	617645	NA	NA	NA	NA
SNOT-22	35486	137698	452739	82495	604952	586207
CT (LMS, SSS)	NA	NA	NA	NA	NA	NA
UPSIT	791320	427593	NA	NA	NA	NA
ACQ-6	401612	42430	NA	NA	NA	NA

8.10 Efficacy Estimands

Statistical Category	Estimand[1]			Section
	Endpoint (Population[2])	**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 (FAS) • CFB in mean NBS (FAS)	Included in analysis regardless of treatment discontinuation; WP after NP surgery, WOCF after SCS_NP[3] (primary estimand)	Mean difference between interventions (LSMD from CFB ANCOVA following hybrid WP/WOCF, MI at Week 40)[2]	4.2.6.1
Sensitivity	• CFB in endoscopic total NPS (FAS) CFB in mean NBS (FAS)	• Included in analysis regardless of treatment discontinuation or SCS_NP; WP after NP surgery[4] (primary estimand)	• Mean difference between interventions (LSMD from CFB ANCOVA following hybrid WP, MI at Week 40)	Appendix 8.2
	• CFB in endoscopic total NPS (FAS) CFB in mean NBS (FAS)	• Disc of treatment assumed based on DRMI[3] (primary estimand)	• Mean difference between interventions (LSMD from CFB ANCOVA following hybrid WP/WOCF, DRMI at Week 40)	Appendix 8.3
	• CFB in endoscopic total NPS (FAS) CFB in mean NBS (FAS)	• Remained adherent to treatment without surgery or SCS_NP[5] (effectiveness)	• Mean difference between interventions (LSMD from CFB MMRM at Week 40)	Appendix 8.4
	• CFB in endoscopic total NPS (FAS) CFB in mean NBS (FAS)	• Included in analysis regardless of treatment discontinuation, NP surgery, or SCS_NP[6] (treatment policy)	• Mean difference between interventions (LSMD from CFB MMRM at Week 40)	Appendix 8.5
Secondary Objective: To evaluate the effect of benralizumab on disease specific health-related quality of life				
Secondary/MCP	CFB in SNOT-22 score (FAS)	Included in analysis regardless of treatment discontinuation; WP after NP surgery, WOCF after SCS_NP[3] (primary estimand)	Mean difference between interventions (LSMD from CFB ANCOVA following hybrid WP/WOCF, MI at Week 40)[2]	4.2.7.1
Sensitivity	• CFB in SNOT-22 score (FAS)	• Included in analysis regardless of treatment discontinuation or SCS_NP; WP after NP surgery[4]	• Mean difference between interventions (LSMD from CFB ANCOVA following	Appendix 8.2

Statistical Category	Estimand[1]			Section
	Endpoint (Population[2])	**Intercurrent Event Strategy[1]**	**Population Level Summary[1] (Analysis)**	
		(primary estimand)	hybrid WP, MI at Week 40)	
	• CFB in SNOT-22 score (FAS)	• Disc of treatment assumed based on DRMI[3] (primary estimand)	• Mean difference between interventions (LSMD from CFB ANCOVA following hybrid WP/WOCF, DRMI at Week 40)	Appendix 8.3
	• CFB in SNOT-22 score (FAS)	• Remained adherent to treatment without surgery or SCS_NP[5] (effectiveness)		Appendix 8.4
	• CFB in SNOT-22 score (FAS)	• Included in analysis regardless of treatment discontinuation, NP surgery, or SCS_NP[6] (treatment policy)	• Mean difference between interventions (LSMD from CFB MMRM at Week 40) • Mean difference between interventions (LSMD from CFB MMRM at Week 40)	Appendix 8.5
Secondary Objective: To evaluate the effect of benralizumab on NP Surgery and/or SCS_NP				
Secondary/MCP	• Time to first NP Surgery and/or SCS NP (FAS) • Time to first NP Surgery (FAS)	Included in analysis regardless of treatment discontinuation[6] (treatment policy)	Hazard ratio from Cox proportional hazards model	4.2.7.2 4.2.7.3
Secondary Objective: To evaluate the effect of benralizumab on sense of smell				
Secondary/MCP	CFB in mean DSS (FAS)	Included in analysis regardless of treatment discontinuation; WP after NP surgery, WOCF after SCS_NP[3] (primary estimand)	Mean difference between interventions (LSMD from CFB ANCOVA following hybrid WP/WOCF, MI at Week 40)[2]	4.2.7.4
Sensitivity	• CFB in mean DSS (FAS)	• Included in analysis regardless of treatment discontinuation or SCS_NP; WP after NP surgery[4] (primary estimand)	• Mean difference between interventions (LSMD from CFB ANCOVA following hybrid WP, MI at Week 40)	Appendix 8.2
	• CFB in mean DSS (FAS)	• Disc of treatment assumed based on DRMI[3] (primary estimand) • Remained adherent to treatment	• Mean difference between interventions (LSMD from CFB ANCOVA following	Appendix 8.3

Statistical Category	Estimand[1]			Section
	Endpoint (Population[2])	**Intercurrent Event Strategy[1]**	**Population Level Summary[1] (Analysis)**	
	• CFB in mean DSS (FAS)	without surgery or SCS_NP[5] (effectiveness)	hybrid WP/WOCF, DRMI at Week 40)	Appendix 8.4
	• CFB in mean DSS (FAS)	• Included in analysis regardless of treatment discontinuation, NP surgery, or SCS_NP[6] (treatment policy)	• Mean difference between interventions (LSMD from CFB MMRM at Week 40) • Mean difference between interventions (LSMD from CFB MMRM at Week 40)	Appendix 8.5
Secondary Objective: To evaluate the effect of benralizumab on sinus opacification				
Secondary/MCP	CFB in mean LMS (FAS CT subset)	Included in analysis regardless of treatment discontinuation or SCS_NP; WP after NP surgery[4] (primary estimand)	Mean difference between interventions (LSMD from CFB ANCOVA following WP at EOT/IPD)	4.2.7.5
Sensitivity	• CFB in mean LMS (FAS CT subset)	• Included in analysis regardless of treatment discontinuation or SCS_NP; WP after NP surgery[3] (primary estimand)	• Mean difference between interventions (LSMD from CFB ANCOVA following WP/WOCF at EOT/IPD)	4.2.7.5
Safety Objective: To evaluate the safety and tolerability of benralizumab				
Safety	• Presence of TEAEs during main-study period (Safety) • Presence of serious TEAEs during main-study period (Safety) • Presence of TEAEs during on-study period (Safety) • Presence of serious TEAEs during on-study period (Safety) • Presence of TEAEs during on-study period (Safety – EFU Cohort) • Presence of serious TEAEs during on-study period (Safety – EFU Cohort)	Included in analysis regardless of treatment discontinuation[6] (treatment policy)	Categorical descriptive	4.2.10.1

FAS = Full Analysis Set; MCP = Multiple comparisons procedure; NP = Nasal Polyposis; SCS_NP = Systemic Corticosteroids used for Nasal Polyps; NPS = Nasal Polyps Score; NBS = Nasal Blockage Score; DSS = Difficulty with Sense of Smell; WP = Worst Possible; WOCF = Worst Observation Carried Forward;

Statistical Category	Estimand[1]			Section
	Endpoint (Population[2])	Intercurrent Event Strategy[1]	Population Level Summary[1] (Analysis)	

EOT = End of treatment; LSMD = Least Squares Mean Difference; CFB = Change from baseline; ANCOVA = Analysis of Covariance; MMRM = Mixed model for repeated measures; TEAE = Treatment emergent adverse event.

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

[2] Week 56 is also a multiplicity protected timepoint for co-primary and key secondary repeated measures endpoints

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

[4] Treatment Condition: Treatment with benralizumab versus placebo, regardless of compliance, where rescue with NP surgery indicates treatment failure.

[5] Treatment Condition: Treatment with benralizumab versus placebo, while on treatment and without rescue by NP surgery and/or SCS_NP.

[6] Treatment Condition: Treatment with benralizumab versus placebo, regardless of compliance or occurrence of rescue by NP surgery and/or SCS_NP.